



SBM CDT 2020

Computational Module

Day 3

Fernanda Duarte

Assoc. Prof Computational Organic Chemistry
University of Oxford, UK



Twitter @duarte_group

<http://fduartegroup.org>

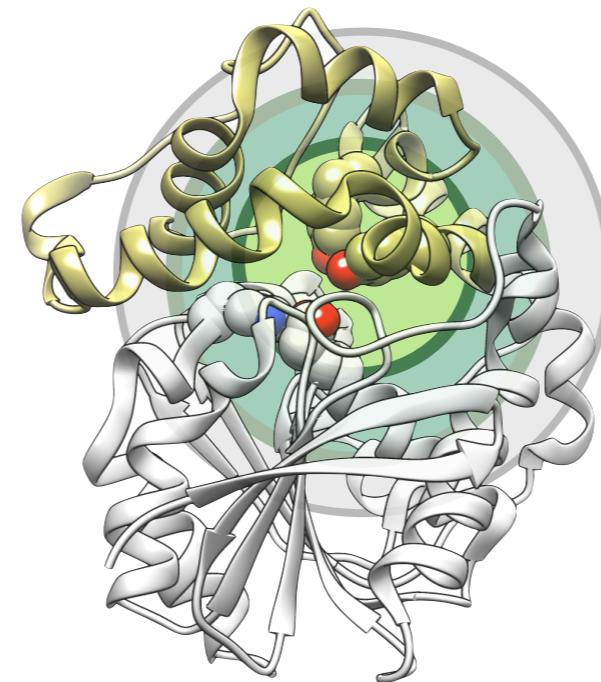


Github: duartegroup

Lecture 3

Outline

- Recap of Day 2
- Applications & Challenges in DFT
- MD and QM/MM



Basis set
Pople
Dunning
ECPs

Functional
GGA
Meta-GGA
Hybrid

Dispersion
Correction
DX



Solvent Models: Explicit vs. Implicit Solvation

Inaccuracies in DFT

... It all goes back to deficiencies in the treatment of exchange-correlation

$$E[\rho] = T[\rho] + E_{Ne}[\rho] + E_H[\rho] + E_x[\rho] + E_c[\rho]$$

- Dispersion Interactions
- Self-interaction Error
- Multi-reference problems

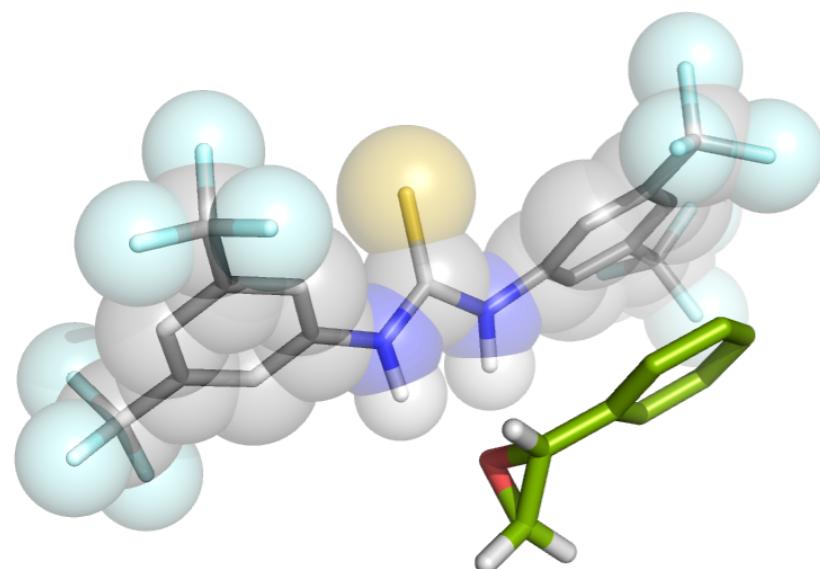
B3LYP seems to (luckily) cancel out the first two

Inaccuracies in DFT

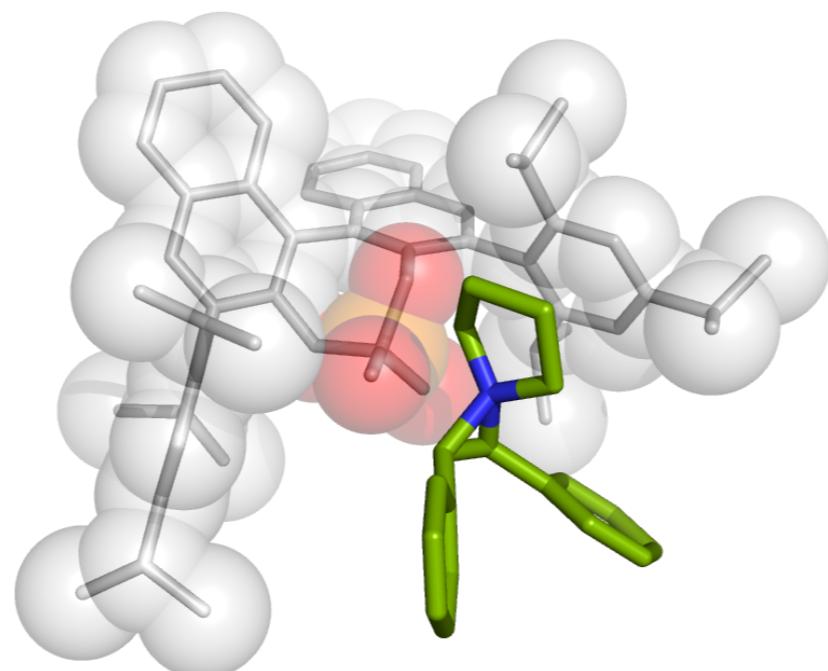
Non-covalent interactions

Dispersion arises from simultaneous excitation of electrons on different fragments – correlation problem

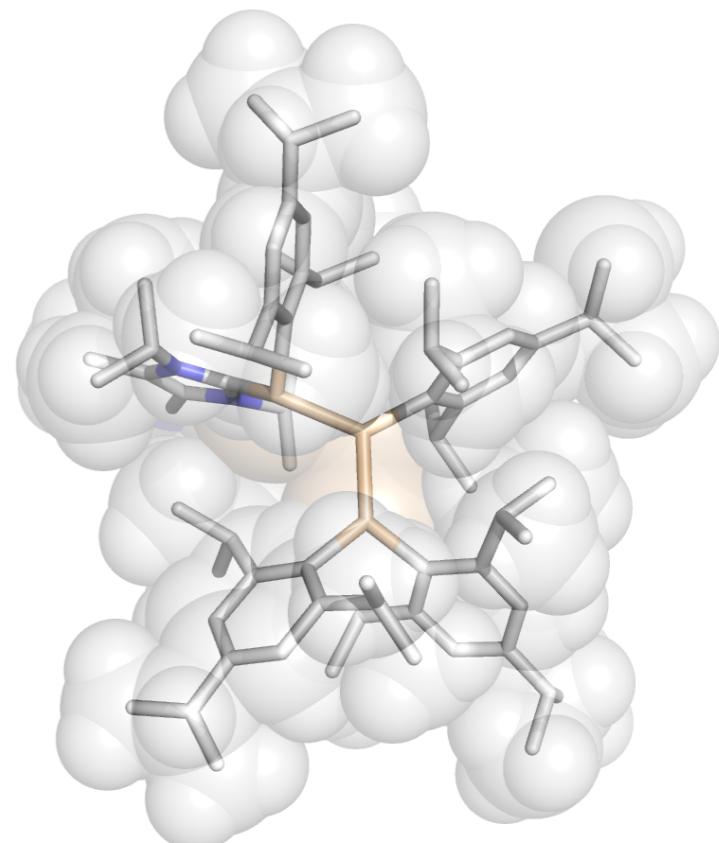
$$E_{\text{DFT-D}} = E_{\text{DFT}} + E_{\text{disp}}$$



Hydrogen Bonding &
cation - π



Ion-pairing



Dispersion

Inaccuracies in DFT

D3 correction: includes both two- and three-body components

$$E_{\text{disp}} = E^{(2)} + E^{(3)}$$

$$E^{(2)} = -S_6 \sum_{A < B} V_6^{\text{PP}}(Z_A, Z_B, R_{AB}) - S_8 \sum_{A < B} V_8^{\text{PP}}(Z_A, Z_B, R_{AB})$$

$$E^{(3)} = \sum_{A < B < C} V^{\text{TBP}}(Z_A, Z_B, Z_C, R_{AB}, R_{AC}, R_{BC}, \cos \theta_a, \cos \theta_b, \cos \theta_c)$$

$$V^{\text{PP}} = \frac{\sqrt{C_6^A C_6^B}}{R_{AB}^6} f(R_{AB})$$

Inaccuracies in DFT

- Self-interaction Error

$$J[\rho] = \frac{1}{2} \int d\vec{r}_1 \int d\vec{r}_2 \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}}$$

It includes interaction of each electron with all – including itself!

In HF theory the self-interaction is exactly cancelled by exchange

Dissociation of radicals (H^{+2})

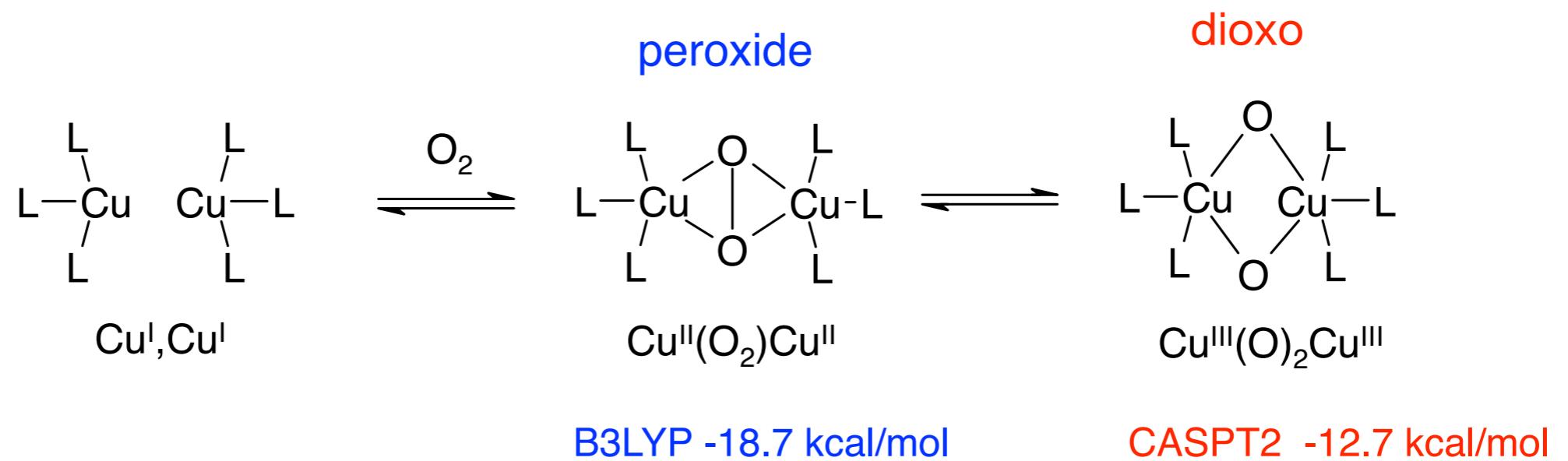
Activation Energies

Size-dependent

Inaccuracies in DFT

Multi-reference treatment

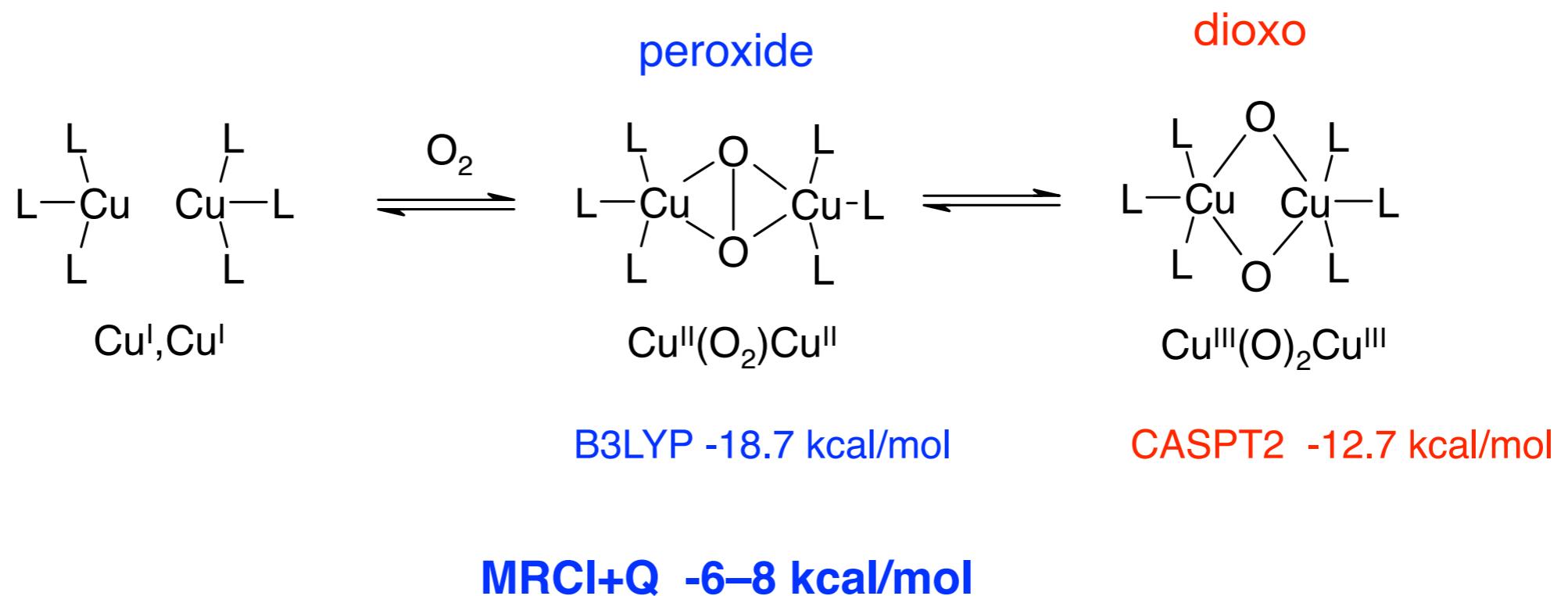
dicopper complexes: Using B3LYP two minima are found, the peroxide form is significantly more stable



Inaccuracies in DFT

Multi-reference treatment

This system has been reinvestigated using a more accurate multi-reference method and large basis sets



B3LYP gives qualitatively the correct answer

Inaccuracies in DFT

Specific Solvation: short range interactions of solvent molecules with solute.

Approach: adding solvent molecules into the calculation:

Where do you place the solvent molecules?

How many solvent molecules are needed?

Long Range Effects: screening of charges by solvent polarisation.

Approach: Self-Consistent Reaction Field (SCRF) Methods. the solute is placed in a cavity defined by atomic van der Waals radii.

Different models (PCM, COSMO...) differ in e.g. how the charge distribution of the solute or the cavity surface is calculated, or in the way the cavity is defined.

To keep in Mind

- DFT is very useful for exploring many aspects of chemistry, but be careful!

Assess the methodology (functionals and basis sets)

Assess conformational flexibility

- DFT is not perfect:

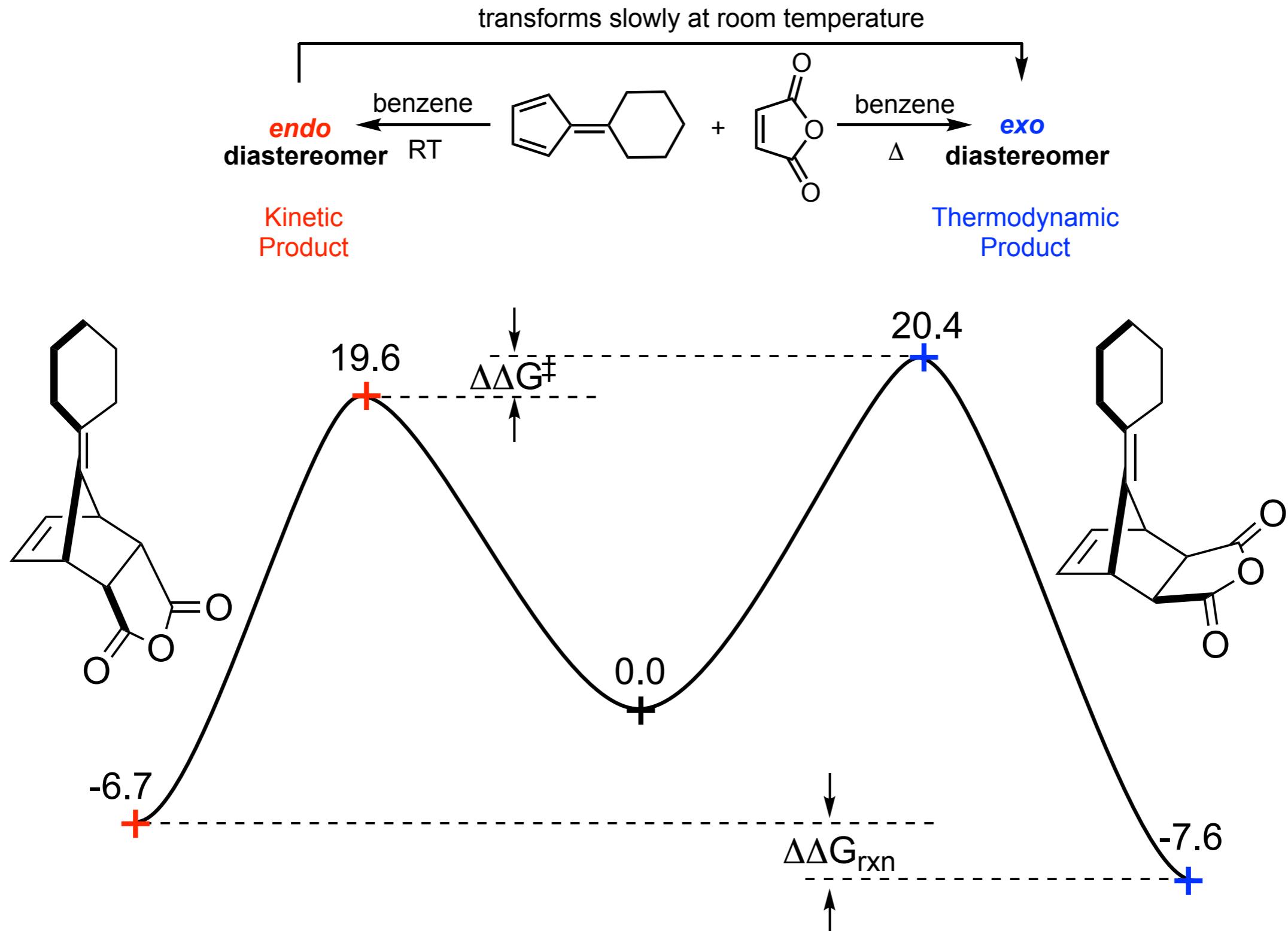
Assumes the system is well described by a single electronic configuration

Does not describe long-range dispersion effects well

- The field is moving rapidly

New functionals and approaches being developed all the time

Computational Organic Chemistry



Computational Organic Chemistry

Kinetics

$$k = \frac{k_B T}{h} e^{-\Delta G^\ddagger / RT}$$

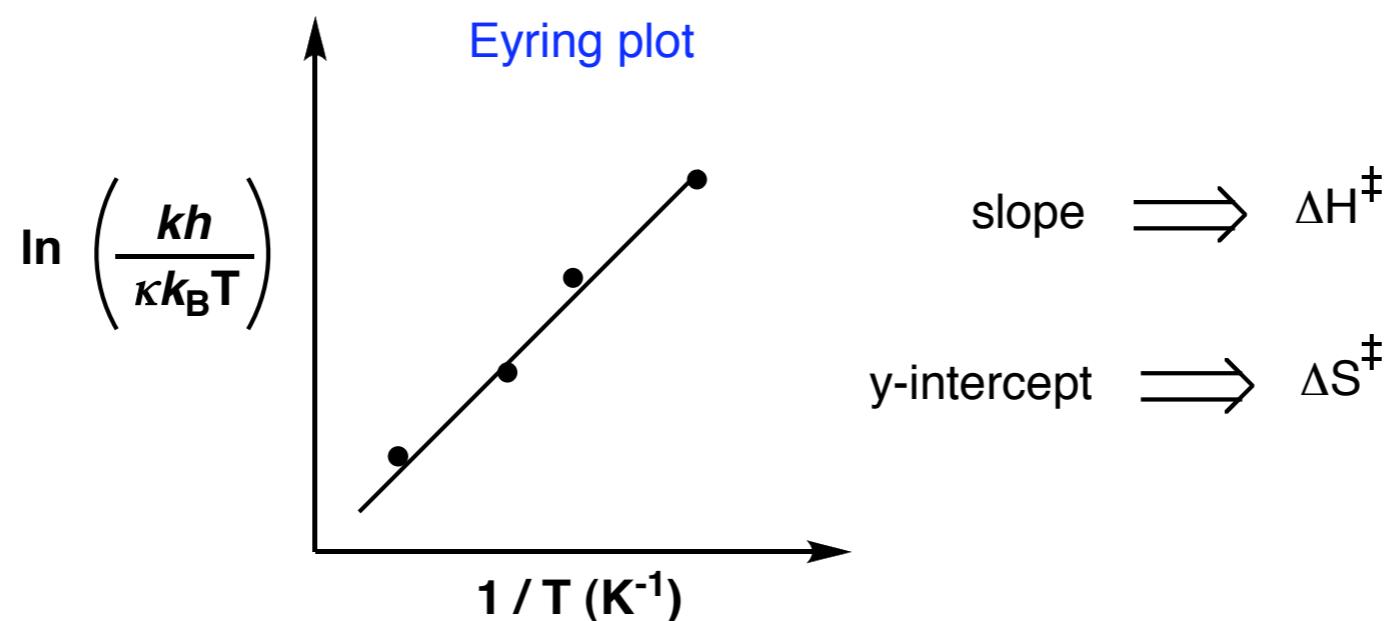
ΔG^\ddagger (kcal mol ⁻¹)	k (s ⁻¹)	$t_{1/2}$ (s ⁻¹)	$t_{1/2}$
12	9.8×10^3	7.1×10^{-5}	70.5 μs
17	2.11	3.3×10^{-1}	327 ms
22	4.5×10^{-4}	1.5×10^3	25 min
27	9.8×10^{-8}	7.1×10^6	81.1 days
30	6.2×10^{-10}	2.4×10^{-6}	35.5 years

$$T(K) = 298 ; R \text{ (kcal mol}^{-1}\text{)} = 0.00831 ; R \text{ (kJ K}^{-1} \text{ mol}^{-1}\text{)} = 8.3144598 \times 10^{-3}$$
$$h(J^*s) = 6.6262 \times 10^{-34} \quad k_b(J/K) = 1.3807 \times 10^{-23}$$

Computational Organic Chemistry

Experimental Determinations of Activation Parameters

$$\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger$$

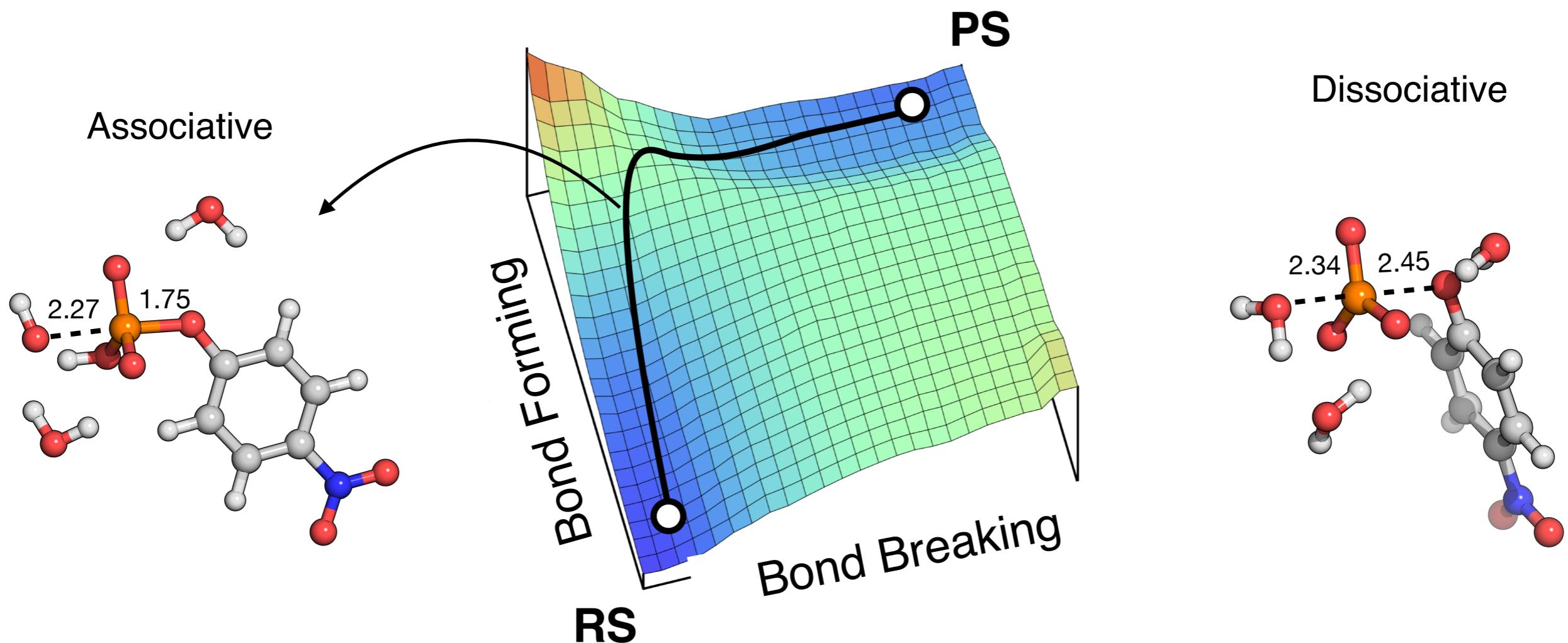


Entropy: energy associated with conformation, bond strength, vibrational states and how changes in these properties affect the overall energy of the system.

Enthalpy: energy related to the height of the surface

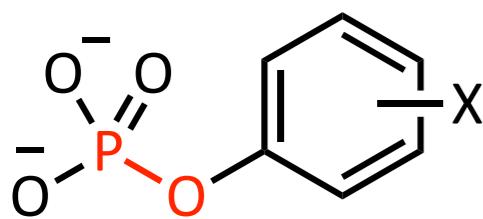
Computational Organic Chemistry

Phosphate/sulfate hydrolysis

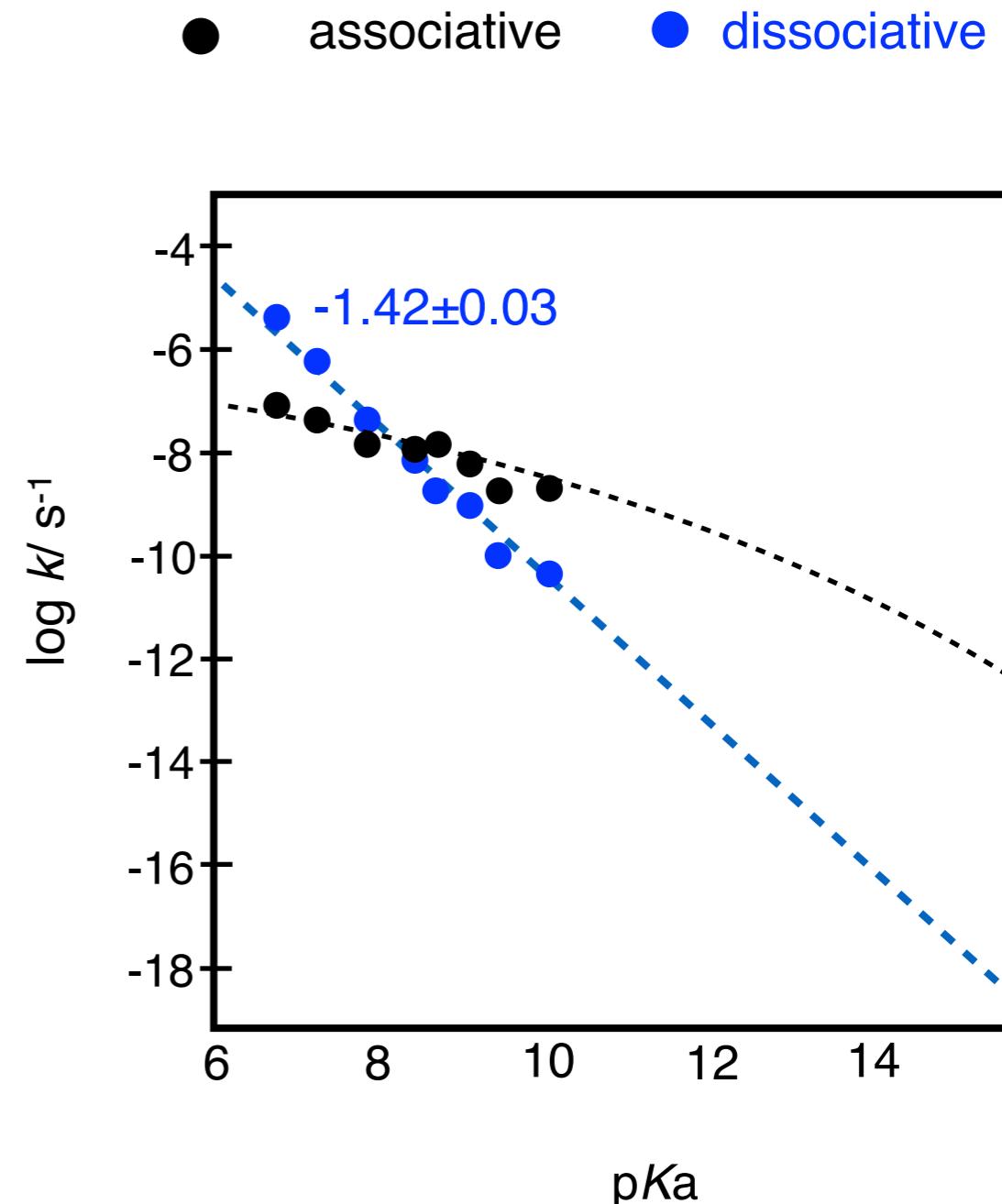
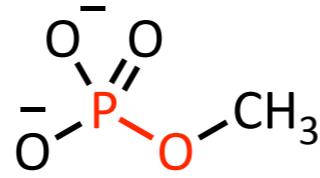


Computational Organic Chemistry

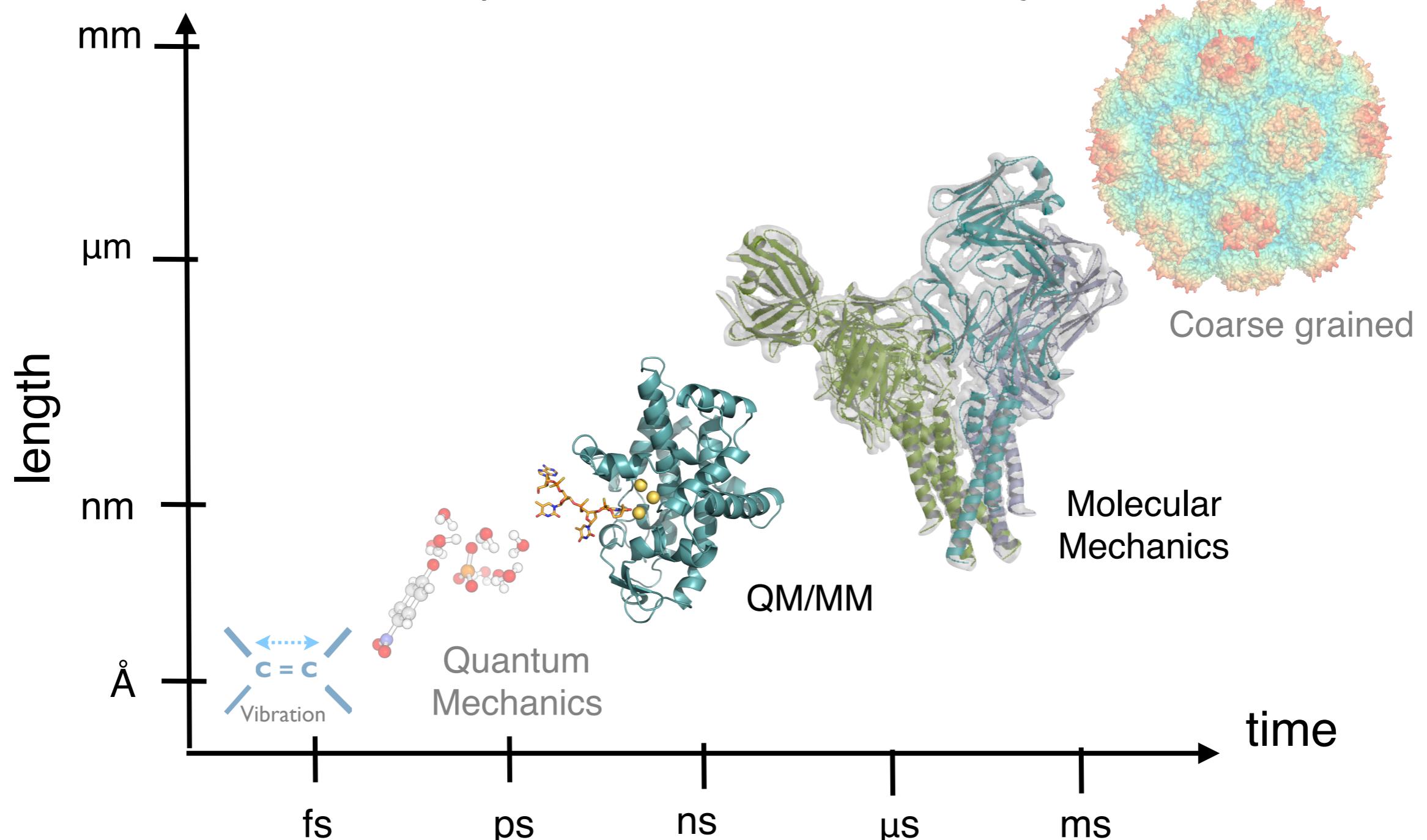
Phosphate/sulfate hydrolysis



- a. 3,5- NO_2
- b. 4- NO_2
- c. 3- NO_2 -4-Cl
- d. 3- NO_2
- e. 3,4-Cl
- f. 3-Cl
- g. 4-Cl
- h. H

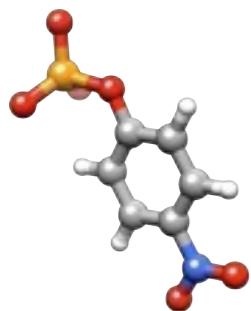


Computational Chemistry



“Everything should be made as simple as possible, but not simpler.”

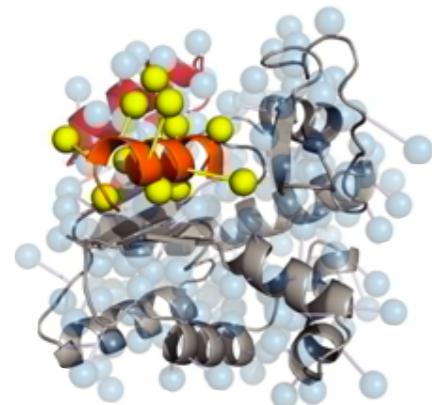
Computational Chemistry



Quantum
Mechanics

Electronic Structure
(Schrödinger equation)

Accurate \Rightarrow Expensive (N^4)

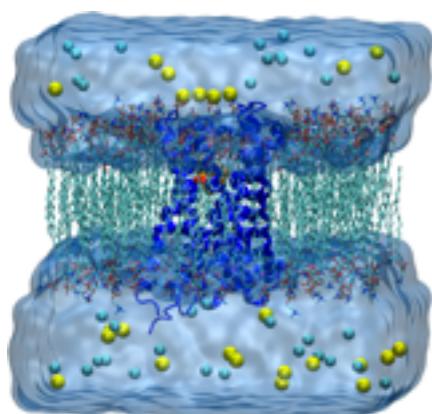


Molecular
Dynamics

Monte
Carlo

Empirical Forces
(Newton's equations/
Random move)

LESS accurate \Rightarrow Faster ($N \log N$ /
 N^2)



Coarse grain

Reduction of number of
degrees of freedom
(Molecular/Langevin
Dynamics)

Interaction details
eliminated

Which System Do I Have?

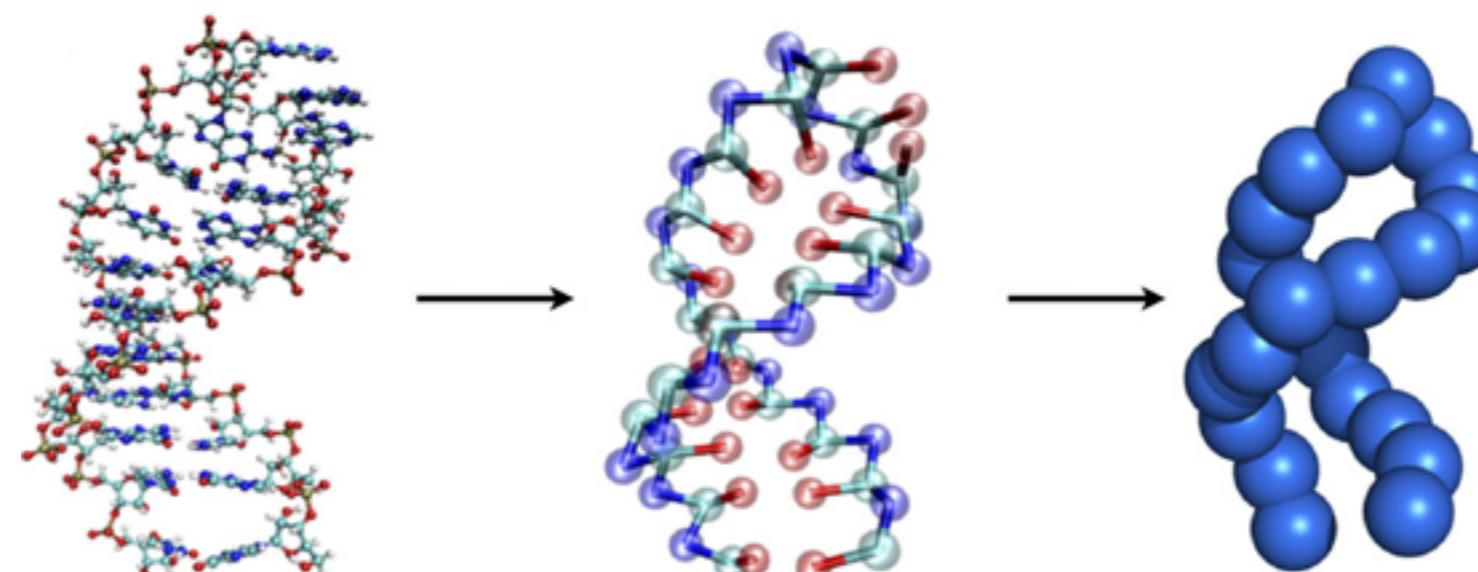


What Do You Want to Compute (and Why)?

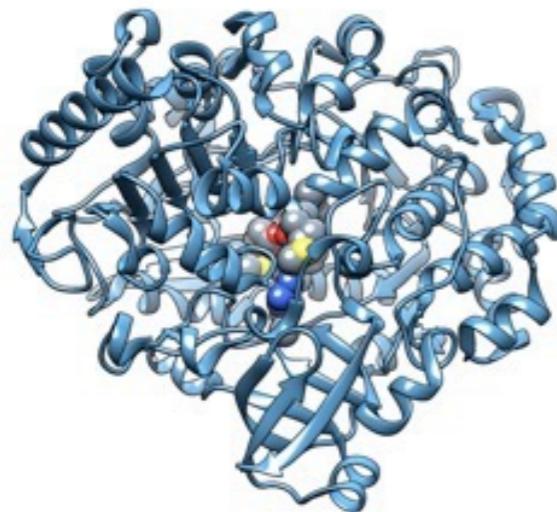


How much detail we need to describe the phenomenon of interest?

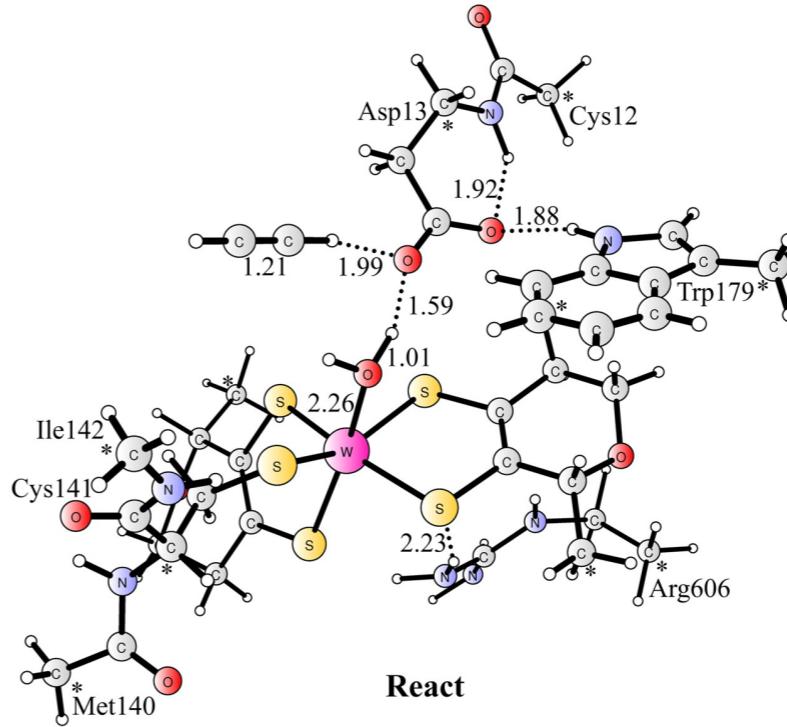
What is the level of computational resources available?



Quantum Mechanics(Cluster Model)



Tungsten-dependent Acetylene Hydratase



Liao R et al. PNAS 2010;107:22523-22527

Position (r)



Crystal coordinates.

Model of the active site (100-200 atoms!). Scale N^4

Charge (q)

Bond information

Electronic structure

Energy/Geometry

Schrödinger eq.
Minimization Algorithms

Environment

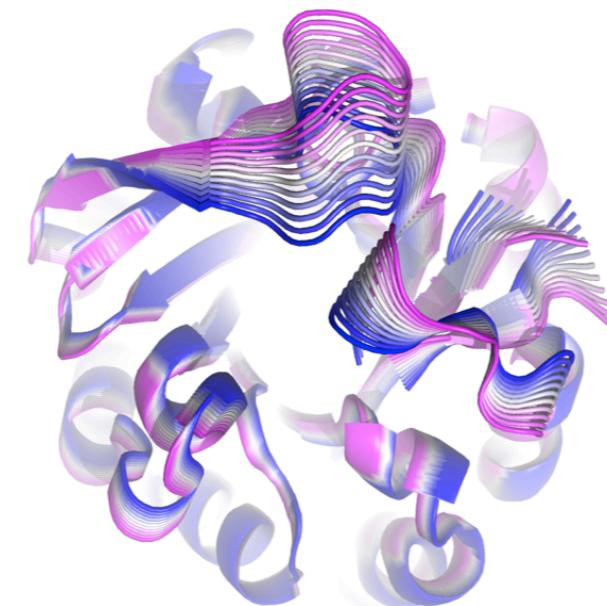


Steric / electrostatic effects from environment
are implicitly considered

Quantum Mechanics(Cluster Model)

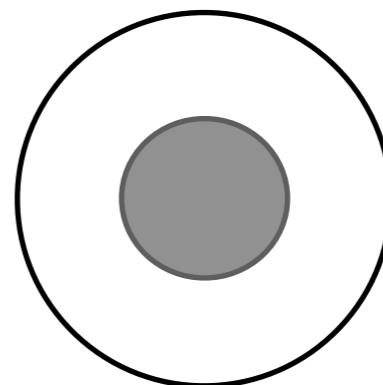
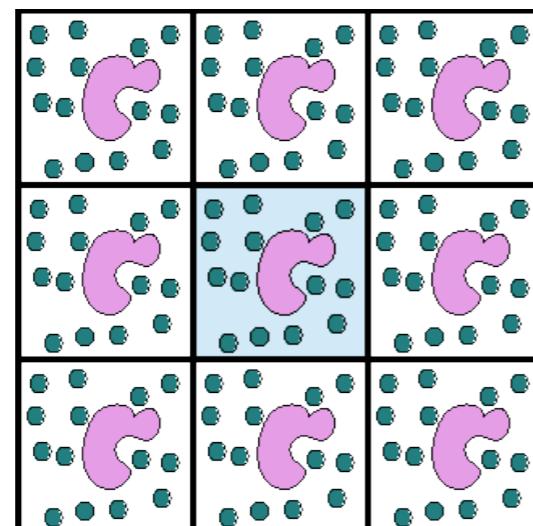
.... Just a matter of time, and computer power?

- There is No Such Thing as A Structure!
- Is the crystal structure the biologically relevant conformer?
- Molecules are Dynamic, Not Static (Conformational Ensemble)

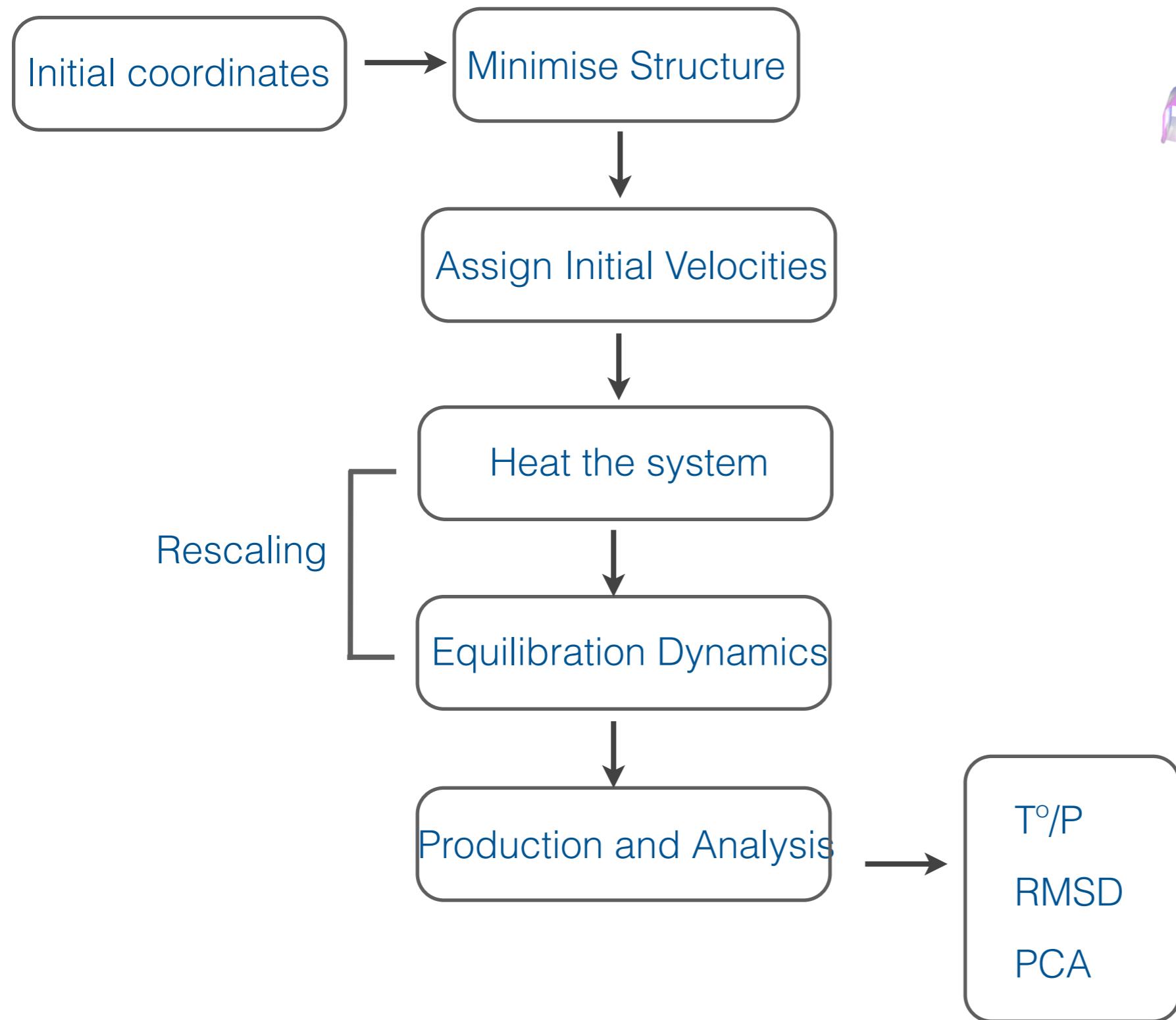


Molecular Dynamics

Position (r)	→	Crystal coordinates/homology models
Charge (q) Bond information	→	Empirical energy function(CHARMM,OPLSAA,AMBER)
Momentum ($m + v$)	→	Newton's equations
Environment	→	Solvent (Implicit/Explicit:TIP3P, SPC) Spherical/Periodical boundaries

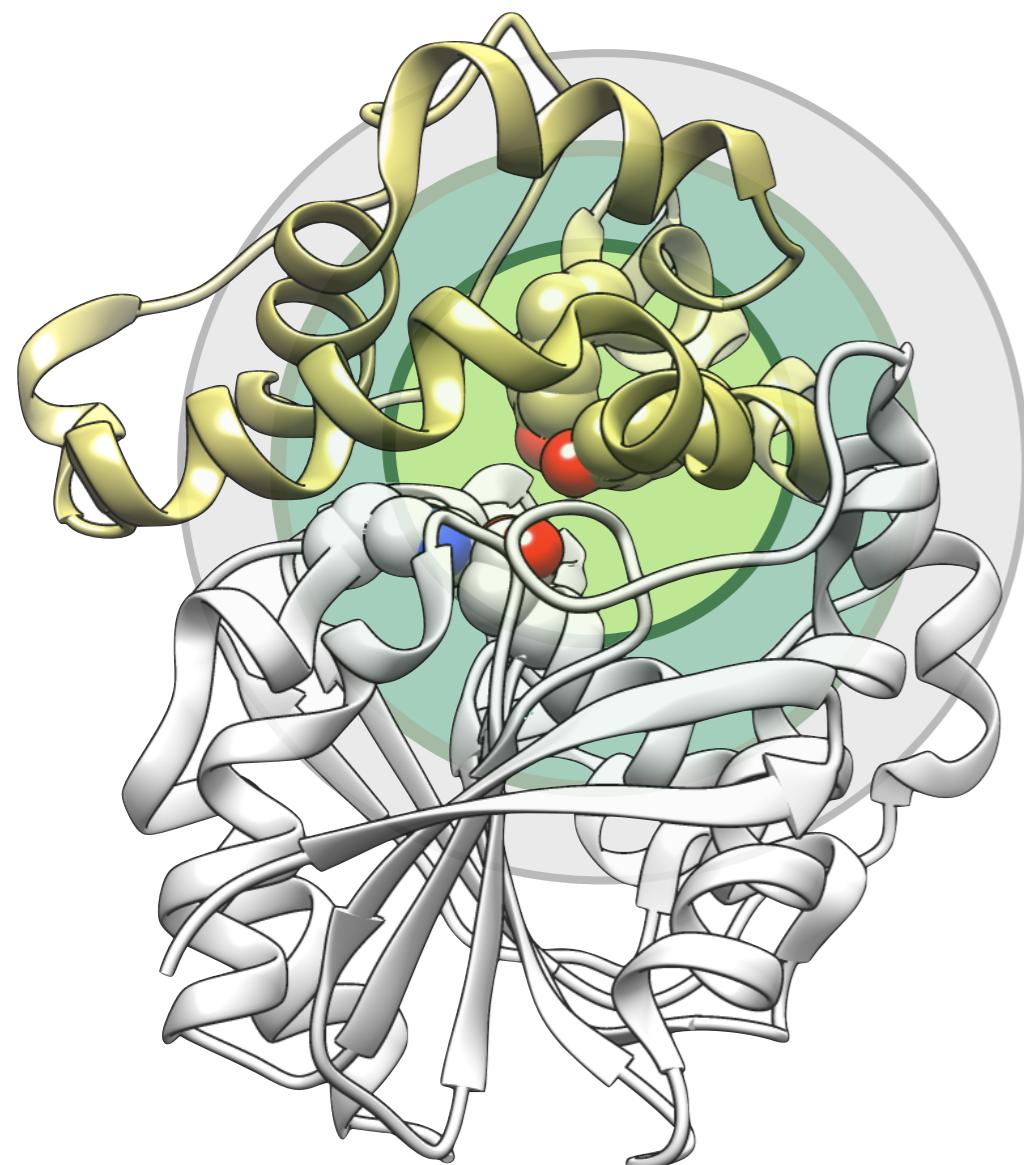


Molecular Dynamics



QM/MM methods to Address Enzymatic Reactions

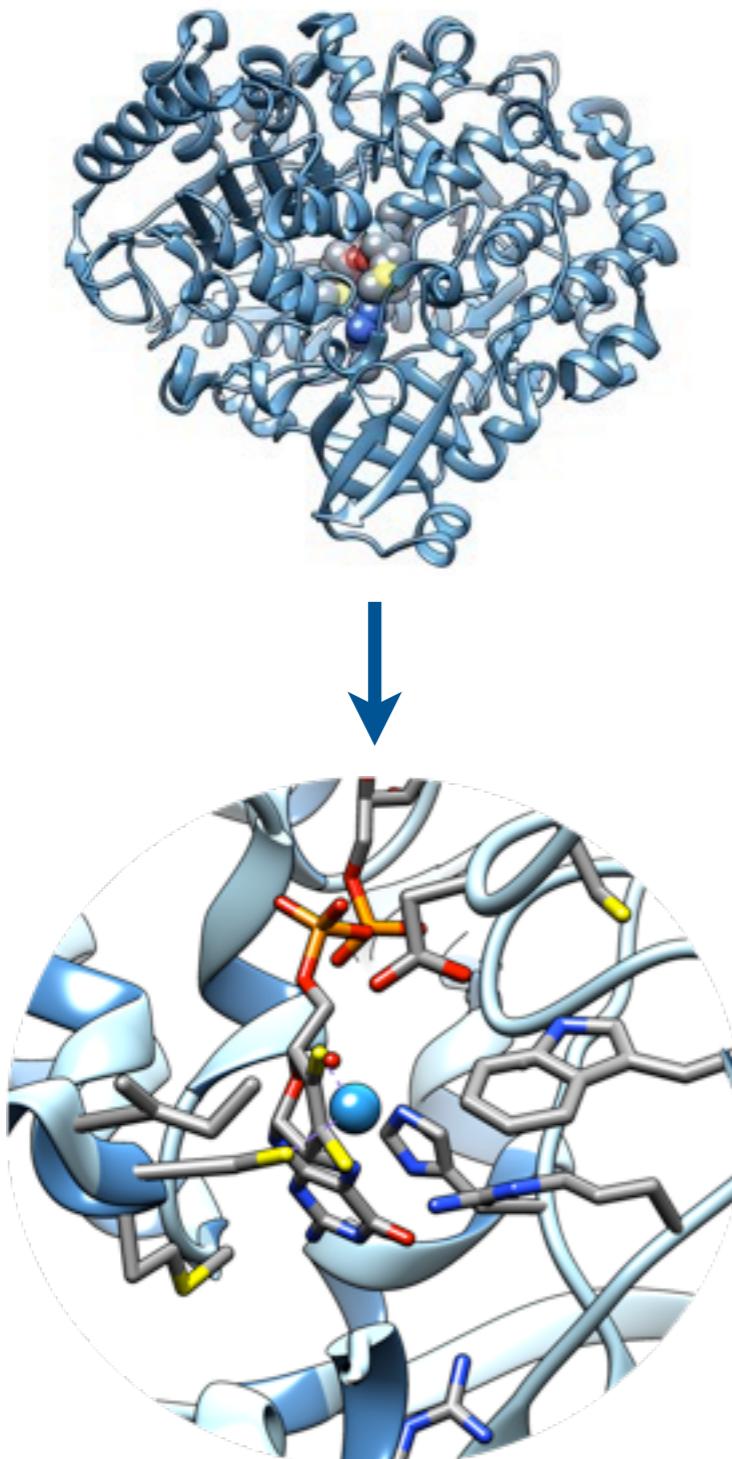
Combine the best of each world...



- ✓ The whole solvated protein is chosen as the model.
- ✓ Complete description of electrons breaking/forming bond processes/exited state reactions, etc
- ✓ Explicitly incorporating the steric and electrostatic effects of the environment
- *How to handle QM/MM boundary?*
- *Is the sampling problem solved?*

QM/MM methods to Address Enzymatic Reactions

Case Study: Acetylene Hydratase



Conclusion:

- QM-only (32 atoms): two-step reaction profile.
- QM(116 atoms)/MM: one-step mechanism (very high barrier). Influence of the key groups neglected in the QM-only
- QM(157 atoms)/MM: two-step mechanism, reasonable overall barrier. Change in the rate limiting step.

QM/MM methods to Address Enzymatic Reactions

Advantages & Challenges

- ✓ The region of most interest in the system is treated with more accurate methods.
 - ✓ Bond being broken and form can be explicitly described using QM methods.
 - ✓ The environment can be taken into account at a lower level of theory.
-
- Large computer time in order to evaluate the QM energy.
 - Limited configurational sampling. Proteins are not flexible entities.
 - Reaction pathways will strongly depend on protein configurations.

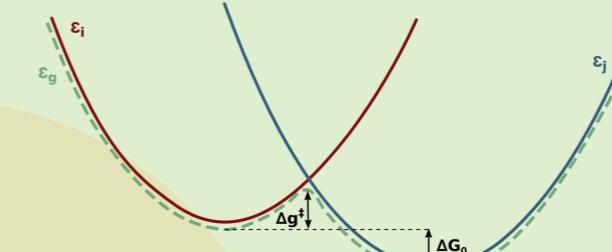
Computational Chemistry

Theory

$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = \frac{-\hbar^2}{2m} \sum_i^{\text{electrons}} \nabla_i^2 - \sum_i^{\text{electrons}} \sum_A^{\text{nuclei}} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{\text{electrons}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

Modelling



Experiments

*Synthesis
Kinetics
spectroscopy*



... Back to the Lab with some new knowledge ...