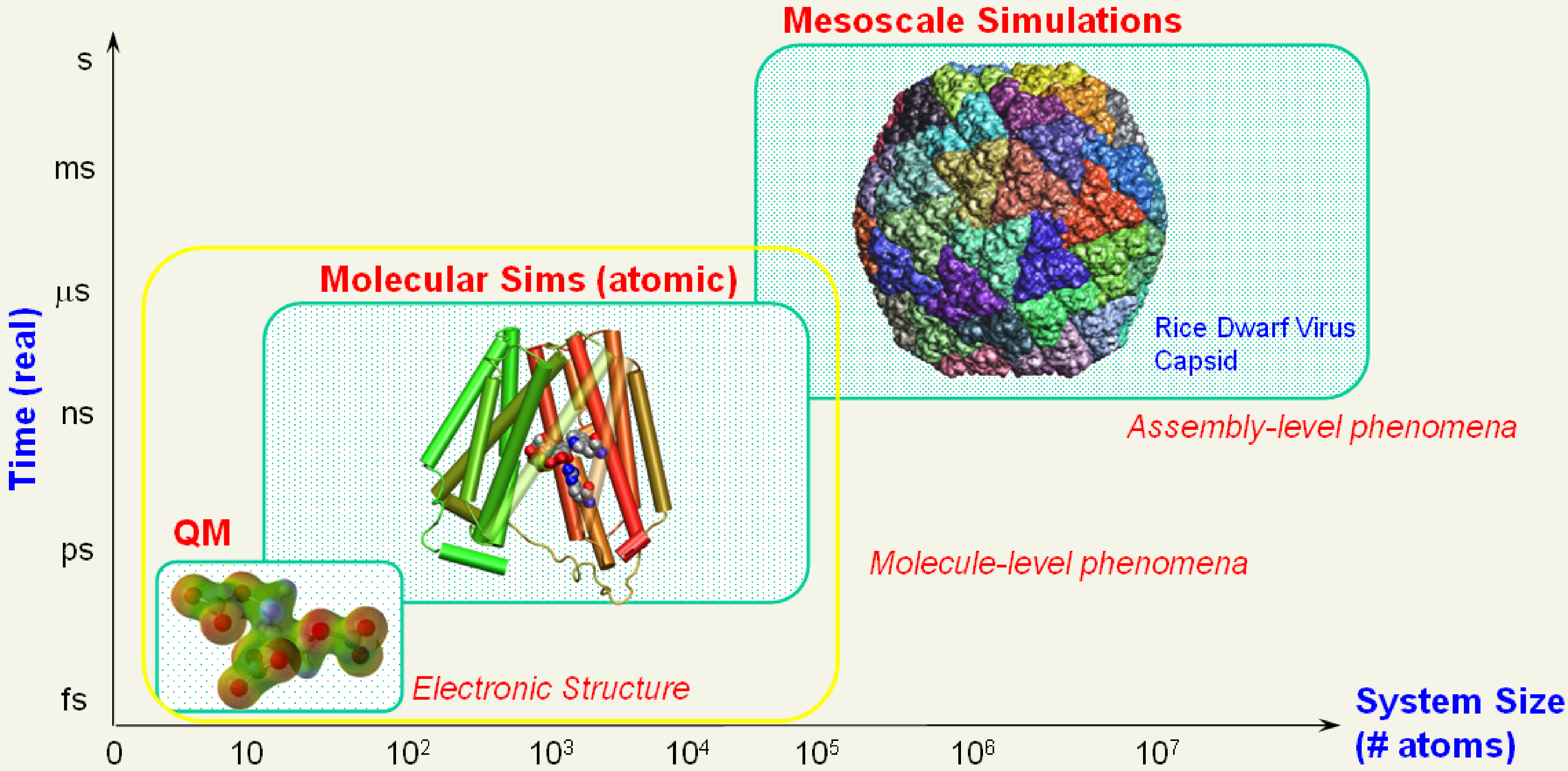
Quantum Mechanics/ Molecular Mechanics (QM/MM)

Modeling Approaches



- Research Design

ab initio Quantum Mechanics → Molecular Simulations & Models → Macroscopic Properties

Quantum Mechanics

- In theory, a very accurate treatment of the system
- Largely *ab initio*, i.e. parameter-free
- Very expensive — typically scales as $\mathcal{O}(N^4)$ or worse
- Limited to very small systems at high accuracy (e.g. DFT)
- Can be used for larger systems at lower accuracy (e.g. semi-empirical)
- Entire proteins cannot be simulated without enormous supercomputer power

Molecular Mechanics

- Treats the electrons implicitly — no handling of polarization or electron transfer
- Bonds, angles, and dihedrals are held by a parameterized force field

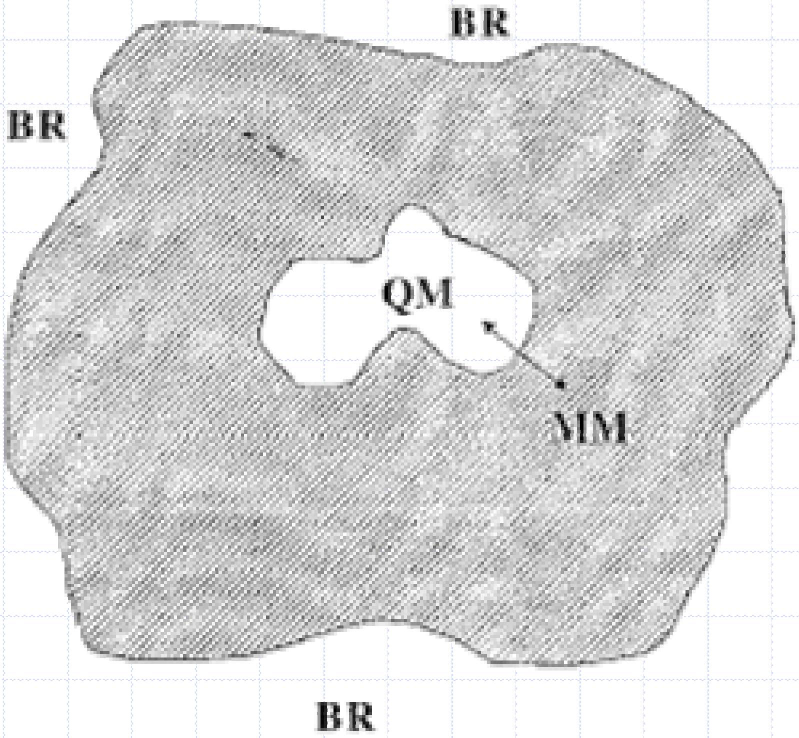
$$V_{total} = \sum_{\text{bonds}} V_{bond} + \sum_{\text{angles}} V_{angle} + \sum_{\text{dihed.}} V_{dihed}$$

$$+ \sum_{\text{impr.}} V_{impr} + \sum_{i \neq j} (V_{vdW} + V_{elec})$$

- Can be used to simulate very large systems — e.g. transmembrane proteins
- Cannot handle bond breaking or formation, so cannot be used to simulate chemical reactions

The QM/MM Idea

Multi-layered method

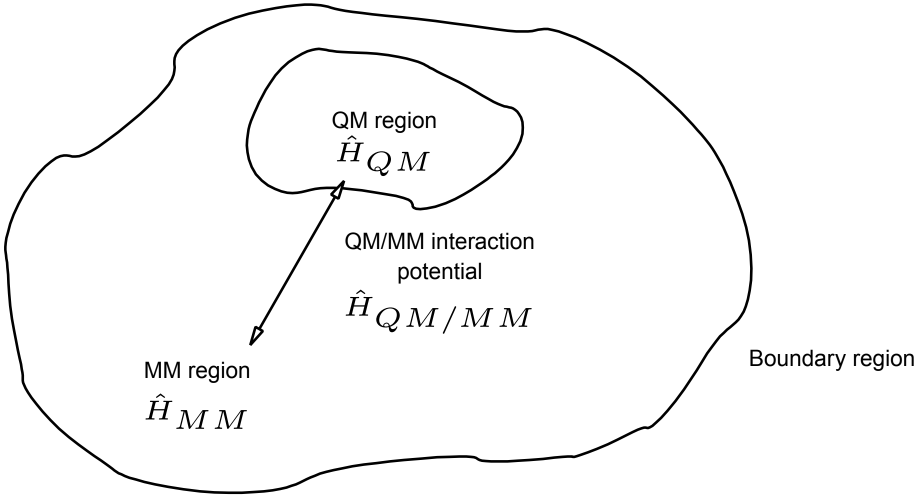


Monard, et. al, Acc. Chem. Res., **32**, 904 (1999)

Hybrid QM/MM

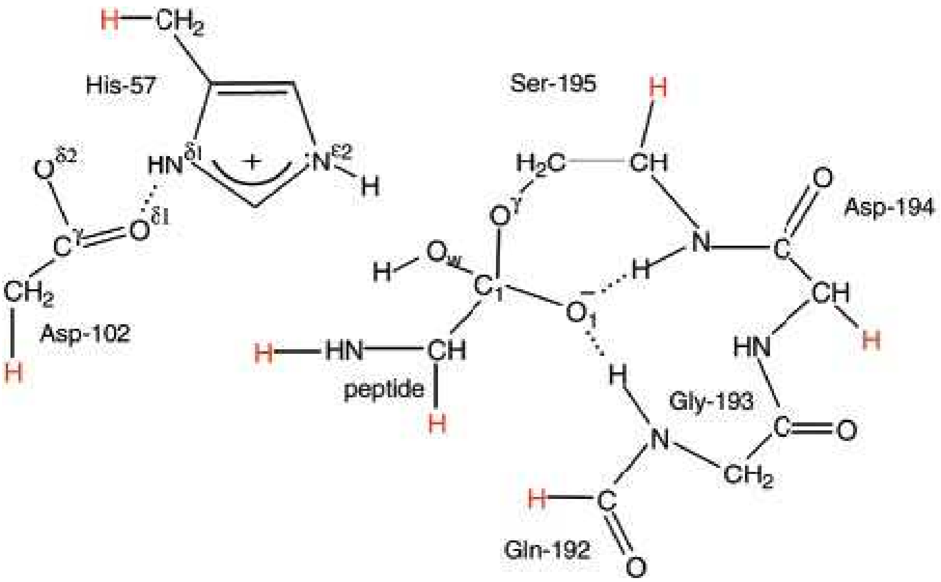
- Combines quantum mechanical and molecular mechanical methods
- Treats just the reacting part of the system quantum mechanically, and uses MM for the surroundings
- Uses a combined Hamiltonian for the system:

$$\hat{H}_{total} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}$$



Example Systems

- Study of serine protease deacylation reaction [a]



- Catalyzed isomerization of methylmalonyl-CoA [b]

---

[a] Topf, M., Várnai, P. and Richards, W. G. *J. Am. Chem. Soc.* **2002**, *124*, 14780

[b] Loferer, M., Webb, B., Grant, G, and Liedl, K. *J. Am. Chem. Soc.* **2003**, *125*, 1072

QM/MM Partitioning

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

The tough part – how do QM and MM interact?

Energy of MM subsystem

$$E_{QM} = \frac{\langle \psi \hat{H} \psi \rangle}{\langle \psi \psi \rangle}$$

Warshel and Levitt, J. Mol. Biol.
Field, Bash and Karplus, J. Comp. Chem.

QM Region

- What should be used in the QM region?
  - Ab Initio
  - DFT
  - Semiempirical
- Usually, the answer to this is dictated by cost. Most QM/MM simulations to date have used semiempirical QM regions
- Why? QM/MM interaction term can be problematic – it is not good to have this boundary close to the chemistry of interest…

Pitfalls in QM/MM

$$V(\vec{R}) = \sum_{i \in bonds} k_{i,bond}\left(r_{i,bond} - r_{i,bond}^{eq}\right) + \sum_{i \in angles} k_{i,angle}\left(\theta_{i,angle} - \theta_{i,angle}^{eq}\right) + \cdots$$

$$+ \tfrac{1}{2} \sum_{\substack{i \neq j \\ i,j \in atoms}} \frac{q_i q_j}{\left|r_i - r_j\right|} + \tfrac{1}{2} \sum_{\substack{i \neq j \\ i,j \in atoms}} V_{LJ}^{ij}\left(\left|r_i - r_j\right|\right)$$

Not clear which force fields to use – much experience with expected accuracy of *ab initio* methods alone and MM methods alone, but not much with QM/MM

No direct map from wavefunction to parameters

Mechanical Embedding

- Crudest level of QM/MM
- Include only Van der Waals in $E_{QM/MM}$
- Useful to impose only steric constraints
- Can take advantage of this to isolate effects…

$$H_{QM/MM} = \sum_{\substack{i \in MM \\ j \in QM}} V_{ij}^{VdW}(r_i, r_j)$$

Electrostatic Embedding

- Include electrostatic interaction in $H_{QM/MM}$
- Many possible implementations – best is to evaluate integrals over continuous QM charge density and discrete MM charge density

$$H_{QM/MM} = H_{QM/MM}^{mechanical} + \sum_{i \in MM} q_i \int \frac{\rho_{QM}(r)}{|r - r_i|} dr$$

Oft-used approximation (questionable):

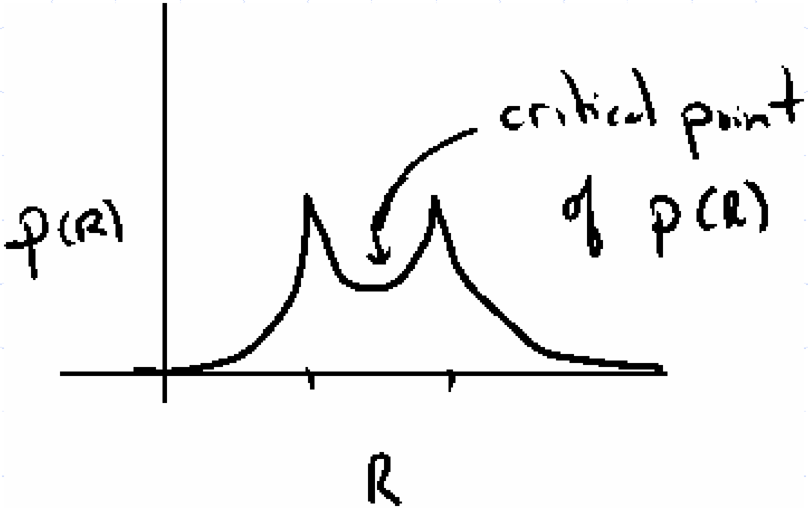
$$H_{QM/MM} = H_{QM/MM}^{mechanical} + \sum_{\substack{i \in MM \\ j \in QM}} q_i q_j (\rho_{QM})$$

Atomic Charge Schemes

- "Atoms" are not well-defined in molecules – there is no quantum mechanical operator corresponding to an atom.
- This leads to ambiguity in the definition of an atomic charge
- Population Analysis Schemes
  - Basically, sum over all electrons using the basis functions of a given atom
  - Depends on the atom-centered nature of the basis set
  - Breaks down as the basis functions become more delocalized – results do not usually converge with increasing basis set!

Charge Schemes

- Atoms-in-molecules
  - Atoms are defined by “critical points” of the charge density
  - More stable than Mulliken/Lowdin schemes with respect to basis set expansion
  - Implemented in Gaussian
  - Not clear whether stable=“correct”



Charge Schemes

- ESP-Fitting
  - Determine charges which reproduce the electrostatic potential generated by the molecule
  - If using charges in an MM potential, this appears to be the right way
  - But, equations have many solutions, especially when molecule has an “interior”

Charge for solvated ion will be essentially undetermined

Charge Schemes

- Restricted ESP-Fitting (RESP)
  - Attempts to avoid unphysical solutions of ESP-charges
  - Requires user guidance in imposing “reasonable” values of charges

Boundary Treatment

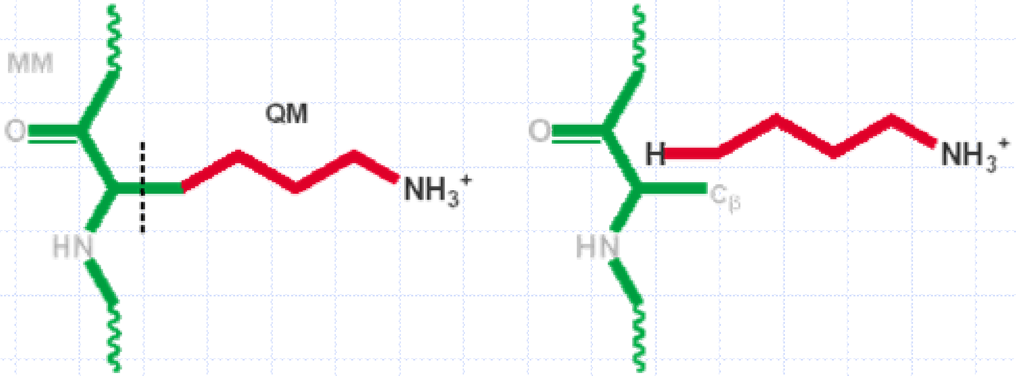
- How do we deal with bonds between the QM and MM regions?
  - The valence of the QM region must be satisfied
  - MM bond, angle, dihedral terms need a partner atom to act on, in order to maintain the geometry of the system
- QM/MM is often used to simulate a solute quantum mechanically, with explicit solvent treated with MM — in this instance, the problem of QM-MM bonds is avoided

Covalent Embedding

- Most difficult embedding – cutting across covalent bonds
- Almost always required in biological context
- Many strategies; still not clear which is best or whether any of them "work"

Singh & Kollman (1986)

E_{QM}: HF E_{MM}: AMBER $E_{QM/MM}$: Link atom



Link Atoms

- Conventional solution: ‘link atoms’ (usually hydrogen atoms, but sometimes halogens or even methyl groups) are added along the bond [a]
  - The link atom satisfies the valence of the QM region
  - The QM atom is used for calculation of all MM bond terms
  - For nonbond (electrostatic terms), originally the link atom did not interact with any MM atom (termed a ‘QQ’ link in CHARMM parlance)
  - Better properties are usually obtained if the link atom interacts with the entire MM region (‘HQ’ link)
  - Poor handling of electron density

---

[a] Singh, U. and Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718

Covalent Embedding

- Potential Problems with Link Atom Idea
  - Extra degrees of freedom which somehow need to be removed; i.e. the link atom somehow needs to be connected to the MM part of the simulation
  - Electronic structure at boundary will be very different if H and the atom it replaces do not have similar electronegativities

Covalent Embedding

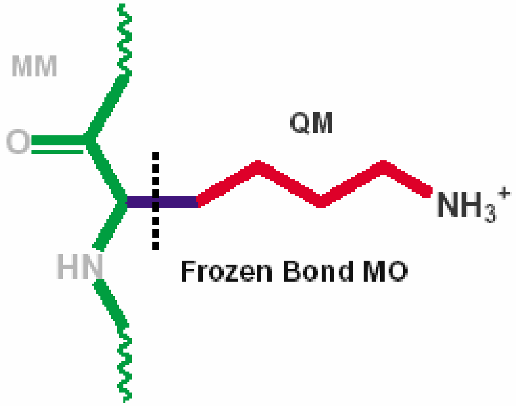
- Thiel
  - Adjust electronegativity of link atom to be equivalent to target atom. Also adjust size of atom
  - Can only do this easily with semiempirical models
  - Still can cause problems, especially with electronically excited states – the 2s-3s transition of H-like atom is much lower than the 1s-2s transition!

Covalent Embedding

- Frozen orbital ideas:

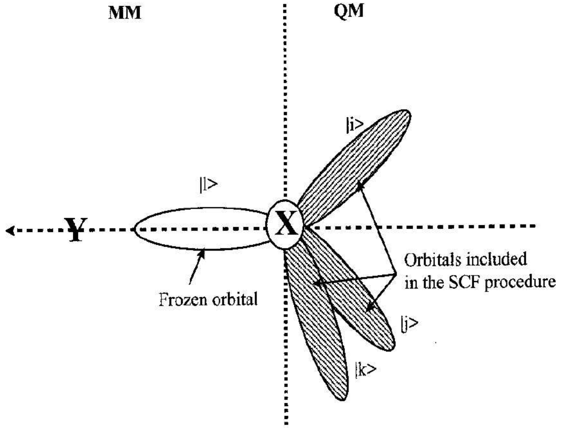
Rivail & co-workers (1994)

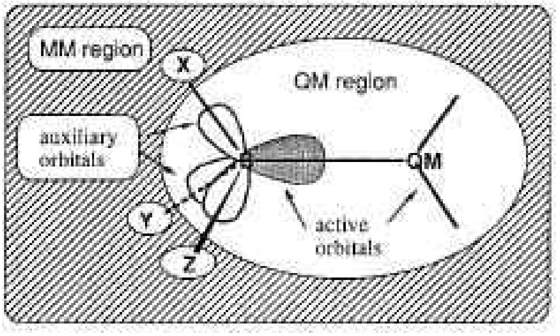
E_{QM}: AM1 E_{MM}: AMBER $E_{QM/MM}$: **Hybrid MO**



Improved Bond Treatments

- Local Self-Consistent Field (LSCF) [a] uses a parameterized frozen orbital along the QM-MM bond, which is not optimized in the SCF
- Generalized Hybrid Orbital (GHO) [b] includes the QM-MM orbitals in the SCF





---

[a] Warshel, A. and Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227

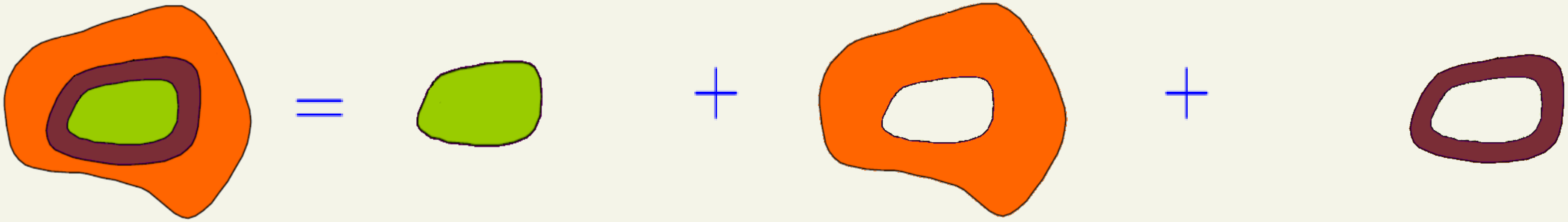
[b] Gao, J. *et. al. J. Phys. Chem. A* **1998**, *102*, 4714

General QM/MM Methodology

Two main strategies:

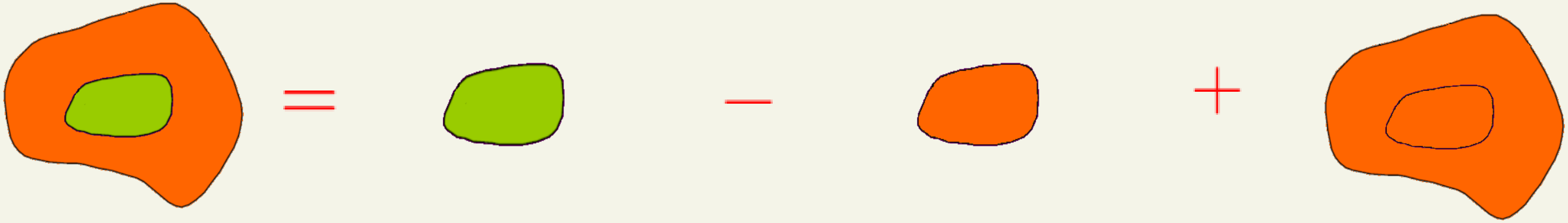
- Additive Method

$$H_{tot} = H_{QM}(QM) + H_{MM}(MM) + H_{QM/MM}(Inter.)$$



- Subtractive Method

$$E_{tot} = E_{QM}(QM) - E_{MM}(QM) + E_{MM}(All)$$



Other Approaches

- ONIOM [a] divides the system into the ‘real’ (full) system and the ‘model’ (subset) and treats the model at high level, and the real at low level, giving the total energy as

  $$E(high, real) \simeq E(low, real) + E(high, model) - E(low, model)$$

  which relies on the approximation

  $$E(high, model) - E(low, model) \simeq E(high, real) - E(low, real)$$

- The ‘model’ system still has to be properly terminated
- Extension to three level systems is relatively straightforward (e.g. *ab initio* core, semi-empirical boundary, MM surroundings)

---

[a] Svensson, M. *et. al. J. Phys. Chem.* **1996**, *100*, 19357

Other Approaches

- Empirical Valence Bond method [a] treats any point on a reaction surface as a combination of two or more valence bond structures
    - Parameterization is made from QM or experimental data
    - An effective method, but must be carefully set up for each system
- Effective Fragment Potential [b] adds ‘fragments’ to a standard QM treatment, which are fully polarizable and are ‘parameterized’ from separate *ab initio* calculations
    - Treatment of bonds between the ‘true’ QM region and the fragments is still problematic

---

[a] Warshel, A. and Weiss, M. *J. Am. Chem. Soc.* **1980**, *102*, 6218

[b] Webb, P. and Gordon, M. *J. Phys. Chem.* **1999**, *103*, 1265

Summary of Current Approaches

Karplus & co-workers

E_{QM}: DFT, HF, AM1 E_{MM}: CHARMM $E_{QM/MM}$: Link atoms
CHARMM interface with GAMESS or CADPAC

Friesner & co-workers

E_{QM}: DFT, HF E_{MM}: OPLS-AA $E_{QM/MM}$: Hybrid Orbital
Qsite (Macromodel interface with Jaguar)

Gao & co-workers

E_{QM}: AM1 E_{MM}: CHARMM $E_{QM/MM}$: Hybrid Orbital*

Yang & co-workers

E_{QM}: HF, DFT E_{MM}: CHARMM $E_{QM/MM}$: Link atom*

*Specific parameterization allows bond length change

Dynamics

- Chemical reactions are often simulated by molecular dynamics, e.g. with umbrella sampling
- Dynamics of a QM/MM system are almost identical to those of an MM system:
  - Forces are calculated from first derivatives on each atom
  - The QM nuclei are treated identically to the MM partial charges
  - The system is propagated by standard Newtonian dynamics

Monte Carlo

- QM/MM can also be used in conjunction with Monte Carlo methods
- A complication: the MM atoms affect the QM electron density, so an SCF is required for every Monte Carlo move
- Workaround: approximate the energy change of the QM region by first-order perturbation theory ('Perturbative QM/MC') as long as moves are far enough away from the QM region [a]

---

[a] Truong, T. and Stefanovich, E. *Chem. Phys. Lett.* **1996**, *256*, 348

Drawbacks of QM/MM

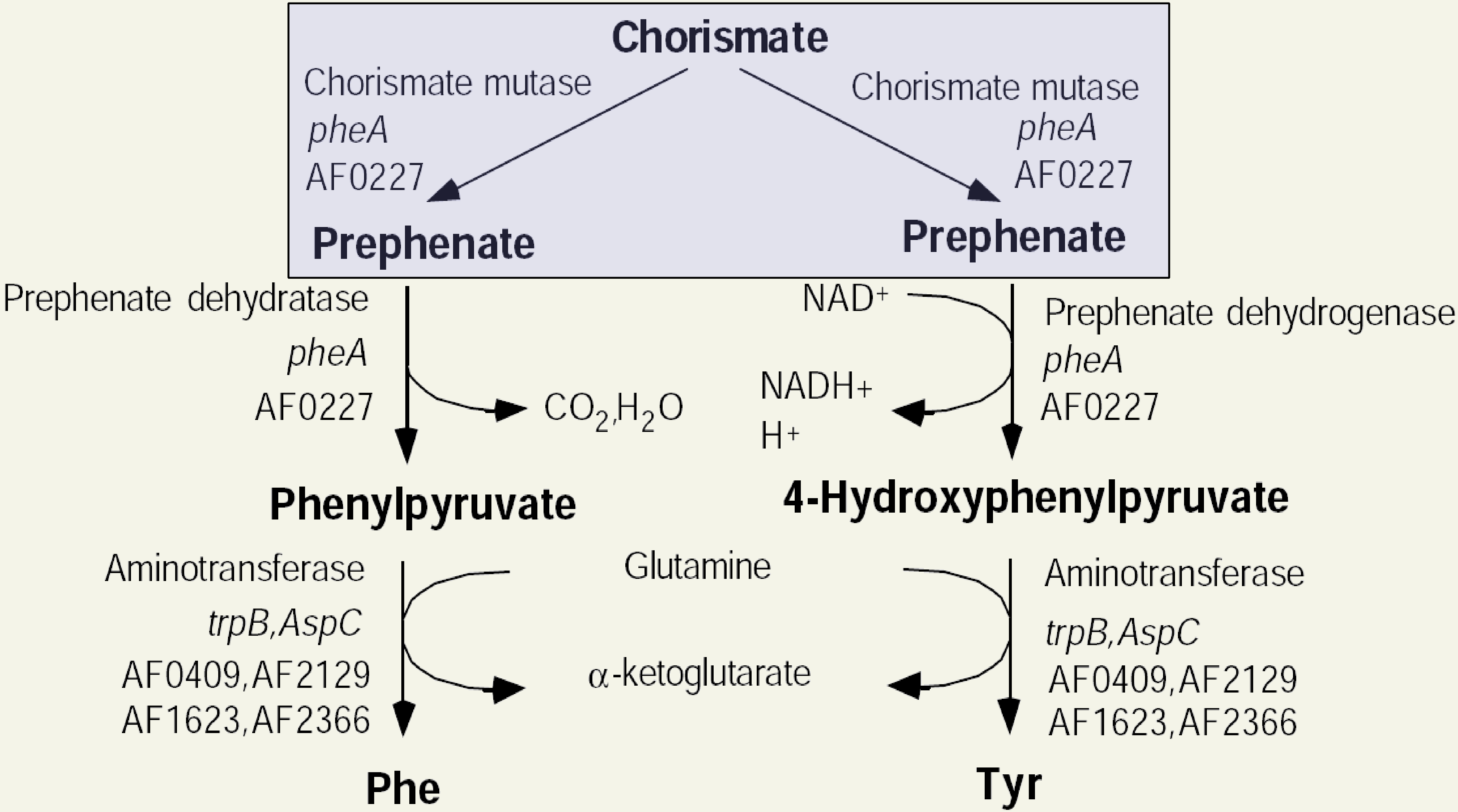
- Some parameterization is still required for the boundary treatment
- The choice of the size of the QM region is still something of an art
- Although the QM region polarizes in response to the MM partial charges, the reverse is not also true (although fully polarizable QM/MM methods are being developed)
- The free energy of a QM system can be determined via frequency calculation; however, this is rather inaccurate when applied to QM/MM systems (second derivatives are poorly determined, e.g. due to the harmonic approximation)

Cautions

- Most force fields do not include polarizability, but QM region will
- This can lead to imbalance and amplification of errors
- All covalent embedding schemes should be treated with caution – it is surely possible to break almost every implemented scheme
- One needs to test carefully the dependence of the results on the QM/MM partitioning

Chorismate Mutase

Plays a key role in the shikimate pathway of bacteria, fungi, and other higher plants



Chorismate

Chorismate mutase
pheA
AF0227

Chorismate mutase
pheA
AF0227

Prephenate

Prephenate

Prephenate dehydratase
pheA
AF0227

CO_2,H_2O

NAD^+

NADH+
H^+

Prephenate dehydrogenase
pheA
AF0227

Phenylpyruvate

4-Hydroxyphenylpyruvate

Aminotransferase
trpB,AspC
AF0409,AF2129
AF1623,AF2366

Glutamine

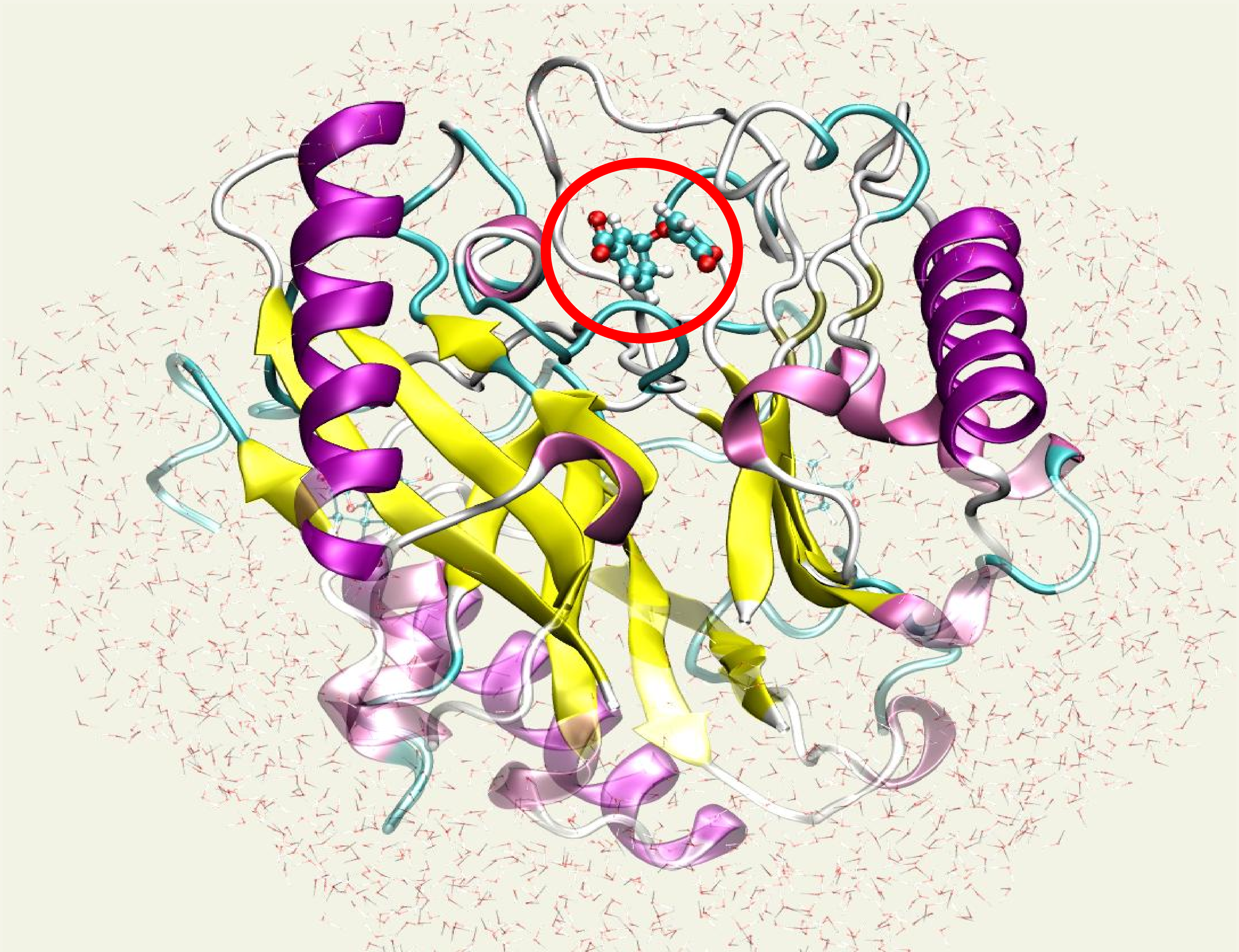
α-ketoglutarate

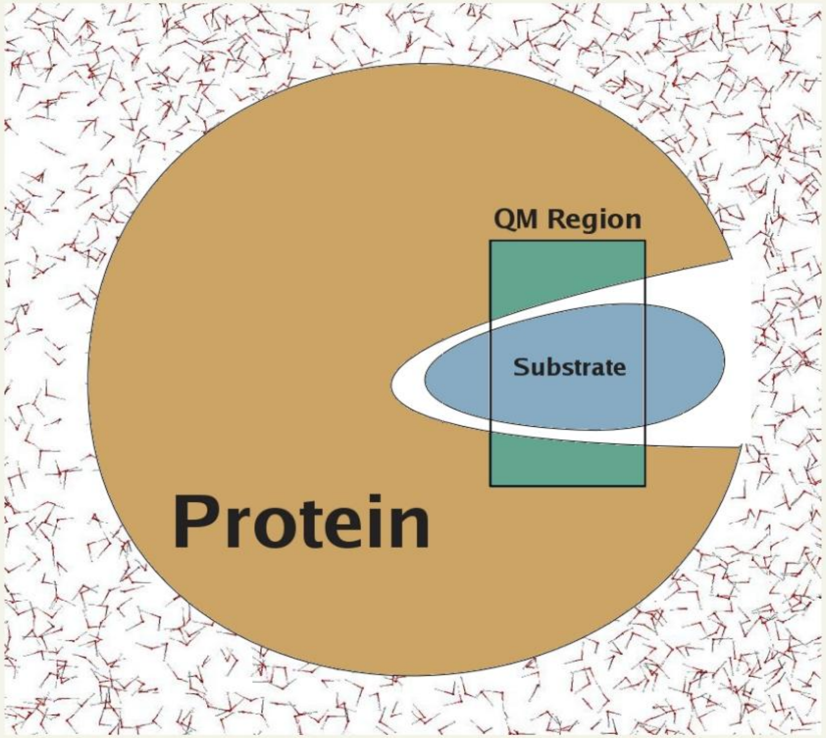
Aminotransferase
trpB,AspC
AF0409,AF2129
AF1623,AF2366

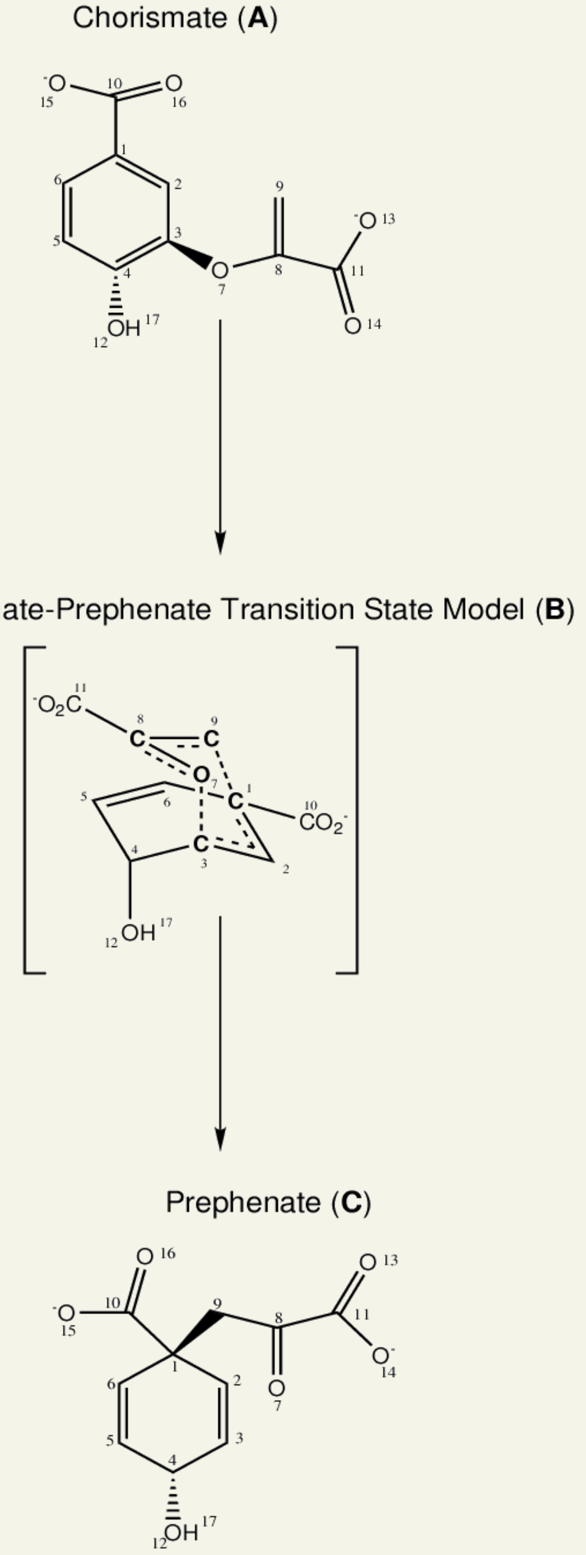
Phe

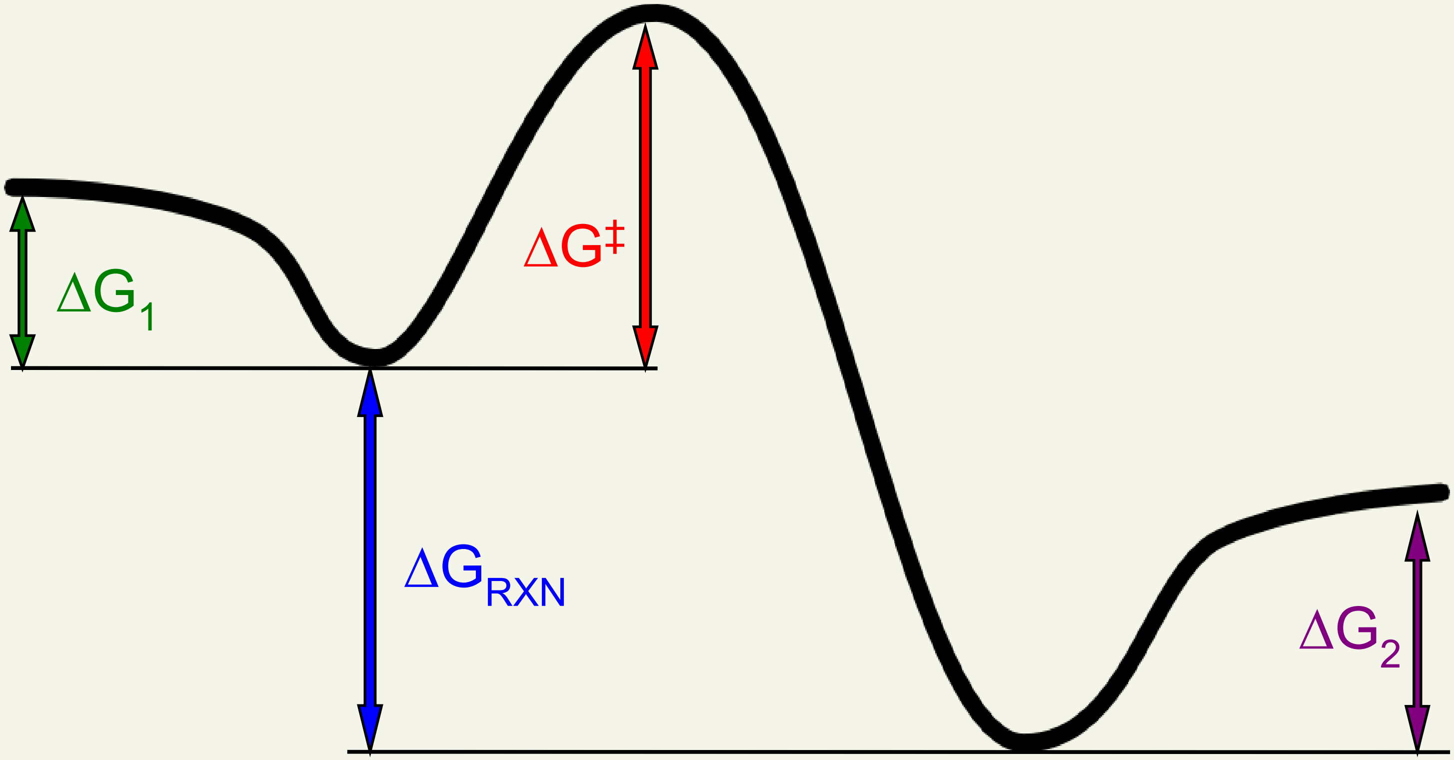
Tyr

Chorismate Mutase

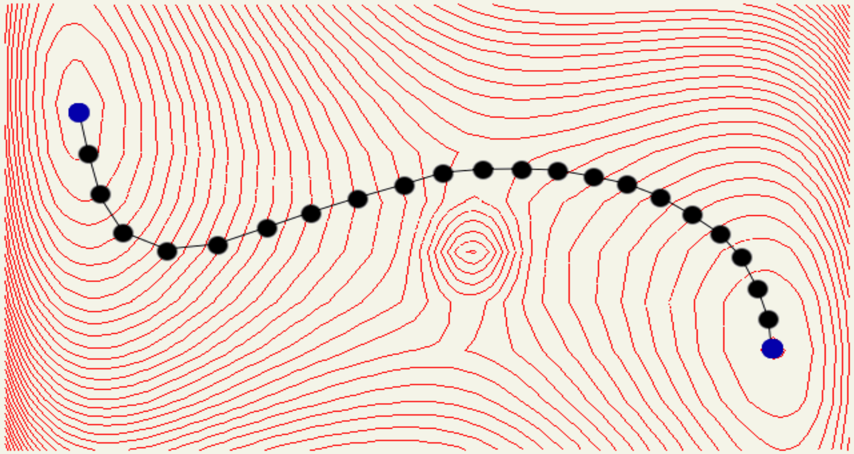



QM Region
Substrate
Protein


Chorismate (**A**)
Chorismate-Prephenate Transition State Model (**B**)
Prephenate (**C**)


ΔG_1
$\Delta G^{\ddagger}$
ΔG_{RXN}
ΔG_2

Reaction Path Methods

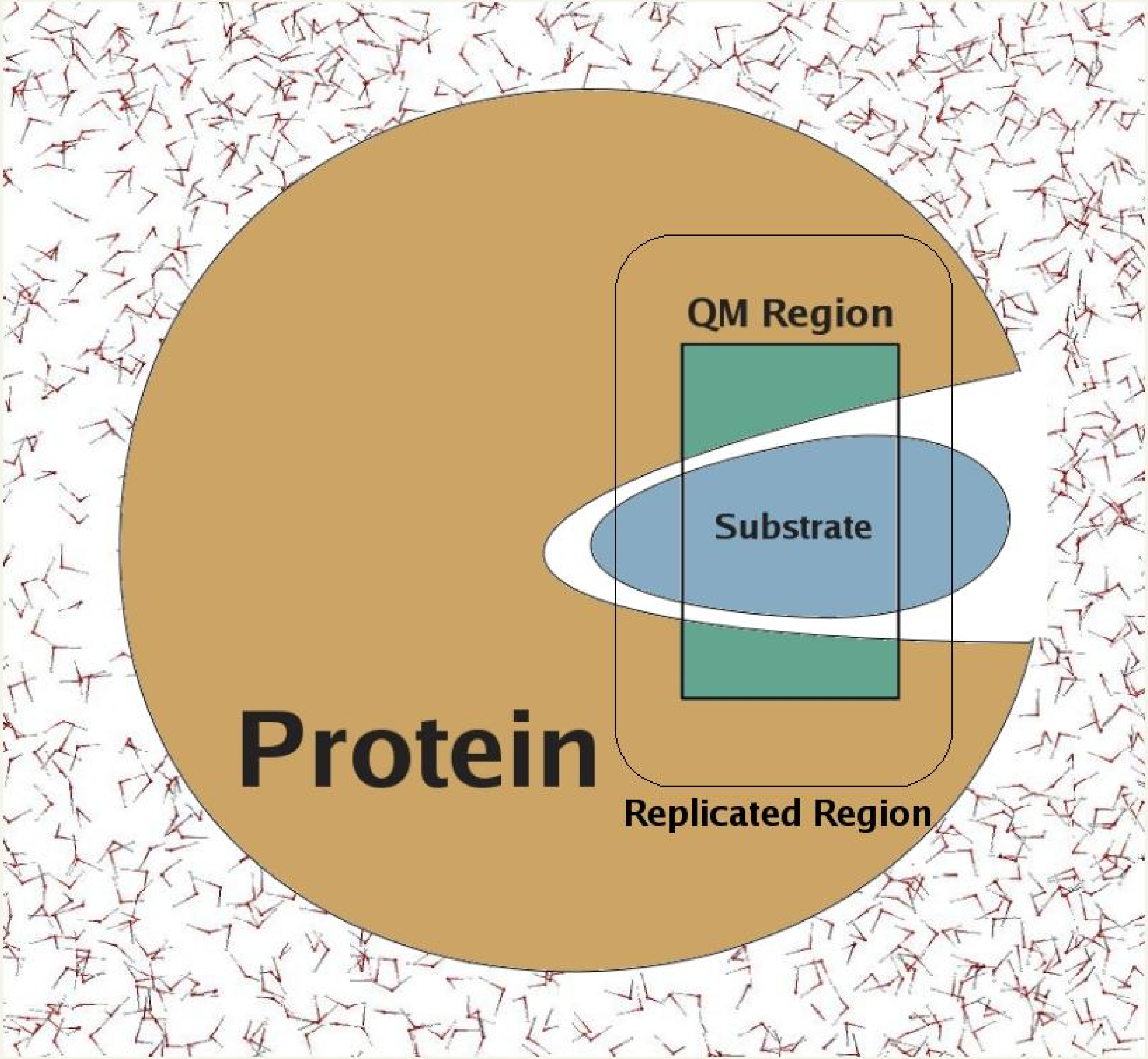


- Reaction Coordinate Driving:
  - Predetermined reaction coordinate
  - Usually some linear combination of distances
  - Gradually changed
- Cons:
  - Difficult or impossible to define reaction coordinate
  - Hysteresis: requires repeated walks to resolve
  - Sequential method: inefficient use of modern computational resources

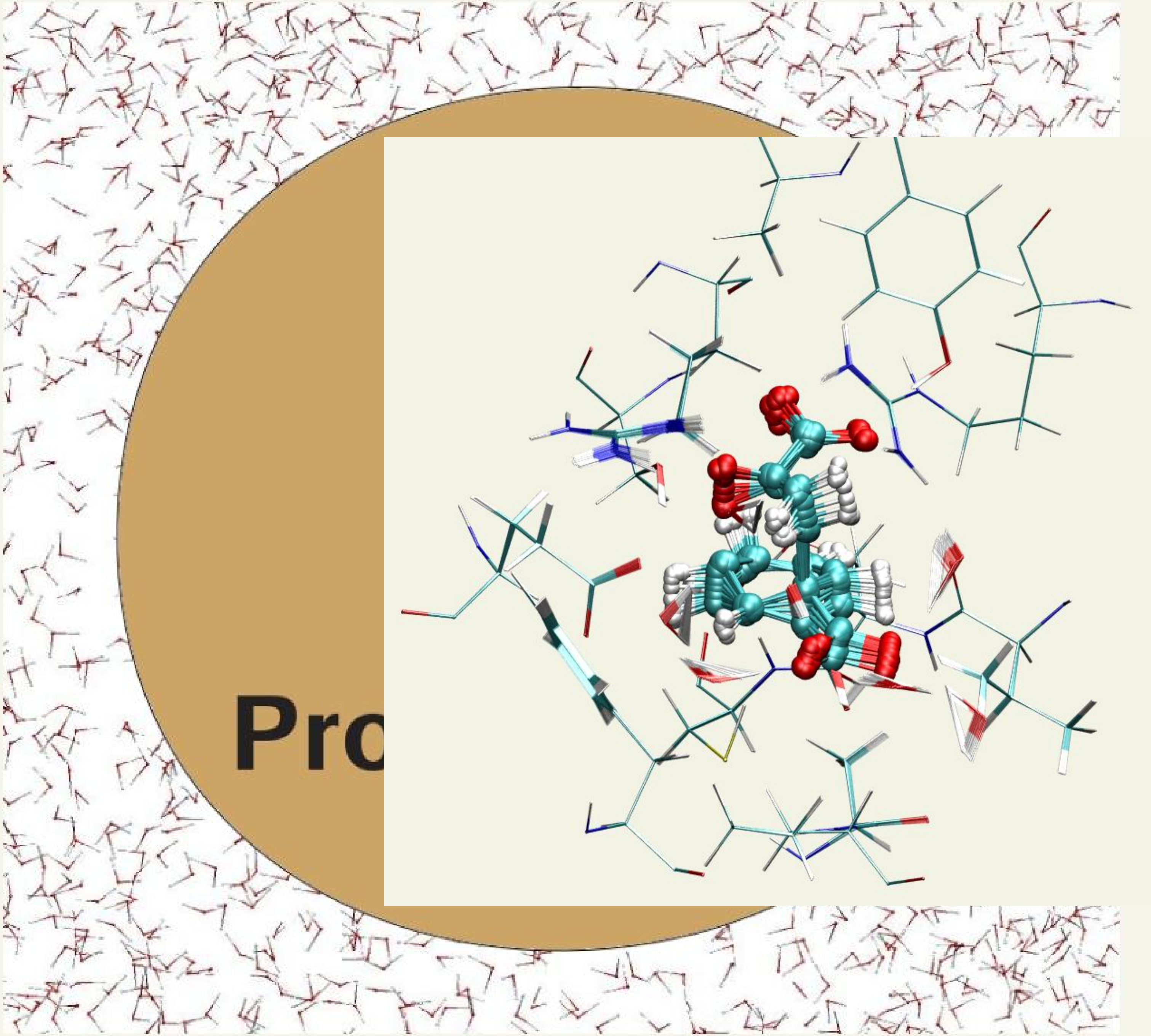
- Eigenvector Following Methods:
  - Typically require transition state to be known a priori
  - Too expensive for high dimensional systems

- Chain-of-replica Methods:
  - Path is defined as discrete structures from reactant to product
  - Removes predetermination of reaction coordinate
  - Restraints are applied to force points to be minima in all directions except path
  - Can take advantage of parallel computers (i.e. Beowulf cluster)add an outline

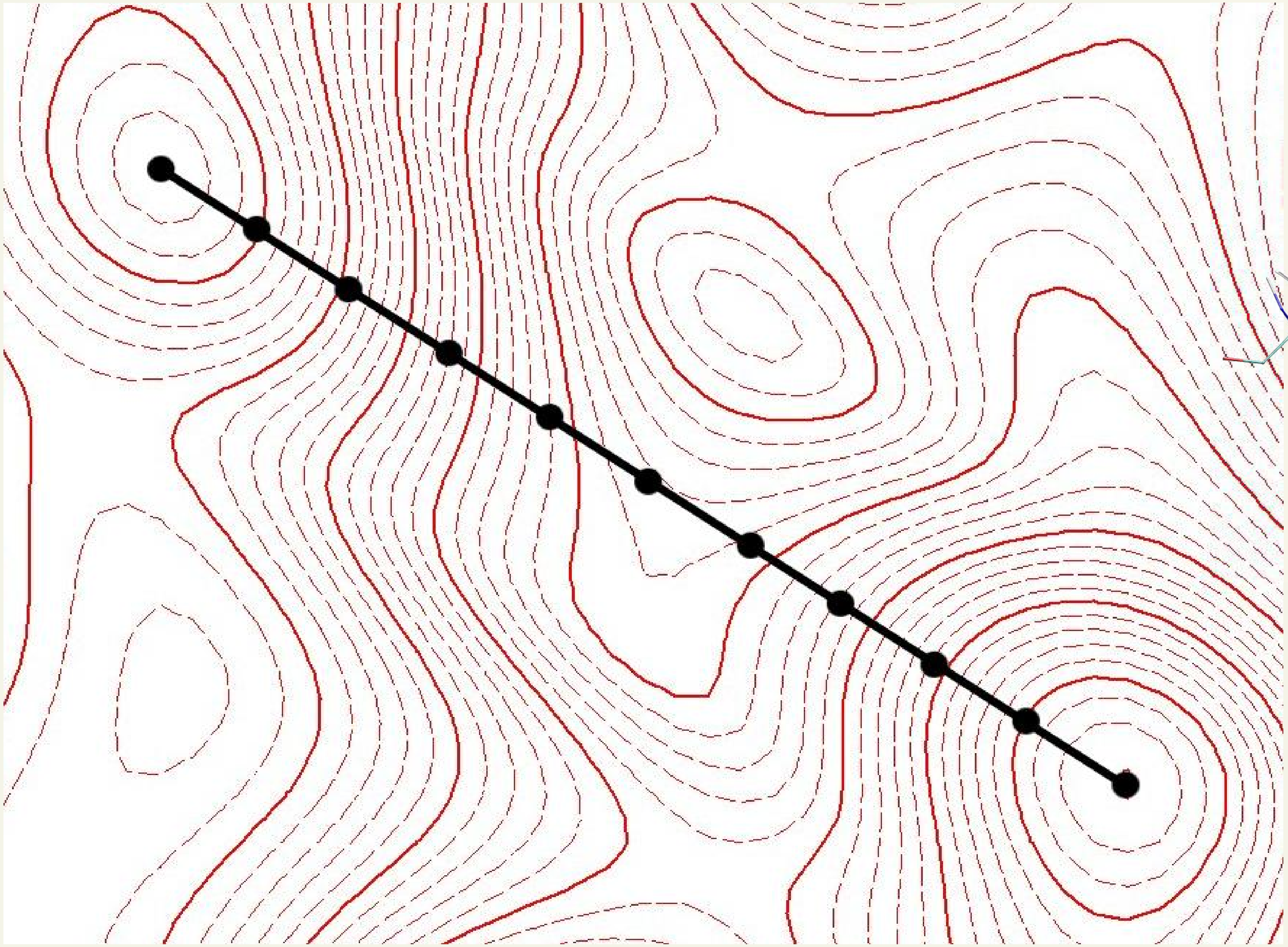
The Replica Path Method

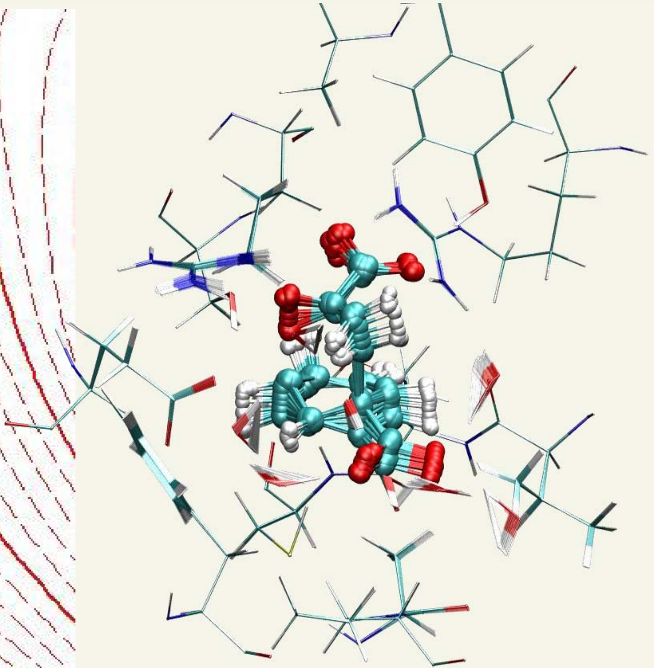


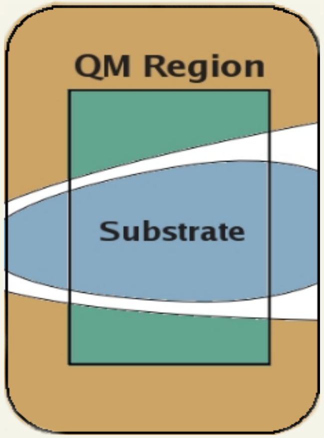
The Replica Path Method



The Replica Path Method

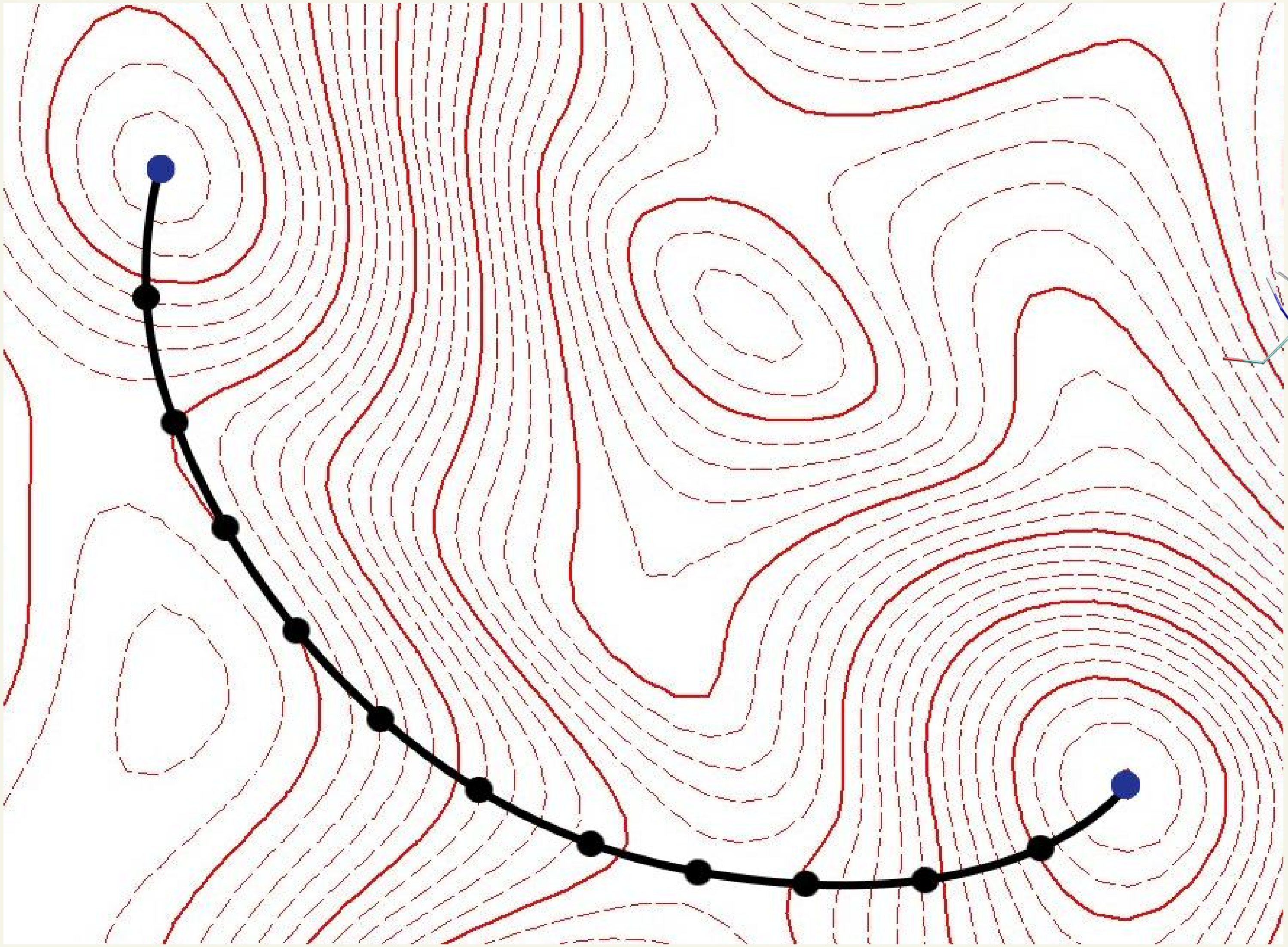


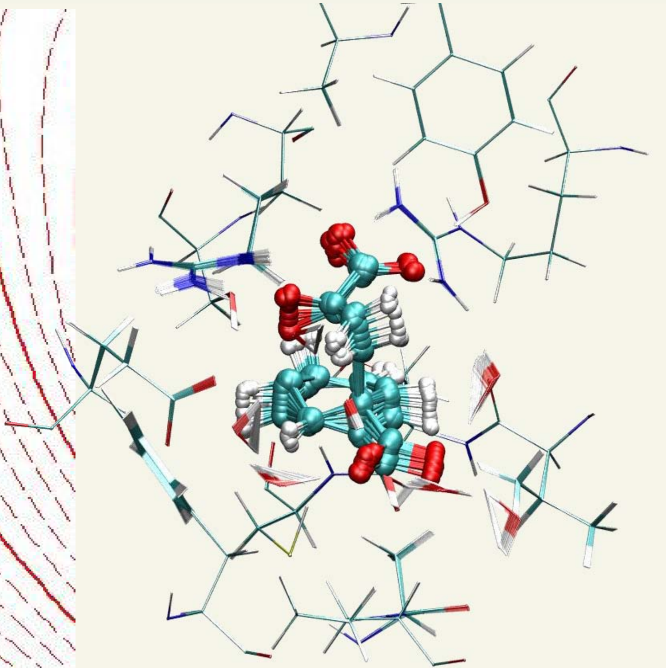


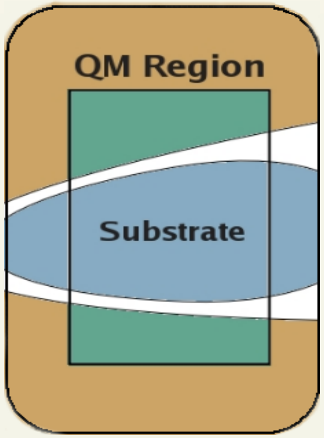


Define **X** number of steps to describe the pathway of interest

The Replica Path Method





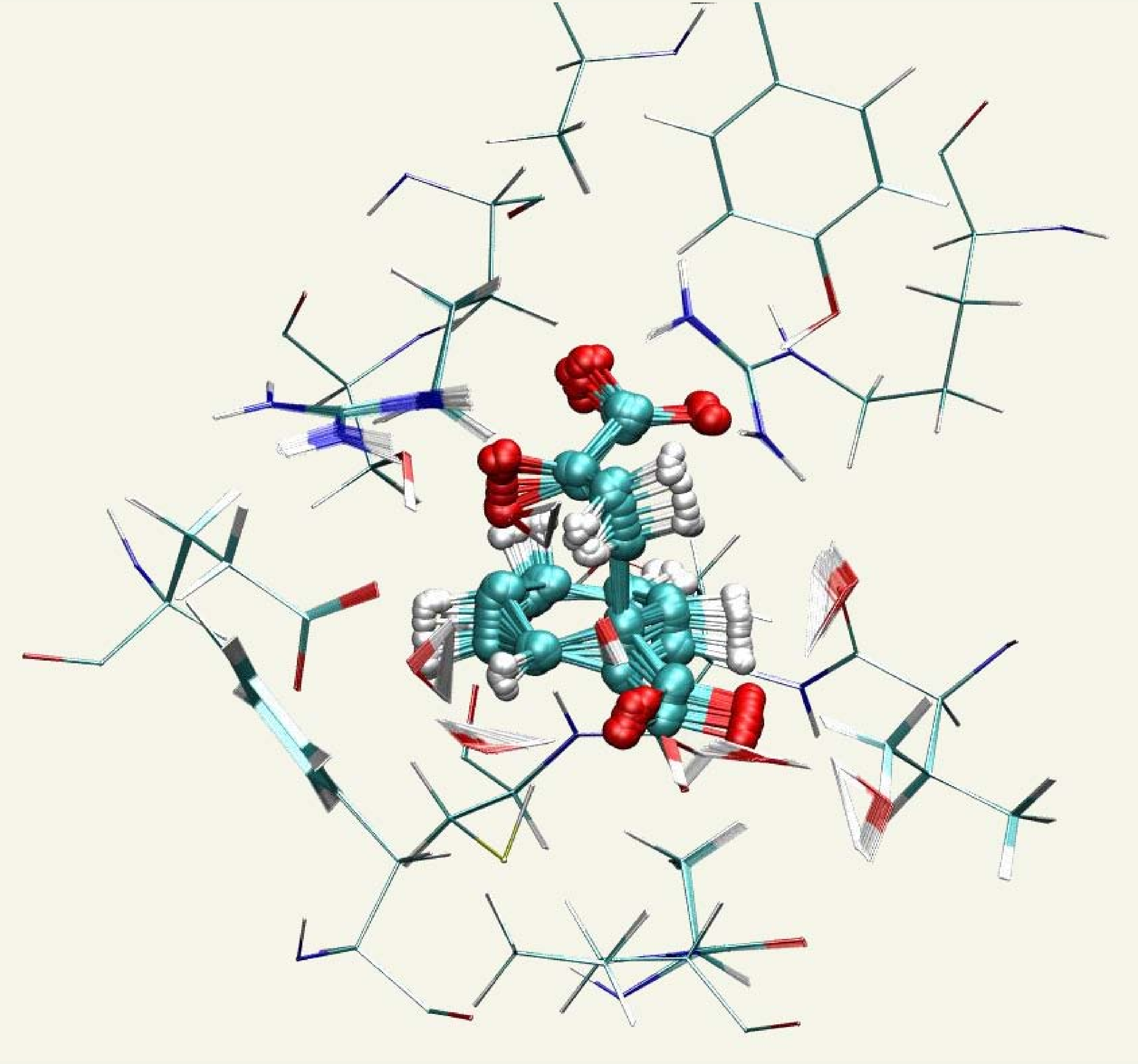


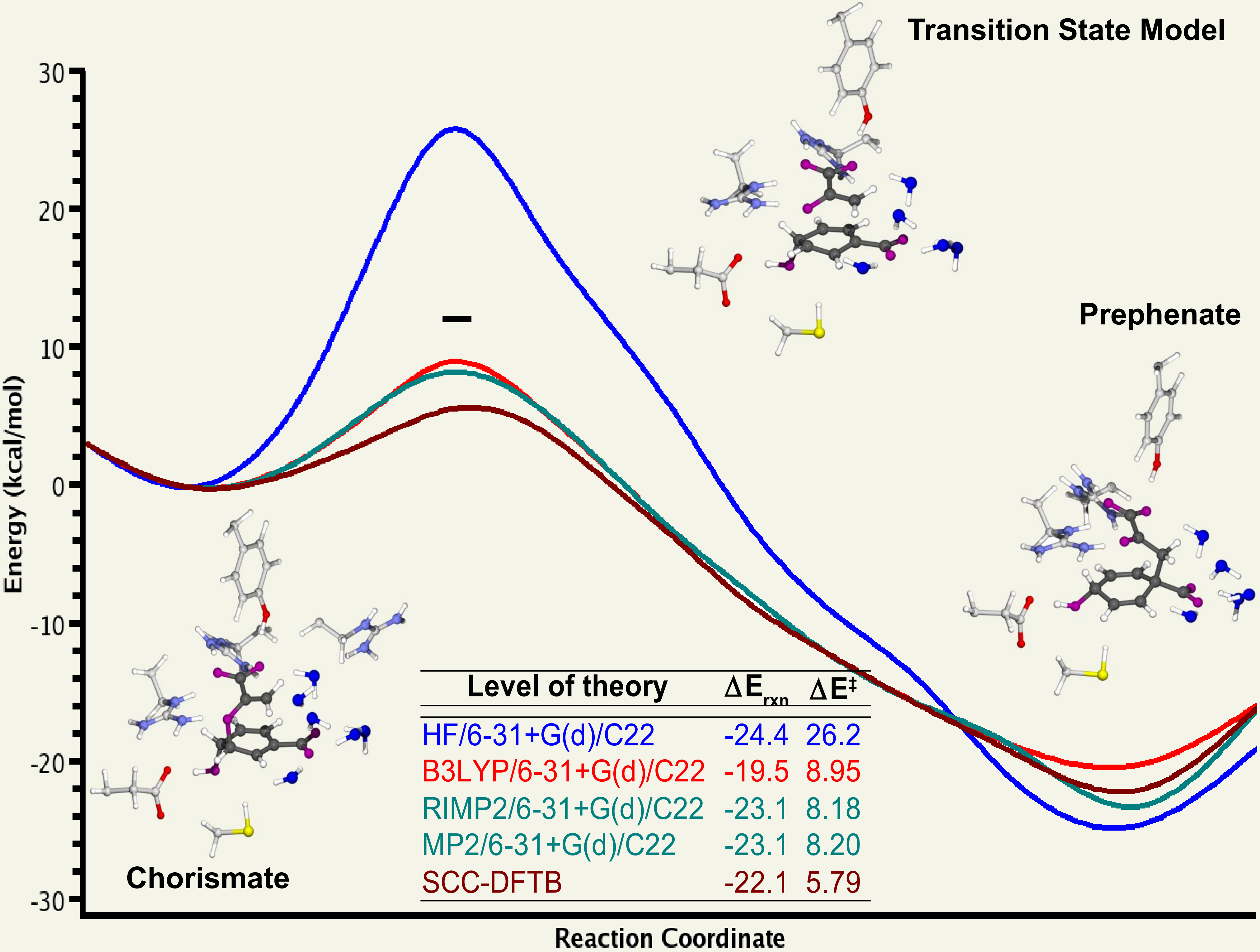
$$E_{\mathrm{RMS}} = \sum_{1}^{\mathrm{N}} \frac{1}{2} K_{\mathrm{r}} (r_{\mathrm{i}} - \bar{r})^2 \quad r_{\mathrm{i}} = RMSd_{\mathrm{bestfit}}(i, i+1) \; , \; \bar{r} = \sum_{1}^{\mathrm{N}} \frac{r_{\mathrm{i}}}{N}$$

$$RMSD(R_i, R_{ref}) = \sqrt{\frac{\sum_{\mathrm{i}}^{\mathrm{N}} (R_i - R_{ref})^2 m_i w_i}{\sum_{\mathrm{i}}^{\mathrm{N}} m_i w_i}}$$

$$E_{\mathrm{ANGLE}} = \sum_{1}^{\mathrm{N}} \frac{1}{2} K_{\phi} (\mathrm{cosmax} - \cos{(\phi_{\mathrm{i}})})^2 \quad \text{If } \mathrm{cosmax} > \cos{(\phi_{\mathrm{i}})}$$

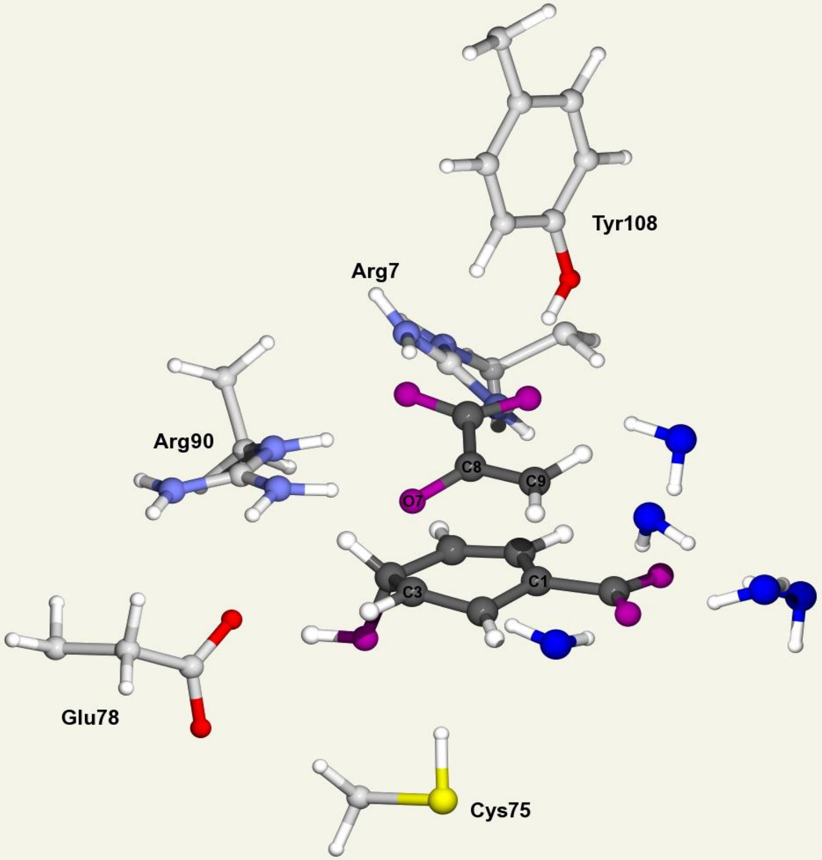
Chorismate Mutase

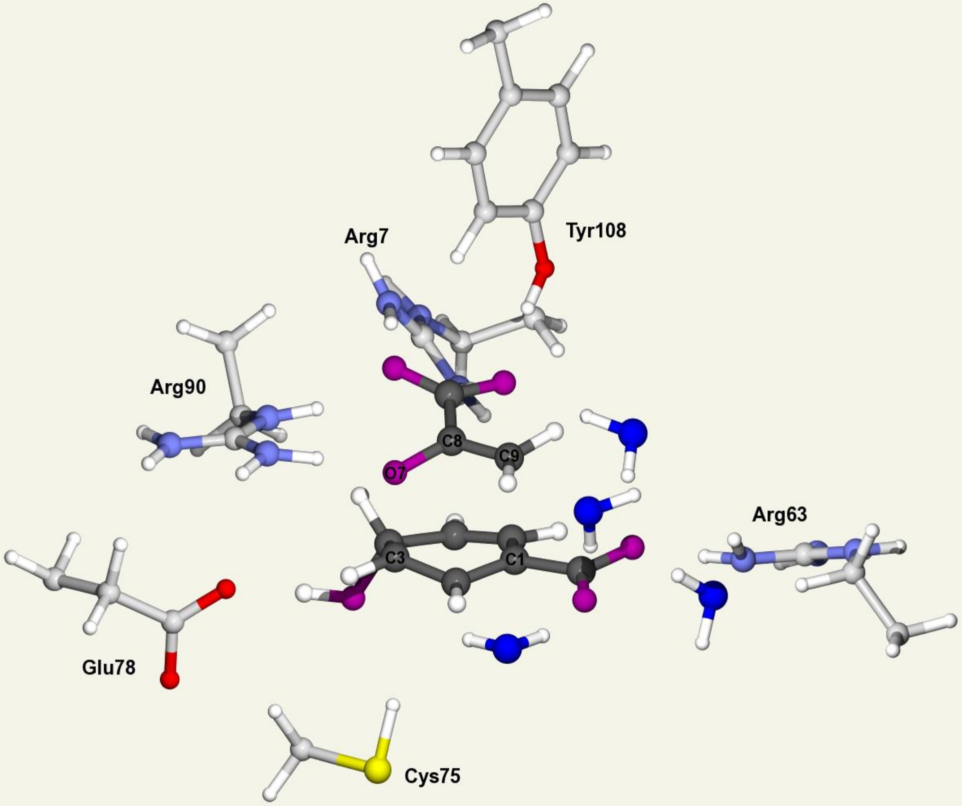


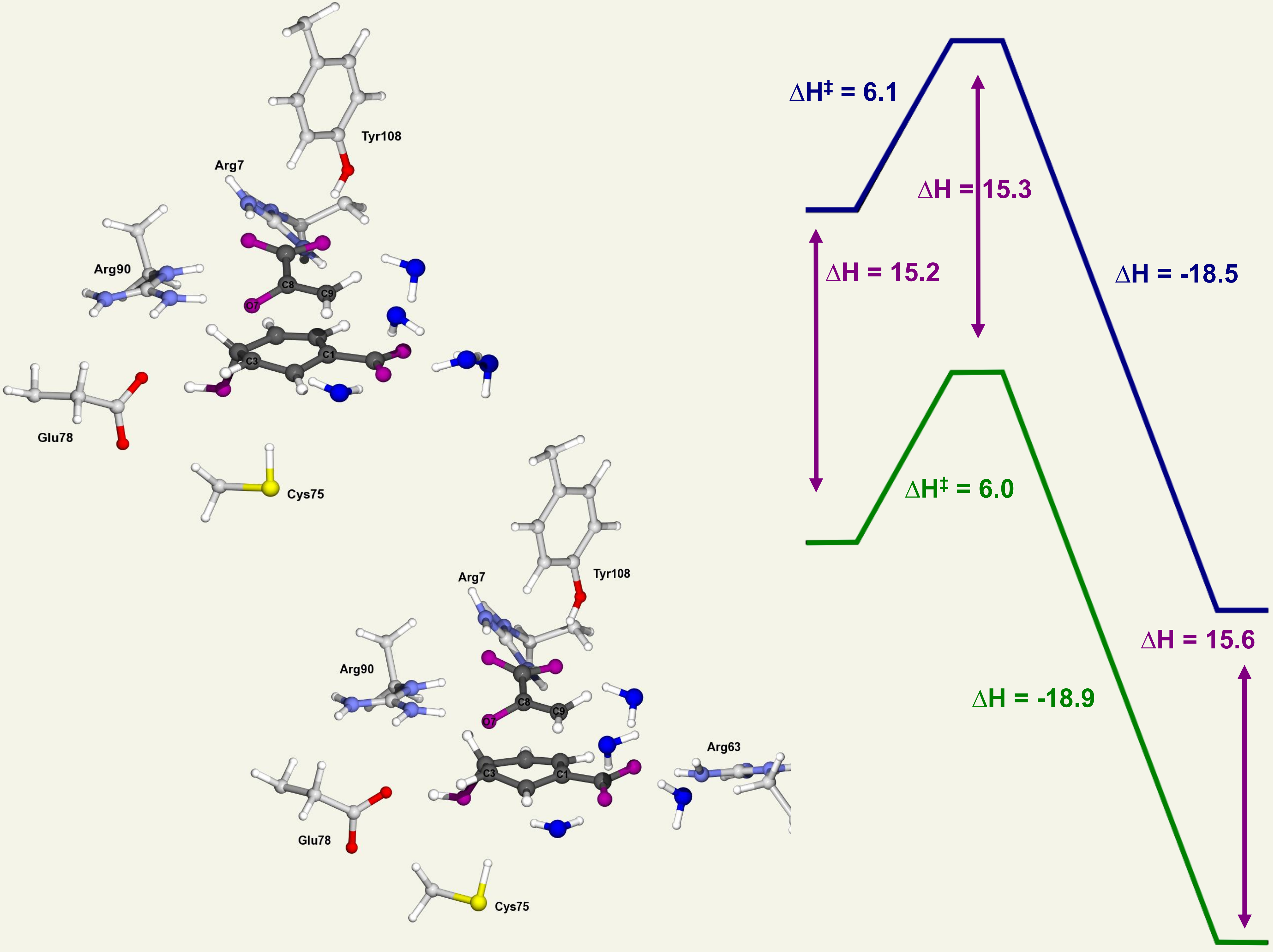


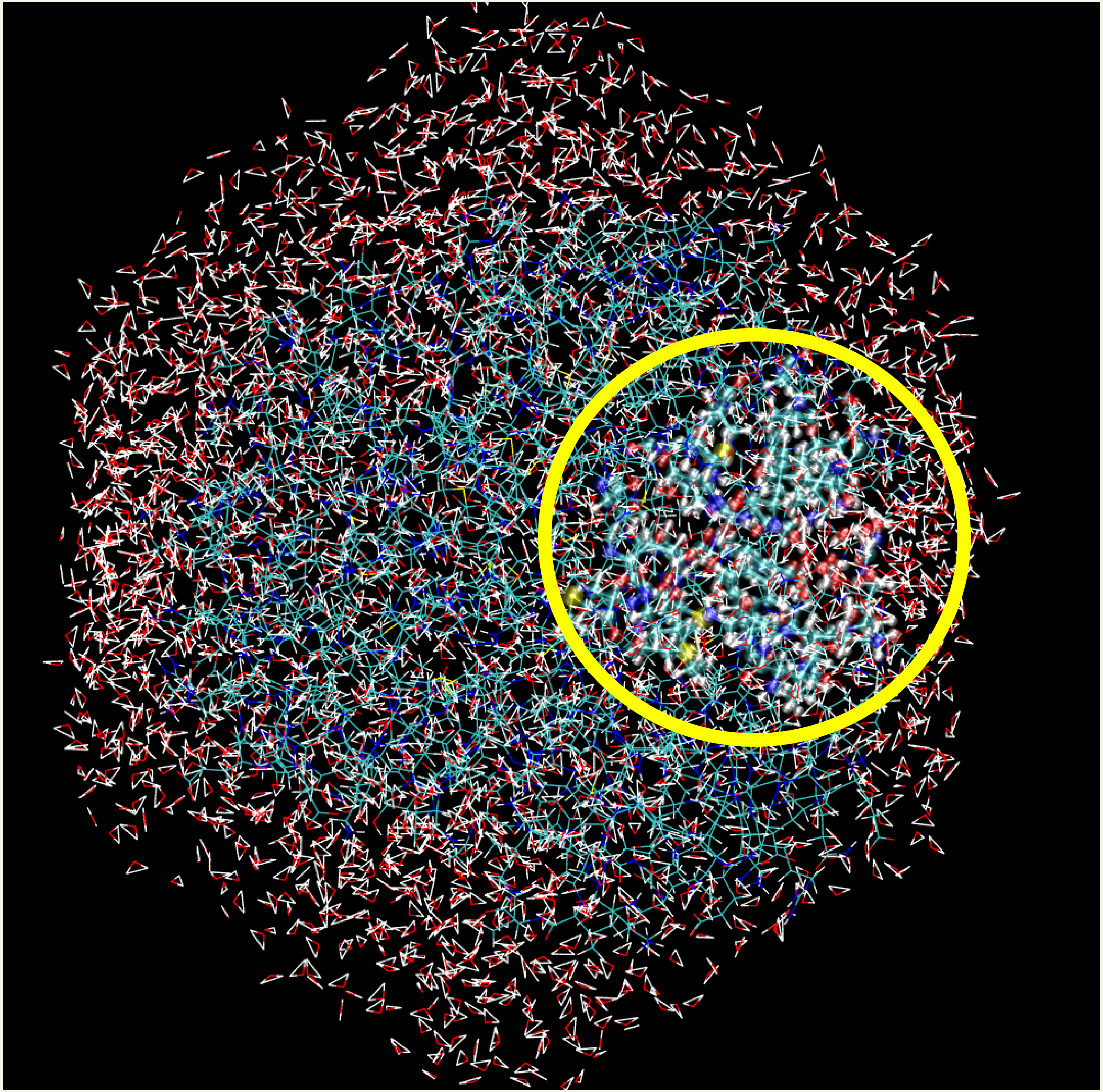
Level of theory	ΔE_{rxn}	$\Delta E^{\ddagger}$
HF/6-31+G(d)/C22	-24.4	26.2
B3LYP/6-31+G(d)/C22	-19.5	8.95
RIMP2/6-31+G(d)/C22	-23.1	8.18
MP2/6-31+G(d)/C22	-23.1	8.20
SCC-DFTB	-22.1	5.79

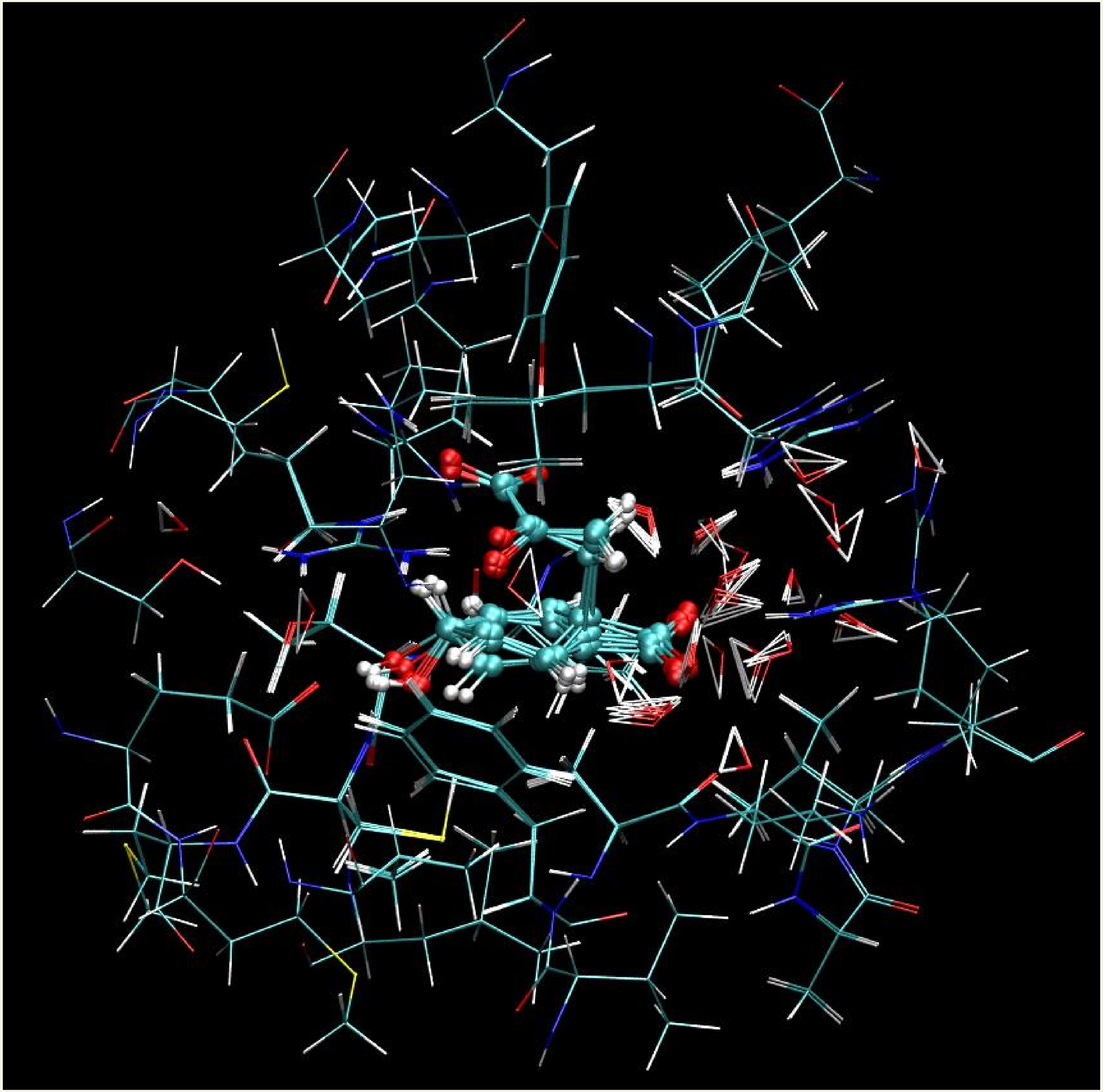
What role does Arg63 play in the reaction?

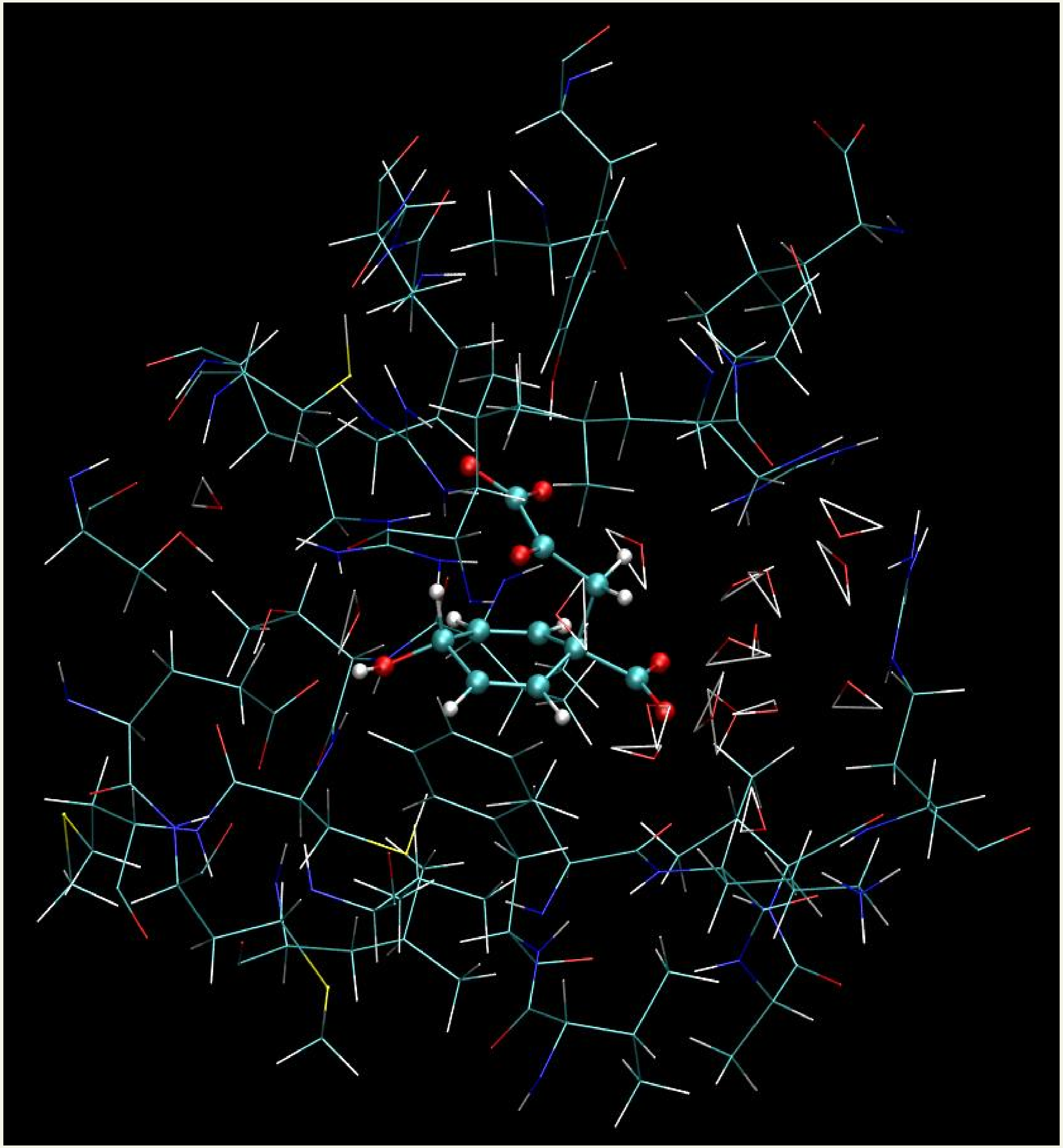


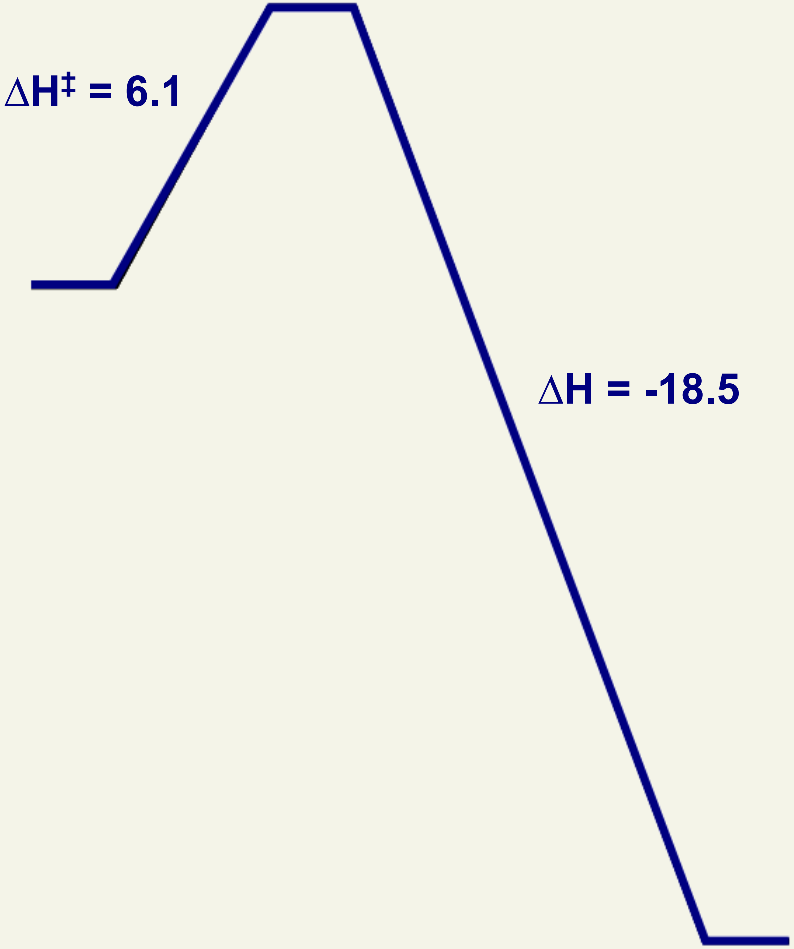


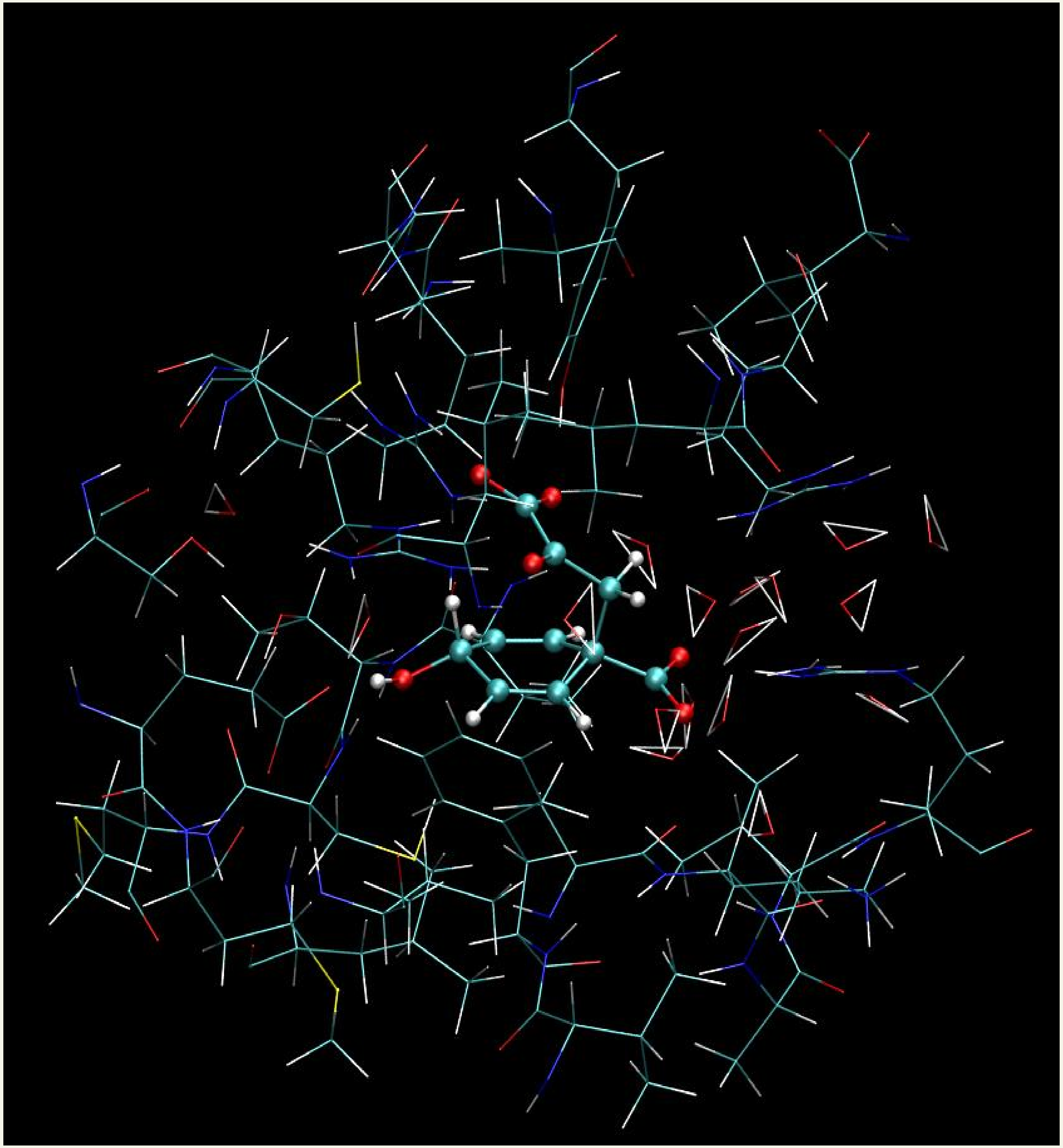

Tyr108
Arg7
Arg90
C8
C9
O7
C1
C3
Glu78
Cys75
Tyr108
Arg7
Arg90
C8
C9
O7
Arg63
C3
C1
Glu78
Cys75
ΔH‡ = 6.1
ΔH = 15.3
ΔH = 15.2
ΔH = -18.5
ΔH‡ = 6.0
ΔH = 15.6
ΔH = -18.9

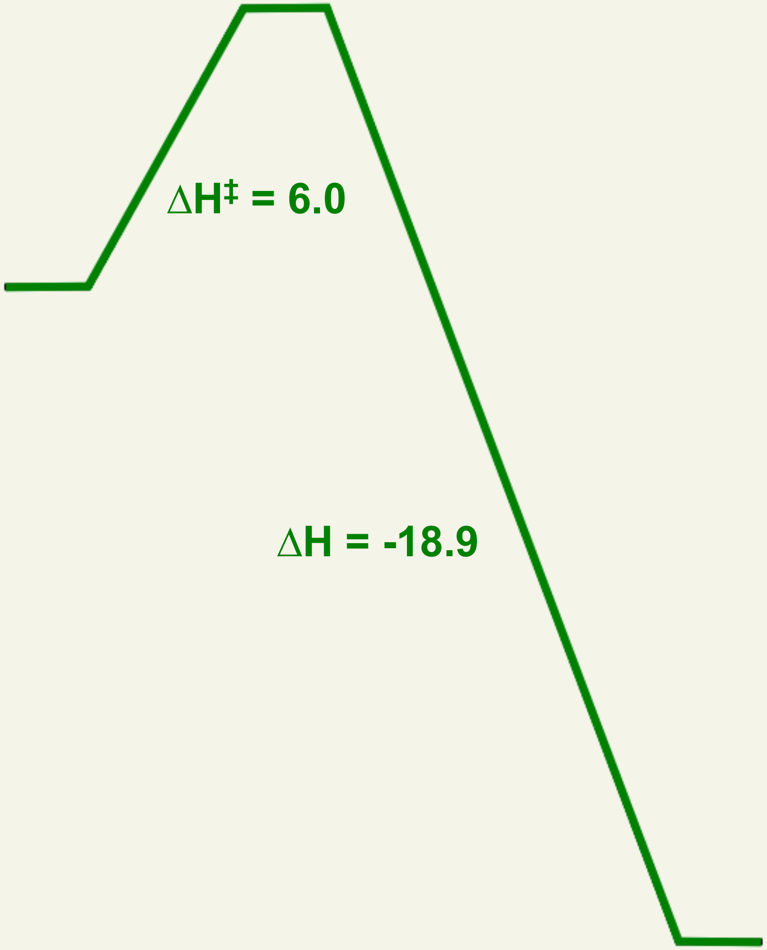







$\Delta H^{\ddagger} = 6.1$
$\Delta H = -18.5$



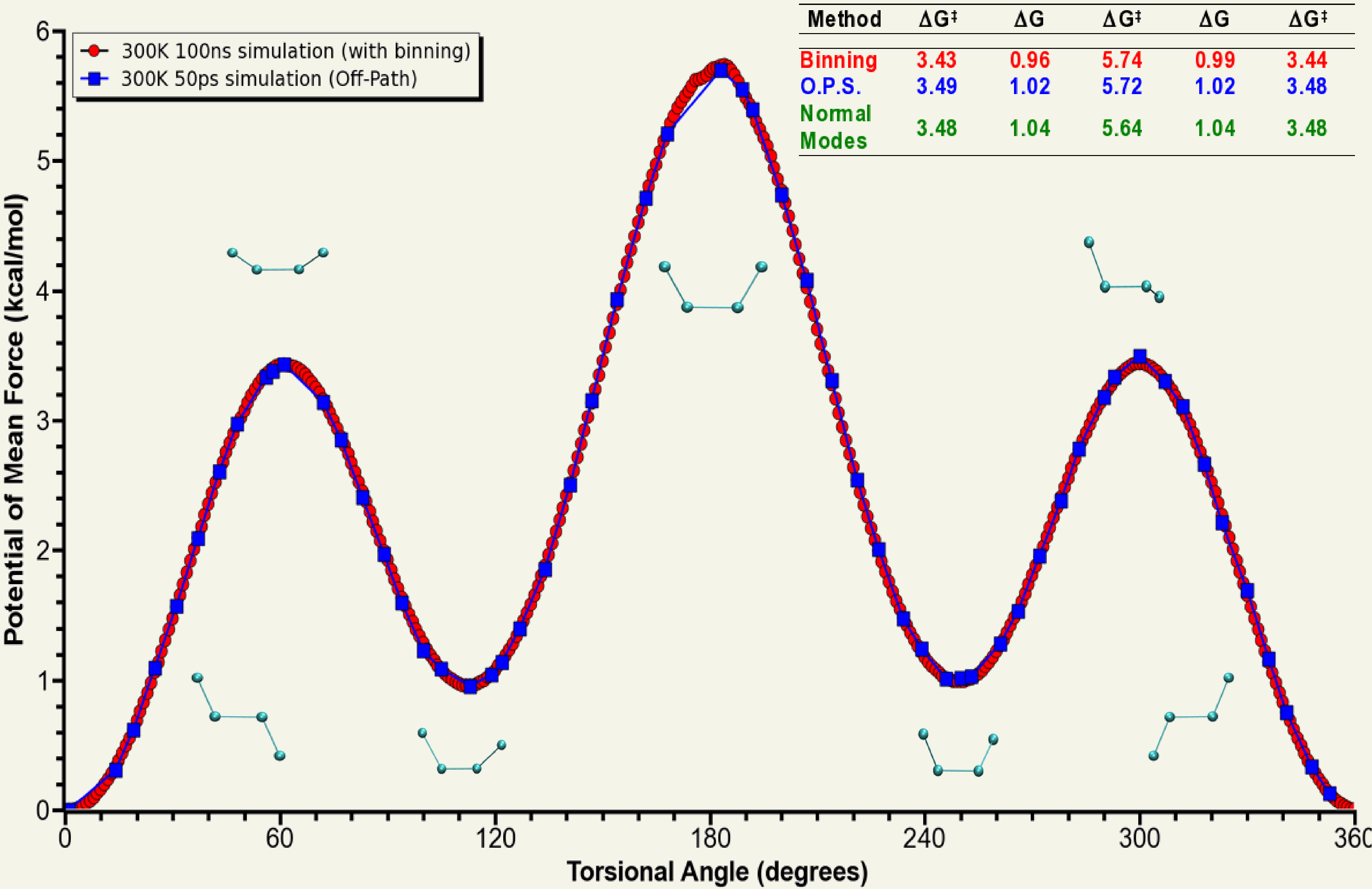

$\Delta H^{\ddagger} = 6.0$
$\Delta H = -18.9$

What Next?

- Need to compute free energies!
- Methodology?
  - Can we use the Replica Path Method?
  - Simulation methods?
  - Harmonic methods?

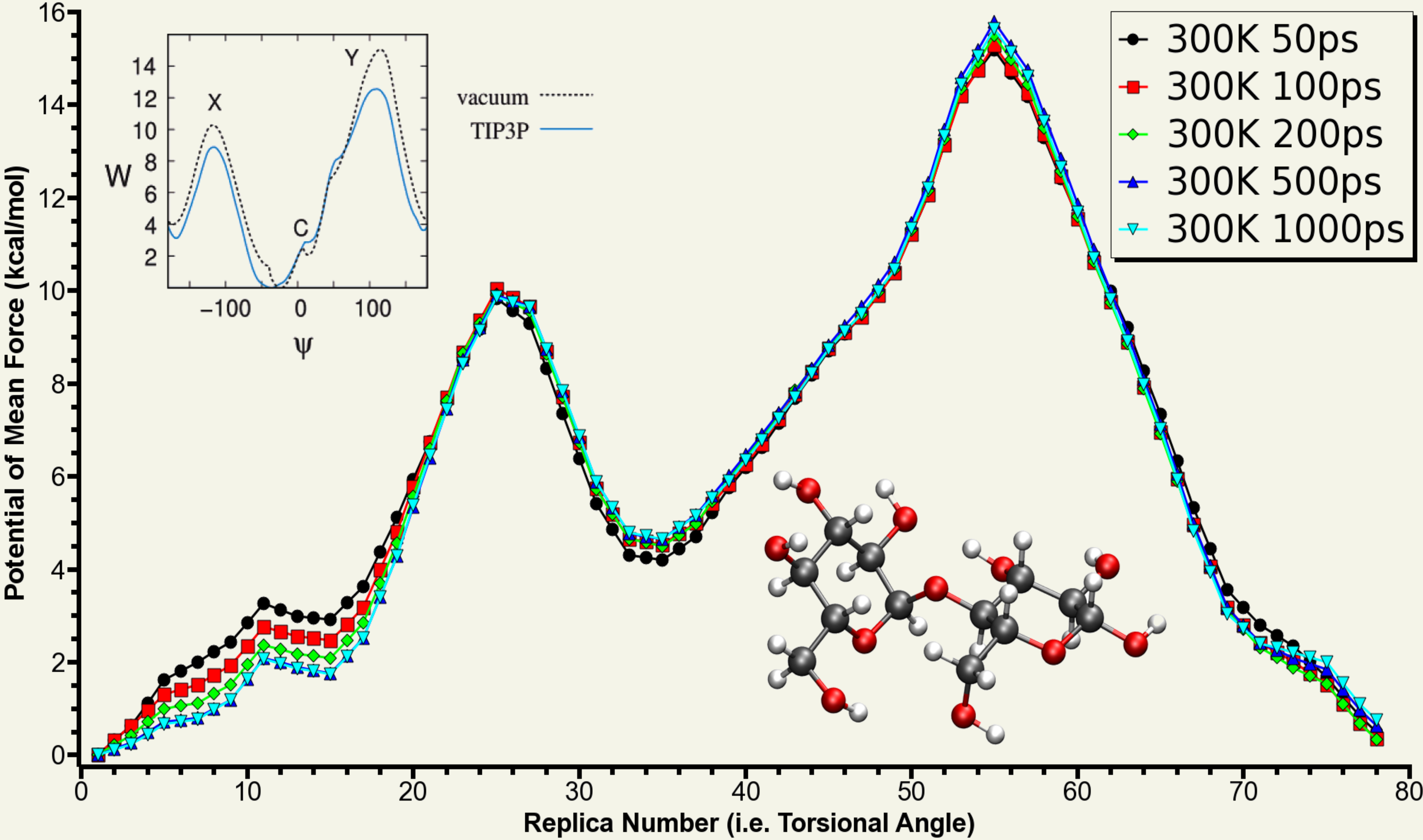
new methods to explore this...

Off-Path Simulation Results: Butane at 300K

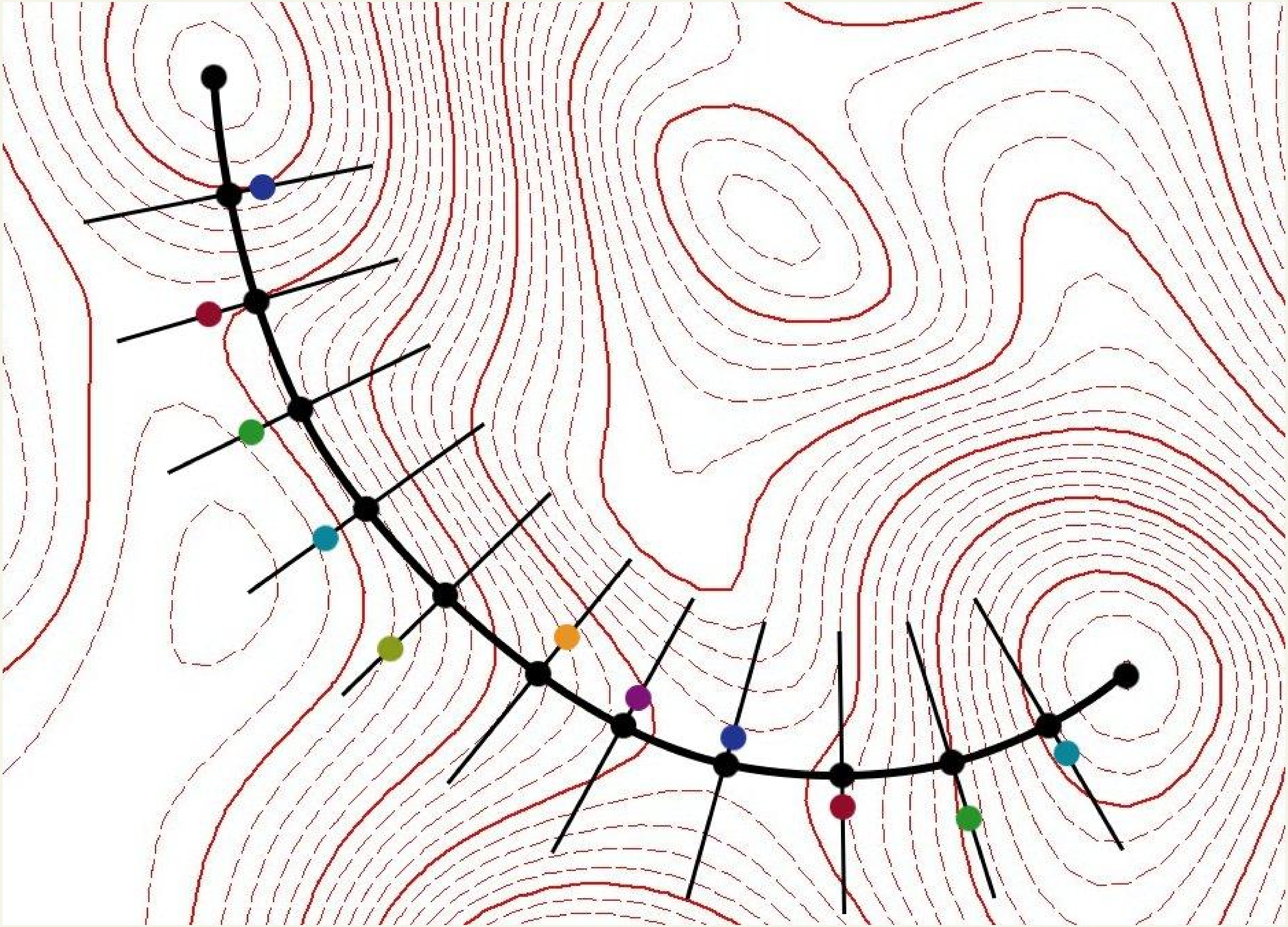


Method	ΔG‡	ΔG	ΔG‡	ΔG	ΔG‡
Binning	3.43	0.96	5.74	0.99	3.44
O.P.S.	3.49	1.02	5.72	1.02	3.48
Normal Modes	3.48	1.04	5.64	1.04	3.48

Off-Path Simulation Results: Maltose at 300K



Off-Path Simulation Method for Computing Free Energy Barriers



Conclusions

- Replica Path Method
  - Chorismate Mutase reaction profile
  - Examined methodological dependence
- Showed the role of Arg63 in Chorismate Mutase is NOT catalytic
  - Same Environment, Multiple State Method (SEMS)
- Vibrational Subsystem Analysis (VSA)
- Off-Path Simulation Method
  - Butane: quantitative agreement between OPS PMF and brute force PMF
  - Maltose: Good agreement between OPS and umbrella sampling
- Additional Developments...