

AICS Cafe (2012/01/13)

Recent developments of QM/MM

Jaewoon Jung
(Computational Biophysics Research Team)



Outline

- 1. MM – Forcefield description**
- 2. Importance of QM/MM**
- 3. How to deal with QM/MM boundary**
- 4. Recent developments related with QM/MM**
- 5. Application results**

Molecular Mechanics (Force field)

1. The basic functional form of a force field encapsulates both bonded terms relating to atoms that are linked by covalent bonds, and nonbonded (also called "noncovalent") terms describing the long-range electrostatic and van der Waals forces

$$\begin{aligned}E_{total} &= E_{bonded} + E_{nonbonded} \\E_{bonded} &= E_{bond} + E_{angle} + E_{dihedral} \\E_{nonbonded} &= E_{vanderWalls} + E_{electrostatic}\end{aligned}$$

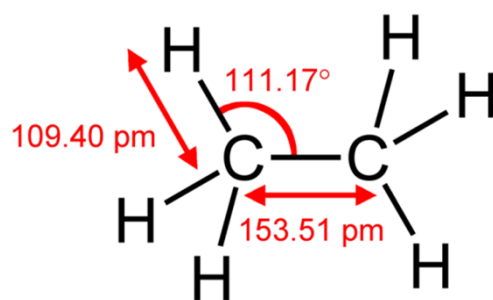
2. A force field defines a set of parameters for each type of atom. For example, Amber has a functional form

$$\begin{aligned}E_{total} &= \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\&+ \sum_{j=1}^{N-1} \sum_{i=j+1}^N \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon r_{ij}} \right\}\end{aligned}$$

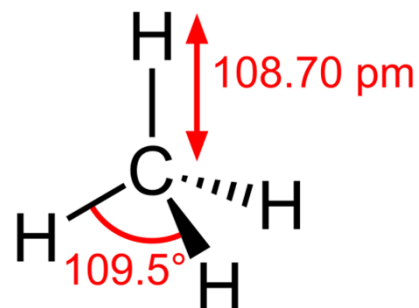
In Amber force field, parameters k_b , b_0 , k_a , θ_0 , V_n, γ , ϵ_{ij} , r_{0ij} , and q_i are defined according to each atom type.

Problem of a Force field - transferability

Even the same atom, parameters can be changed according to the atom type.



ethane

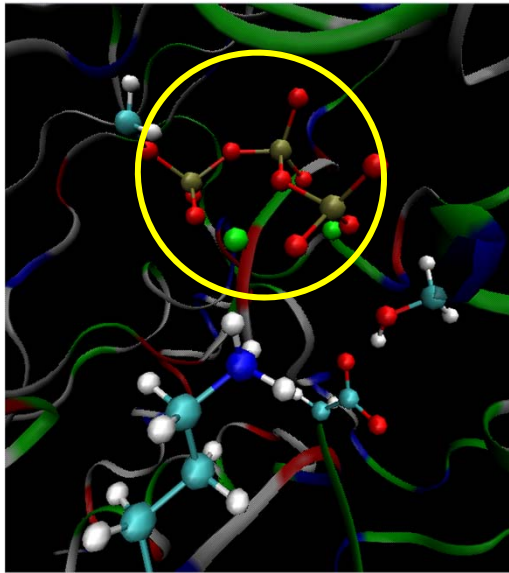


methane

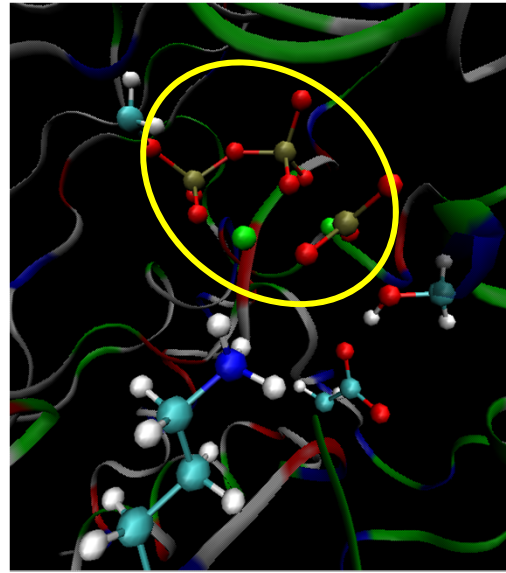
Thus, parameters are not transferable to new ligands or unusual functional groups.

Problem in a Force field – description of chemical reaction

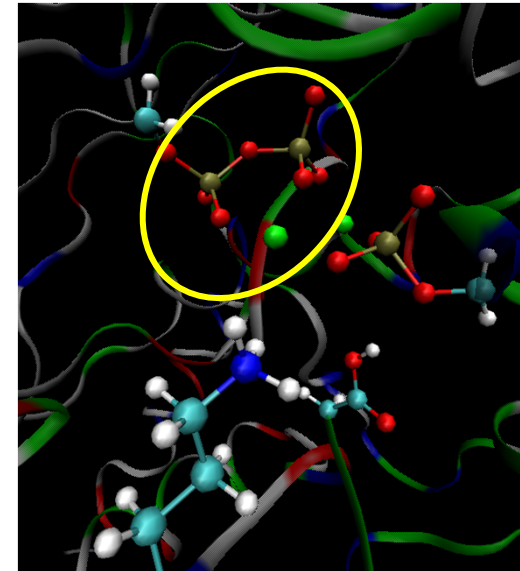
1. Chemical reactions (bond breaking or bond forming) requires rearrangements of electrons.
2. Force field couldn't be used for such chemical reactions because electrons are excluded from force fields.



Force field of ATP



??



Force field of ADP

Example : force field of ATP during chemical reaction



QM (Molecular orbital theory)

1. Solving Schrödinger equation with the electron molecular Hamiltonian

$$E\psi = \hat{H}\psi$$
$$\hat{H} = -\sum_i \frac{1}{2} \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{i,A}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} - \sum_A \frac{1}{2M_A} \nabla_A^2$$

2. The variational solution is assumed to be a linear combination of a finite number of basis functions.
3. Each energy eigenfunction is assumed to be a single Slater determinant, an antisymmetrized product of one-electron wave function.

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \cdots & \chi_1(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix}$$

4. This approach is the conceptual basis of the Hartree-Fock method and further post Hartree-Fock methods.

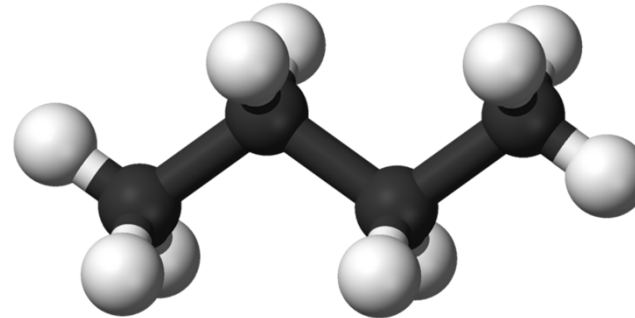
Problem of a Molecular orbital theory – Computational cost

1. Computation time for energy/gradient calculation of butane (Force field : CHARMM22, QM basis : cc-pVDZ)

Force field : <0.001 sec

QM Hartree Fock : 7 sec

QM HF+MP2 : 8.5 sec

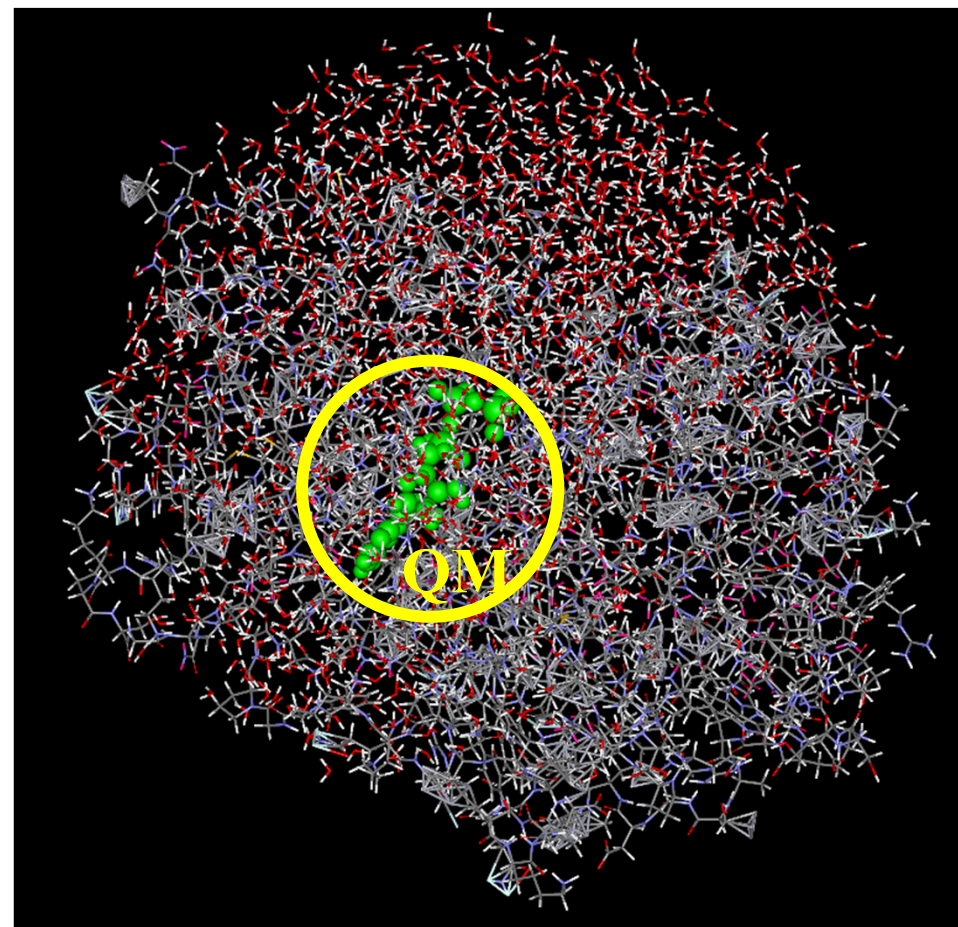


2. MM calculation with cutoff is $O(N \log N)$ with PME, but HF is $O(N^4)$, MP2 is $O(N^5)$, and CCSD(T) is $O(N^7)$. Thus, QM calculation is available only for very small systems.

➡ QM/MM could be a solution!!

What is QM/MM ?

1. **Bond-forming and – breaking processes are treated by QM**
2. **The remaining part is described by MM**
3. **QM and MM regions are separated by link atoms or frozen orbitals**

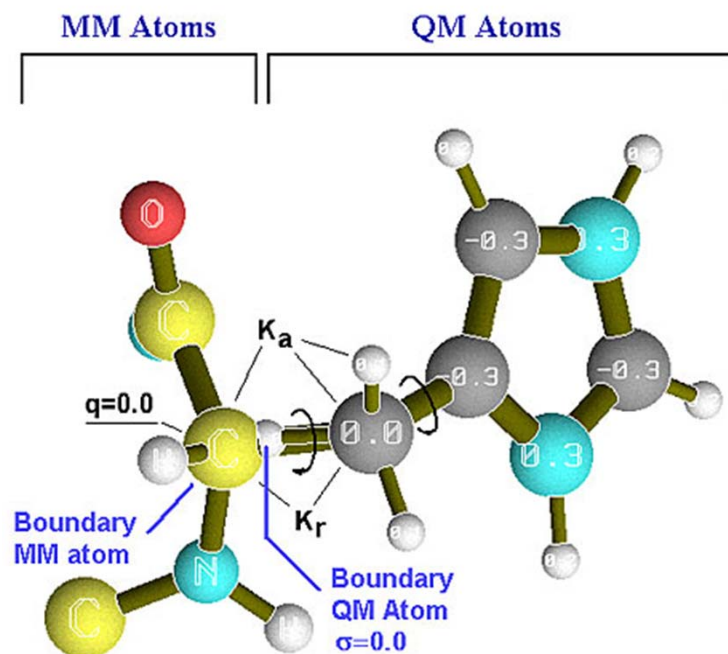


Previous QM/MM Methods

QM/MM	Year	Main author	Method
Link Atom	1990	M.J. Field <i>et al.</i>	A link atom (hydrogen) is used as a boundary
ONIOM	1996	M. Svensson <i>et al.</i>	Different layer according to the different level of calculation is introduced
	2005	T.Vreven <i>et al.</i>	Extension to Electronic Embedding
Connection atom	1999	I. Antes and W. Thiel	Parametrized atom instead of the link atom is used for QM/MM boundary
Pseudobond	1999	Y. Zhang <i>et al.</i>	One-free-valence atom with an effective core potential instead of the link atom
Quantum Capping Potential	2002	G.A. DiLabio <i>et al.</i>	quantum capping potentials are formed by replacing excess valence electrons
Local self-consistent filed (LSCF)	1994	V. Thery <i>et al.</i>	The boundary is represented as localized orbitals (semiempirical)
	1999	X. Assfeld <i>et al.</i>	Extended to the ab initio level
Generalized Hybrid Orbital (GHO)	1998	J. Gao <i>et al.</i>	A set of hybrid orbitals is set on the boundary
	2005	J. Pu <i>et al.</i>	Extended to the ab initio level
	2007	J.Jung <i>et al.</i>	New implementation

GHO scheme provides an electronically stable connection between QM and MM

Link atom approach for QM/MM boundary

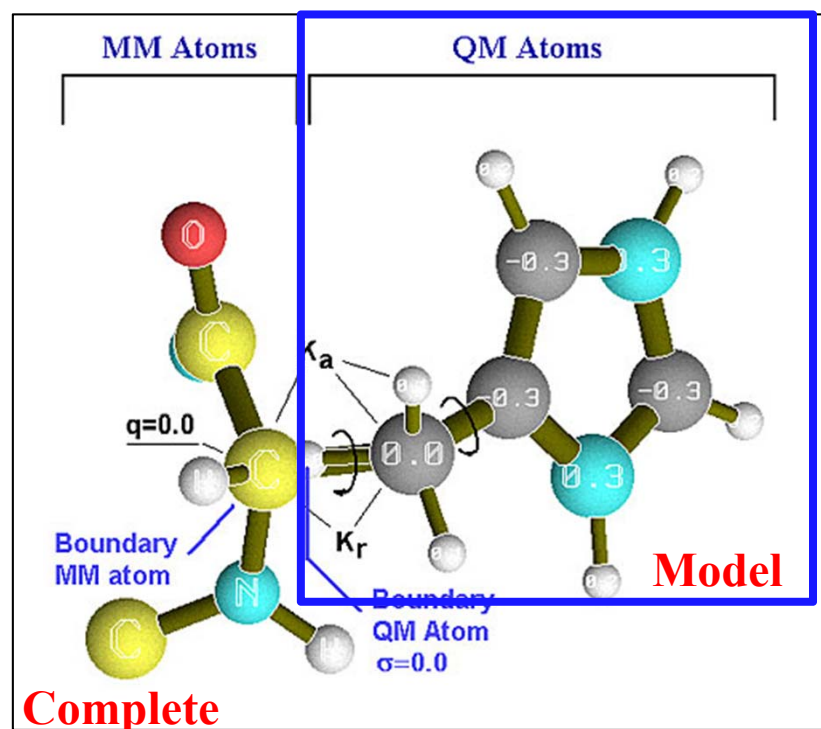


1. Link atoms are used to cap the electron density
2. For QM region, the link atom is hydrogen. It interacts with MM region only electrostatically
3. Bonded terms on the QM/MM boundary are considered if at least one MM atoms are involved
4. How to deal with the interaction between MM and the link atom is a problem.

$$E = E_{QM} + E_{MM} + E_{QM/MM}$$

$$E_{QM/MM} = \left\langle \psi \left| - \sum_i^{elec} \sum_A^{MM} \frac{q_A}{r_{iA}} \right| \psi \right\rangle + E_{QM,MM(vdw)} + E_{QM,MM(bond)}$$

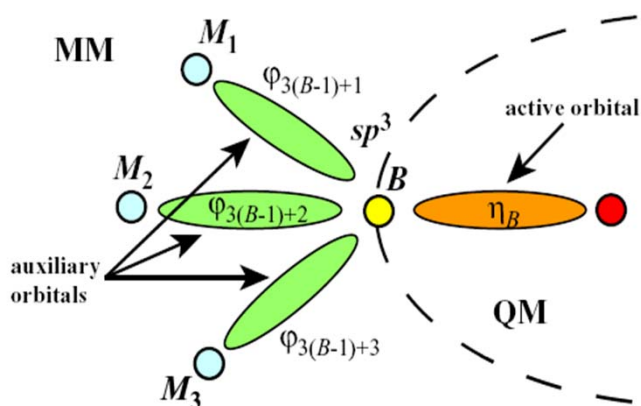
Subtraction approach of the Link atom (ONIOM)



1. **Model system : QM + Link atom**
2. **Do not require a parametrized expression to describe the interaction between two regions**
3. **Avoids any over counting of the non-bonded interactions.**
4. **How to deal with the electrostatic interaction between two is not straightforward.**

$$E = E_{MM,complete} + E_{QM,model} - E_{MM,model}$$

Overview of the GHO scheme



$$\begin{pmatrix} \eta_B \\ \varphi_{3(B-1)+1} \\ \varphi_{3(B-1)+2} \\ \varphi_{3(B-1)+3} \end{pmatrix} = \mathbf{t}_B^+ \begin{pmatrix} s_B \\ p_{Bx} \\ p_{By} \\ p_{Bz} \end{pmatrix}$$

1. The covalent bonding connecting the QM and MM fragment is assumed to be a σ bond
2. The Boundary atom B is chosen as sp³ carbon, which is connected to M₁, M₂ and M₃.
3. The valence s and p basis functions of B are transformed into a set of orthogonal hybrid orbitals.
4. Only one of the orbital lying along the Q-B bond will be optimized, whereas the other three will act as a set of auxiliary orbitals

Recent developments in the GHO method (1)

1. Occupation numbers of the auxiliary orbitals are determined according to the neighboring MM atom types

$$n_{\alpha} = 1 + \frac{\text{EN}_{C_{sp^3}} - \text{EN}_M}{2(\text{H}_{C_{sp^3}} + \text{H}_M)} \quad \text{EN: electronegativity, H: hardness}$$

Previous : $n_{\alpha} = 1 - q_B/3$ (Do not dependent upon the neighboring MM atom type)

2. Multiply a factor to the occupation numbers such that the formal charge condition is fulfilled

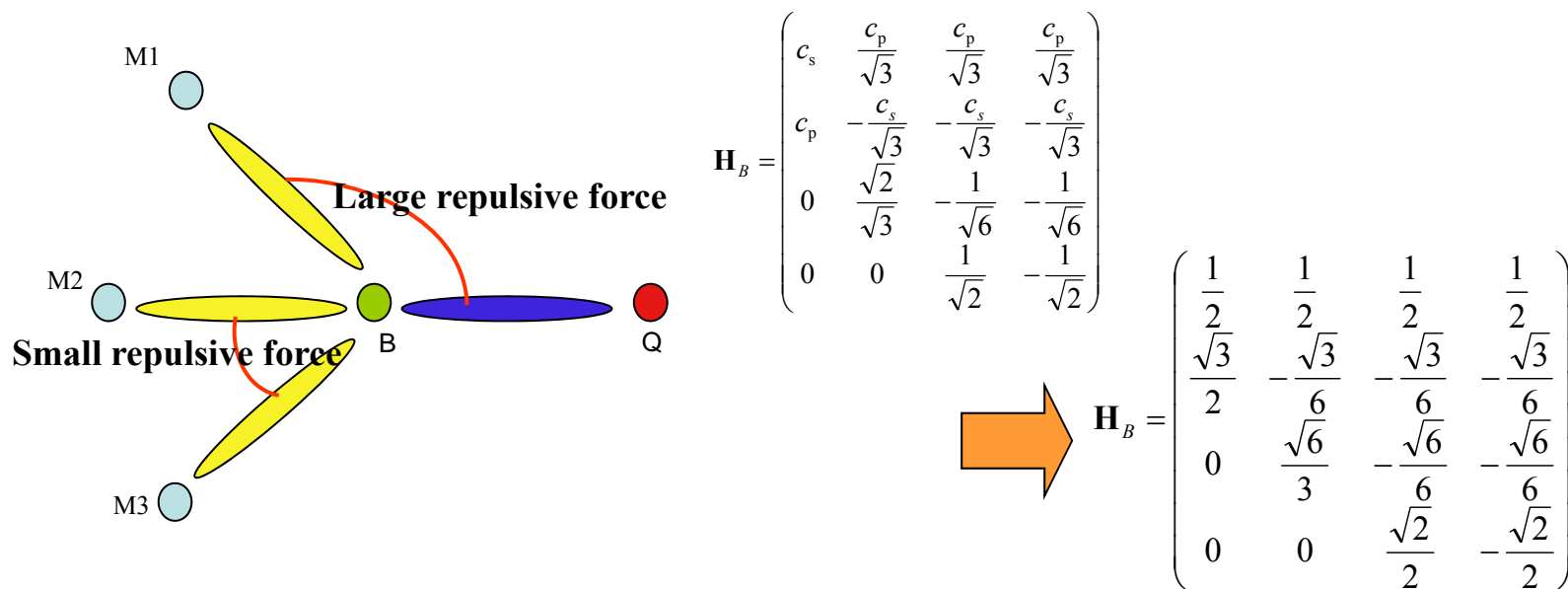
$$3N_B + \sum_M^{\text{pure MM}} q_M - \sum_{\alpha}^{\text{aux.}} n'_{\alpha} = 0$$

3. Make auxiliary orbitals orthogonal to each other before projection

$$\begin{pmatrix} \bar{\varphi}_1 \\ \bar{\varphi}_2 \\ \vdots \\ \bar{\varphi}_{3n} \end{pmatrix} = \mathbf{s}^{-\frac{1}{2}} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{3n} \end{pmatrix}$$

J. Jung et al. *J. Chem. Phys.* **127**, 204102 (2007)

Recent developments in the GHO method (2)



1. In the GHO method, the Pauli repulsion forces over the auxiliary orbitals are underestimated due to their partial occupations.
2. This fact causes systematic errors concerning the neighboring atoms
3. Simple way to reduce errors is to use a restrained hybridization (RH) matrix with all the amplitudes of s-orbital components fixed to the value in the ideal sp^3 hybridization limit.

MP2 Gradient in the GHO method (1)

1. Derivatives of molecular orbitals

$$C_{\mu i}^x = \sum_p^{\text{M.O.}} C_{\mu p} U_{pi}^x + \sum_{\alpha}^{\text{aux.}} A_{\mu\alpha} R_{\alpha i}^x$$

U_{pi}^x Standard coupled-perturbed HF coefficients

$R_{\alpha i}^x$ Expansion coefficients of auxiliary orbitals

$$\mathbf{R}^x = -(\mathbf{A}^+ \mathbf{S} \mathbf{A})^{-1} (\mathbf{A}^+ \mathbf{S}^x \mathbf{C} + (\mathbf{A}^x)^+ \mathbf{S} \mathbf{C})$$

2. Density matrix

$$D_{\mu\nu} = 2 \sum_i^{\text{d.o.}} C_{\mu i} C_{\nu i} + \sum_{\alpha}^{\text{aux.}} n_{\alpha} A_{\mu\alpha} A_{\nu\alpha}$$

3. Hartree Fock equation

$$\sum_{\nu} F_{\mu\nu} C_{\nu p} = \varepsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu p} + \sum_{\alpha}^{\text{aux.}} \sum_{\nu} S_{\mu\nu} A_{\nu\alpha} \omega_{\alpha p}$$

$$\varepsilon_p = \sum_{\mu\nu} C_{\mu p} F_{\mu\nu} C_{\nu p}$$

$$\boldsymbol{\omega} = (\mathbf{A}^+ \mathbf{S} \mathbf{A})^{-1} \mathbf{A}^+ \mathbf{F} \mathbf{C}$$

MP2 Gradient in the GHO method (2)

3. Final form of the MP2 gradient

$$E^x = E_{\text{MP2}}^x + E_{\text{GHO-HF}}^x + E_{\text{MA}}^x + E_{\text{AA}}^x + E_{\text{O}}^x + E_{\text{T}}^x$$

$$E_{\text{GHO-HF}}^x = 4 \sum_i^{\text{d.o.}} \sum_{\alpha}^{\text{aux.}} \sum_{\mu\nu} C_{\mu i} S_{\mu\nu} A_{\nu\alpha}^x \omega_{\alpha i} + 2 \sum_{\mu\nu} \sum_{\alpha}^{\text{aux.}} n_{\alpha} A_{\mu\alpha}^x F_{\mu\nu} A_{\nu\alpha}$$

HF modification due to GHO

$$E_{\text{MA}}^x = 2 \sum_i^{\text{d.o.}} \sum_{\alpha}^{\text{aux.}} \sum_{\mu\nu} R_{\alpha i}^x A_{\mu\alpha} C_{\nu i} \sum_{\lambda\sigma} P_{\lambda\sigma}^{(2)} \{2(\mu\nu | \lambda\sigma) - (\mu\lambda | \nu\sigma)\}$$

Density matrix modification due to GHO

$$E_{\text{AA}}^x = \sum_{\alpha}^{\text{aux.}} \sum_{\mu\nu} n_{\alpha} A_{\mu\alpha}^x A_{\nu\alpha} \sum_{\lambda\sigma} P_{\lambda\sigma}^{(2)} \{2(\mu\nu | \lambda\sigma) - (\mu\lambda | \nu\sigma)\}$$

$$D_{\mu\nu} = \underbrace{2 \sum_i^{\text{d.o.}} C_{\mu i} C_{\nu i}}_{\text{MA}} + \underbrace{\sum_{\alpha}^{\text{aux.}} n_{\alpha} A_{\mu\alpha} A_{\nu\alpha}}_{\text{AA}}$$

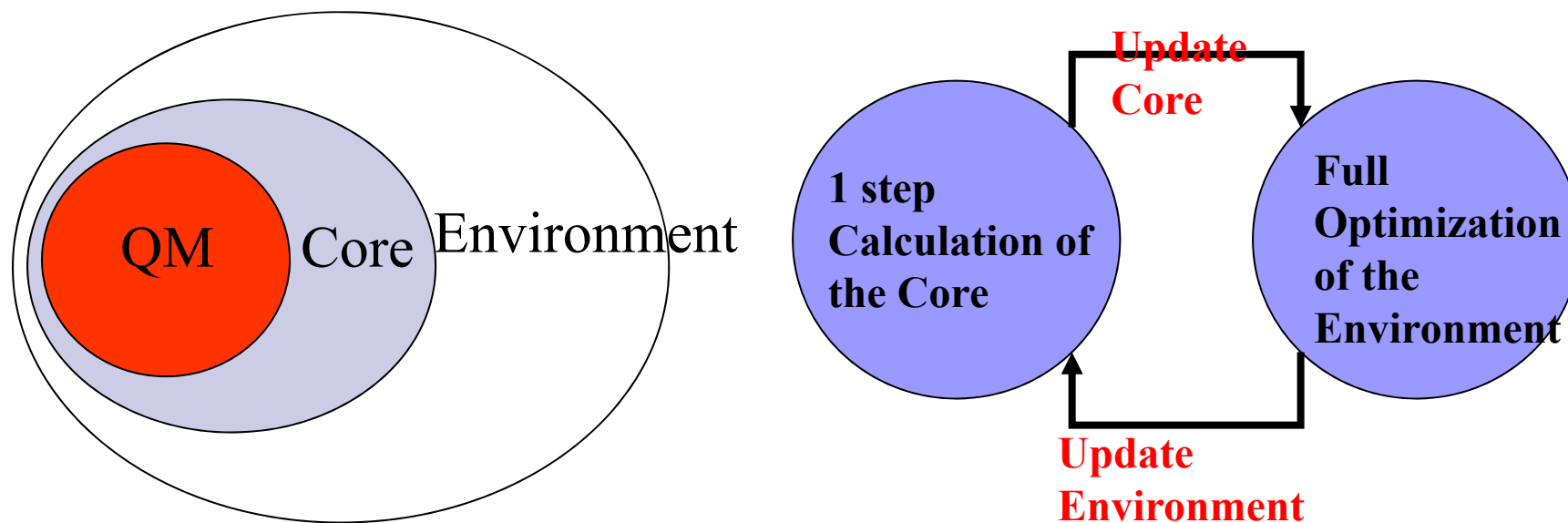
$$E_{\text{O}}^x = 2 \sum_{pq}^{\text{M.O.}} P_{pq}^{(2)} \sum_{\alpha\beta}^{\text{aux.}} R_{\alpha p}^x \tilde{S}_{\alpha\beta} \omega_{\beta q}$$

Fock matrix modification

$$E_{\text{T}}^x = 4 \sum_{ij}^{\text{oact.}} \sum_{ab}^{\text{vir.}} \sum_{\alpha}^{\text{aux.}} \frac{2(ia | jb) - (ib | ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \{R_{\alpha i}^x (\alpha a | jb) + R_{\alpha a}^x (i\alpha | jb)\}$$

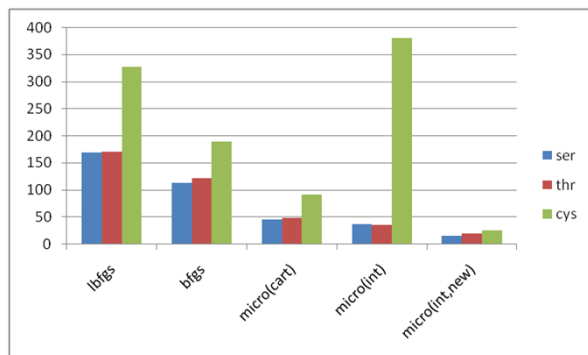
Derivative of two electron integrals

Optimization of GHO QM/MM :Adiabatic micro-iterative scheme

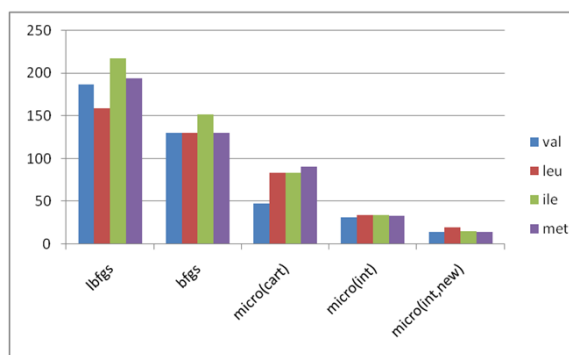


This way reduced the number of QM calculation

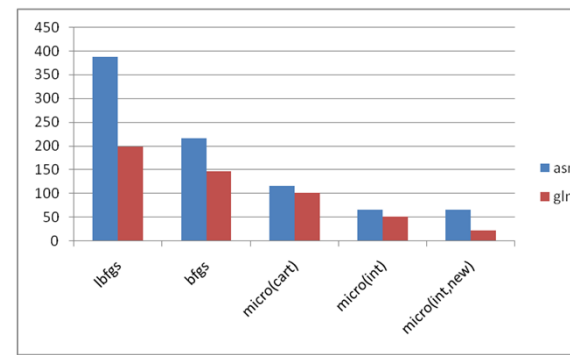
QM/MM optimization for 20 amino acid dipeptide (QM: side chain)



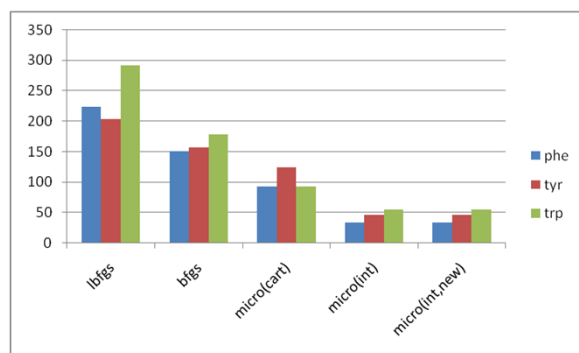
Nucleophilic



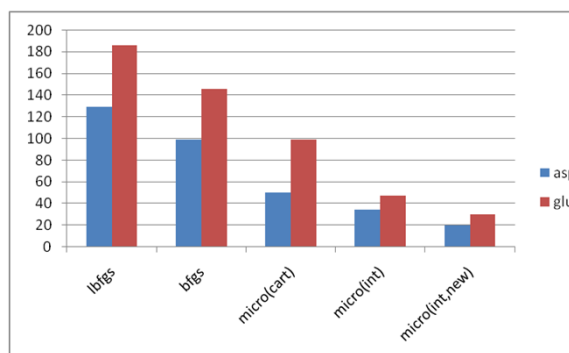
Hydrophobic



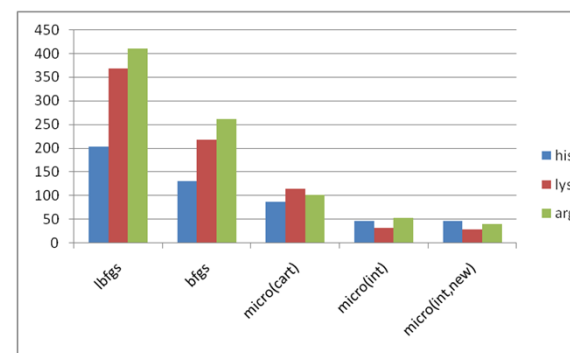
Amide



Aromatic



Acidic

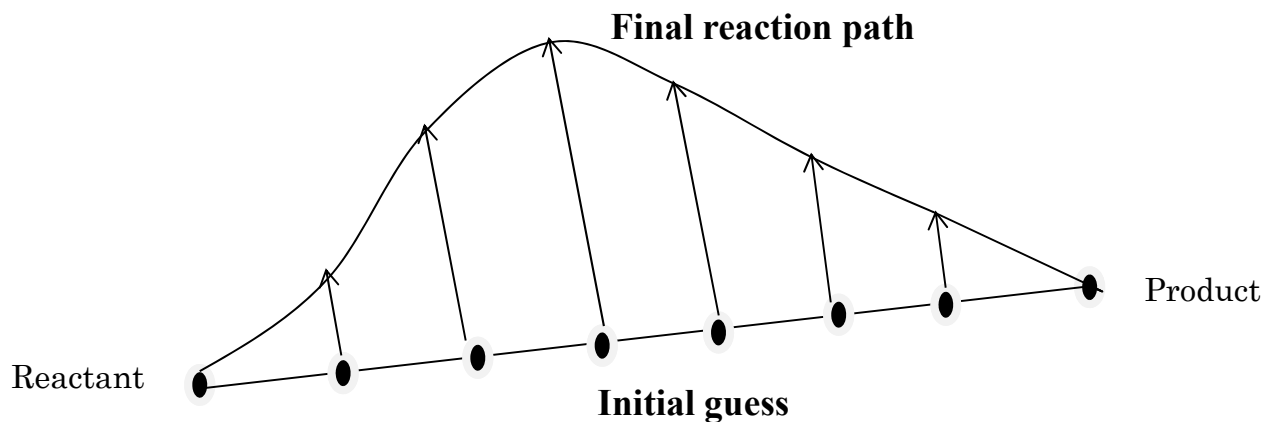


Basic



**How to find the chemical reaction path with
QM/MM ?**

Constrained optimization scheme



1. Calculate a minimum of the potential in a hyperplane perpendicular to the straight line connecting reactant and product
2. Fix the rigid body translation and rotation during minimization => Add constraint force

1. R. Czerminski and R. Elber, PNAS, 86, 6963 (1989)

2. R. Czerminski and R. Elber, JCP, 92, 5580 (1990)

Constrained optimization scheme

3. Target function includes 7 constraint forces

$$C_1(\mathbf{r}_i) = (\mathbf{r}_i - \mathbf{r}_{i0}) \cdot (\mathbf{r}_0 - \mathbf{r}_1) = 0$$

\mathbf{r}_{i0} is the coordinates of the initial structure of i -th image

\mathbf{r}_0 and \mathbf{r}_1 are two local minima structures of the reaction path

$$\sum_{i=1}^N m_i (\mathbf{r}_i - \mathbf{r}_i^o) = \mathbf{0} \quad \Rightarrow \quad C_2(\mathbf{r}_i), C_3(\mathbf{r}_i), C_4(\mathbf{r}_i)$$

$$\sum_{i=1}^N m_i \mathbf{r}_i \times (\mathbf{r}_i^o - \mathbf{r}_i) = \mathbf{0} \quad \Rightarrow \quad C_5(\mathbf{r}_i), C_6(\mathbf{r}_i), C_7(\mathbf{r}_i)$$

$$T(\mathbf{r}_i) = V(\mathbf{r}_i) + \sum_{k=1}^7 \sigma_k(\mathbf{r}_i)$$

σ_k : independent linear combination of 7 constraint force

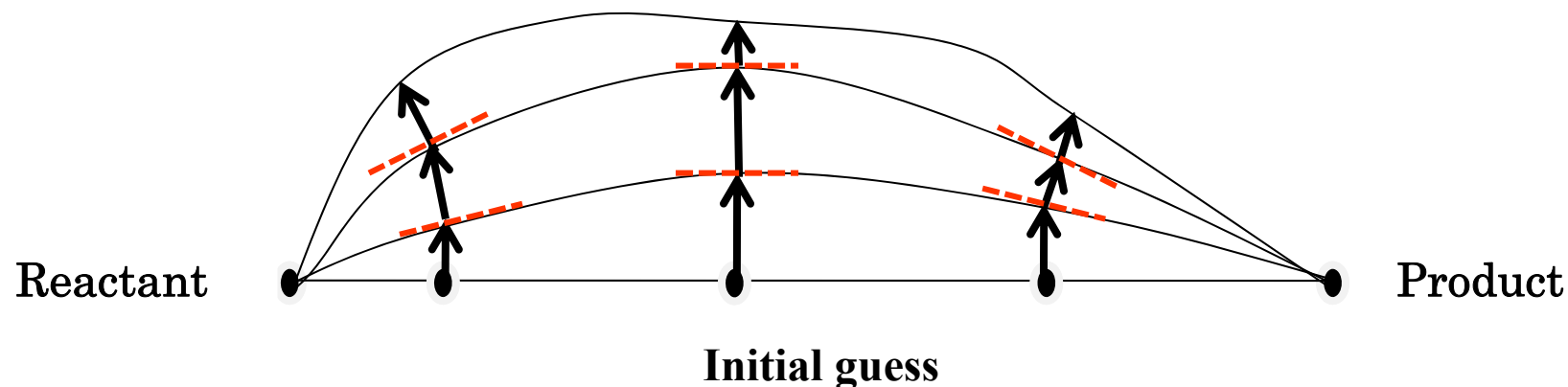
4. Optimization is used to get the minum of the target function

1. R. Czerminski and R. Elber, PNAS, 86, 6963 (1989)

2. R. Czerminski and R. Elber, JCP, 92, 5580 (1990)

Constrained optimization with locally updated planes (CO-LUP) scheme

1. In this scheme, tangent is updated after finite step of minimization¹



2. Tangent is determined by the two neighboring images

$$\boldsymbol{\tau}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_{i-1}}{|\mathbf{r}_{i+1} - \mathbf{r}_{i-1}|}$$

3. The first constraint is expressed by

$$C_1(\mathbf{r}_i) = (\mathbf{r}_i - \mathbf{r}_{i0}) \cdot \boldsymbol{\tau}_i = 0$$

New consideration – 1) Tangent estimation

1. Only the image with higher energy and the image i are used in the estimate

$$\boldsymbol{\tau}_i = \begin{cases} \mathbf{R}_{i+1} - \mathbf{R}_i & \text{if } V_{i+1} > V_i > V_{i-1} \\ \mathbf{R}_i - \mathbf{R}_{i-1} & \text{if } V_{i+1} < V_i < V_{i-1} \end{cases}$$

2. If image i is at a minimum or at a maximum, the tangent estimate becomes

$$\boldsymbol{\tau}_i = \begin{cases} \Delta V^{\max} (\mathbf{R}_{i+1} - \mathbf{R}_i) + \Delta V^{\min} (\mathbf{R}_i - \mathbf{R}_{i-1}) & \text{if } V_{i+1} > V_{i-1} \\ \Delta V^{\min} (\mathbf{R}_{i+1} - \mathbf{R}_i) + \Delta V^{\max} (\mathbf{R}_i - \mathbf{R}_{i-1}) & \text{if } V_{i+1} < V_{i-1} \end{cases}$$

where

$$\Delta V^{\max} = \max(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$$

$$\Delta V^{\min} = \min(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$$

=> Kinks are eliminated by the new tangent estimation => less number of calculations

New consideration – 2) Arc length parametrization

To make all images are equally distributed to each other, we use an arc length parametrization before tangent update:

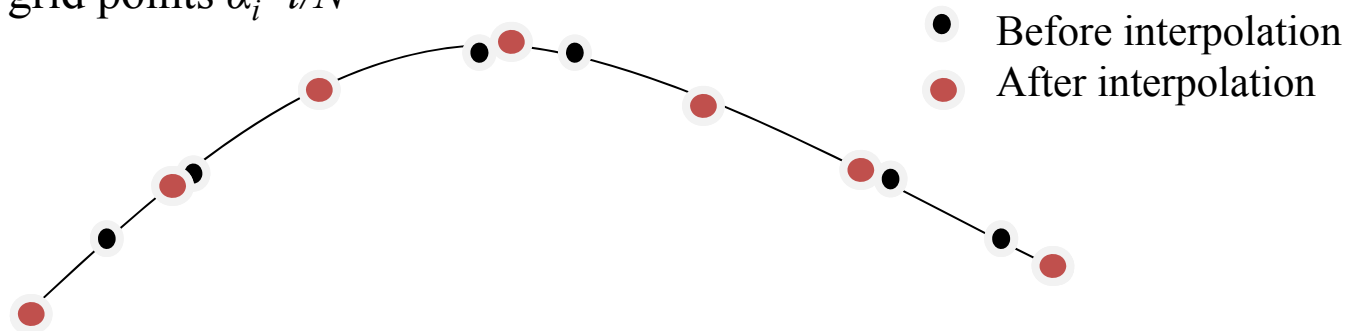
1. Calculate the arc length corresponding to the current images,

$$s_0 = 0, \quad s_i = s_{i-1} + |\varphi_i - \varphi_{i-1}|$$

2. The mesh is obtained by normalizing $\{s_i\}$

$$\alpha_i = \frac{s_i}{s_N}$$

3. Paths are expressed in terms of the mesh and the new points are obtained at the uniform grid points $\alpha_i = i/N$



$$W_{i,i+1} = E_{\text{high}} - E_{\text{low}} + (V_i + V_{i+1})/2 - E_{\text{low}}$$

New consideration – 2) Arc length parametrization

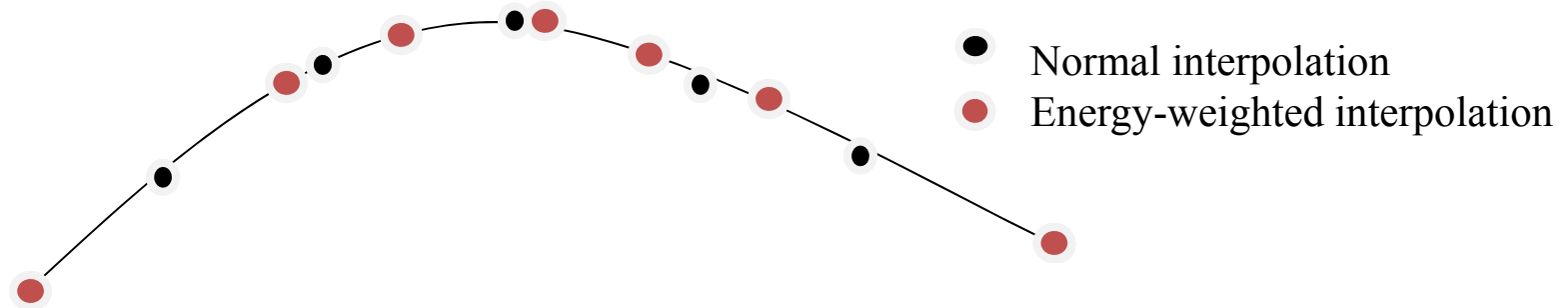
4. To get a finer resolution around the saddle point, we use the energy-weighted arc parametrization


$$W_{i,i+1} = E_{\text{high}} - E_{\text{low}} + (V_i + V_{i+1})/2 - E_{\text{low}}$$

E_{high} : Higher local minimum energy

E_{low} : Lower local minimum energy

$$d(i, i+1) = |\varphi_i - \varphi_{i+1}| \times W_{i,i+1}$$





New consideration – 3) Removing translation/rotation constraints

1. With an optimizer permitting translation and rotation, 6 constraints are necessary because translation/rotation can affect the frame and eventually the first constraint.
2. With an optimizer that do not allow translation/rotation, these constraints are not necessary any longer.
3. Translation/rotation constraints increase the total number of energy/gradient calculations
4. Target function is

$$T(\mathbf{r}_i) = V(\mathbf{r}_i) + \lambda C(\mathbf{r}_i)$$

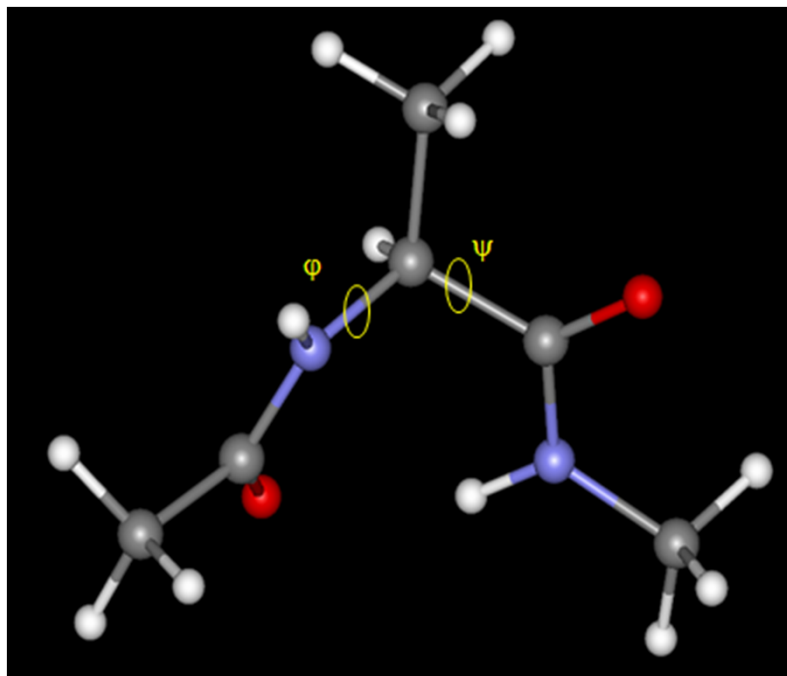
The corresponding gradient is

$$-\nabla T = -\nabla V - \lambda \nabla C = \mathbf{F} - \lambda \boldsymbol{\tau}_i$$

where

$$\lambda = \mathbf{F} \cdot \boldsymbol{\tau}_i$$

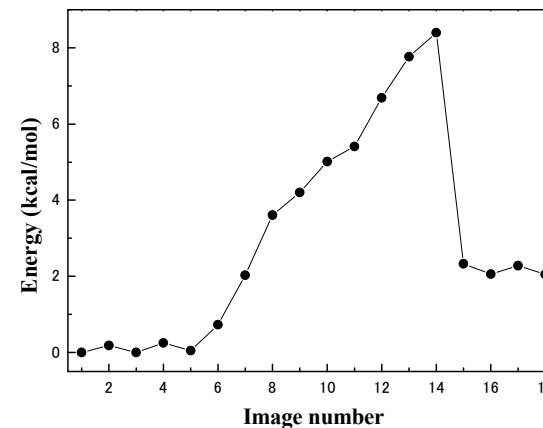
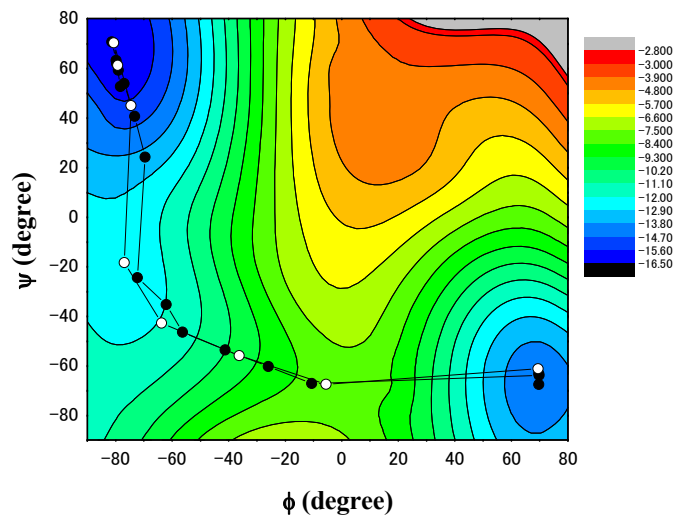
Test example – Alanine dipeptide (C7eq \rightarrow C7ax reaction with CHARMM22 force field)



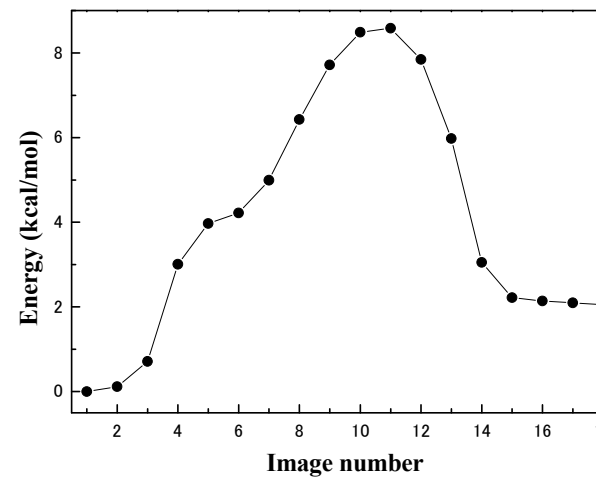
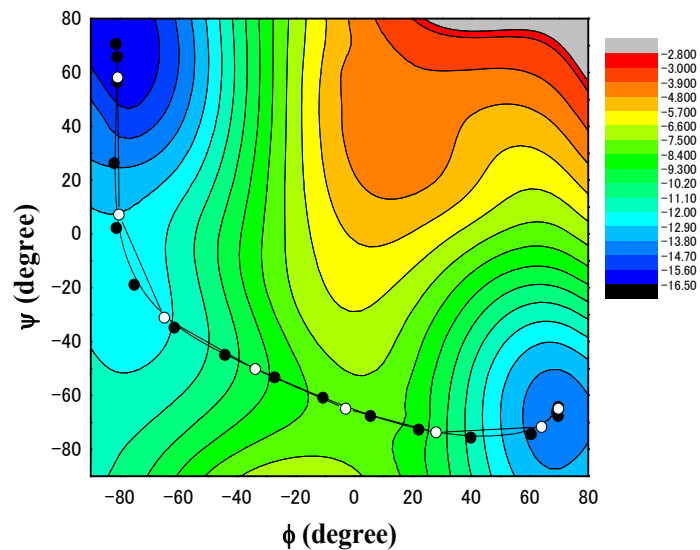
C7eq and C7ax conformers are two minimum conformers obtained by changing two dihedral angles ϕ and ψ .

Reaction path between C7eq and C7ax by CO (constrained optimization) and CO-LUP

Original CO

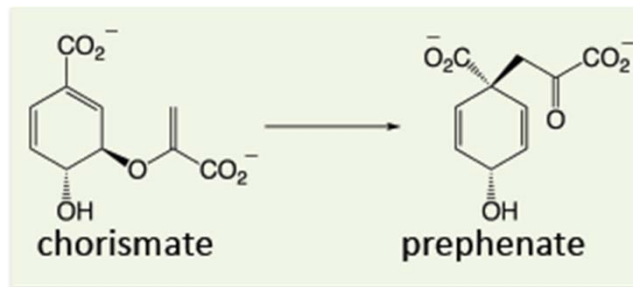


New Scheme

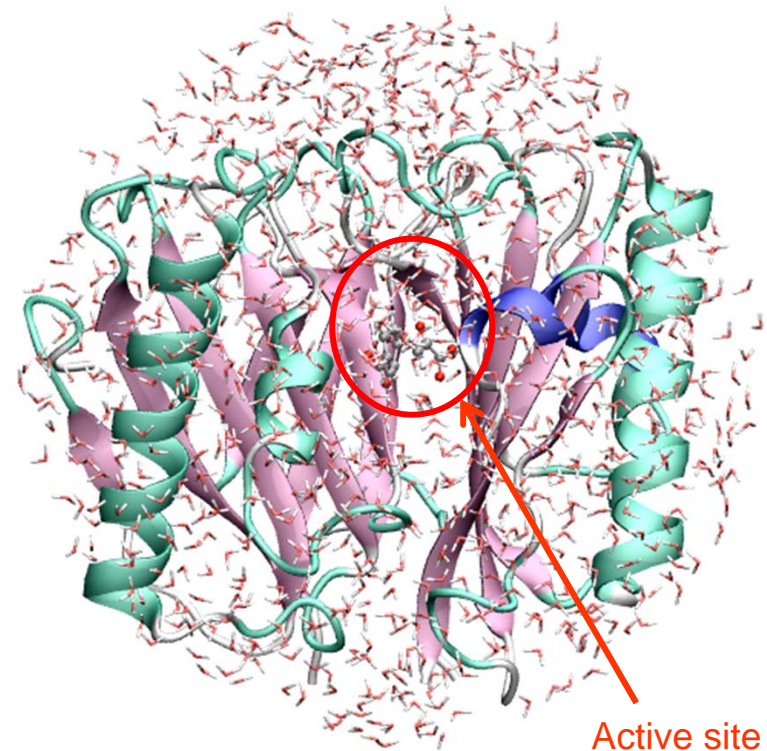


Chorismate mutase

1. **Catalyzed the chemical reaction for the conversion of chorismate to prephenate (Claisen rearrangement).**

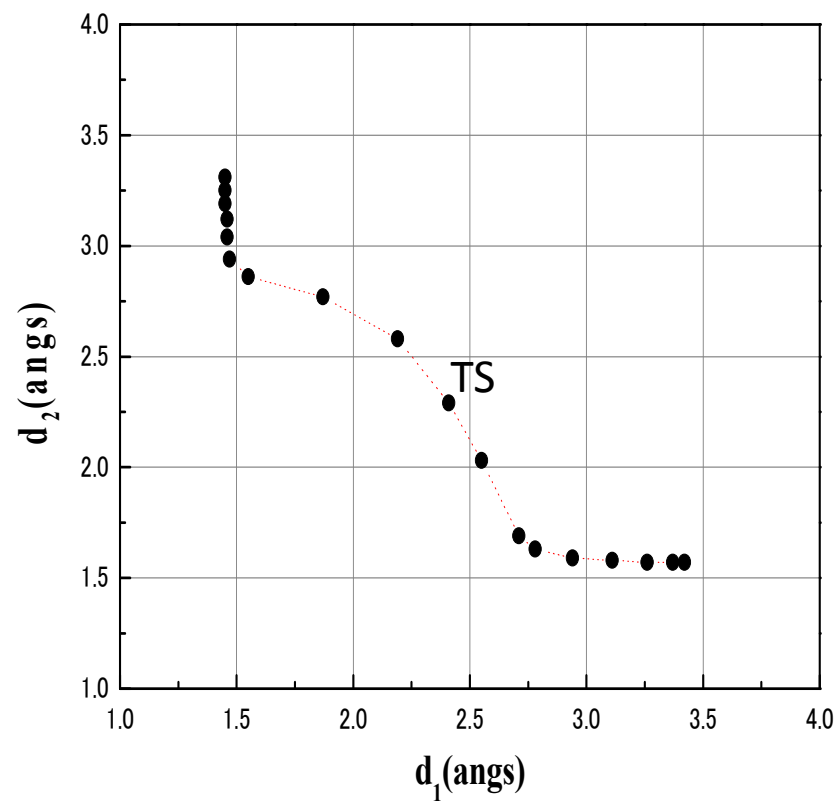
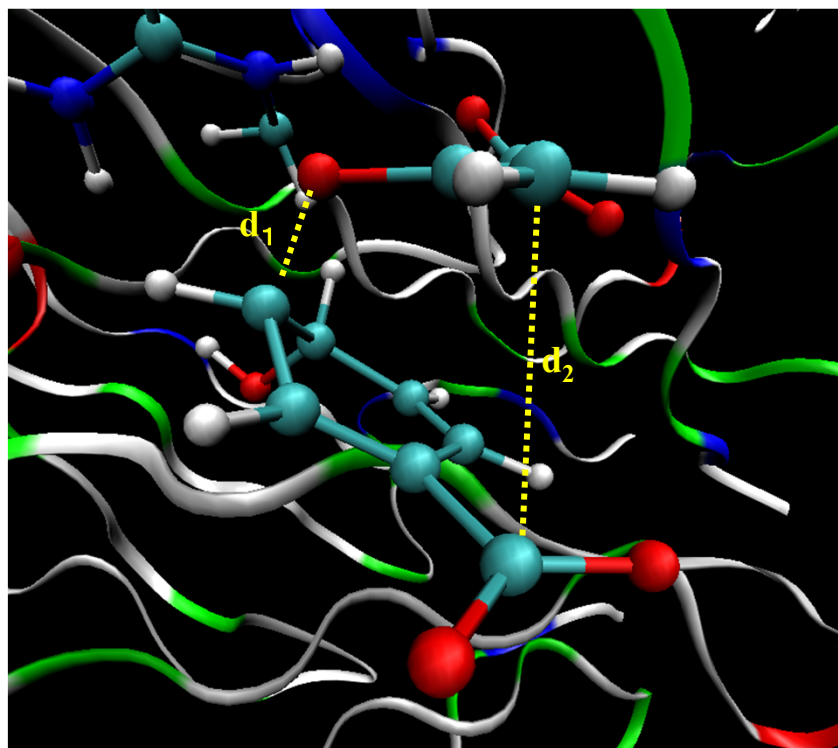


2. **Maintain the balance of aromatic amino acids in the cell.**

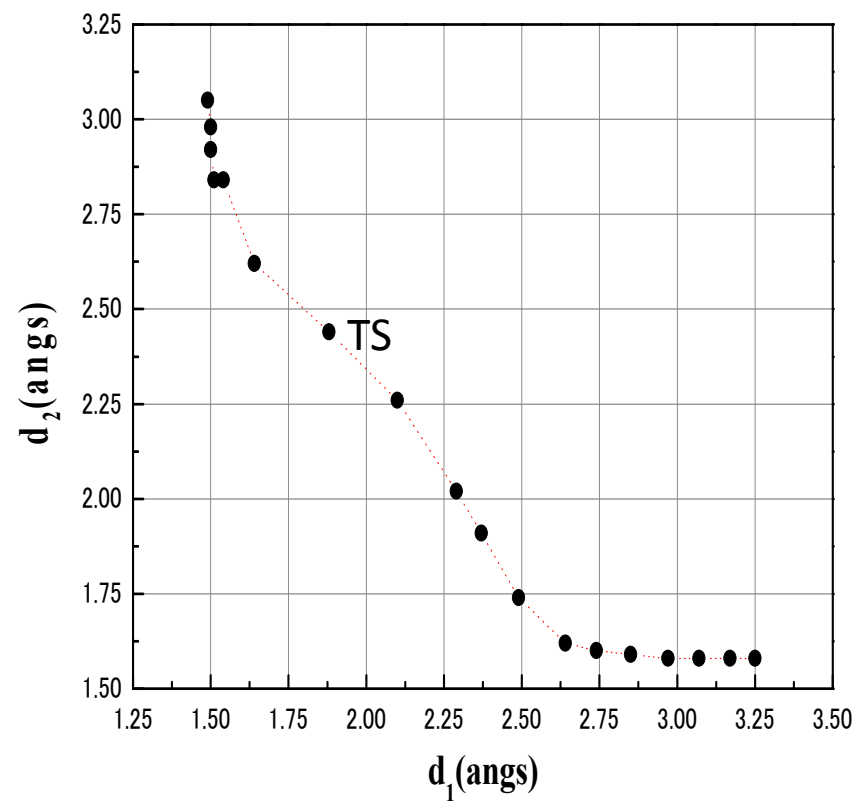
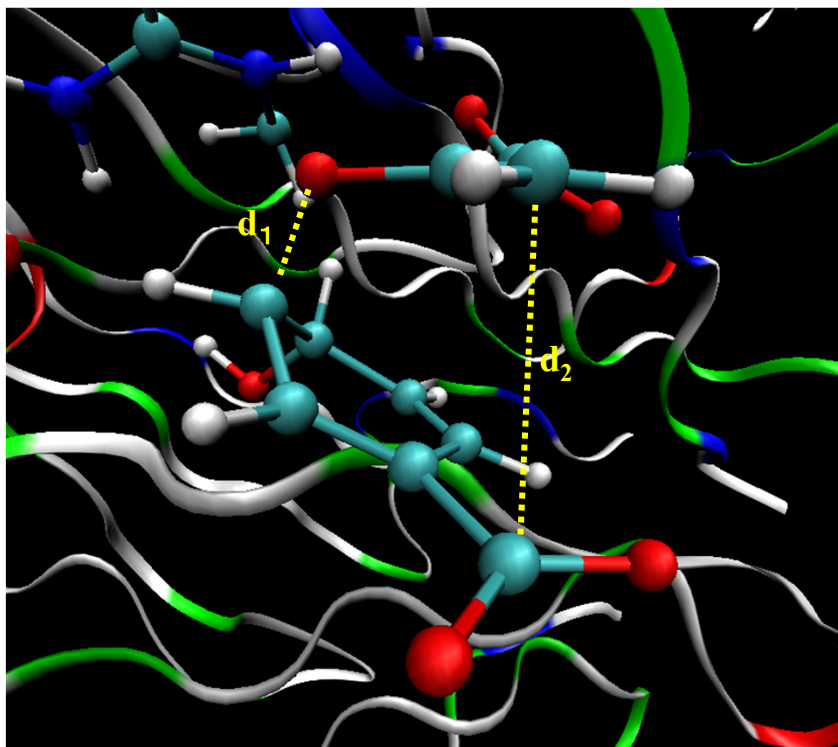


7076 atoms (947 H₂O_s)

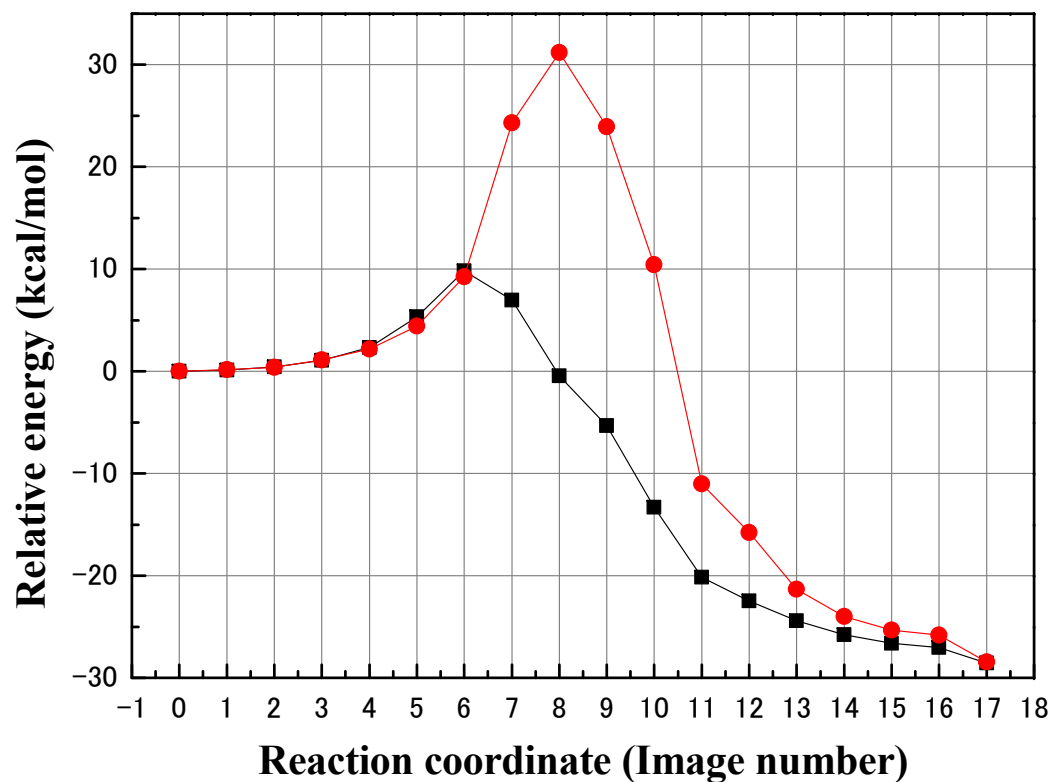
Reaction path of Chorismate mutase (HF)



Reaction path of Chorismate mutase (MP2)



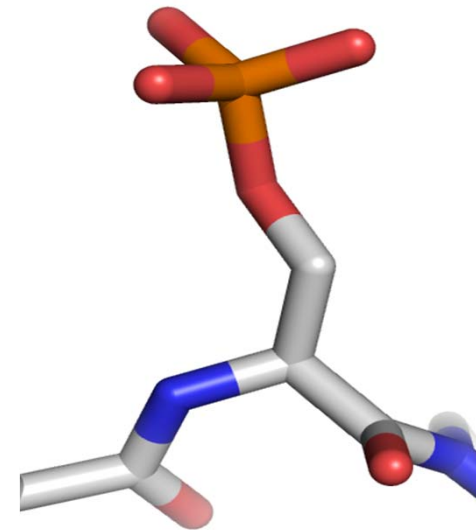
Potential energy curve of chorismate mutase along the reaction



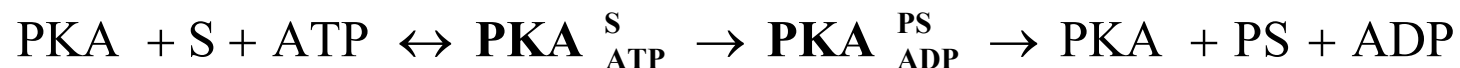
Reaction energy barrier: **GH0-HF ~31 kcal/mol**
GH0-MP2 ~10 kcal/mol
(Experiment = 12.7kcal/mol)

Phosphorylation Reaction by cAMP-dependent Protein Kinase (PKA)

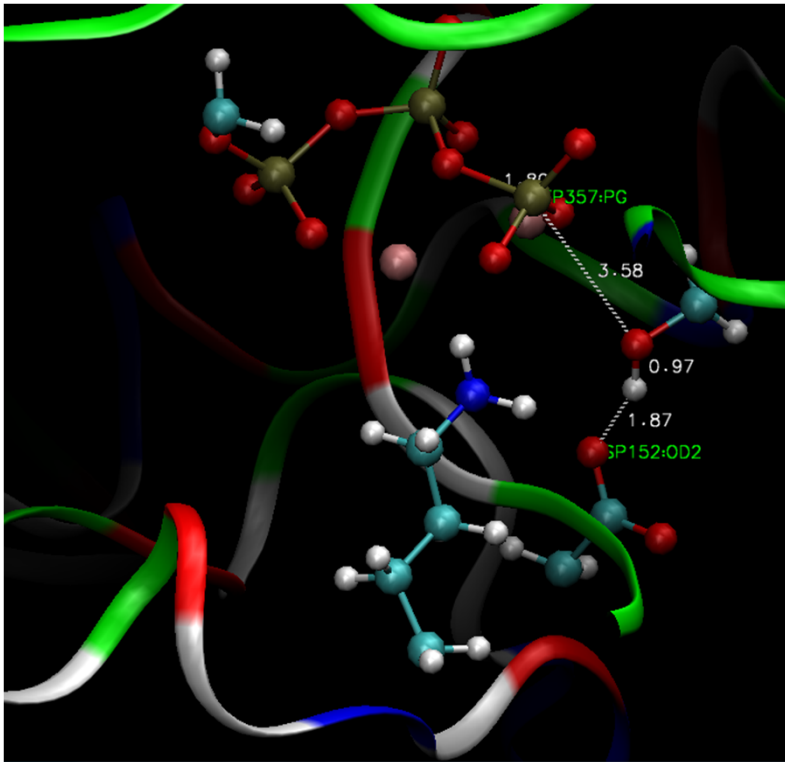
1. **Phosphorylation** : Addition of a phosphate (PO_4) to a protein plays a critical role in cellular regulation and signal transduction.
2. **Function of PKA** : Catalysis of the γ -phosphate transfer of ATP to specific serine or threonine residues.



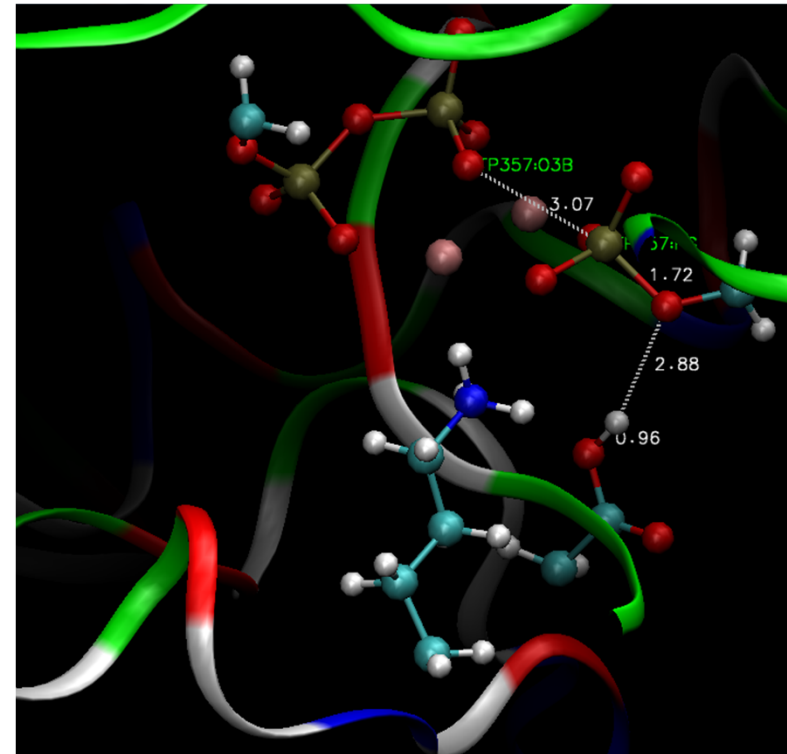
A phosphorylated serine residue



Optimized structure of PKA (QM=HF, bas=6-31G)



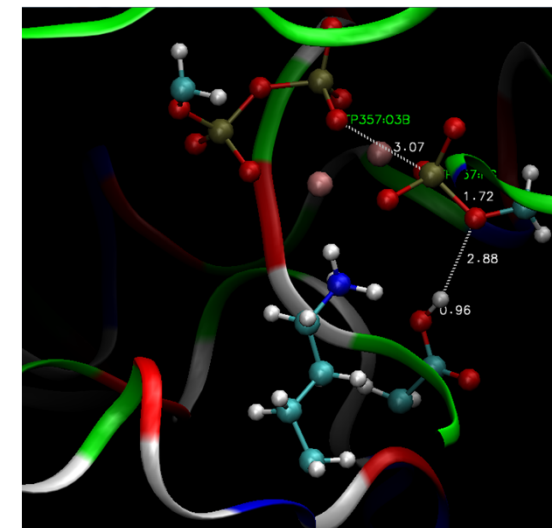
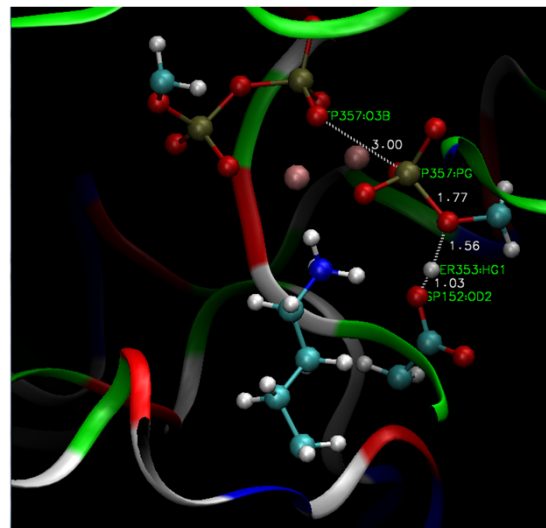
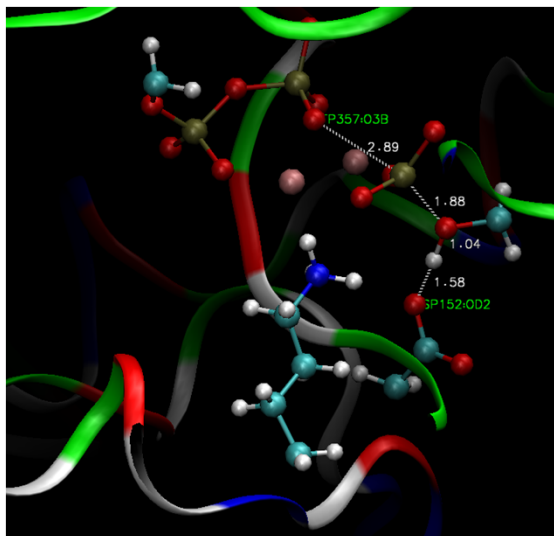
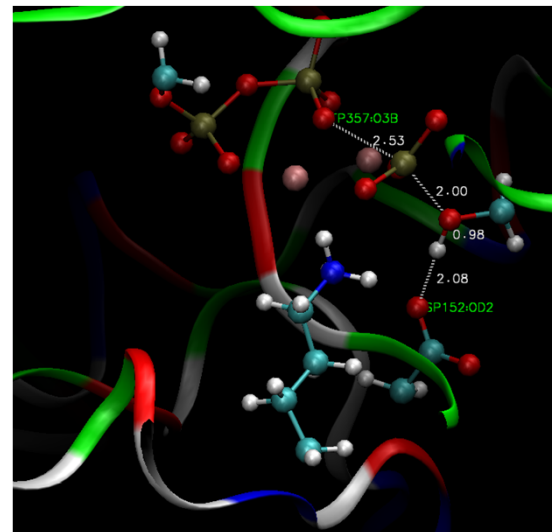
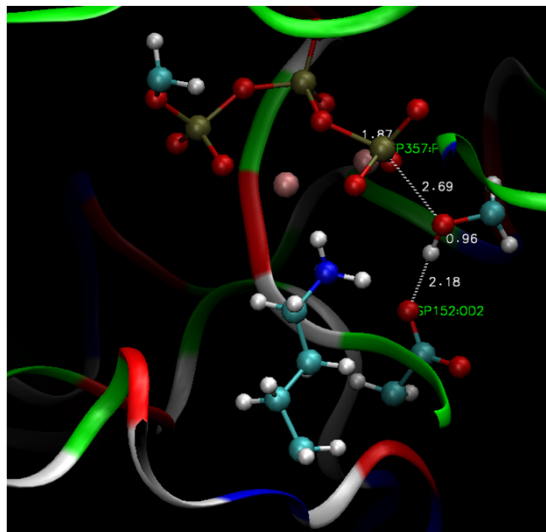
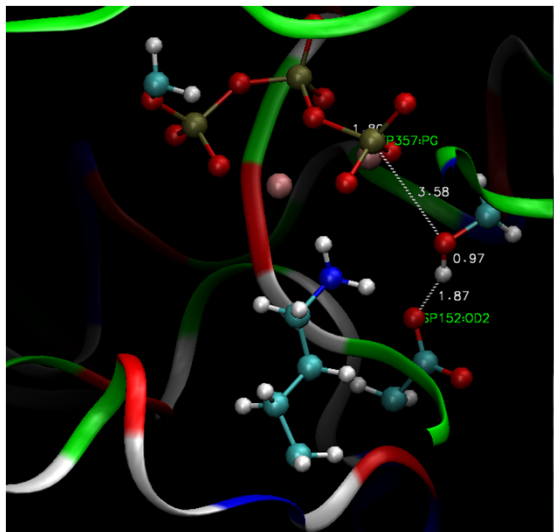
Reactant



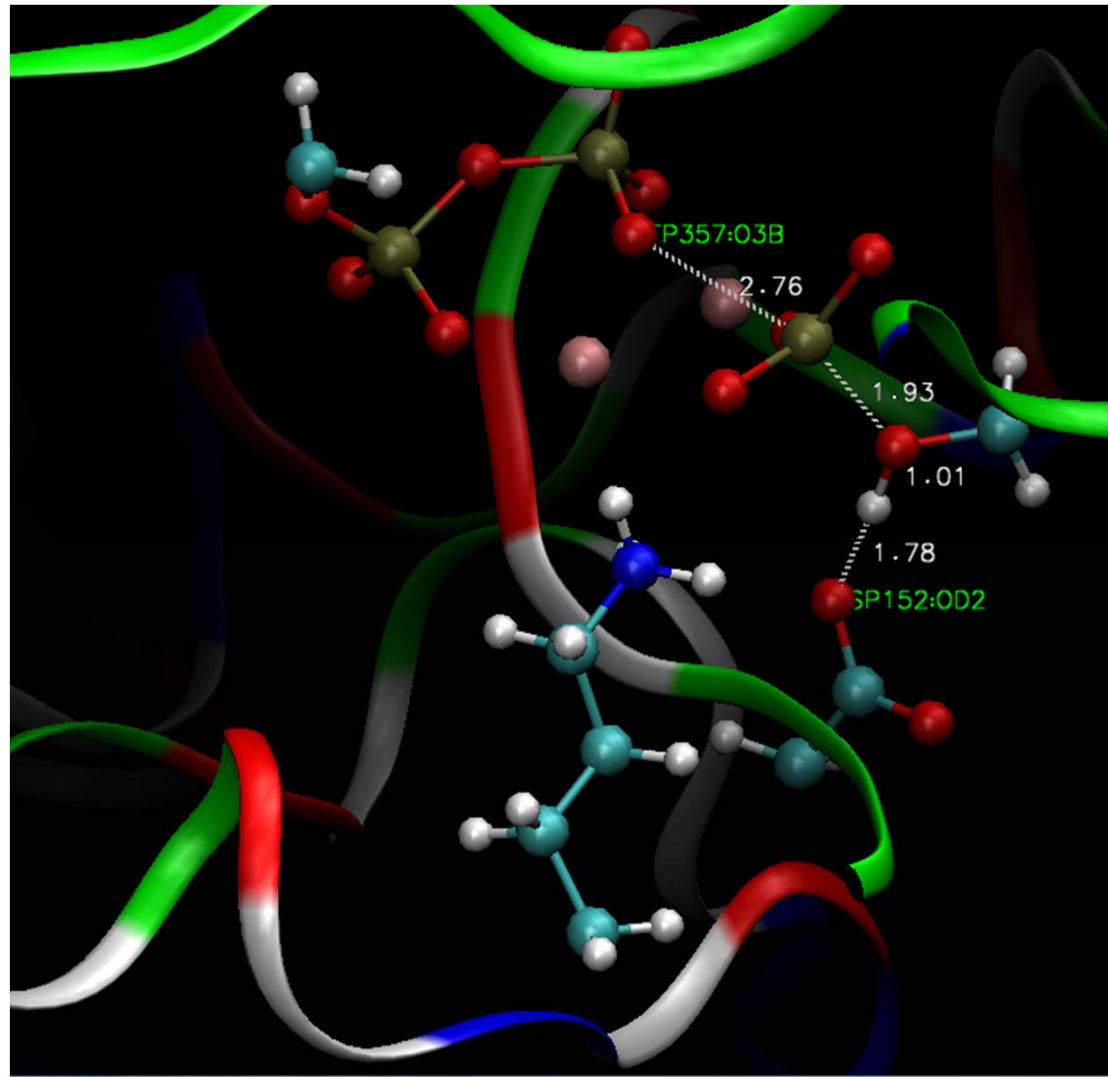
Product

1. PO_3 group is moved from ATP to SER residue (Ser353 is phospholylated).
2. Asp152 serves as the catalytic base to accept the proton from phospholylated Ser353.

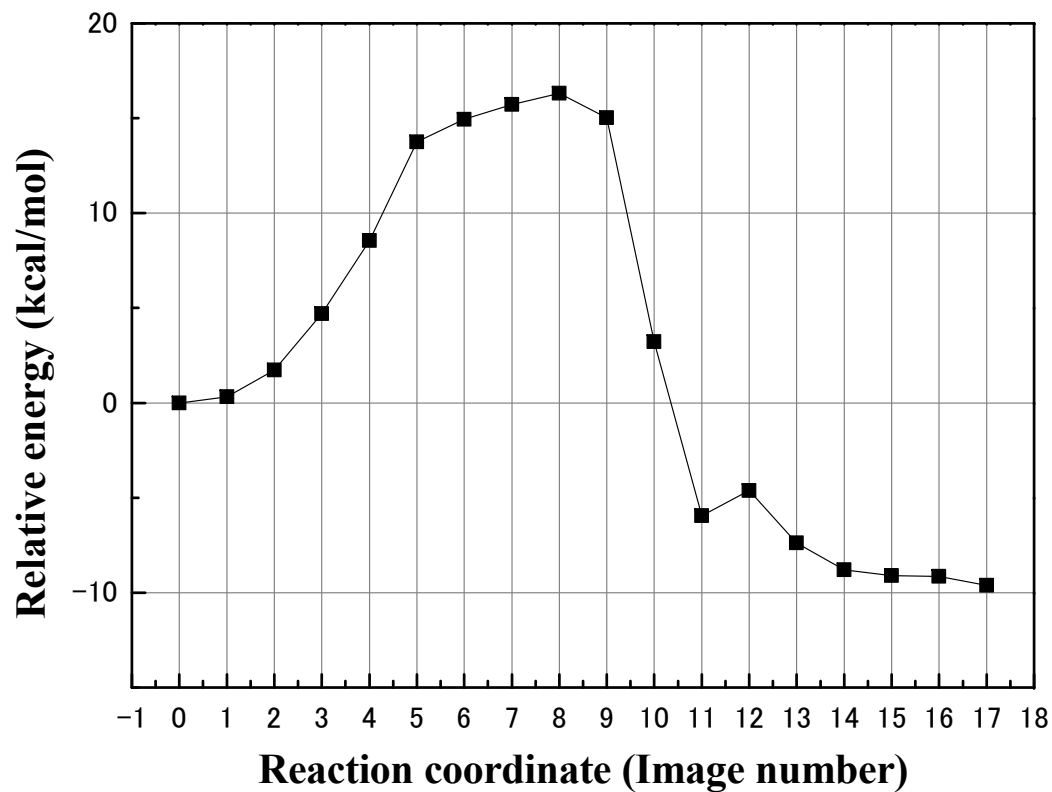
Reaction procedure of PKA



Transition state structure of PKA



Potential energy curve of PKA along the reaction



Reaction energy barrier ~ 17 kcal/mol (Experiment = 13.8kcal/mol)



Summary

1. QM/MM is important for the description of chemical reaction reaction.
2. Out of many treatment of QM/MM boundary, Generalized Hybrid Orbital (GHO) scheme is considered and developed for the stable electronic connection between QM and MM.
3. For the accurate molecular structure, MP2 is developed.
4. Reaction path determination scheme is developed for the systematic understanding of the chemical reaction with protein environment.