Free energies and mechanisms of chemical reactions in solution and in enzymes with DFT QM/MM method

Weitao Yang, Duke University

DFT Theory
Biological
Nano
Energy

Funding
NSF
NIH
ONR
DOE (EFRC-UNC)

LONI, Baton Rouge, LA
July 2011

Studies of Biological Systems

-Development

-Enzyme/Solution Reaction Mechanism

-Design

Yingkai Zhang (NYU)
Haiyan Liu (USTC, China)
Andres Cisneros (Wayne State U)

Hao Hu
(HKU)

Jerry Parks
(ORNL)

Ganglong Cui

X. "Fox" Zeng

Pan Wu

Xiangqiang Hu

Steven Burger

Takumi Hori

Mingliang Wang

Kazuhiro Fujimoto

Zhenyu Lu

The Schrodinger Equation for N electrons

$$\hat{H}\Psi(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N) = E\Psi(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)$$

$$\mathbf{x} = \mathbf{r}, s$$

$$H = \hat{T} + \hat{V}_{ee} + \sum_i^N v_{ext}(\mathbf{r}_i)$$

$$\hat{V}_{ee} = \sum_{i<j}^N \frac{1}{r_{ij}}$$

$$\hat{T} = \sum_i^N -\frac{1}{2}\nabla^2$$

$$v_{ext}(\mathbf{r}) = -\sum_A -\frac{Z_A}{r_{Ai}}$$

How many data points are needed

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N)$$

$\mathbf{x}_1$	10	*points*
$\mathbf{x}_2$	10	*points*
...		...
$\mathbf{x}_N$	10	*points*

We need 10^N exponential growth with N

---the curse of dimensionality

The curse of dimensionality in QM

The problem is caused by the exponential increase in volume associated with adding extra dimensions to a (mathematical) space.

Another look at the problem

For N non-interacting electrons, the wavefunction is a Slater determinant

$$\Phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\mathrm{x}_1) \chi_j(\mathrm{x}_1) \cdots\cdots\cdots \chi_N(\mathrm{x}_1) \\ \chi_i(\mathrm{x}_2) \chi_j(\mathrm{x}_2) \cdots\cdots\cdots \chi_N(\mathrm{x}_2) \\ \vdots \qquad\qquad\qquad \vdots \\ \chi_i(\mathrm{x}_N) \chi_j(\mathrm{x}_N) \cdots\cdots \chi_N(\mathrm{x}_N) \end{vmatrix}$$

$$|\Phi_0\rangle = |\chi_1 \chi_2 \cdots \chi_i \chi_j \cdots \chi_N\rangle$$

How many determinants ?

$$|\Phi_0\rangle = |\chi_1\chi_2...\chi_i\chi_j...\chi_N\rangle$$

$$|\Phi_i^a\rangle = |\chi_1\chi_2...\chi_a\chi_j...\chi_N\rangle$$

$$\left|\Phi_{ij}^{ab}\right\rangle = |\chi_1\chi_2...\chi_a\chi_b...\chi_N\rangle$$

$$|\Psi\rangle = c_0\,|\Phi_0\rangle + \sum_{ia} c_i^a\,|\Phi_i^a\rangle + \sum_{ijab} c_{ij}^{ab}\,\left|\Phi_{ij}^{ab}\right\rangle + ...$$

For a total M orbitals for N electrons

1. How many n excitation determinants can you construct?

2. How many determinants in total can you construct?

Another look at the Hamiltonian

$$H = \hat{T} + \hat{V}_{ee} + \sum_i^N v_{ext}(\mathbf{r}_i)$$

Only $\sum_i^N v_{ext}(\mathbf{r}_i)$ depends on atoms and molecules

$$\sum_i^N v_{ext}(\mathbf{r}_i) = \int d\mathbf{r}\, v_{ext}(\mathbf{r}) \sum_i^N \delta(\mathbf{r} - \mathbf{r}_i)$$

$$\int d\mathbf{r}\, f(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}_0) = f(\mathbf{r}_0)$$
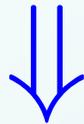
Introducing the electron density

$$\langle \Psi | \sum_i^N v_{ext}(\mathbf{r}_i) | \Psi \rangle \;=\; \int d\mathbf{x}^N \, |\Psi(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)|^2 \sum_i^N v_{ext}(\mathbf{r}_i)$$

$$=\; \int d\mathbf{x}^N \, |\Psi(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)|^2 \int d\mathbf{r} v_{ext}(\mathbf{r}) \sum_i^N \delta(\mathbf{r} - \mathbf{r}_i)$$

$$=\; \int d\mathbf{r} v_{ext}(\mathbf{r}) \int d\mathbf{x}^N \, |\Psi(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)|^2 \sum_i^N \delta(\mathbf{r} - \mathbf{r}_i)$$

$$=\; \int d\mathbf{r} v_{ext}(\mathbf{r}) \rho(\mathbf{r})$$

Exponential growth of information

N, positions and types of atoms $\Longleftrightarrow$ $N,\ v(\mathbf{r})$
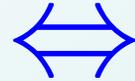
$\Downarrow$

$$H$$ $\Updownarrow$

$\Downarrow$
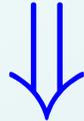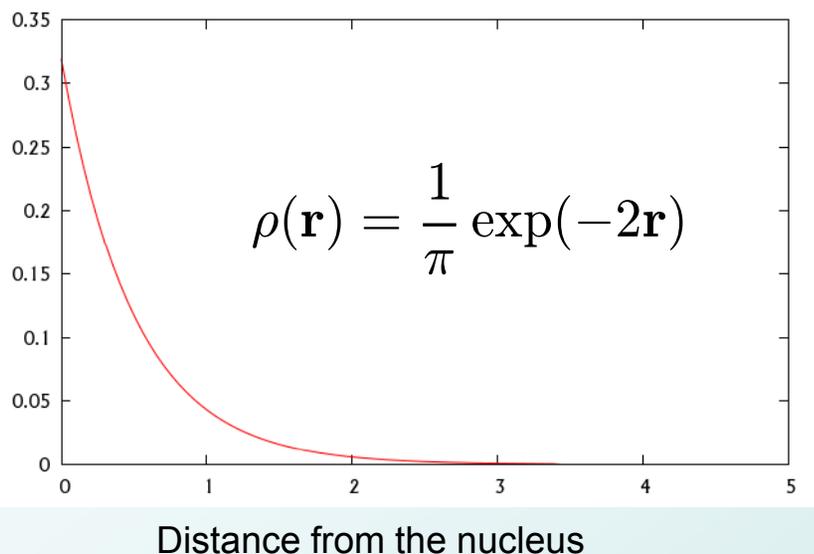
10^N in $\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)$ $\rho(\mathbf{r})$

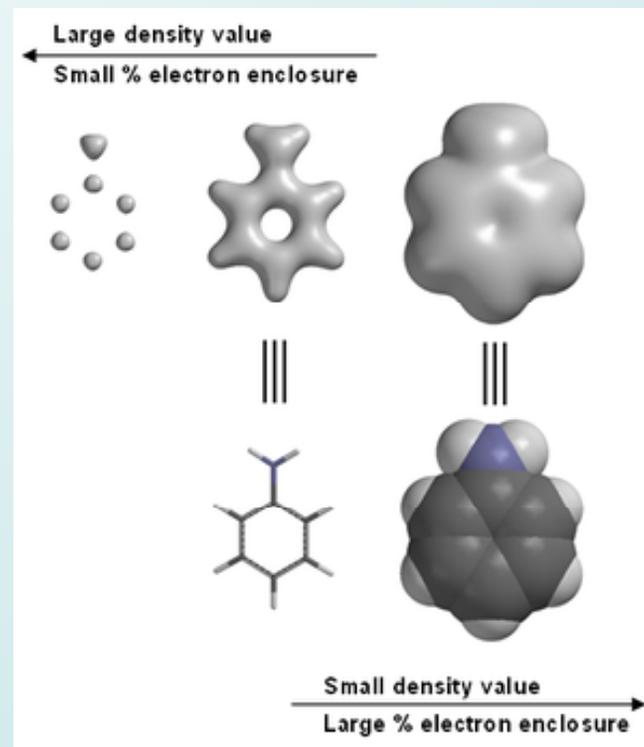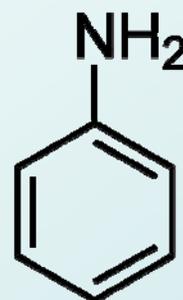1. $v(\mathbf{r})$ is the electrostatic potential from the nuclei

2. $v(\mathbf{r}) = \sum_A^{Atoms} -\frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|}$

Electron density $\rho(\mathbf{r})$ *, 3-dimentional only*

Electron density for the hydrogen atom

Electron density for aniline

$$\rho(\mathbf{r}) = \frac{1}{\pi} \exp(-2\mathbf{r})$$

Distance from the nucleus

NH₂

Large density value
Small % electron enclosure

Small density value
Large % electron enclosure

- **Electron density** $\rho(\mathbf{r})$ is the measure of the probability of an electron being present at a specific location.

- Experimental observables in X-ray diffraction for structural determination of small and large molecules

Density Functional Theory
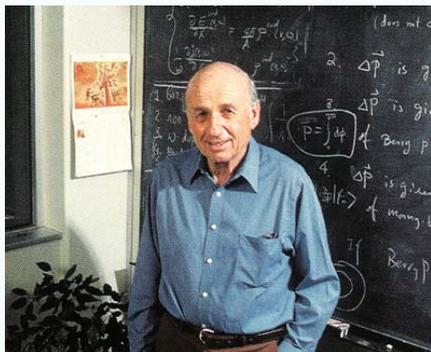
Walter Kohn

Pierre C. Hohenberg
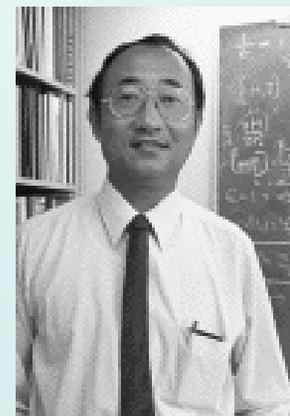
Lu J. Sham

The Nobel Prize in Chemistry 1998

Density functional theory (DFT) (1964, 1965)

1     $\rho(\mathbf{r}) \Longleftrightarrow N, v(\mathbf{r}) \Longleftrightarrow \Psi$

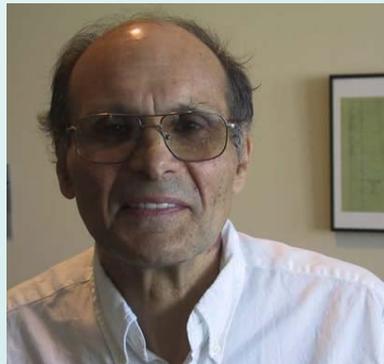2. $\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2$ Just the sum of molecular orbital densities

Wave function theory for ground states

$$
\begin{aligned}
E^0 &= \min_{\Psi} \langle \Psi | H | \Psi \rangle \\[2em]
&= \min_{\Psi} \langle \Psi | T + V_{ee} | \Psi \rangle + \langle \Psi | \sum_i^N v_{ext}(\mathbf{r}_i) | \Psi \rangle \\[2em]
&= \min_{\Psi} \langle \Psi | T + V_{ee} | \Psi \rangle + \int d\mathbf{r}\, v_{ext}(\mathbf{r}) \rho(\mathbf{r})
\end{aligned}
$$

Levy Constrained-Search Formulation (1979)

$$
\begin{aligned}
E^0 &= \min_{\Psi} \left\{ \langle \Psi | T + V_{ee} | \Psi \rangle + \int d\mathbf{r}\, v_{ext}(\mathbf{r}) \rho(\mathbf{r}) \right\} \\
&= \min_{\rho(\mathbf{r})} \min_{\Psi \to \rho(\mathbf{r})} \left\{ \langle \Psi | T + V_{ee} | \Psi \rangle + \int d\mathbf{r}\, v_{ext}(\mathbf{r}) \rho(\mathbf{r}) \right\} \\
&= \min_{\rho(\mathbf{r})} \left\{ \min_{\Psi \to \rho(\mathbf{r})} \langle \Psi | T + V_{ee} | \Psi \rangle + \int d\mathbf{r}\, v_{ext}(\mathbf{r}) \rho(\mathbf{r}) \right\} \\
&= \min_{\rho(\mathbf{r})} \left\{ F[\rho(\mathbf{r})] + \int d\mathbf{r}\, v_{ext}(\mathbf{r}) \rho(\mathbf{r}) \right\} \\
&= \min_{\rho(\mathbf{r})} E_v[\rho(\mathbf{r})]
\end{aligned}
$$

Density Functional Theory: a problem

We need to calculate electronic energy to predict structure of matter

$$\rho(\mathbf{r}) \Longleftrightarrow \Psi(\mathbf{r}_1, \mathbf{r}_2, ... \mathbf{r}_N)$$

$$E = E[\rho(\mathbf{r})] \; ?$$

$$E = \langle \Psi | \, H \, | \Psi \rangle$$

Density Functional Theory

Wavefunction Theory

E is unknown in terms of $\rho(\mathbf{r})$.

But $\rho(\mathbf{r})$ is only 3-dimensional.

E is known in terms of Ψ.

But Ψ grows exponentially!

a major challenge !

15

DFT: Exchange-correlation energy

$$E = E[\rho(\mathbf{r})]$$

E = Kinetic energy + potential energy
+ Coulomb interaction energy
+ $E_{xc}[\rho(\mathbf{r})]$

$E_{xc}[\rho(\mathbf{r})]$

- Exchange-correlation energy

- The only unknown piece in the energy

- About 10% of E

Approximations in exchange-correlation energy

$$E_{xc}[\rho(\mathbf{r})]$$

- 1965: Kohn and Sham, Local Density Approximation (LDA)

- 1980s-1990s: John Perdew   Axel Becke,   Robert Parr
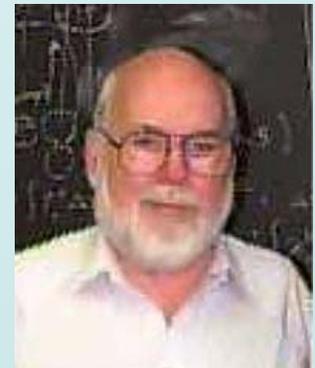
- Generalized Gradient Approximation (GGA)

- Hybrid Functionals (B3LYP, PBE0)

Applications of Density Functional Theory

- Structure of matter: atom, molecule, nano, condensed matter
- Chemical and biological functions
- Electronic
- Vibrational
- Magnetic
- Optical

Publications in Density Functional Theory

- ISI Web of Science search for articles with topic "density functional theory"

52,392 (April 2009)

67,156 (November 2010)
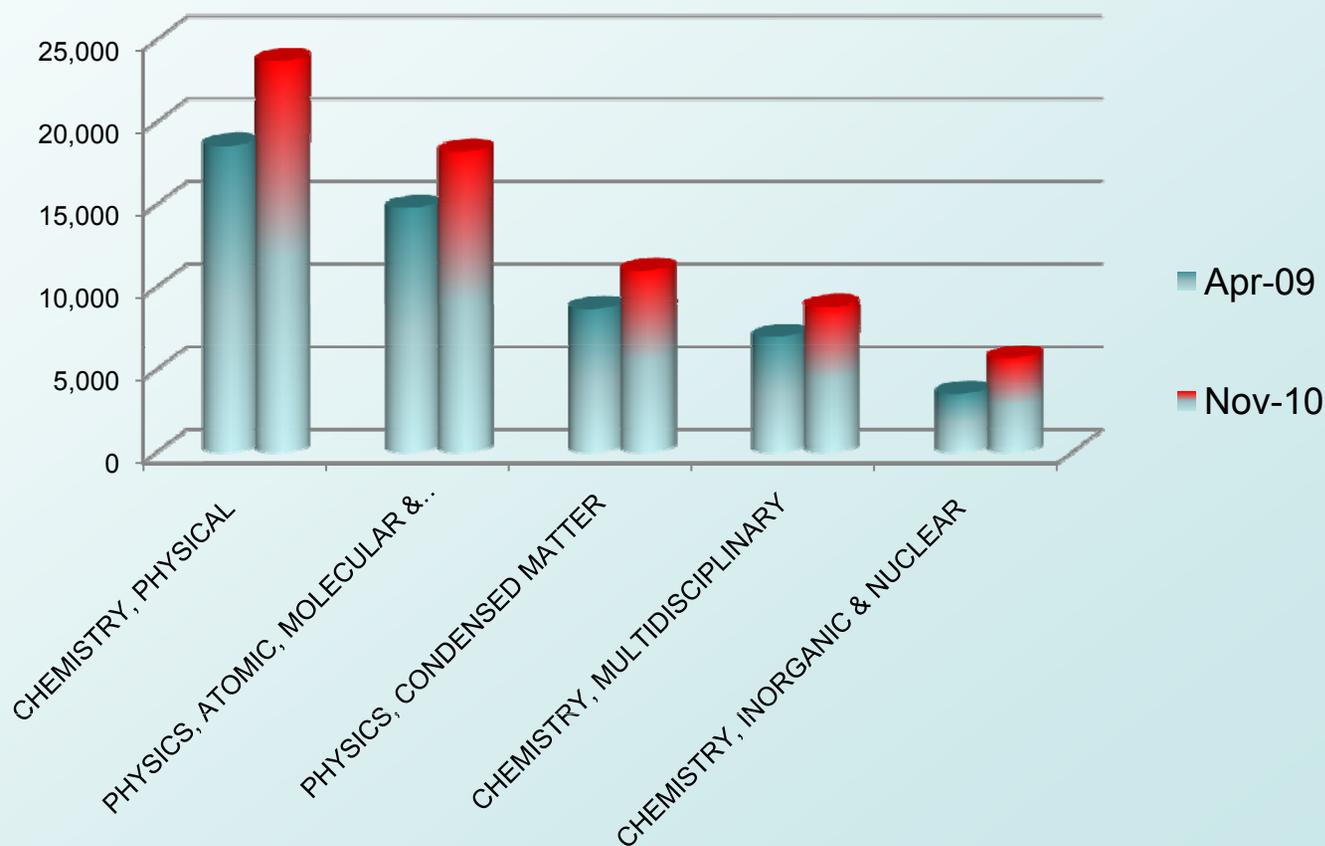
OUTLINE

Reaction Energies and Rates

Free Energies of Reactions with DFT QM/MM
   Minimum Free Energy Path (QM/MM-MFEP)

Mechanisms of Enzymatic Reactions
   4-Oxalocrontonate tautomerase (4OT)
   Orotidine 5′-Monophosphate Decarboxylase (ODCase)
   Pili

Redox Potentials

Chemical Reactions

$$a\,A \;+\; b\,B \;\rightarrow\; c\,C \;+\; d\,D$$

- a process that leads to the transformation of one set of chemical substances to another

- chemistry, materials, biology, life processes

$$A \rightarrow product$$

$$rate = -\frac{d[A]}{dt} = k[A]$$

Reaction Rate Theory

- Reaction is a rare event: reactant molecules have to cross the energy barrier to become product molecules. Direct dynamics simulation is very difficult because of the time scale.

- Reaction rate theory
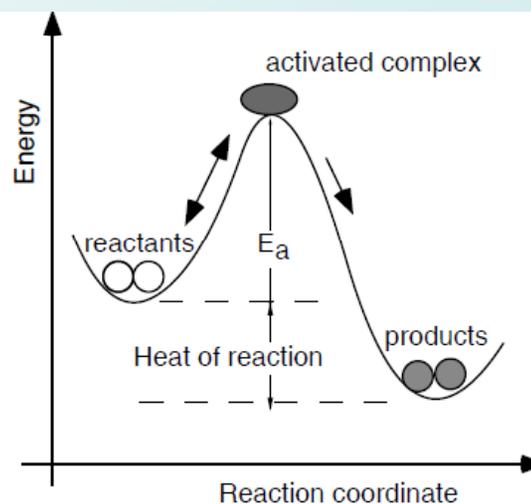
Figure 7 Schematic of a reaction via activated complex formation

$$k = \gamma \frac{kT}{h} \exp(-\Delta G^{\neq} / kT)$$

Transition State Theory

$$\gamma = 1$$

$$k = \gamma \frac{kT}{h} \exp(-\Delta G^{\neq} / kT)$$

Reaction: $HO^- + CH_3Br \rightarrow [HO\text{---}CH_3\text{---}Br]^{\ddagger} \rightarrow CH_3OH + Br^-$
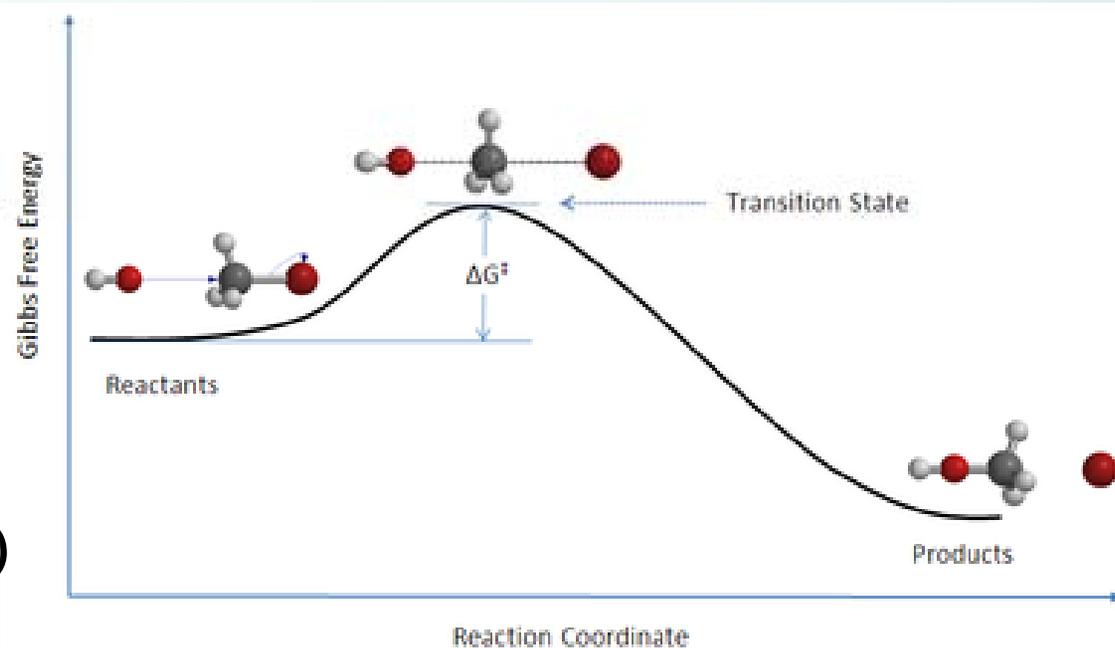
wikimedia.org

assumes a quasi-equilibrium between reactants and activated transition state complexes

Enzymes and Life Processes

The greatest majority of biochemical reactions do not take place spontaneously. The catalysts of biochemical reactions are **enzymes** and are responsible for bringing about almost all of the chemical reactions in living organisms.

Example

The **oxidation of a fatty acid** to carbon dioxide and water is not a gentle process in a test tube - extremes of pH, high temperatures and corrosive chemicals are required. Yet in the body, such a reaction takes place smoothly and rapidly within a narrow range of pH and temperature.

Understanding more about enzyme catalysts - what they are, what they do, and how they do it – leads to many advances in medicine and the life sciences.

How do enzyme work?

solution reaction

TS

TS Stabilization

R

P

enzyme reaction

Steps of Enzyme Reactions

1. $E + S \rightarrow ES$

2. $ES \rightarrow EP \quad (k_{cat})$

3. $EP \rightarrow E + P$

$ES \rightarrow EP \quad (k_{cat})$
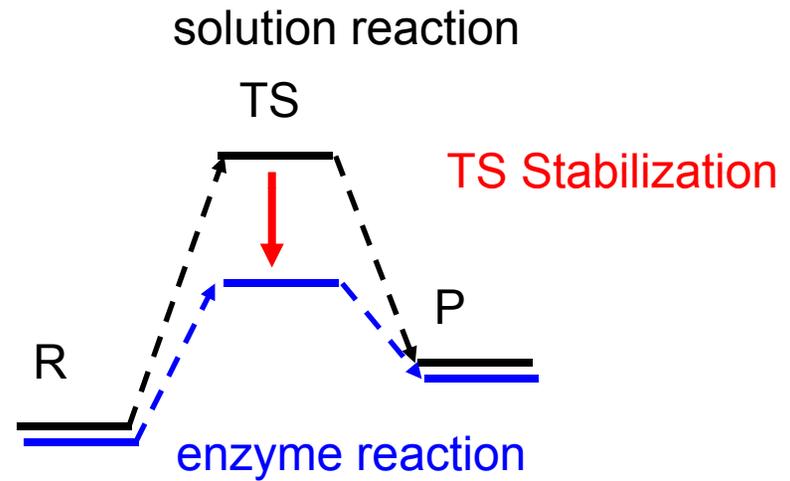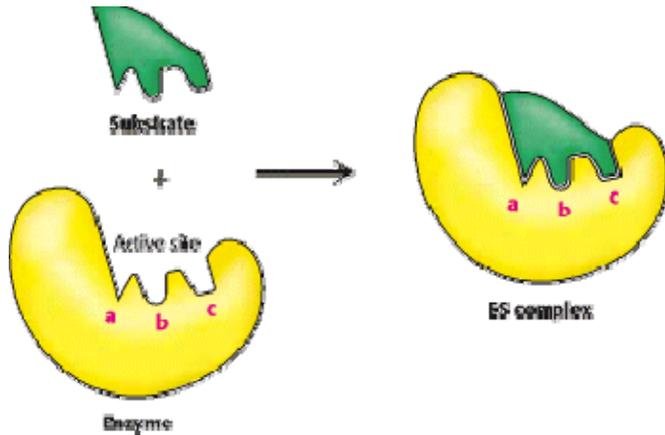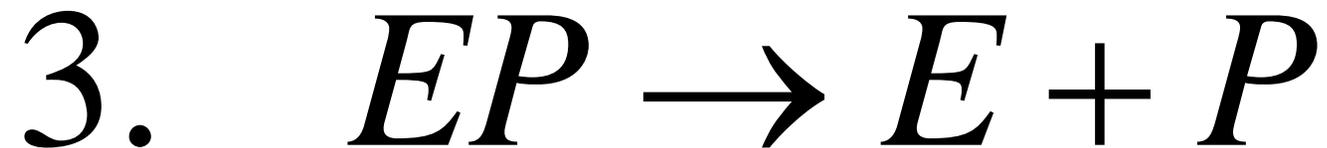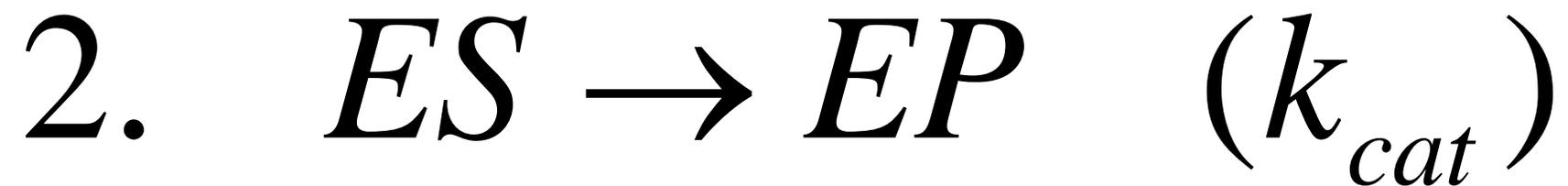
- Chemical Reactions: change of bonding – electron distribution.

- Study of Enzyme Reactions:

    (1) The origin of catalytic power of enzymes

    (2) The mechanism at atomic details

- Experiments: k_{cat}

    Kinetic measurements, mutation, isotope effects,

Challenges in Modeling Enzyme Catalysis

- Accuracy in activation free energies: ~1 or 2 kcal/mol

- Complex environment

Reaction in Gas Phase
    Quantum chemistry for reaction path, transition states

Reactions in Solutions and in Enzymes
    Aim to develop quantum mechanics and molecular dynamics for rate calculation in a robust way

Theory of Reaction Rate

$$k_{cat} = \gamma \frac{kT}{h} \exp(-\Delta G^{\neq}/kT)$$
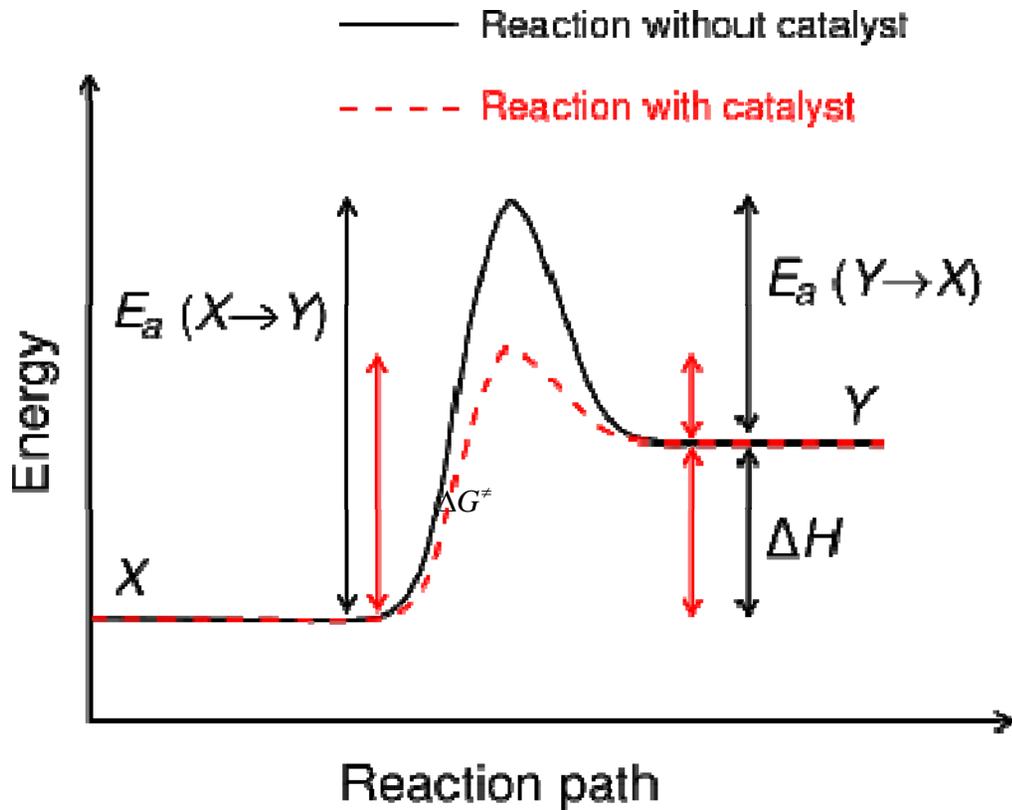
$\Delta G^{\neq}$

--The free energy of activation
--Quasi-equilibrium

γ

--The transmission coefficient
--Dynamics
--Transition State Theory

$$\gamma = 1$$

Potential Energy Surface E(R): QM/MM

Reaction active part: small (QM)

The rest: large (MM)

Warshel and Levitt (1976)

MM Region

QM Region

interaction
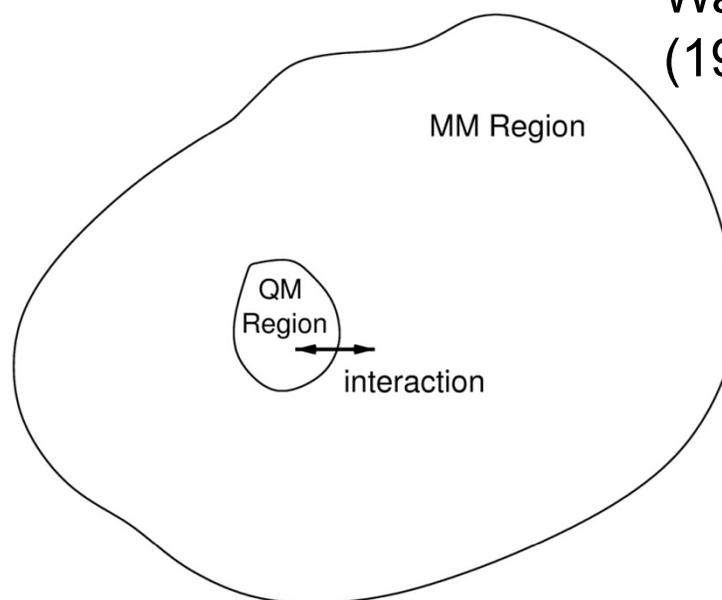
Reviews
Garcia-Viloca, Gao, Karplus, and Truhlar, 2005 *Science*
Friesner RA, Guallar V. 2005. *Annu. Rev. Phys. Chem.*
Mulholland AJ. 2005, *Drug Discov. Today*
Warshel, Sharma, Kato, Xiang, Liu, Olsson. 2006.. *Chem. Rev.*
Senn and Thiel. 2007. *Top Curr Chem*

Illustration of the pseudobond method.
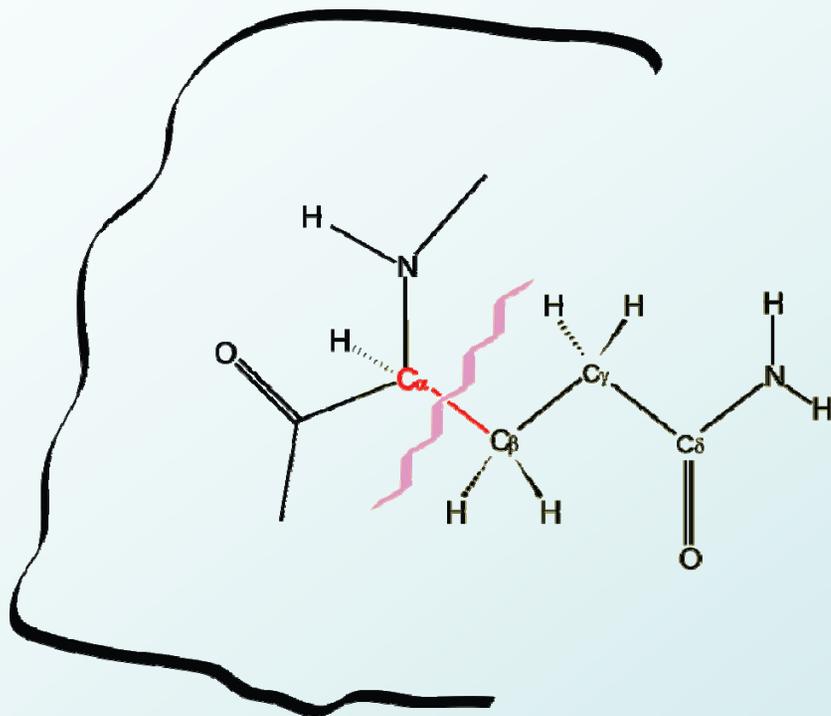
the C_α atom is described by a special basis set and designed effective core potential, with one free valence

Zhang, Lee, Yang. 1999. *J. Chem. Phys.*

Our pseudobond approach

QM region: pseudoatom and all atoms in the active part
MM region: the rest

$$E_{total} = E_{qm} + E_{mm} + E_{qm/mm}$$

$$E_{qm/mm} = E_{qm/mm}^{electrostatics} + E_{qm/mm}^{vdw} + E_{qm/mm}^{MM-bonded}$$

QM Method:

$$E_{qm} + E_{qm/mm}^{electrostatics} = \langle \Psi | H_{eff} | \Psi \rangle$$

MM Method:

$$E_{mm} + E_{qm/mm}^{vdw} + E_{qm/mm}^{MM-bonded}$$

Two types of QM methods

1. **Semiempirical QM methods** (MNDO, AM1, PM3, empirical valence bond, and SCC-DFTB)

   efficient -- direct MD sampling is readily affordable.

2. **Ab Initio QM**

   DFT - optimal balance of efficiency and accuracy.

   demanding -- rigorous statistical mechanics sampling and reaction dynamics calculations with an ab initio QM/MM method are most **challenging**.

Free Energy with ab initio QM/MM

• Jorgensen (*Acc. Chem. Res* 19893) developed a QM-Free Energy method, which uses the reaction path optimized for gas-phase reaction to carry out free-energy simulation in the condensed phase.

• Warshel and coworkers developed a QM(ai)/MM method: sampling with a simplified empirical valence bond potential and then corrected to the ab initio QM level (Rosta, Klahn, Warshel. 2006. *JCP)*

• Gao and Truhlar: Dual level QM -- correcting QM in the reaction coordinate

• ….

• Direct dynamics with DFT QM/MM
        (Rothlisberger, Zhang,)

• QM/MM-FE, reaction path on total potential energy surface (Zhang, Liu and Yang, 2000, JCP)

• QM/MM-MFEP, reaction path on free energy surface (Hu and Yang, 2007-)

Issues with structures and reaction path on total energy surface

- Depend on initial protein/solvent conformations

- Can fail to determine reaction path (defined on total energy) for enzymatic reactions

- Complete failure for reactions in solutions

- Not yet a robust method (like gas phase calculations)

QM/MM Minimum Free Energy Path (QM/MM-MFEP)

Hao Hu, Zhenyu Lu and WY, JCTC 2007

(Duke, now HK Univ,)

- Combine free energy calculations with reaction path determination

- Define the reaction path on the potential of mean force (PMF) surface -- One path represents an ensemble of paths for the whole systems

- Use chain-of-state methods, like the QSM, on the PMF.

- **Eliminate the dependence of the reaction path on the initial solvent, or protein conformations.**

Partition function and Free energy

$$Z_0 = \int \exp(-\beta E(\mathbf{r}_{QM}, \mathbf{r}_{MM})) d\mathbf{r}_{QM} d\mathbf{r}'_{MM}$$

$$A_0 = -\frac{1}{\beta} \ln(Z_0)$$

Potential of Mean Force

$$Z_0(\mathbf{r}_{QM}) = \int \exp(-\beta E(\mathbf{r}_{QM}, \mathbf{r}_{MM})) d\mathbf{r}'_{MM}$$

$$A_0(\mathbf{r}_{QM}) = -\frac{1}{\beta} \ln(Z_0(\mathbf{r}_{QM}))$$

$$\frac{\partial A_0(\mathbf{r}_{QM})}{\partial \mathbf{r}_{QM,i}} = \frac{\partial \left[-\frac{1}{\beta} \ln(Z_0(\mathbf{r}_{QM})) \right]}{\partial \mathbf{r}_{QM,i}} = \left\langle \frac{\partial E(\mathbf{r}_{QM}, \mathbf{r}_{MM})}{\partial \mathbf{r}_{QM,i}} \right\rangle_{\mathbf{r}_{MM}}$$

PMF Surface is normally calculated with free energy perturbation, and is noisy!

$$Z_0(\boldsymbol{r}_{QM}) = \int \exp(-\beta E(\boldsymbol{r}_{QM}, \boldsymbol{r}_{MM}))d\boldsymbol{r}_{MM}$$

$$A_0(\boldsymbol{r}_{QM}) = -\frac{1}{\beta}\ln(Z_0(\boldsymbol{r}_{QM}))$$

Sequential Sampling and Optimization for QM/MM-MFEP
Hao Hu, Zhenyu Lu and WY, JCP 2008,

- Main computational cost of QM/MM-MFEP method is the statistical sampling required for the calculation of the QM PMF and its gradients.

- In the sequential sampling and optimization approach:
  - sampling MM phase space with fixed QM
  - optimizing the QM subsystem in the fixed ensemble of MM
  - iterating until convergence.

- The use of a fixed-size, finite MM conformational ensemble enables the precise evaluation of the QM PMF and its gradients

The idea of sequential sampling and optimization

$$\min_{r_{QM}} A_0(r_{QM}) = \min_{r_{QM}} -\frac{1}{\beta} \ln \sum_{r_{MM}} \exp(-\beta E(r_{QM}, r_{MM}))$$

Iteration:

    a) MM sampling with
       fixed QM geometry.

    b) QM optimization with
       fixed MM ensemble.

$$\min_{r_{QM}} \sum_{r_{MM}}$$

QM/MM-MFEP

t

Animation made by Hao Hu

- Classical S_N2 reaction: $CH_3Cl + Cl^-$
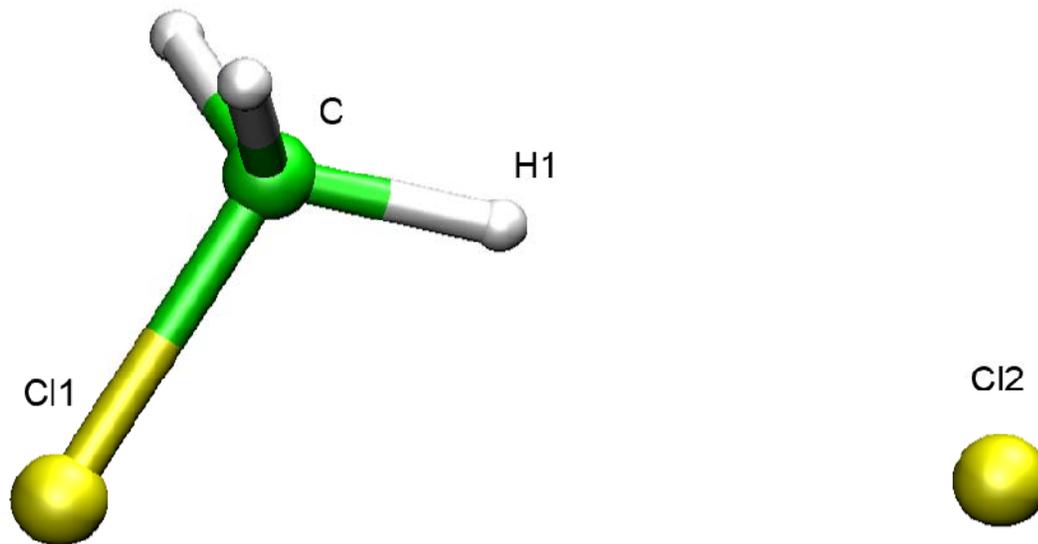
Sequential QM/MM-MFEP Applications: convergence behavior

- ## Convergence of relative free energy (S_N2)
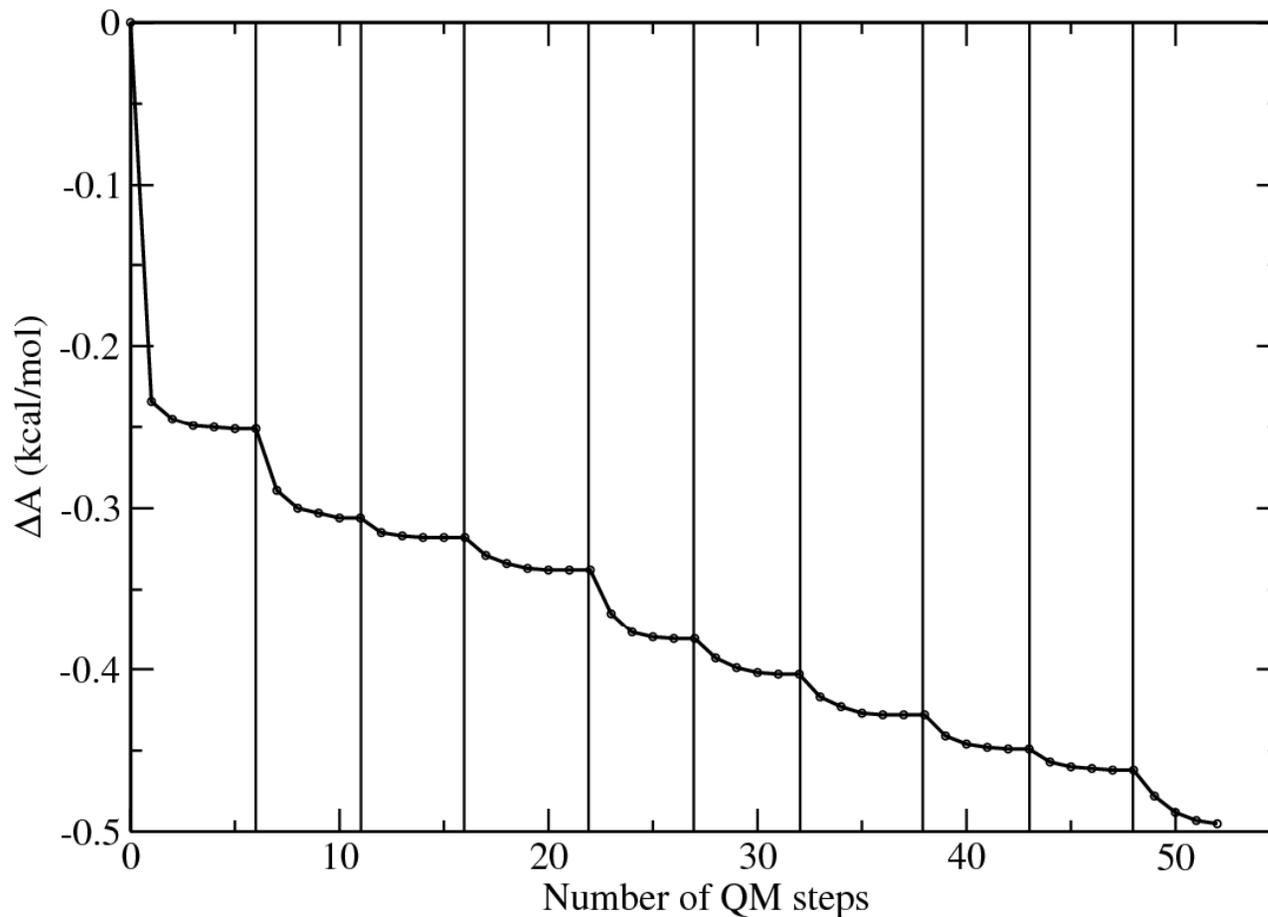
Optimization of the reaction path on the QM PMF

- Can use any path optimization approach

- Quadratic String Method (QSM), Steven Burger and WY, JCP, 2006

Steven Burger

Sequential QM/MM-MFEP Applications: S_N2 reaction

- ## RS, TS structures

- ## Activation free energy

  $\Delta G^{\ddagger}$(MFEP) = 18.7 kcal/mol

  harmonic corrections:

  $\Delta G^{\ddagger}$(MFEP+Freq) = 20.4 kcal/mol

  Additional 3-fold rotational
  multiplicity

  $kT \ln(1/3) = -0.65 \, kcal/mol$

  $\Delta G^{\ddagger}$(MFEP) = 21.1 kcal/mol

v.s.

  $\Delta G^{\ddagger}$(direct dynamics) = 20.7 kcal/mol

index	TS	RS
1	**-298.64**	41.72
2	66.37	49.44
3	95.87	57.76
4	106.46	66.09
5	113.75	77.40
6	125.46	109.97
7	129.38	168.40
8	214.60	175.90
9	217.02	182.38
10	234.07	708.32
11	976.67	1039.97
12	982.33	1052.31
13	1125.74	1408.77
14	1425.90	1497.86
15	1429.94	1514.00
16	3220.21	3146.16

Free Energy Convergence
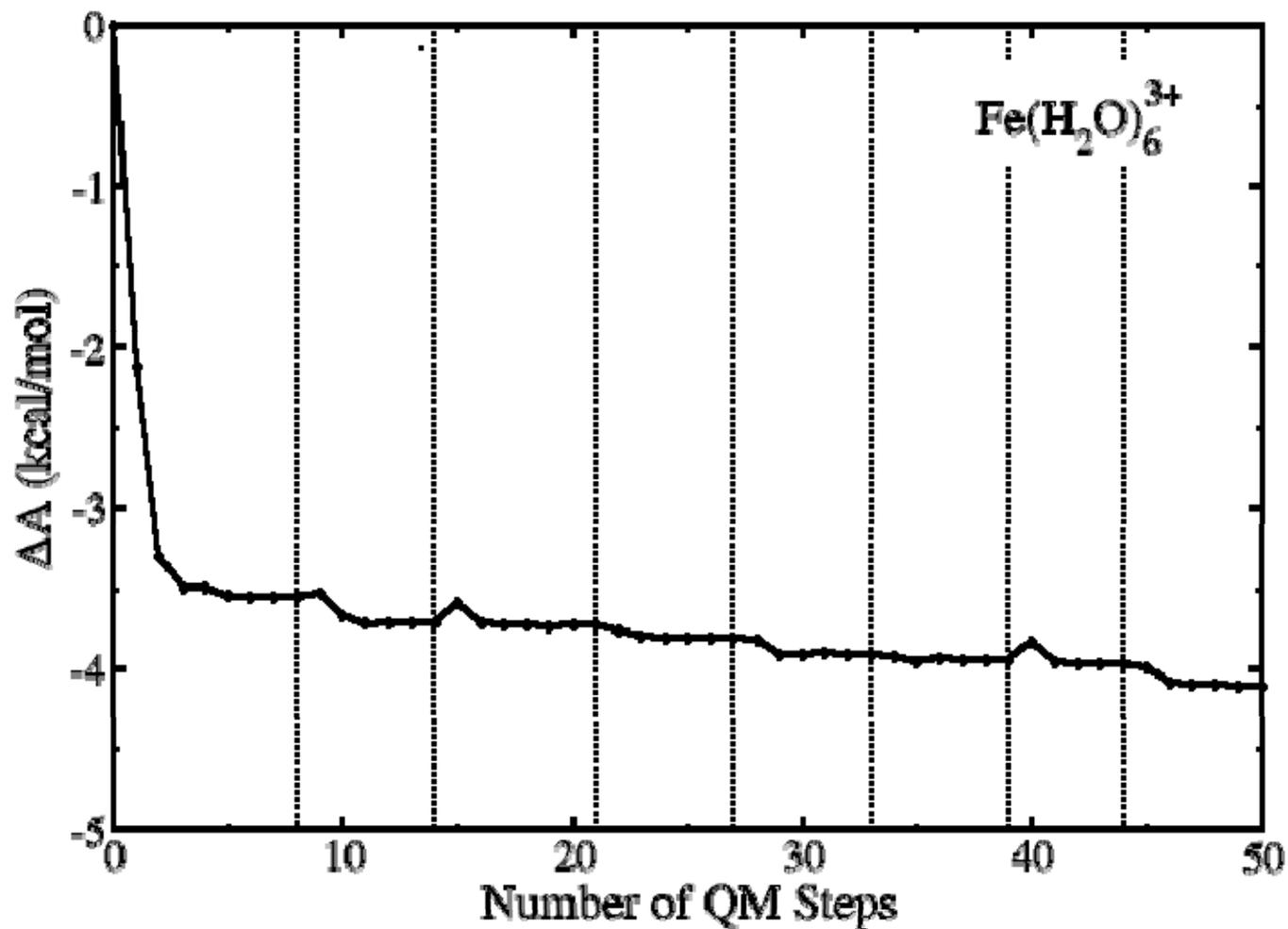
Ab initio QM/MM minimum free energy path

Convergence
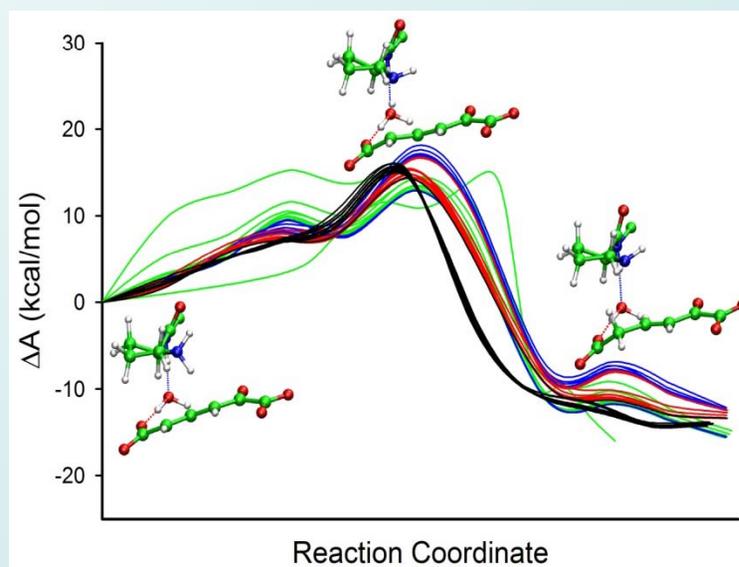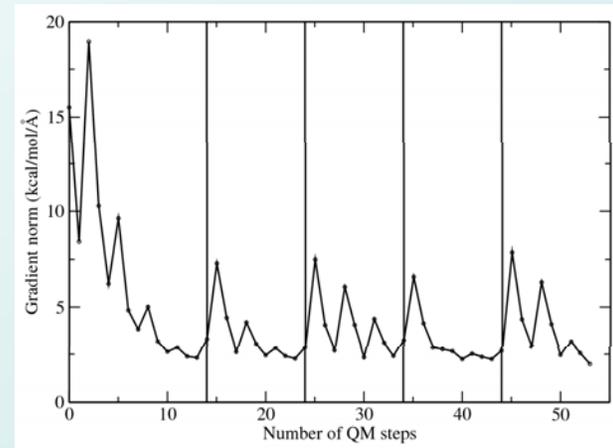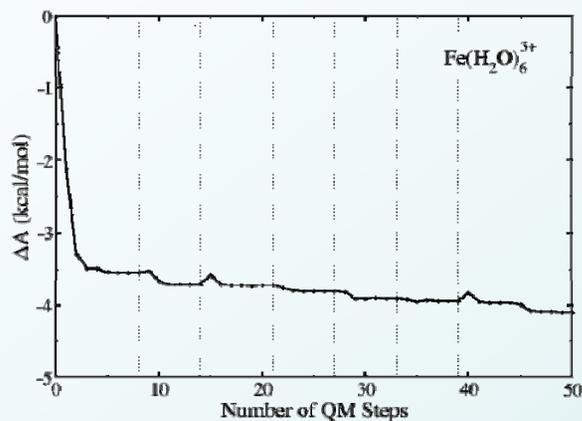
Hu, et al. *J. Chem. Phys.*, 2008; Hu & Yang, *Annu. Rev. Phys. Chem.*, 2008

Ab initio QM/MM minimum free energy path

Advantages

- Complicated solution and enzyme reaction becomes gas-phase alike

- Remove the path-dependence of initial conformation

- Good statistics

- Optimizing reaction path without explicit definition of reaction coordinates

- **A major step toward robust application of ab initio QM/MM to solution and enzymes**.

Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods

Hao Hu and Weitao Yang

Department of Chemistry, Duke University, Durham, North Carolina 27708;
email: hao.hu@duke.edu, weitao.yang@duke.edu

Key Words

enzyme catalysis, solution reaction, enzyme proficiency, multiscale method, potential of mean force, QM/MM

Ann. Rev. Phys. Chem, 2008

4-Oxalocrontonate tautomerase (4OT)
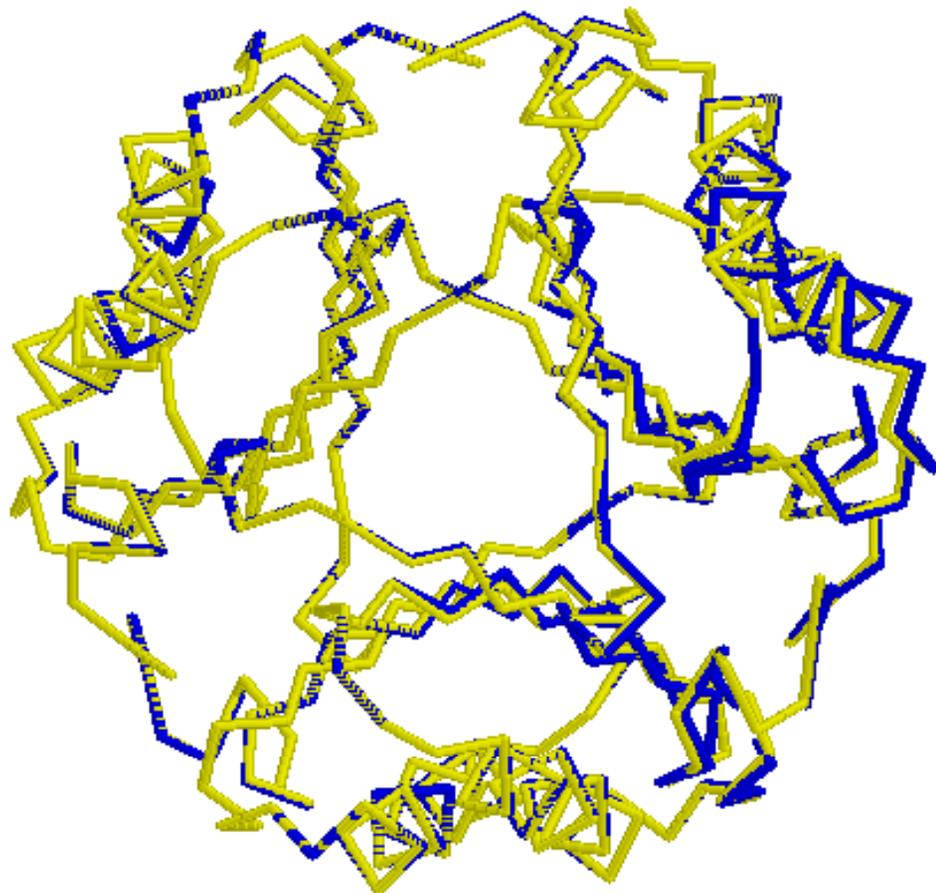
- part of a degradative metabolic pathway

six subunits arranged in dimers

Cisneros, Liu, Zhang, and Yang, JACS, 125, 10384 (2003)

4-Oxalocrontonate tautomerase (4OT)
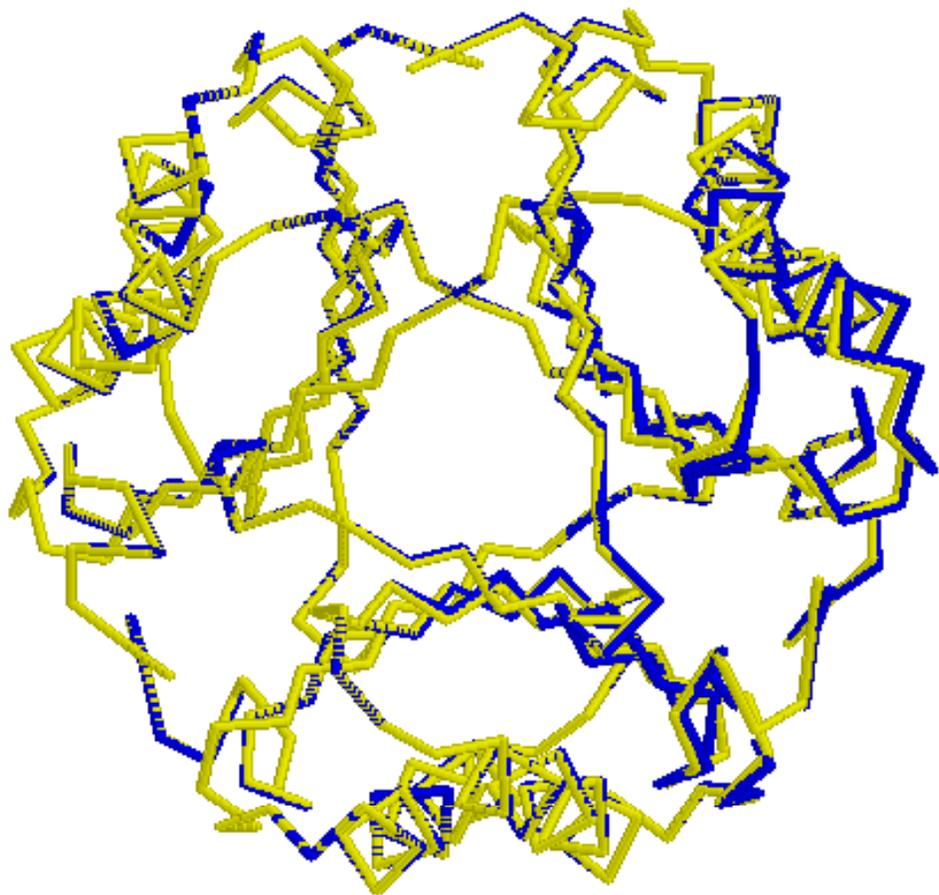
- **Reveal a Mechanism without acid**

- **Water**

- **Arg-39**

- **Prediction on mechanism and mutation effects confirmed by expt later.**

Cisneros, Liu, Zhang, and Yang, JACS, 125, 10384 (2003)

Classic Acid-Basis Catalyzed Reactions?
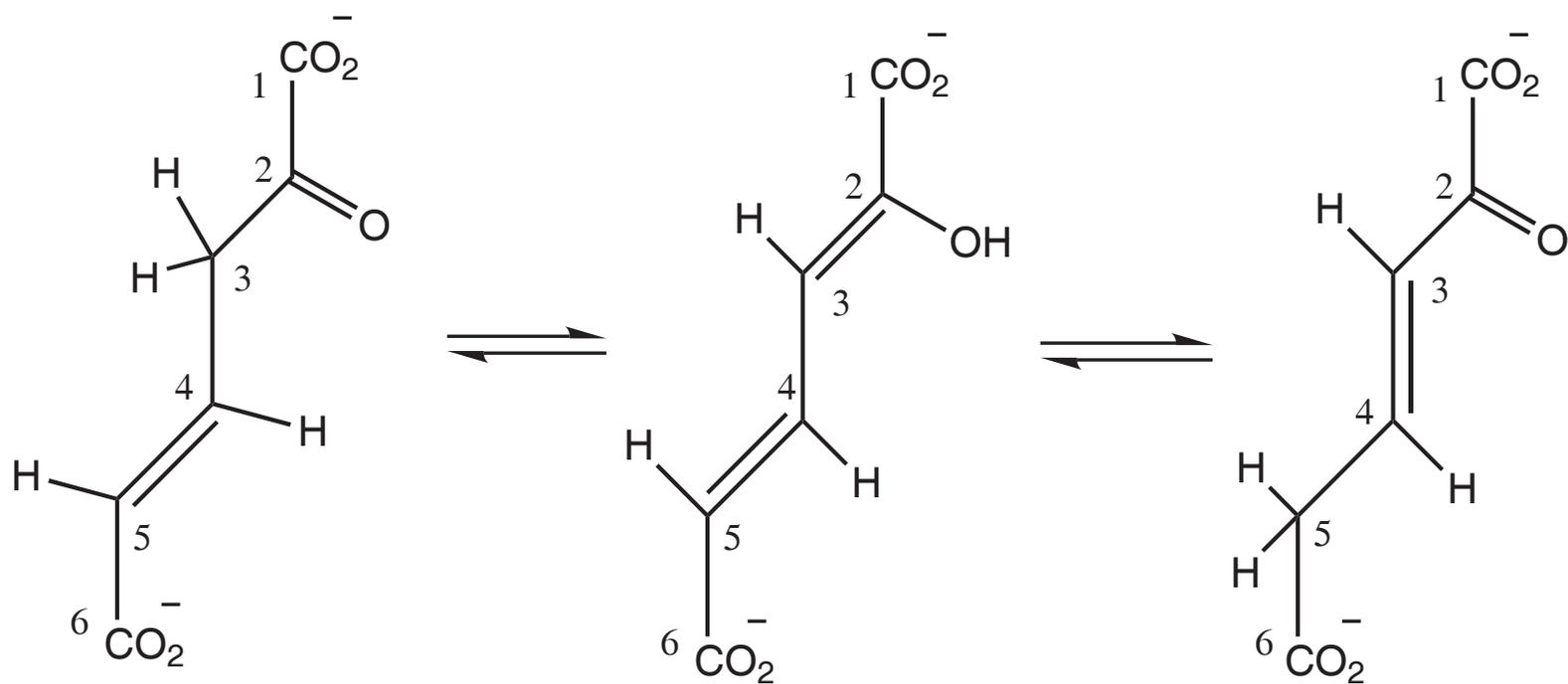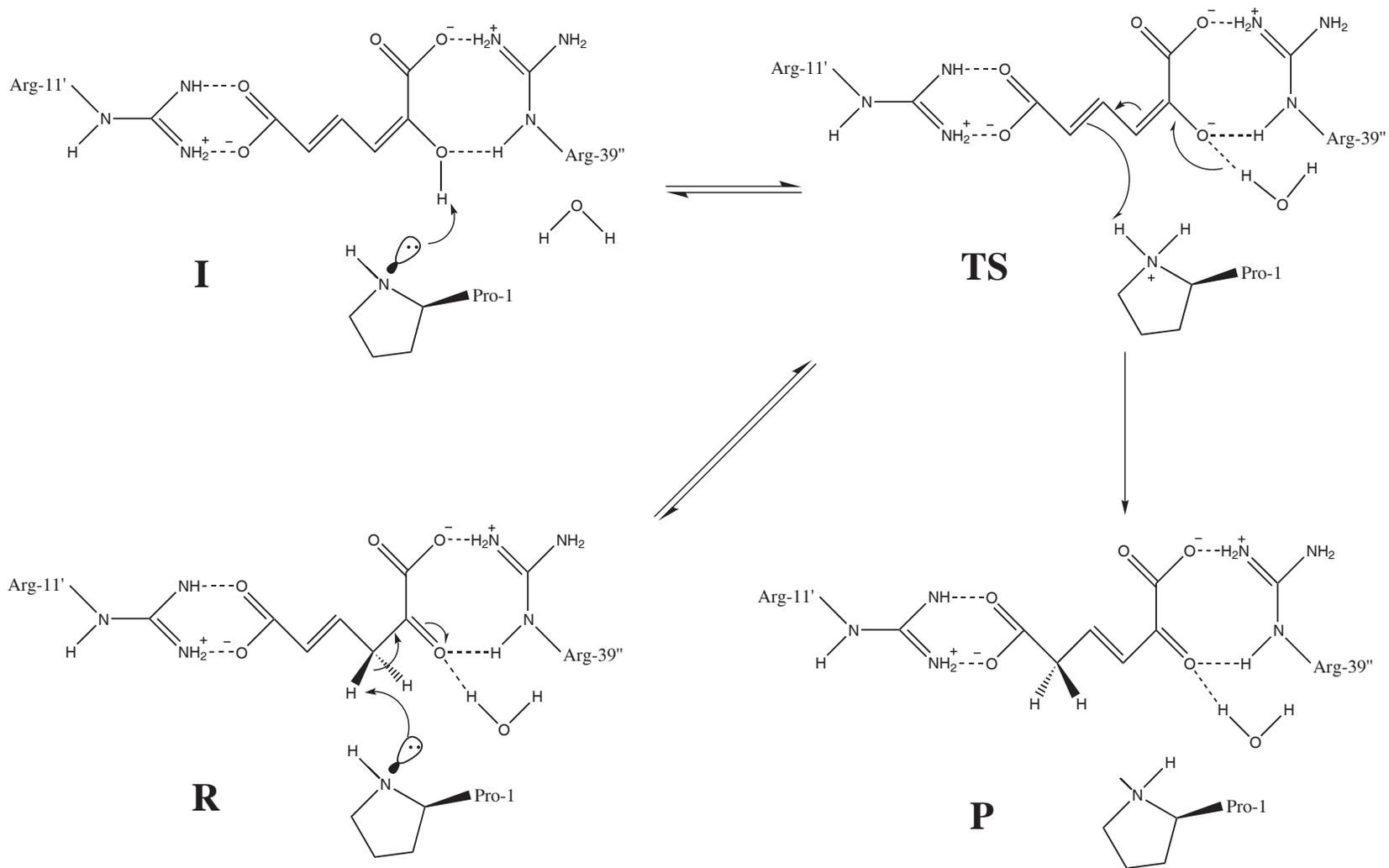
... the Arg39Cit

- reduces k_{cat} by 1660-fold, consistent with Arg39 interacting with the developing negative charge of the ketoacid group in the transition state.

- gives further evidence against the role of Arg39 as a general acid, in agreement with recent **computational studies**.

Protein backbone makes important contributions to 4OT catalysis: Understanding from experiment and theory
Cisneros, Wang, Silinski, Fitzgerald, and Yang, Biochemistry (2004).

Calculations: QM/MM energy calculations,

Experiments: backbone chemical modification of NH to O,
by Wang, Silinski and Fitzgerald at Duke

	Leu8	(OL8)4OT(R61A)
$\Delta\Delta E$ Calculation	1.0	1.5
$\Delta\Delta F$ Expt	1.8	1.2

Motivation for recent work

Pan Wu

➢ Are six potential active sites fully occupied?
➢ How does one active site influence the adjacent one?
➢ What is an appropriate theoretical model for this system?
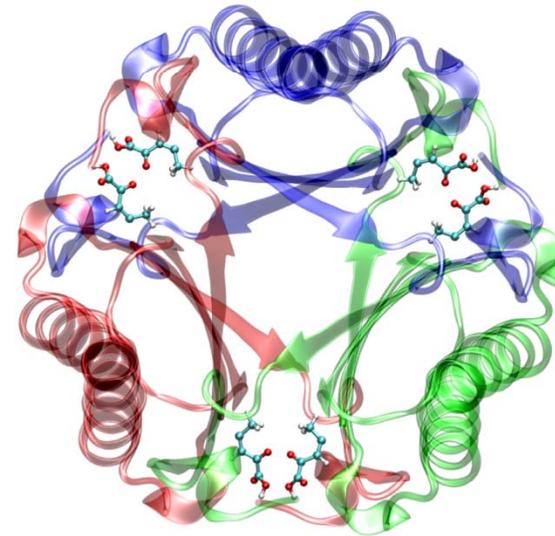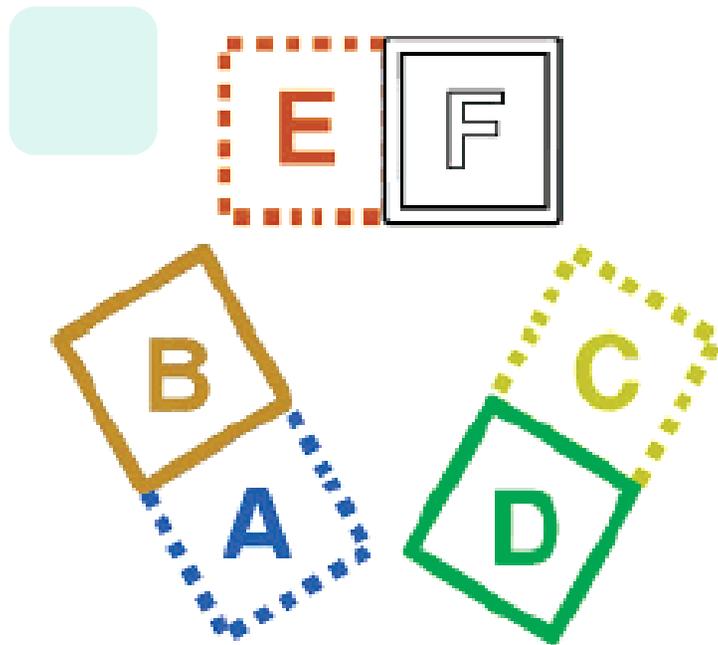
Motivation (con't)

Proline shuttled transfer
(Cisneros, GA et al)
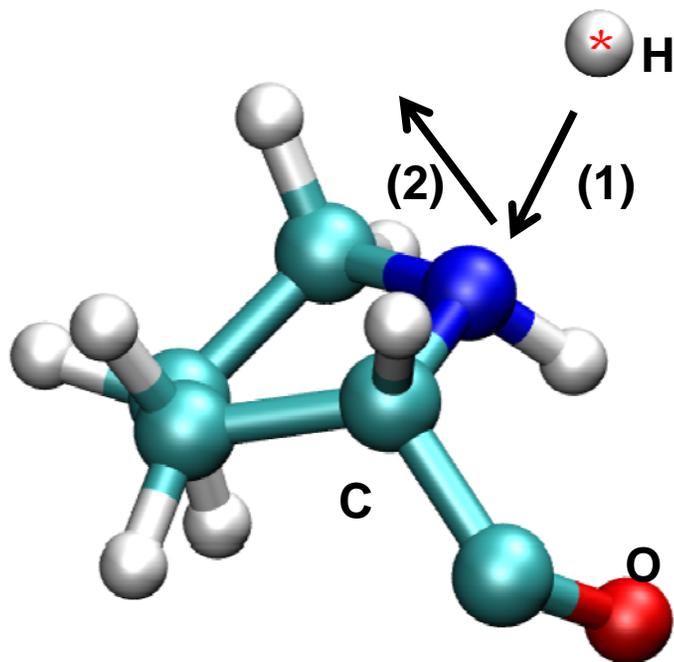
Proline proton involved transfer
(Ruiz-Pernia, JJ et al)
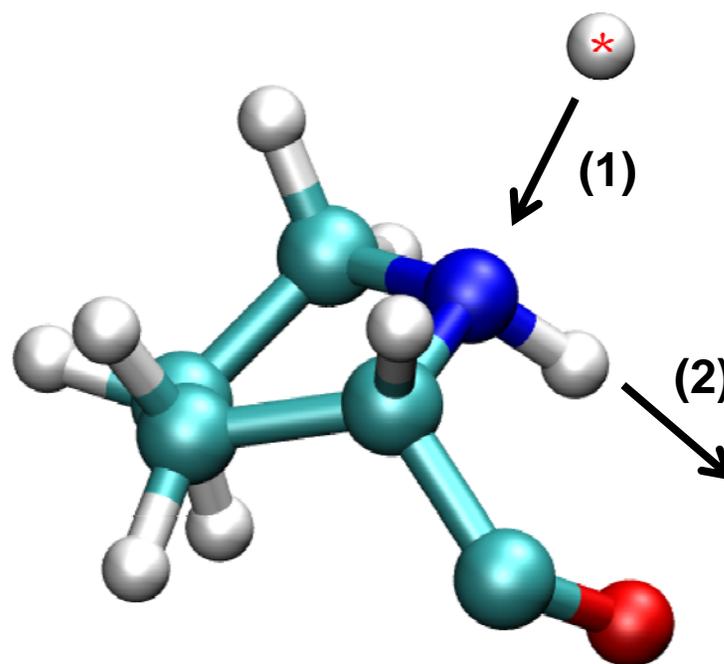
Both have been used in previous theoretical studies; however, which one is the preferred one?

Ruiz-Pernia, JJ et al. JACS 131, 2687 (2009)
Cisneros, GA and WYI, JACS, 125, 10384 (2003)

60

Simulation Protocol

Name	Dimer #	Substrate #	Notes
3D3S	3	3	Hexamer model, half-of-the-sites occupied
3D6S	3	6	Hexamer model, all sites occupied
1D1S	1	1	dimer model

substrate

Pro1

Pseudobond

Ile2
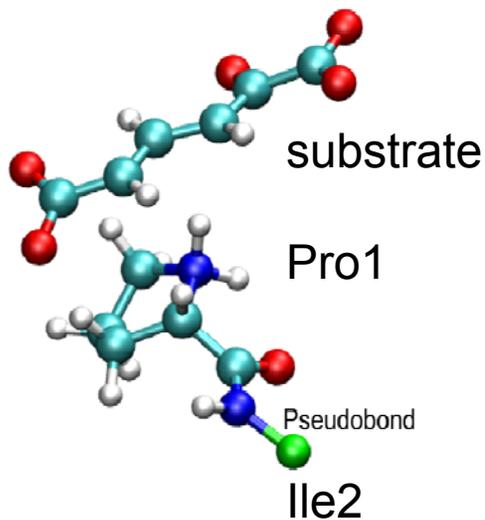
Result 1. dimer model is inappropriate to simulate the reaction
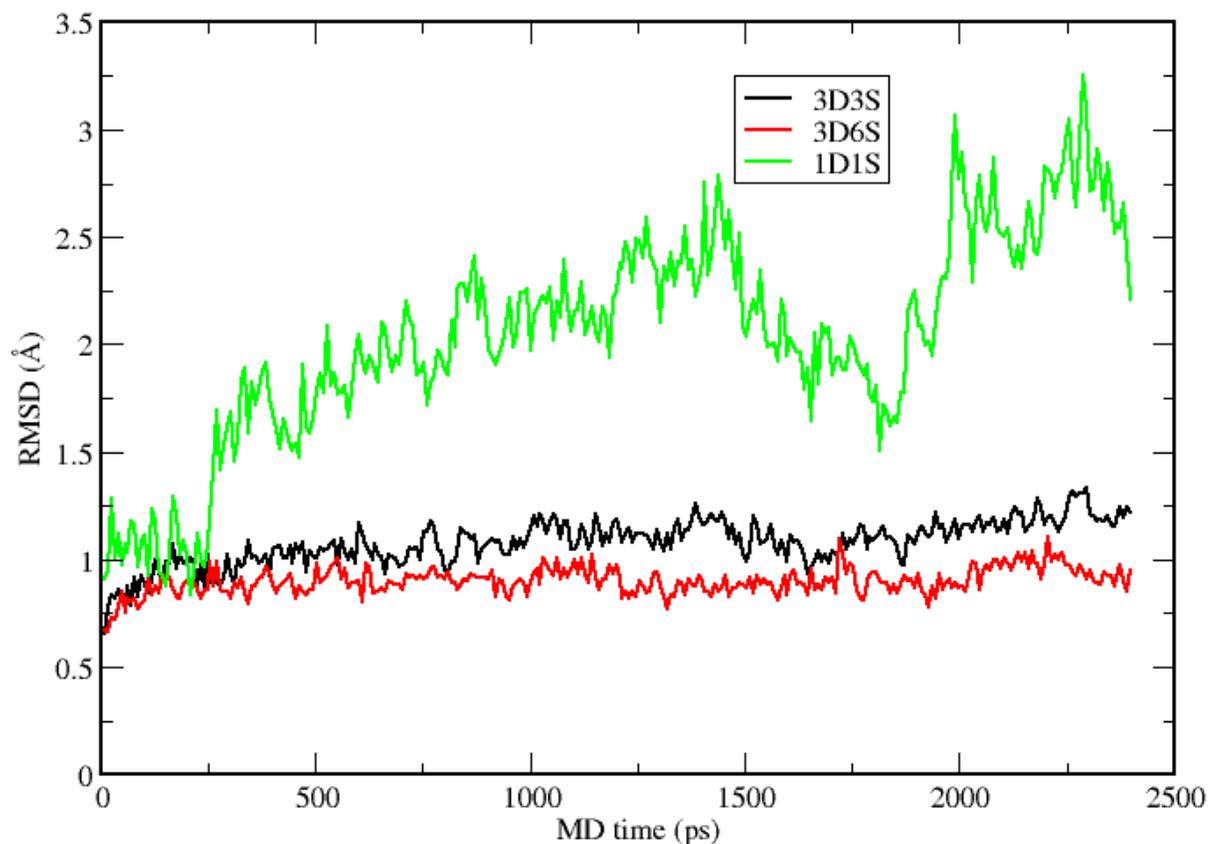
Large Root-Mean-Square-Deviation (RMSD) in dimer model shows highly unstable global structure.

Non-conserved active site structure: Hydrogen bonds between Arg39 – substrate and Arg61' – substrate show high fluctuations in the active site.
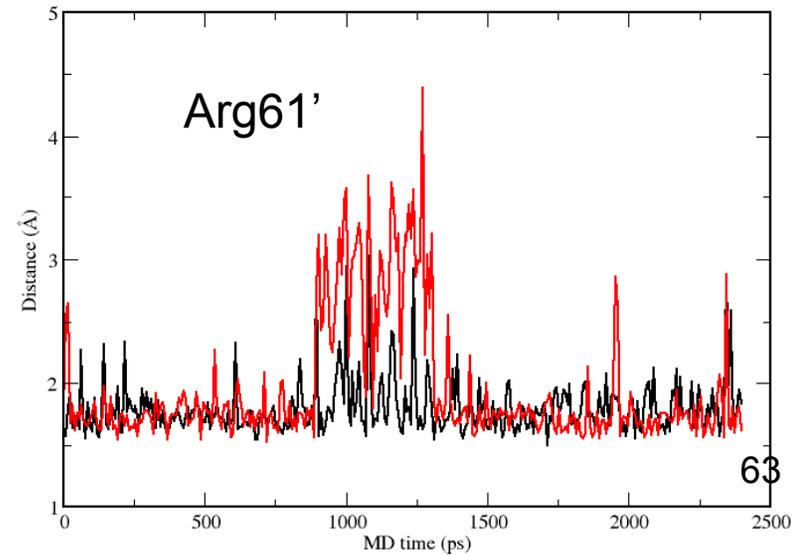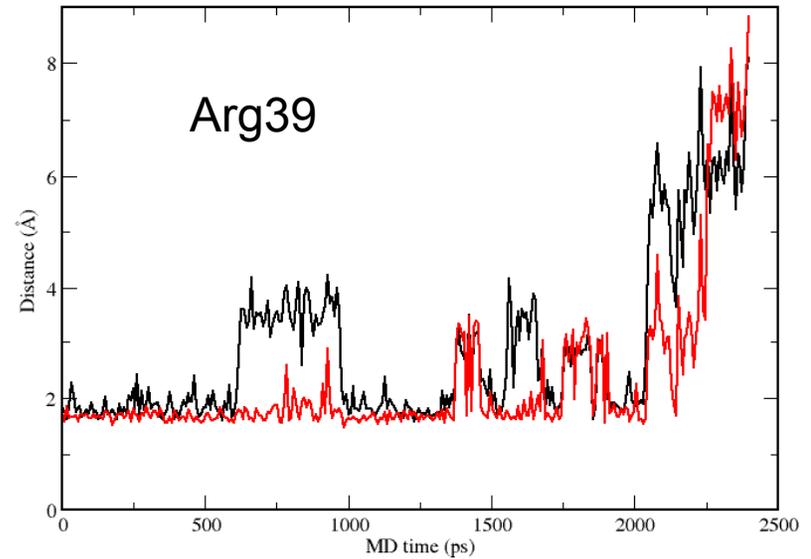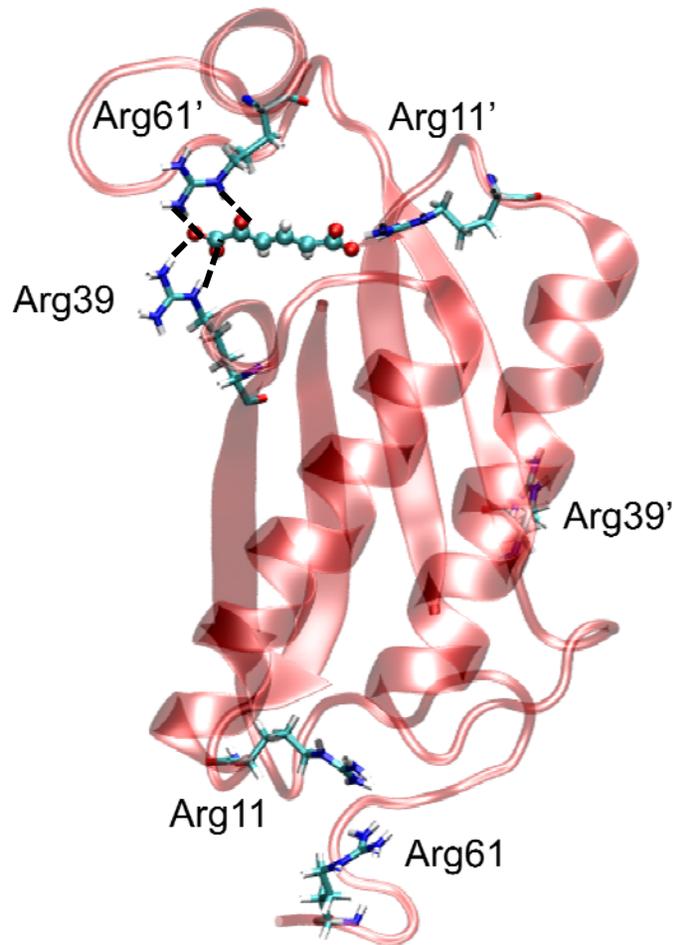
Result 2: One-proton transfer is preferred
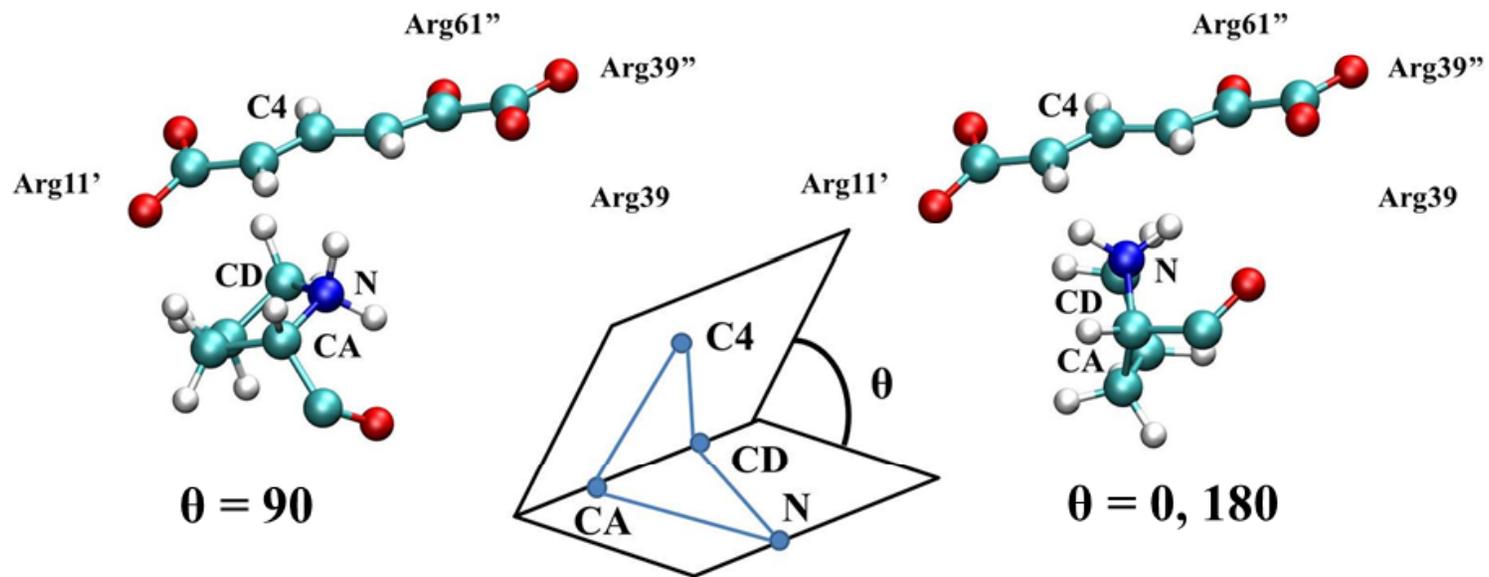
➤ Intermediate MD simulations show the average dihedral angle is 62°.
➤ Direct QM/MM MD simulations on the reactant show the average dihedral angle is 88°.
➤ Both indicate that substrate and Pro1 are parallel, and one proton transfer is preferred.

Result 3. Half-of-the-sites occupation mechanism is supported

1D1S: not realistic
3D6S: $\Delta G^{\ddagger} = 27.1$ kcal/mol (▬ ▬ ▬)
3D3S: $\Delta G^{\ddagger} = 18.5$ kcal/mol (▬▬▬)
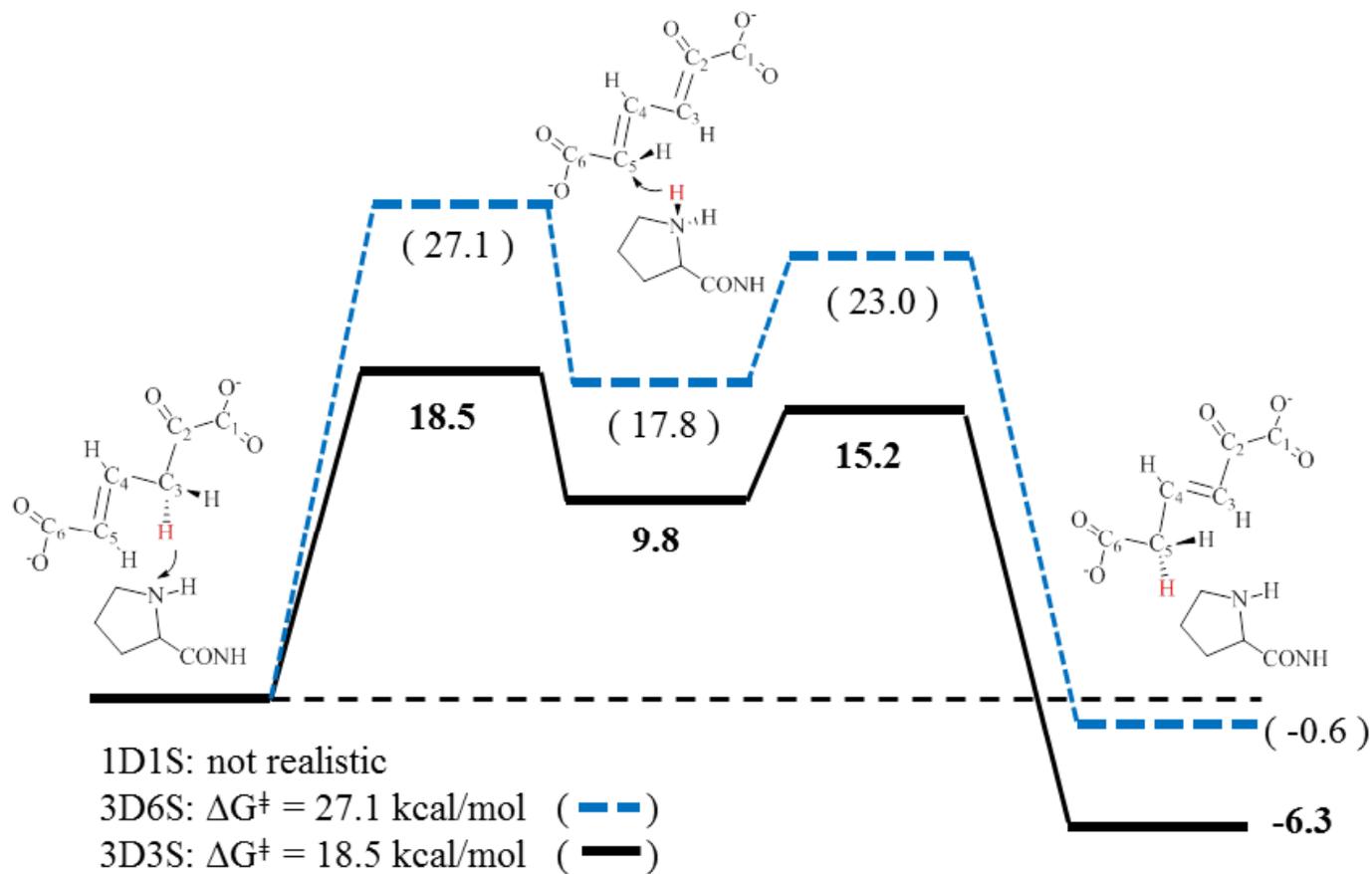
➢ **During the reaction, the active site is observed to shrink to promote the reaction to happen.**

Mechanism of OMP Decarboxylation in Orotidine 5′-Monophosphate Decarboxylase (ODCase)
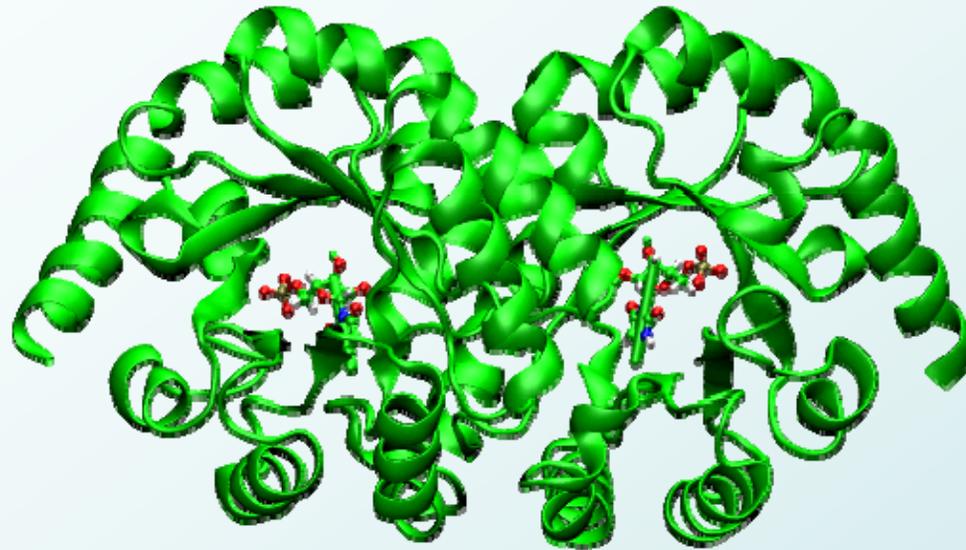
ODCase structure
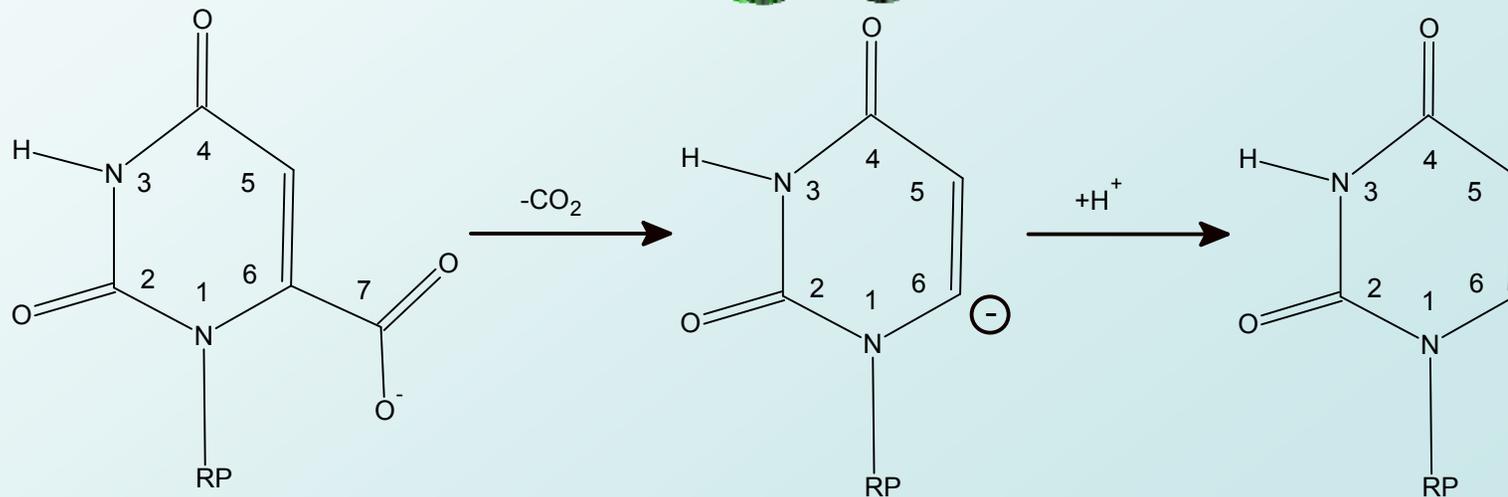
One of the most proficient enzymes known
-- Radzicka, A.; Wolfenden, R. *Science* **1995**

ODCase

- plays a central role in de novo synthesis of uridine-5'-O-monophosphate (UMP), building blocks for the synthesis of RNA and DNA

- present in bacteria, archea, parasites and in humans, i.e. almost every species except in viruses.

- catalyzes the decarboxylation of orotidine monophosphate (OMP) to uridine monophosphate (UMP)

- an extraordinary level of catalytic rate enhancement of over 17 orders of magnitude

ODCase mechanism

Mechanism

- Experiments favor direct-decarboxylation mechanism, but failed to provide definite evidence
- Simulations have been inconclusive too
    - Pai & Gao (AM1/CHARMM27 simulation)
    - Warshel (EVB)
    - Lee & Kollman (QM/MM-FE)
    - Lundberg & Siegbahn (QM)
    - Raugei (CPMD + Jazynski simulation)
    - Houk (CP2K + meta-dynamics)

    - All DFT-based simulations yielded barrier (21 ~ 27 kcal/mol) significantly higher than experimental data (15.3 kcal/mol)

Is DFT inappropriate for this reaction?

ODCase: *gas-phase reaction*
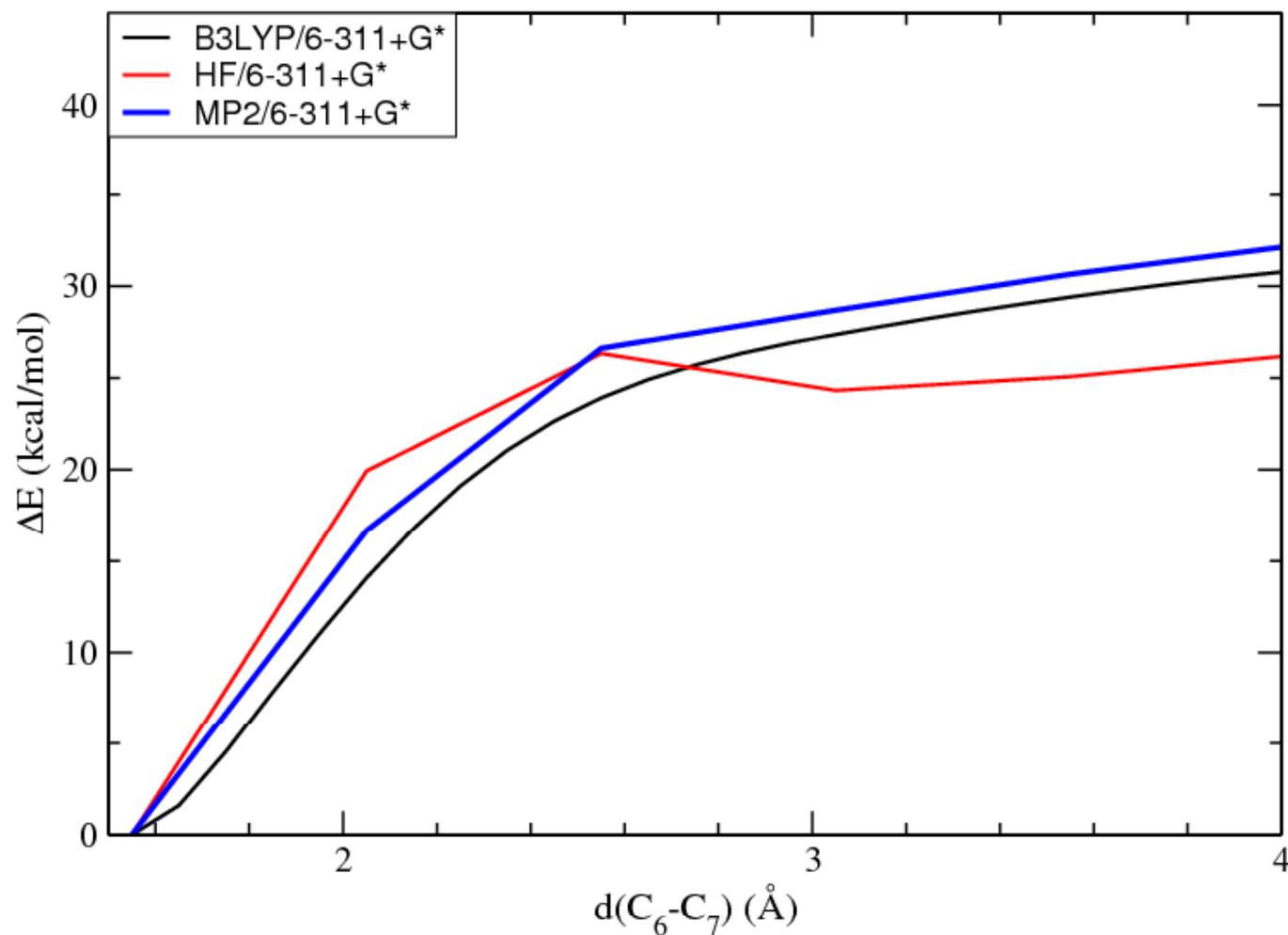
B3LYP/6-311+G* is in good agreement with MP2 method

ODCase: *solution reaction*
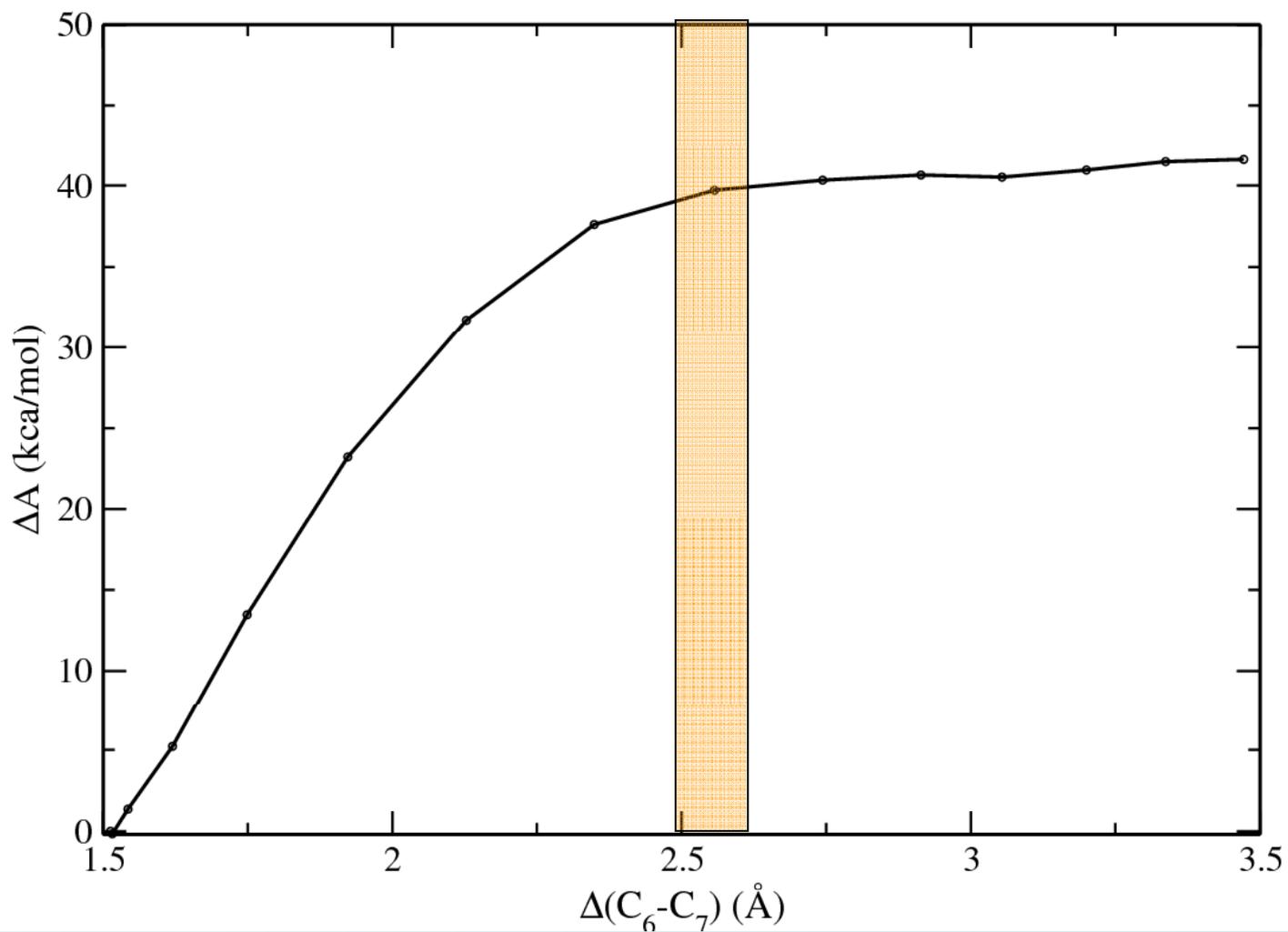
$\Delta G^{\ddagger} \cong 40.2$ kcal/mol   (Expt. $\Delta G^{\ddagger} \cong 38.5$ kcal/mol)
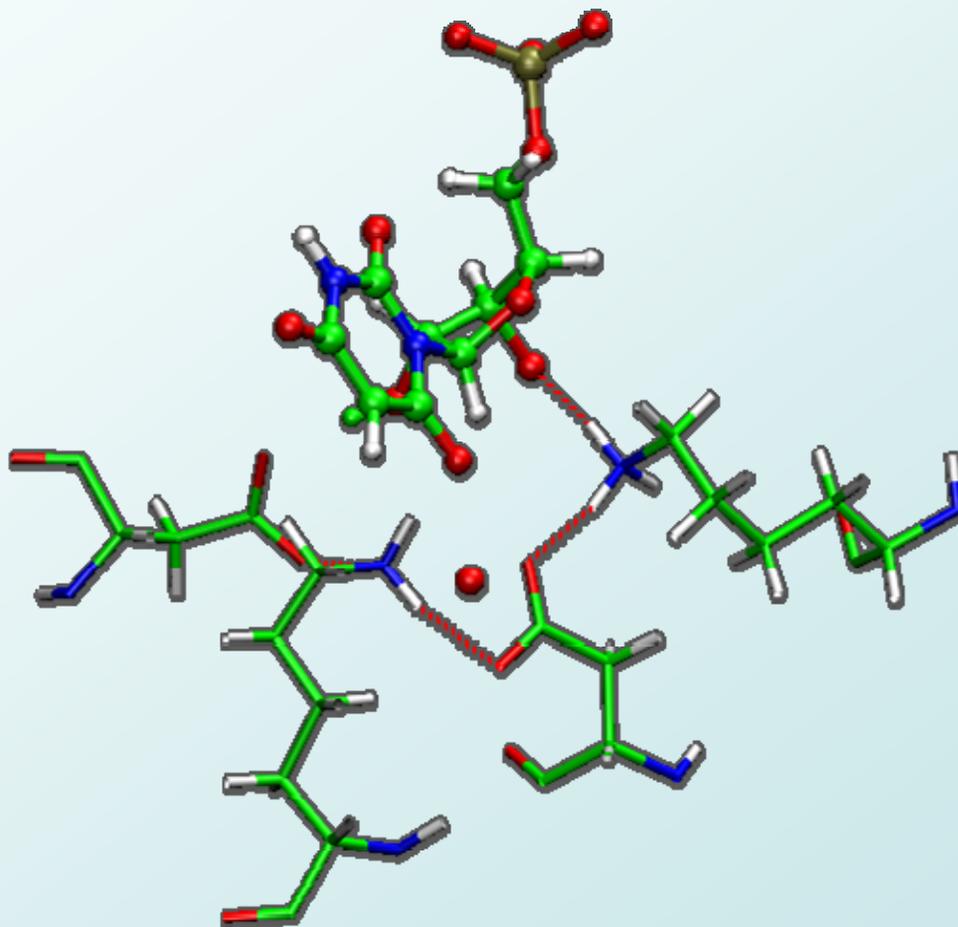
ODCase mechanism

Back to square one
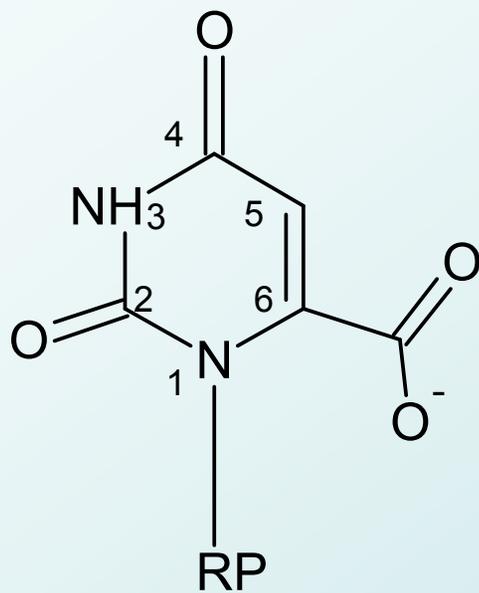
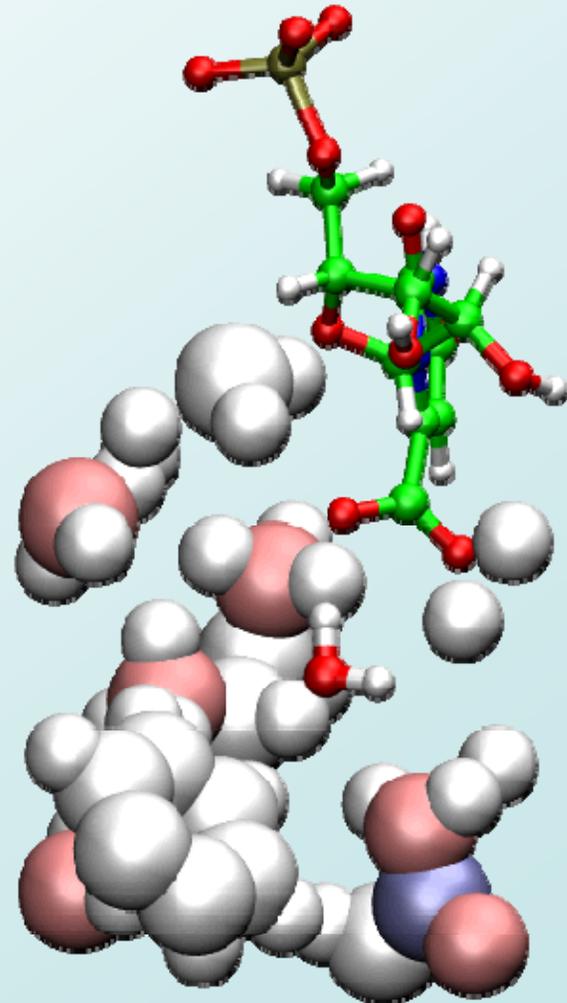OMP   BMP
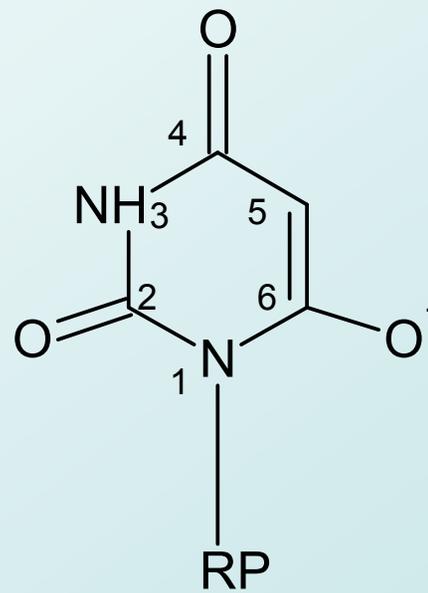
BMP binds to a water molecule

ODCase mechanism
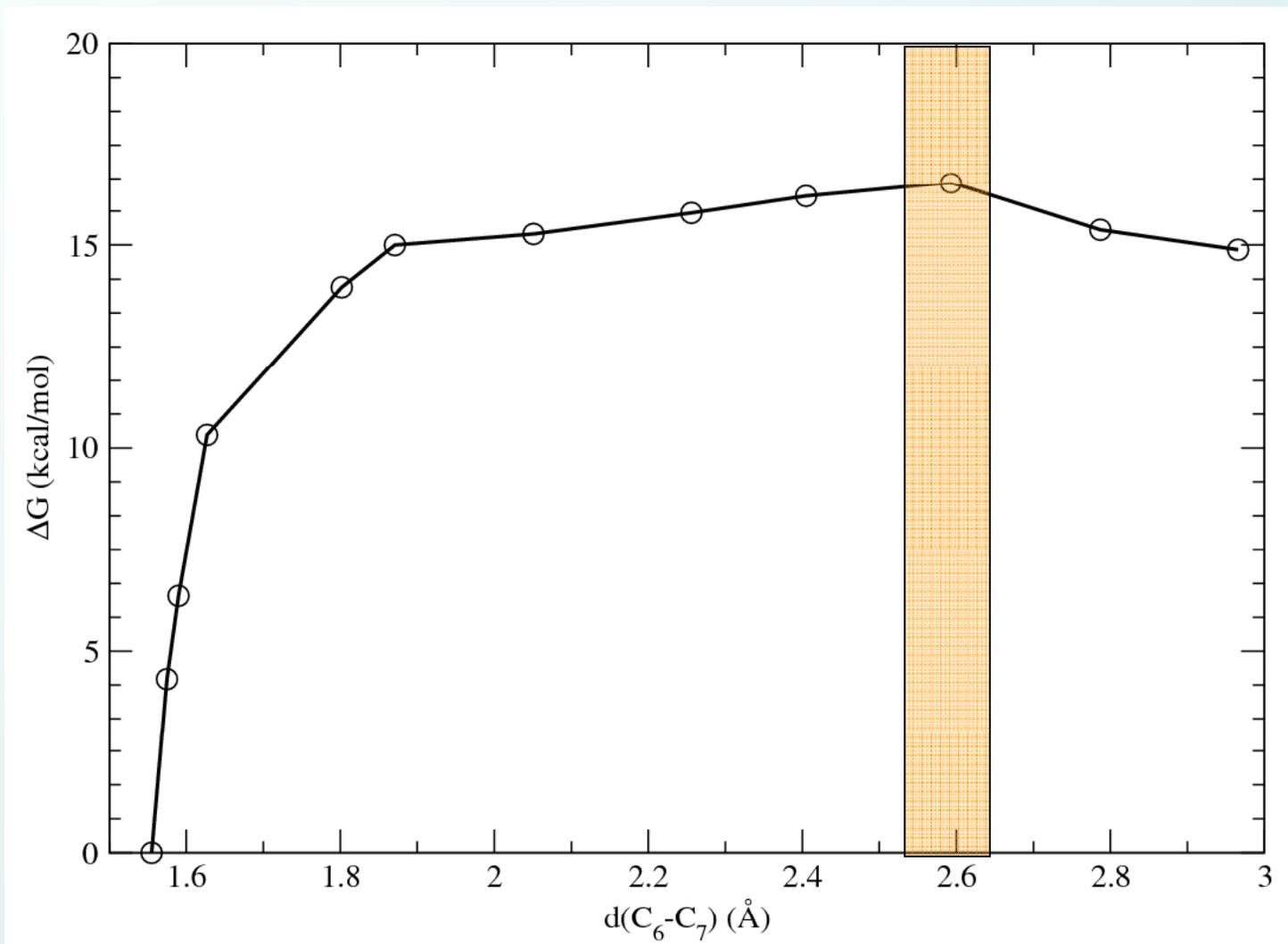
Replace BMP with OMP

The water molecule is surrounded by hydrophobic residues like Val, Leu, ILE, and Phe.

The water is blocking the CO_2 leaving direction.

Our hypothesis:
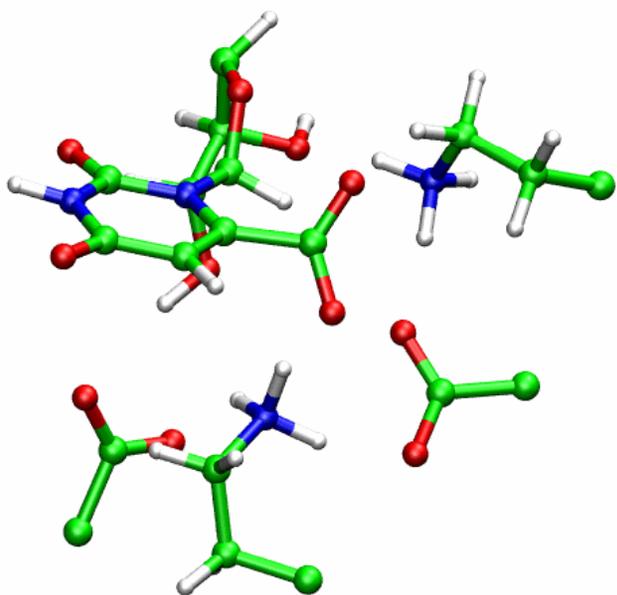There is no water molecule in the active ODCase/OMP complex.

ODCase: enzyme reaction

$\Delta G^{\ddagger} \cong 16.5$ kcal/mol   (Expt. $\Delta G^{\ddagger} \cong 15.3$ kcal/mol)

ODCase: *direct mechanism*

Results

1. Reproduced experimental barrier

2. Supported direct decarboxylation mechanism

3. Highlighted the importance of proper recognition of protein-bound water molecules
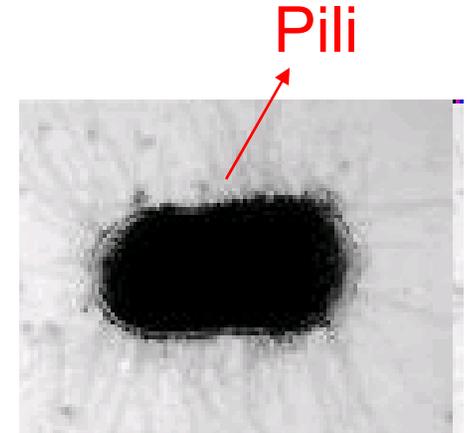
Hao Hu, Amy Boone, and WY, JACS, 2008

Bio-significance of Pili

Xiangqiang Hu

Pili

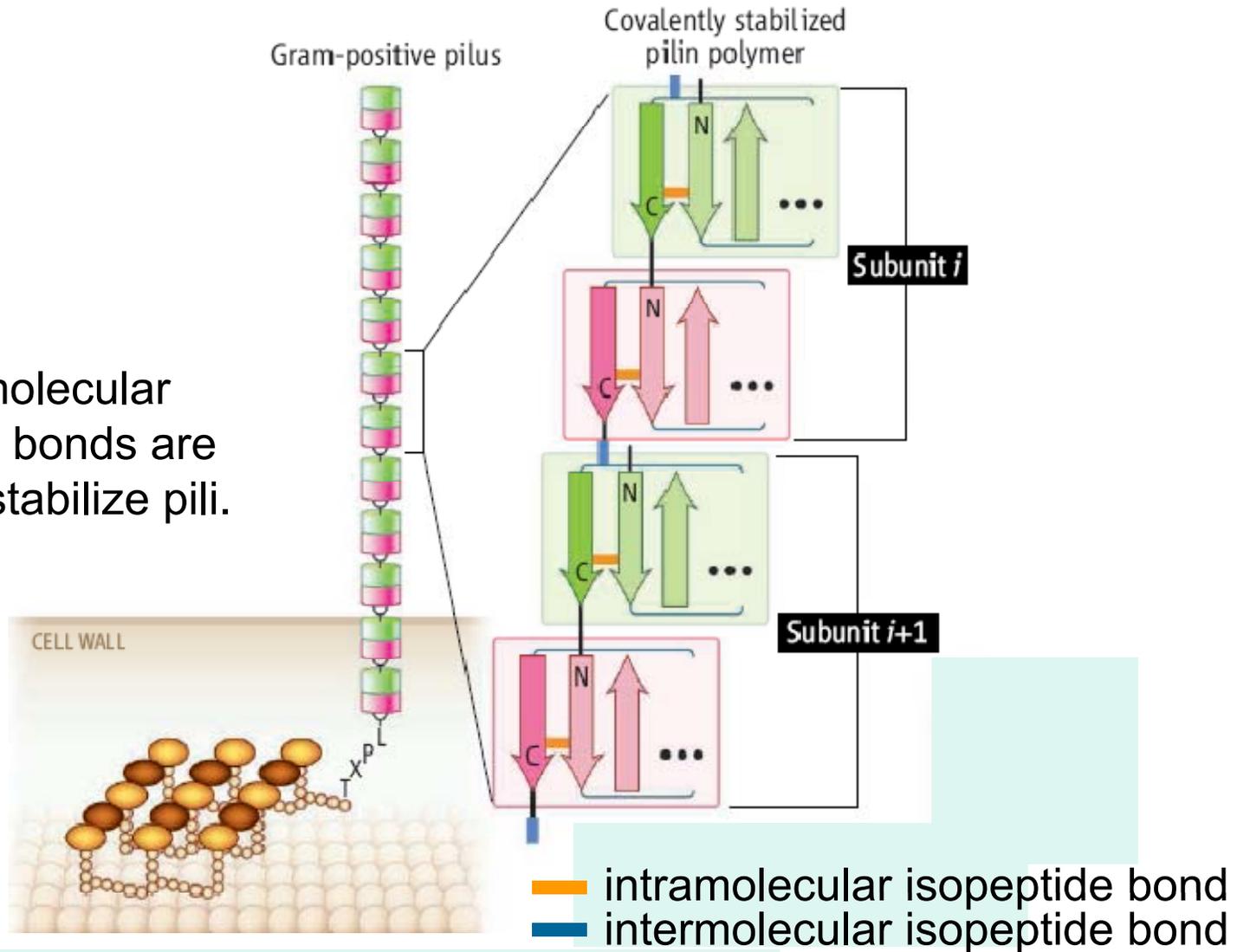Pili are extracellular parts of Gram-positive bacteria:

- ➤ Adhesion to host cell
- ➤ Transfer of genetic material
- ➤ Induction of signaling in host cells
- ➤ Twitching motility
- ➤ Other critical aspects of colonization

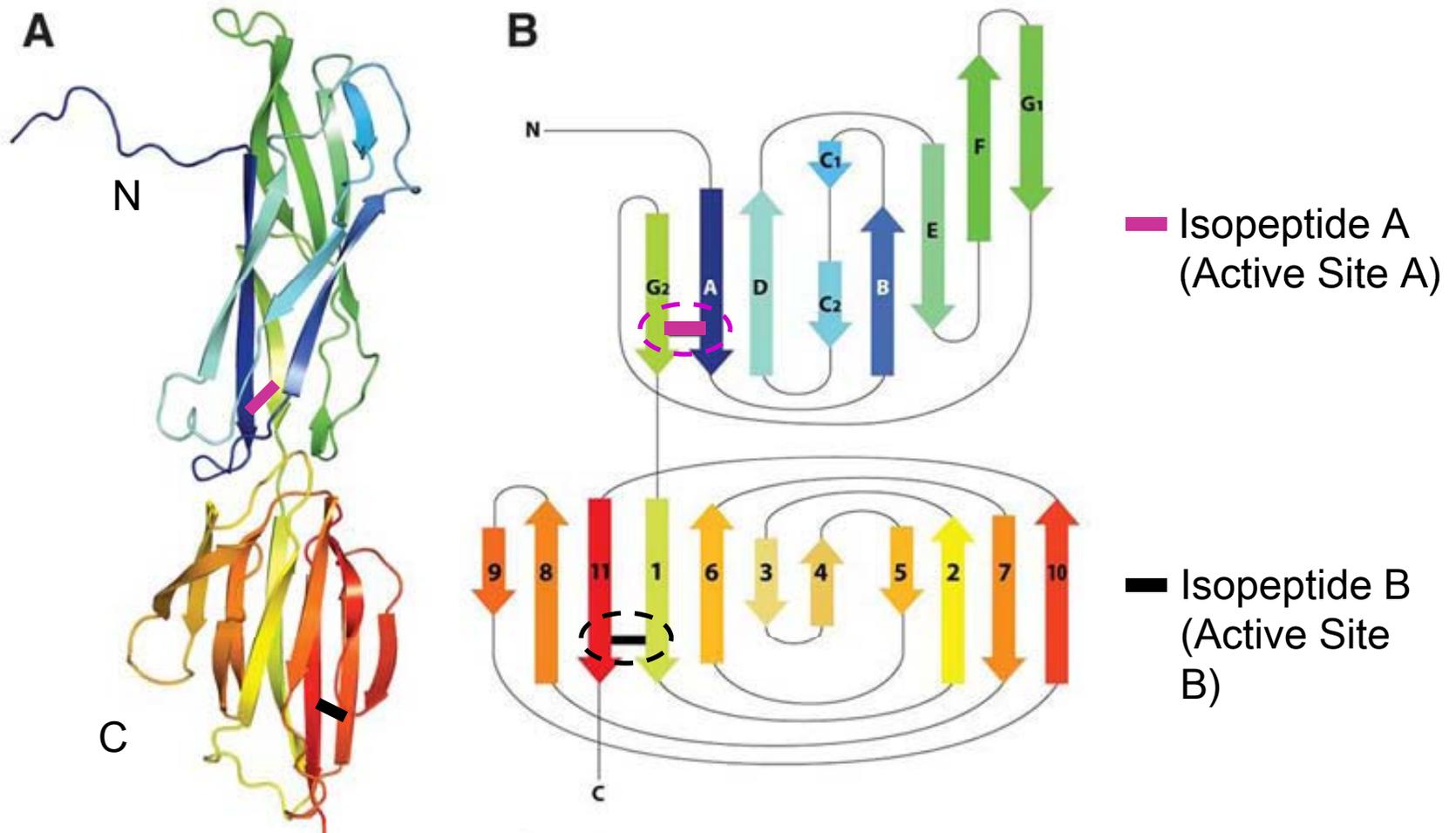Pili are important vaccine candidates for Gram-positive bacteria.

JACS, 2010

Pili in Gram-positive Bacteria

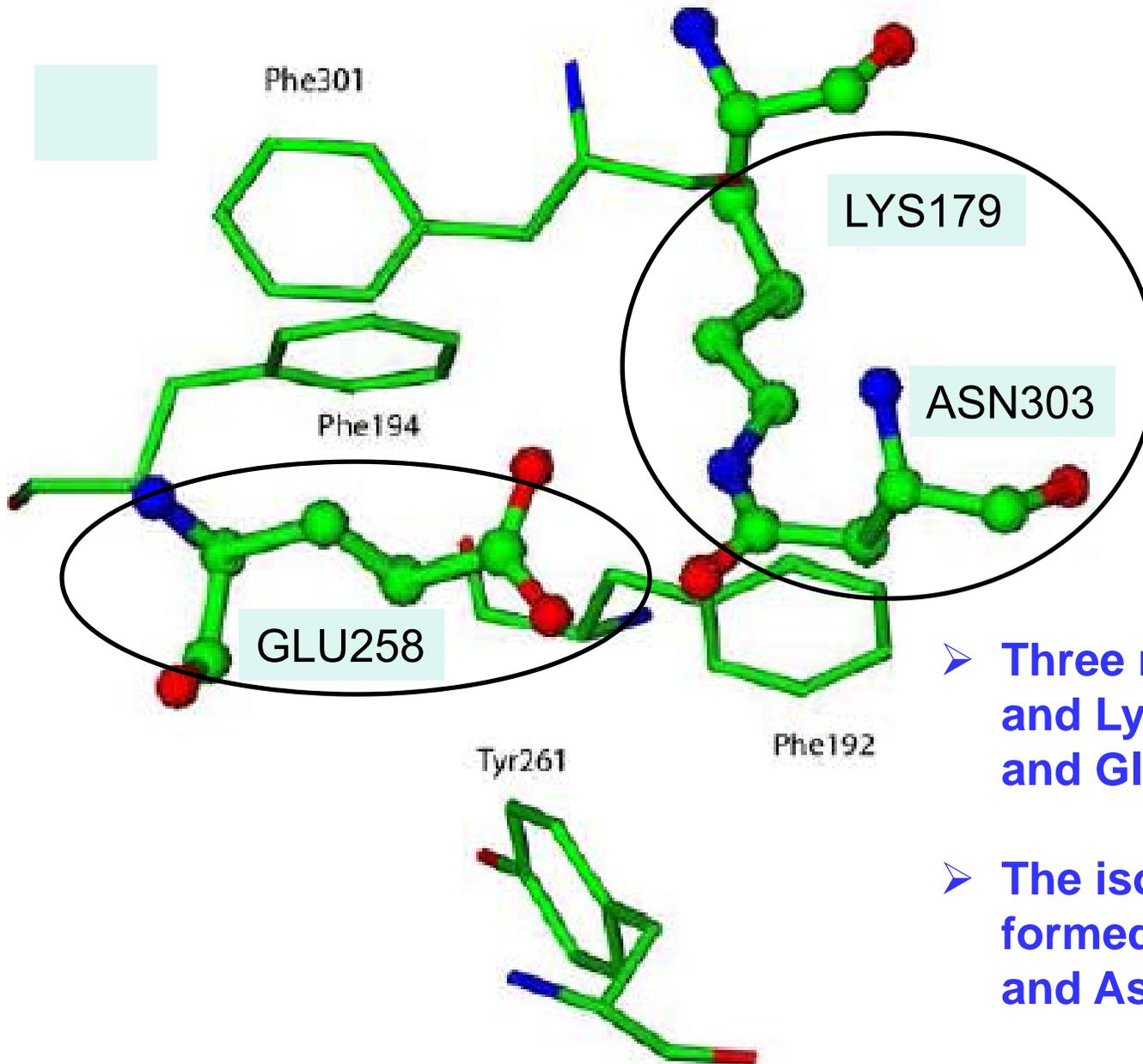The intramolecular isopeptide bonds are critical to stabilize pili.

Gram-positive pilus

Covalently stabilized pilin polymer

Subunit *i*

Subunit *i*+1

CELL WALL

— intramolecular isopeptide bond
— intermolecular isopeptide bond

H. Kang, F. Coulibaly, F. Clow, T. Proft, T., and E. Baker, Science, 2007, *318*, 1558

Crystal structure of the unit pilus

Isopeptide A (Active Site A)

Isopeptide B (Active Site B)

Considering isopeptide bond in Active Site B of the pilus

Phe301

Phe194

LYS179

ASN303

GLU258

Tyr261

Phe192

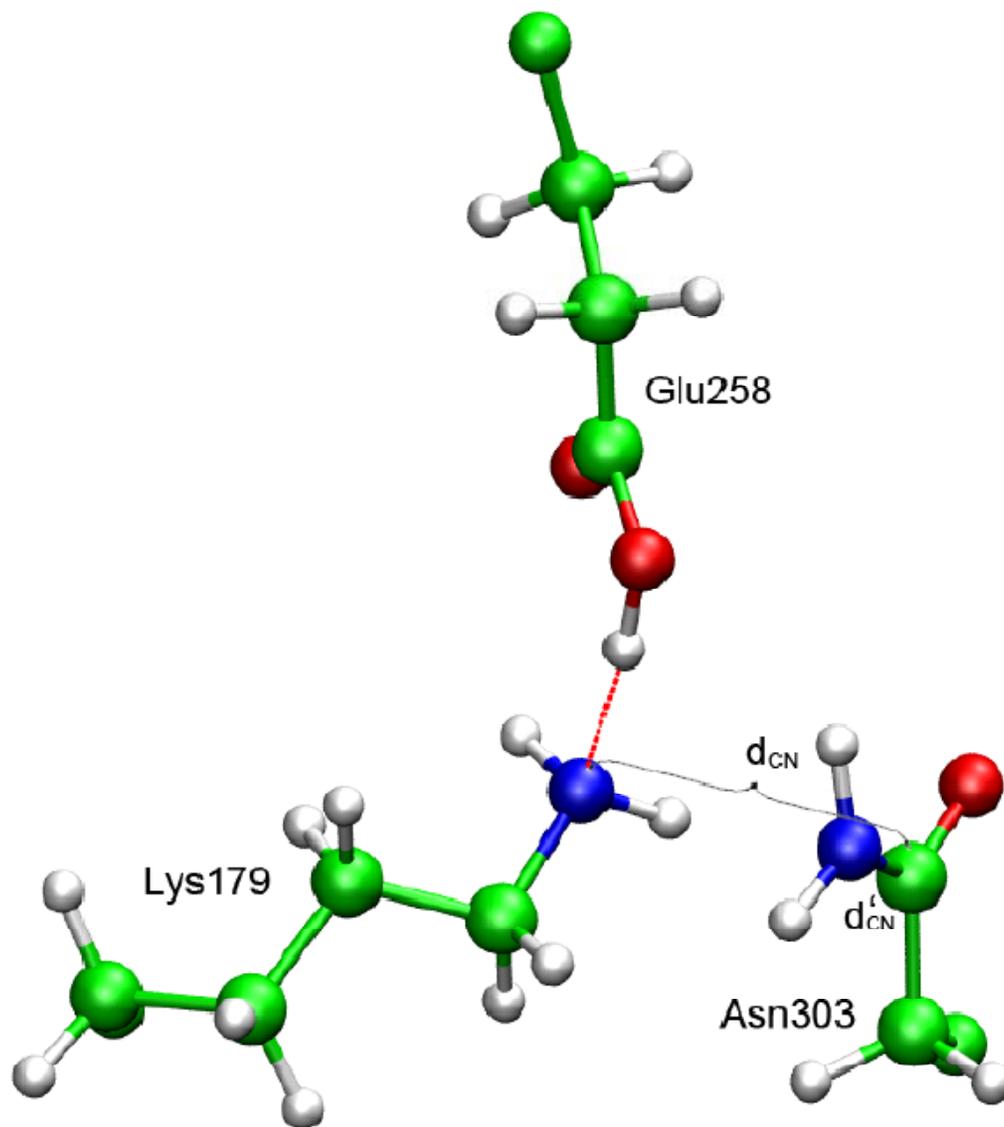> **Three residues Glu, Asn, and Lys are conserved and Glu is essential.**

> **The isopeptide bond is formed between Lys179 and Asn303.**

We proposed this reaction mechanism of intra-isopeptide bond formations:

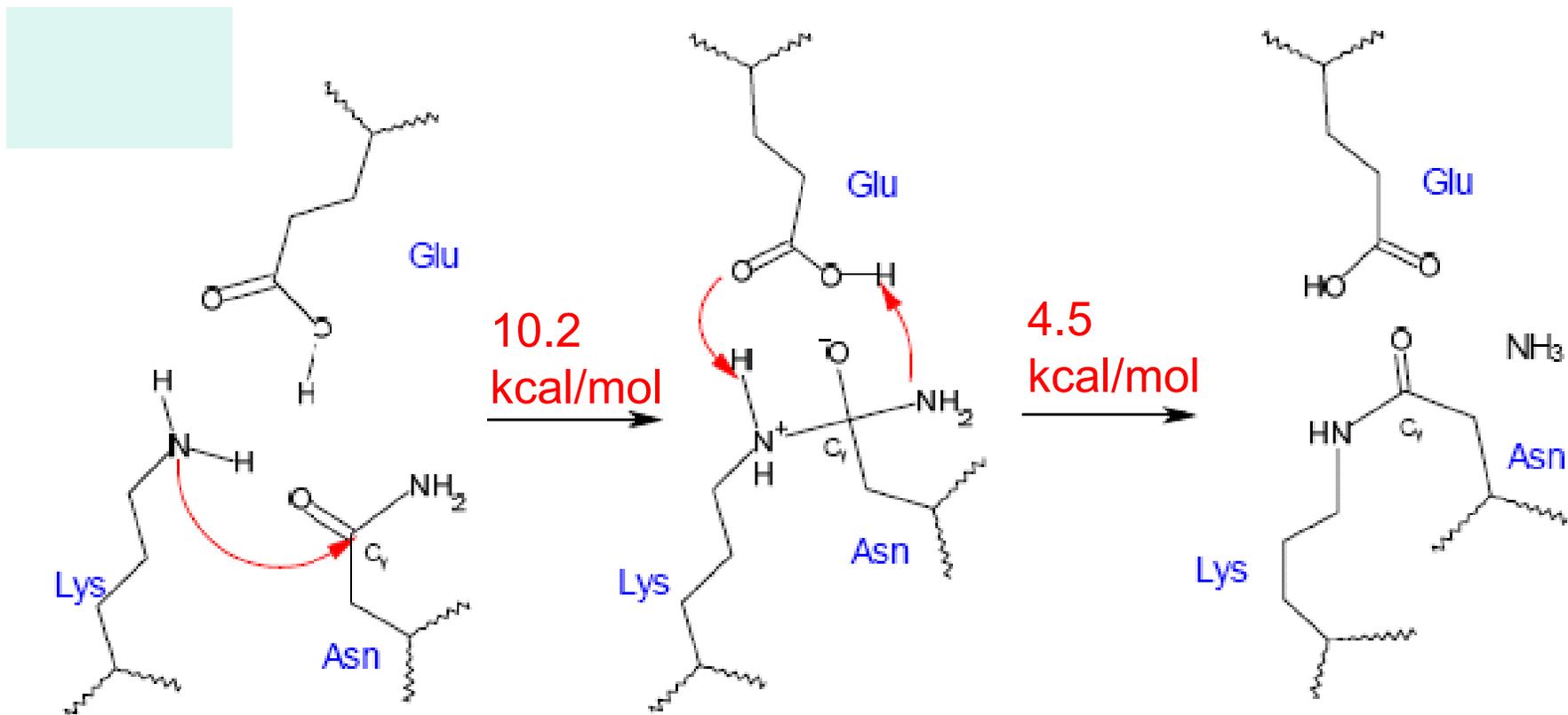Nucleophilic attack   Two proton transfers   NH₃ formation
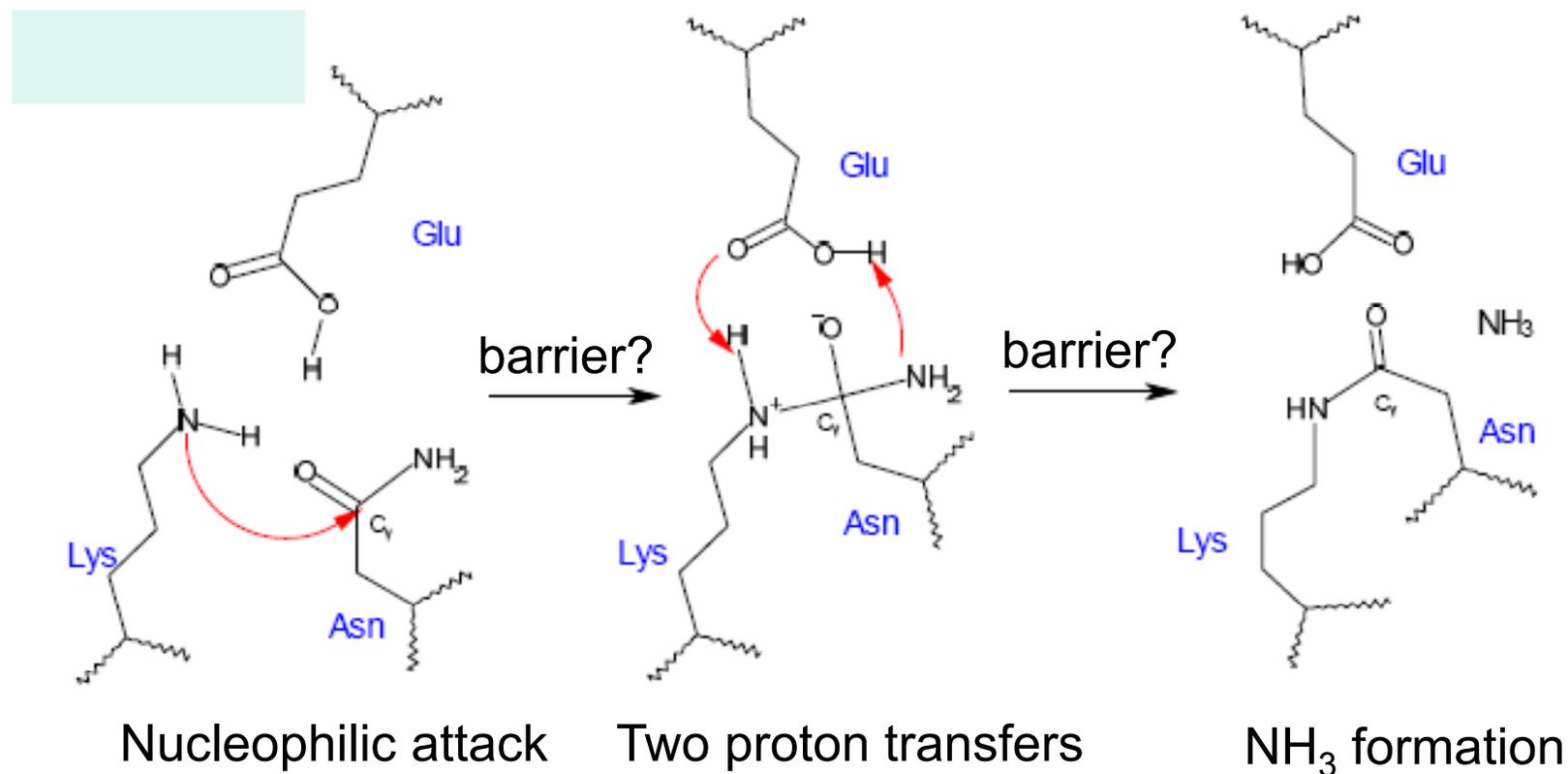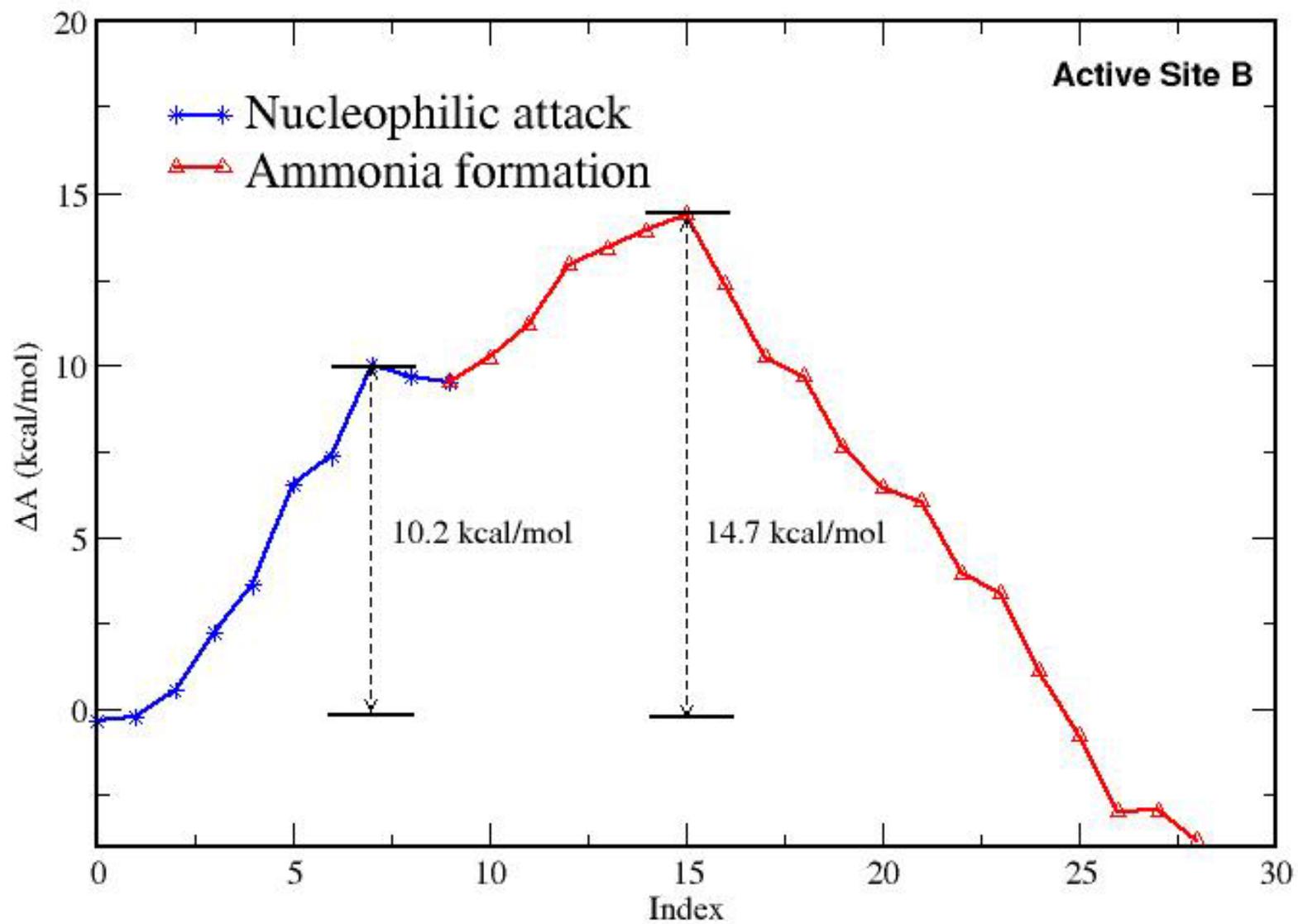
We applied QM/MM-MFEP to validate this mechanism.

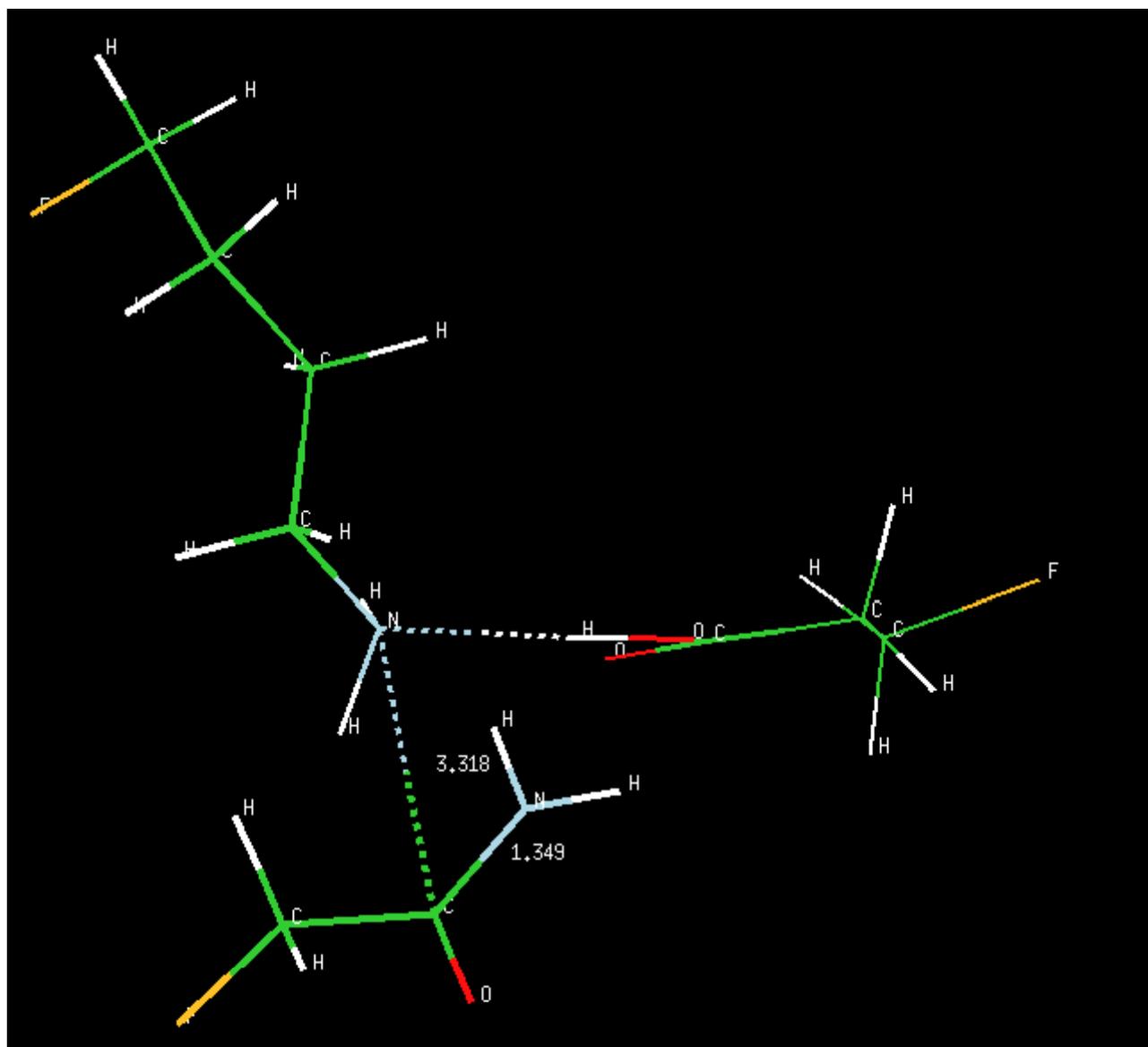1. Reactant geometry optimized by the QM/MM-MFEP approach

The activation barrier (~15 kcal/mol) is sufficiently low to allow reaction at room temperature.

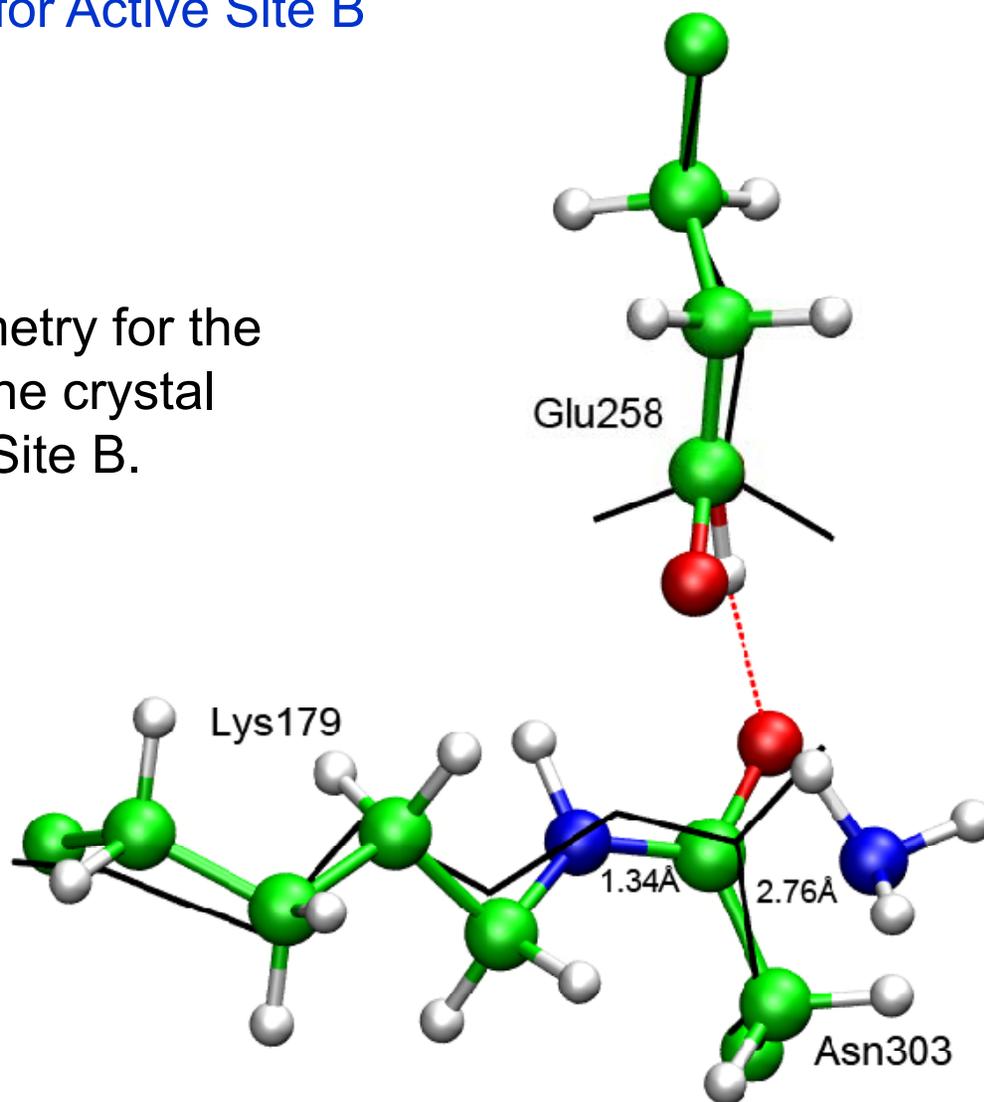## 2. The optimized reaction path by the QM/MM-MFEP approach

3. Product geometry for Active Site B

Our optimized geometry for the product is close to the crystal structure for Active Site B.

Glu258

Lys179

1.34Å   2.76Å

Asn303

(The black line is the crystal structure of Active Site B.)

Summary on Pilli

➢ A definite reaction mechanism for this intramolecular isopeptide bond formation is revealed.

➢ Glu258 is the indispensable proton relay medium during isopeptide bond formation. This agrees with experimental mutation studies.

➢ This demonstrates the power of the QM/MM-MFEP method for determining enzyme reaction mechanisms.

Hu, Hu, Melvin, Clancy, McCafferty, Yang, JACS 2010

Redox Potentials

•The redox potential is a measure (in volts) of the affinity of a substance for electrons — its electronegativity — compared with hydrogen (which is set at 0).

•It is the free energy change of adding/removing an electron.

•In the absence of solvent, it is the electron affinity (A) or ionization potential (I)

•Substances more strongly electronegative than (i.e., capable of oxidizing) hydrogen have positive redox potentials. Substances less electronegative than (i.e., capable of reducing) hydrogen have negative redox potentials.

•Oxidations and reductions always go together. They are called redox reactions.

Redox Free Energy Simulation – Reaction order parameter, JCP 2009

Fox Zeng

- ## Work with Half redox reactions

$$A_{N_e} \to A^+_{N_e-1} + e^-$$

- ## Fractional Number of Electrons (FNE) – redox reaction order parameter
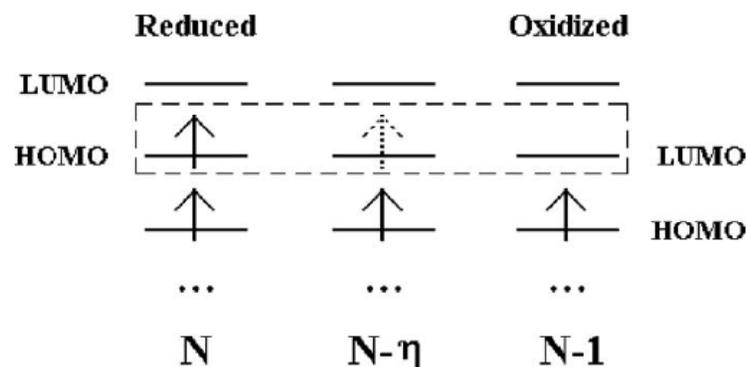

FIG. 2. Scheme of oxidation process. The arrows denote electrons and the horizontal bars are orbitals. In oxidation process, we remove the FNE on the HOMO of the reduced state with N electrons and gradually reaches the oxidized state with $(N-1)$ electrons.

Redox Free Energy Simulation – Sampling

- ## Direct QM/MM-MD + TI

$$\Delta A = \int_0^1 \left\langle \frac{\partial E(\eta)}{\partial \eta} \right\rangle_\eta d\eta = - \int_0^1 \langle \varepsilon_{\mathrm{HOMO}} \rangle_\eta d\eta.$$

- ## QM/MM-Minimum Free Energy Path (MFEP)

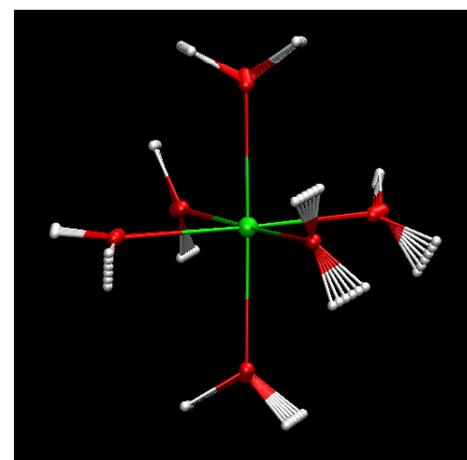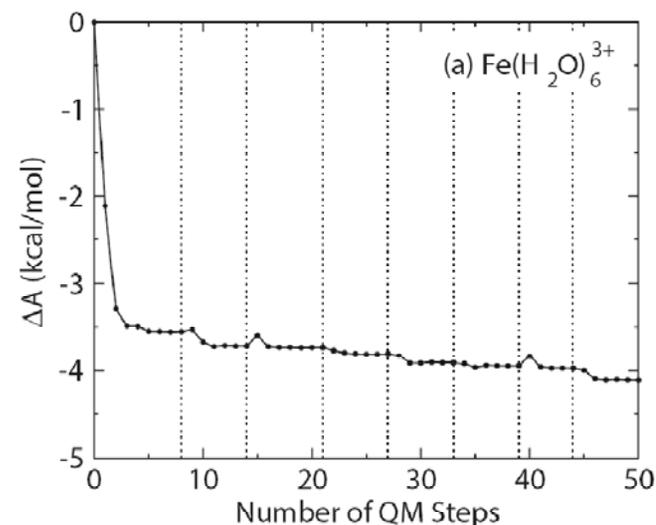 - Represent thermodynamics by the PMF of QM subsystem

$$A(\mathbf{r}_{\mathrm{QM}}) = -\frac{1}{\beta} \ln \left[\int e^{-\beta E(\mathbf{r}_{\mathrm{QM}}, \mathbf{r}_{\mathrm{MM}})} d\mathbf{r}_{\mathrm{MM}} \right],$$

$$\frac{\partial A(\mathbf{r}_{\mathrm{QM}})}{\partial \mathbf{r}_{\mathrm{QM}}} = \left\langle \frac{\partial E(\mathbf{r}_{\mathrm{QM}}, \mathbf{r}_{\mathrm{MM}})}{\partial \mathbf{r}_{\mathrm{QM}}} \right\rangle_{E, \{\mathbf{r}_{\mathrm{MM}}\}},$$

# Redox Free Energy Simulation – Procedure

- **QM/MM-MFEP**
  - Optimize the QM geometry for two end points of the redox process
  - Linearly interpolate the geometries and build path with FNE
 - Perform FEP to calculate the redox free energies

(a) $Fe(H_2O)_6^{3+}$

Redox Free Energy Simulation – Systems

- ## Simple metallic ions
  - $Fe(H_2O)_6^{2/3+}$ and $Ru(H_2O)_6^{2/3+}$
- ## Small bio-organic molecules
  - Lumichrome and riboflavin
- ## Proteins

(a)

Lumichrome

(b)

Riboflavin
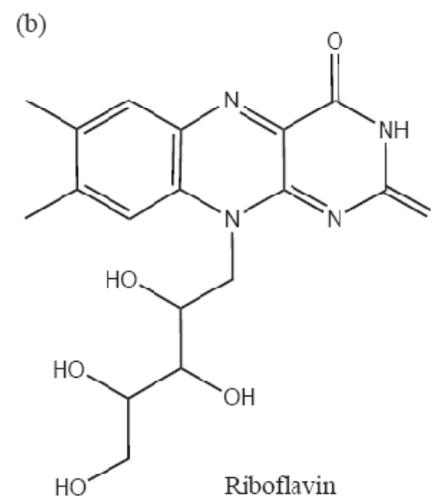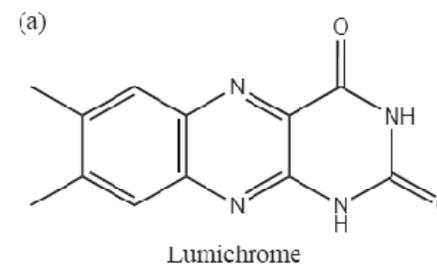
Case 1: Hexaquo Metal Complexes

- $Ru(H_2O)_6^{2+} \rightarrow Ru(H_2O)_6^{3+} + e^-$

  $Fe(H_2O)_6^{2+} \rightarrow Fe(H_2O)_6^{3+} + e^-$

- QM/MM simulation setup
  - G03+Sigma, BLYP/LanL2DZ
  - $64 \times 64 \times 64$ Å^3, ~8600 TIP3P
- Redox potentials (in V)*
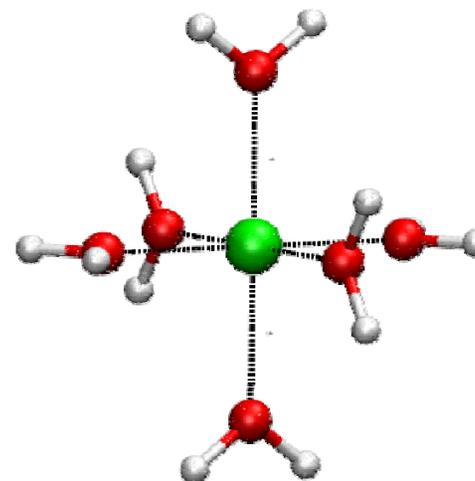
	ϕ_{MFEP}	ϕ_{Direct}	ϕ_{Expt}	$\Delta\phi_{MFEP}$	$\Delta\phi_{Expt}$
Fe$^{2/3+}$	5.8	5.8	0.7+SHE	0.7	0.5
Ru$^{2/3+}$	5.1	5.1	0.2+SHE	-	-

* **Zeng**, Hu, Hu, and Yang, *J. Chem. Phys.* (2009)

System and Computational Details

- ## System – Azurin 4AZU

  - 128 Residues, 43.5×33.1×41.4 Å^3

  - Water box, 90×90×90 Å^3

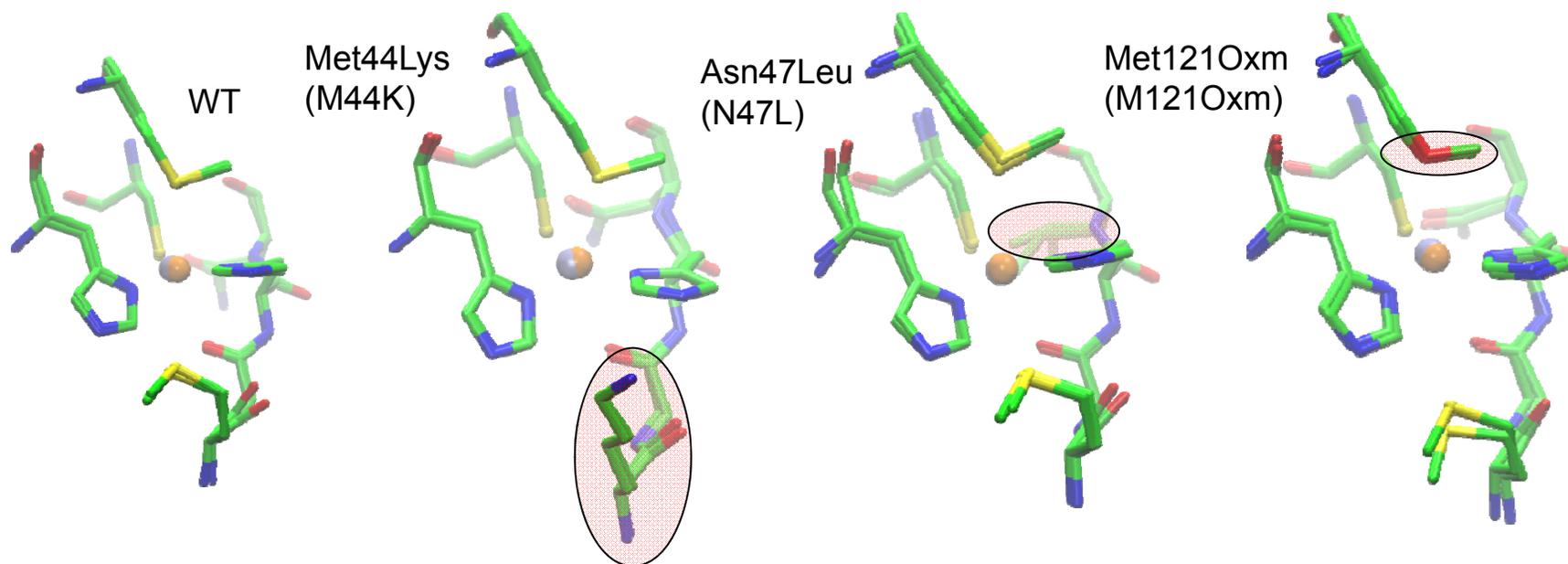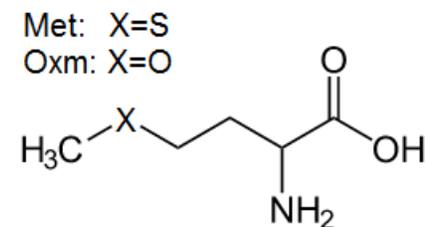 - Mutants: Met44Lys, Asn47Leu, Met121Oxm (artificial)

  - Details

    - G03+Sigma, UBLYP/ccpv-dz

    - QM region: 5 coordination residue + 2 mutation sites (102~107 atoms)

    - MD time: 160 ps / ensemble

    - Cutoff: 20 Å for QM/MM, PME for MD sampling

    - Pseudo bond (Parks *et al.**)

40 Å

Produced with VideoMach
www.videomach.com

Gly 45

Met 44

His 46

His 117

Cu$^{+/2+}$

Asn 47

Cys 112

Met 121
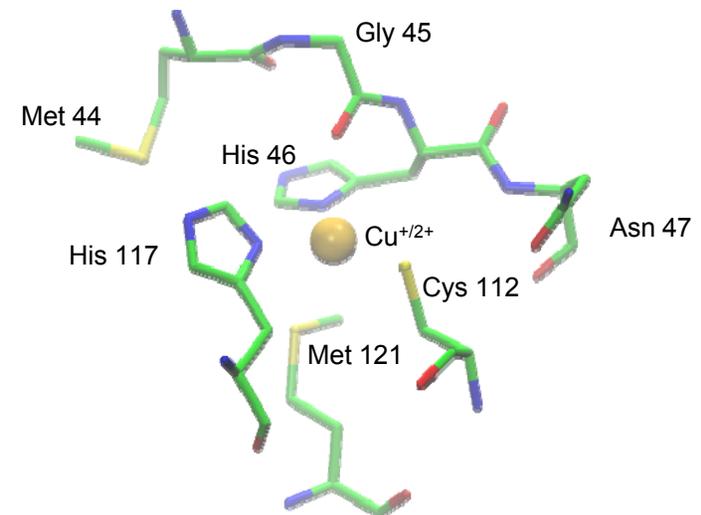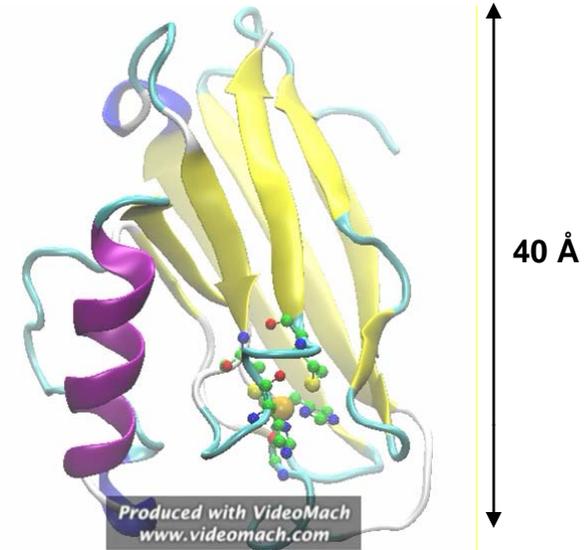
Optimized QM Geometries of Red/Ox States

- RMSDs between red/ox < 0.1 Å

Met: X=S
Oxm: X=O

WT

Met44Lys
(M44K)

Asn47Leu
(N47L)

Met121Oxm
(M121Oxm)

Redox Potentials of Azurin and Mutants

- Relative redox potential (WT as reference)*
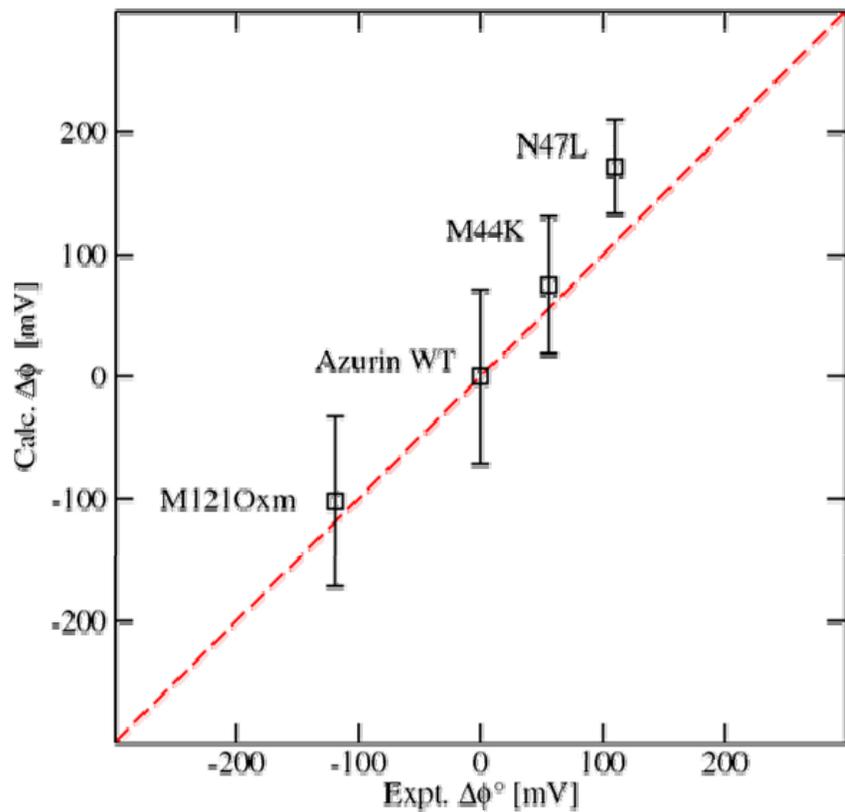
	$\Delta\phi_{Calc}$	$\Delta\phi_{Expt}$	Error[†]
WT	–	–	–
Met44Lys	80	56	20
Asn47Leu	170	110	60
M121Oxm	-100	-119	-20

Units in mV

* **Zeng**, Hu, Hu and Yang, Unpublished

DFT QM/MM

- Multiscale

- DFT + MM

- Quantum Mechanics + Statistical Mechanics

- Complex problems in biocatalysis, material processes, energy research

- Many possibilities