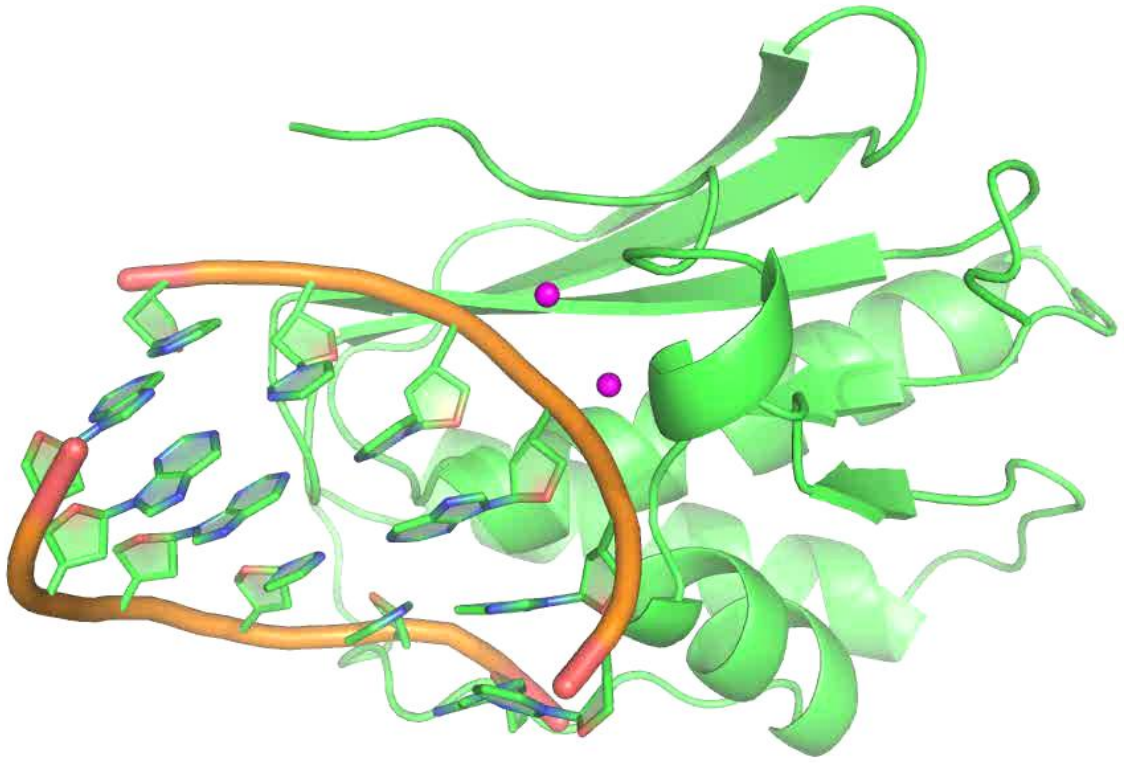
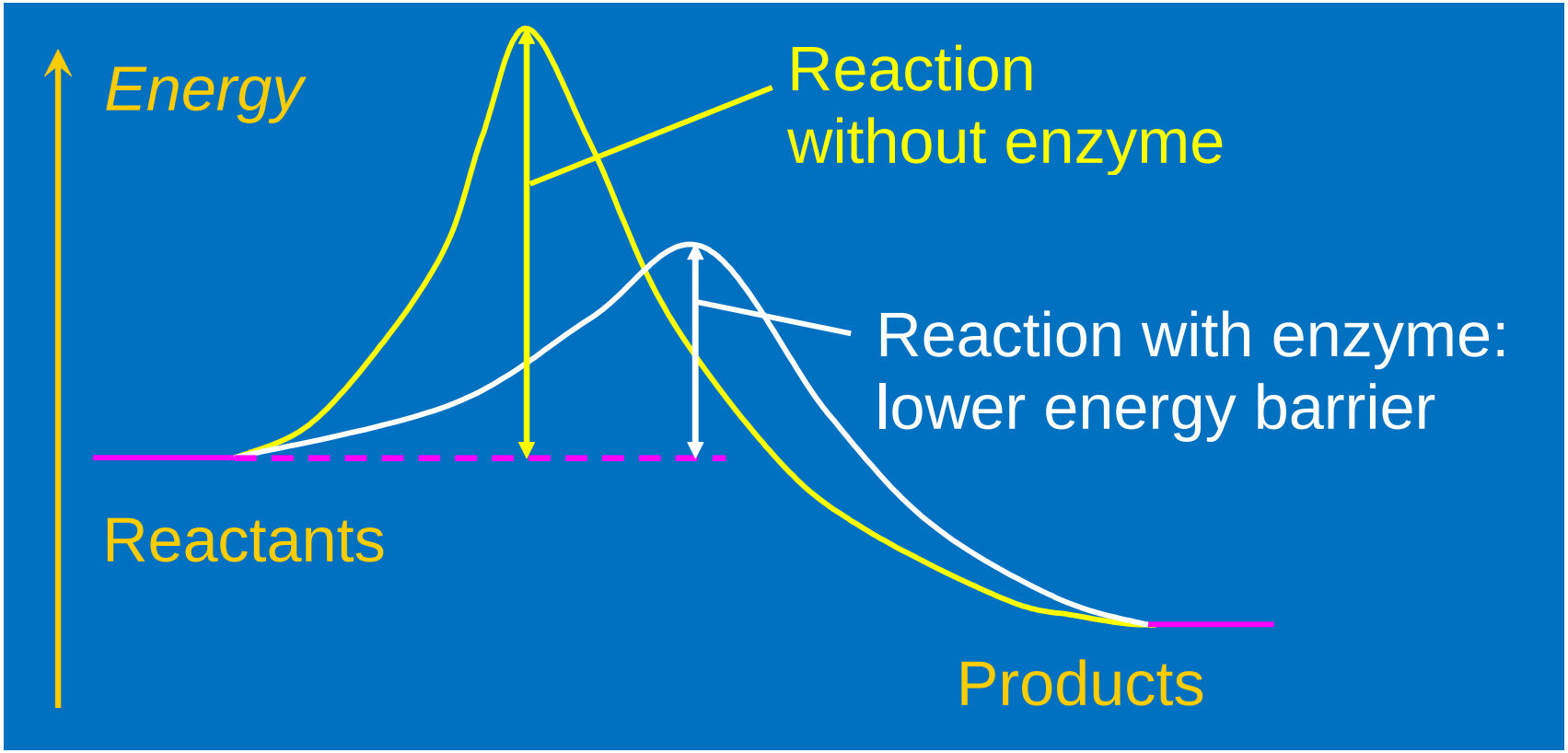
QM/MM methods

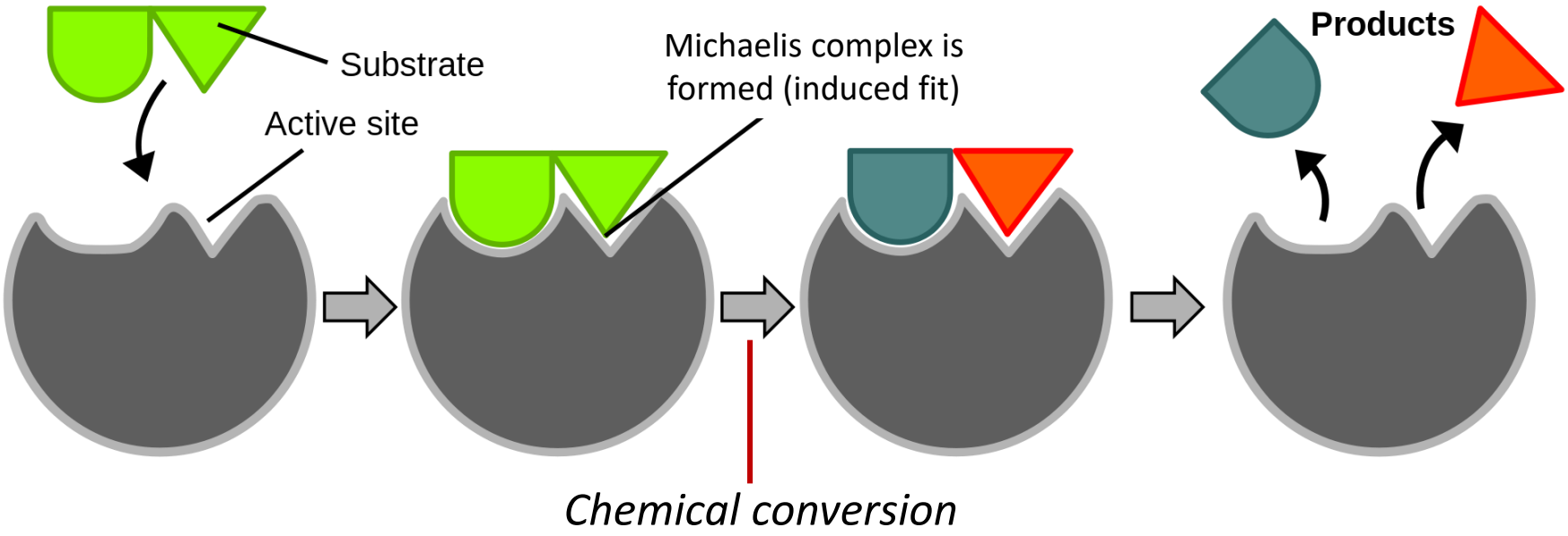
(Slides from Edina Rosta, Todd Martinez, Marc Vanderkamp)



- Enzymes lower the activation energy for reaction



Enzyme catalytic cycles



- Difficult to dissect experimentally: *use computer simulation*
- To simulate chemical change, we need to go beyond molecular mechanics

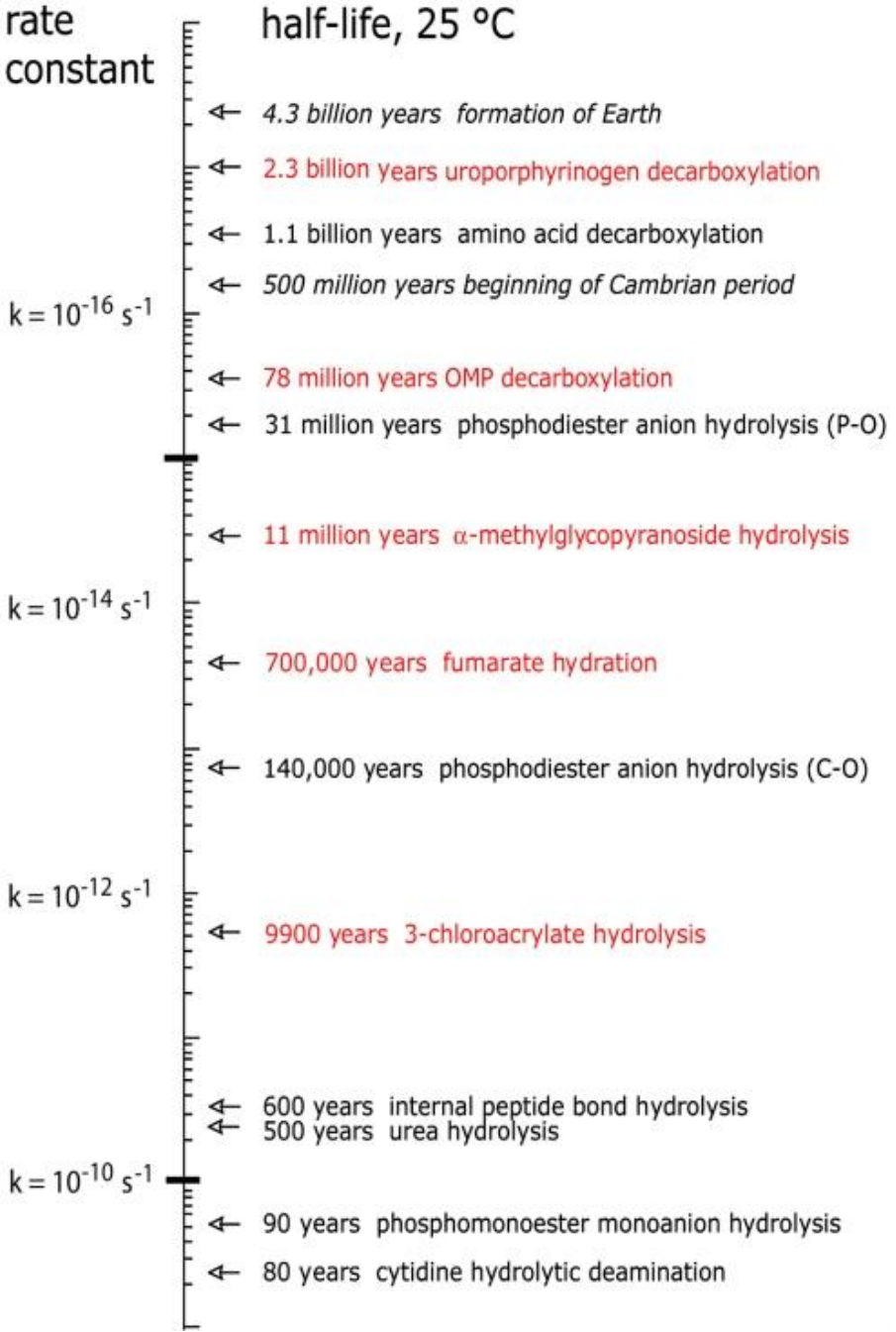
Enzyme Catalysis

- Rate acceleration of 10^{10} is typical

  e.g. protease hydrolysis of peptide bonds

- k_{cat}/k_{uncat} can reach 10^{16}

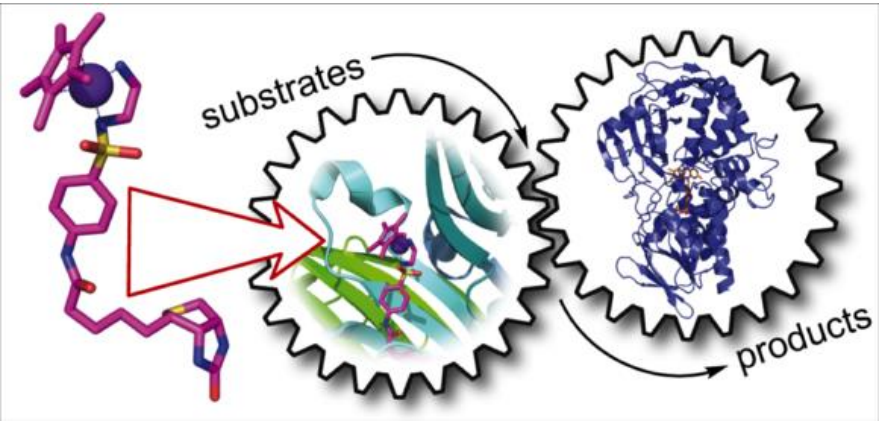
  e.g. *uroporphyrinogen III decarboxylase*



Lewis & Wolfenden PNAS 105, 17328 (2008)

Enzyme engineering

Finding more efficient enzymes for producing biofuels.



jmp
Statistical Discovery™ From SAS.

CHALLENGE

Finding more efficient enzymes for producing biofuels.

SOLUTION

JMP's design of experiment tools, multivariate analysis, the distribution platform, fit Y by X, fit model and the cluster platform.

RESULTS

Novozymes is improving the enzymes that render a more cost-effective and sustainable process for commercially viable biofuels.

MORE INFORMATION

www.novozymes.com

www.jmp.com

Novozymes

Reducing the cost of producing biofuels

Improving processes for its customers with enzyme products–that's the kernel of the issue–and that's Novozymes' business.

"Rethink Tomorrow" is Novozymes' slogan, and one of the areas where it's doing so is in looking for more efficient enzymes for producing biofuels. Novozymes is a Denmark-based world leader in bio-innovation. In cooperation with its customers across a wide range of industries, Novozymes is helping create tomorrow's industrial bio-solutions for a greener planet.

Optimizing enzymes with JMP®

Here in the US–in Novozymes' North American subsidiary in Franklinton, North Carolina–Jesper Frickmann is tasked with supporting customers in implementing and optimizing Novozymes' enzymes for biofuel solutions within their businesses.

JMP statistical discovery software from SAS is an integral tool in Novozymes' research. Frickmann himself began using JMP more than 10 years ago. He's an engineer by training, but before working in customer support, he served as a statistician, providing statistical help to anyone at Novozymes. He used JMP then and still does.

Frickmann says: "Anytime we run an experiment in R&D to find new and better enzyme candidates or to optimize the reaction conditions, we use JMP to design an experiment and analyze the data.

"Or if a customer wants to try a different enzyme in our line, we'll run the two different enzymes in their application, collect the data–for example, the amount of alcohol produced–and determine if there's a significant difference in the two products."

One of the ongoing objectives of Novozymes' researchers is to consistently improve the enzymes that render a more cost-effective and sustainable process for commercially viable biofuels.

Also within the next-generation cellulosic ethanol are biofuels made from feedstocks such as cobs, stalks and leaves from corn, wood chips, paper, switchgrass, etc. "Today all that is mainly left in the field to rot," says Frickmann. "But it could be collected, and the cellulose from that could be converted. We're trying to make all this happen."

STATISTICAL DISCOVERY.™ FROM SAS.

Enzyme engineering

Kendall Houk, UCLA
David Baker, University of Washington

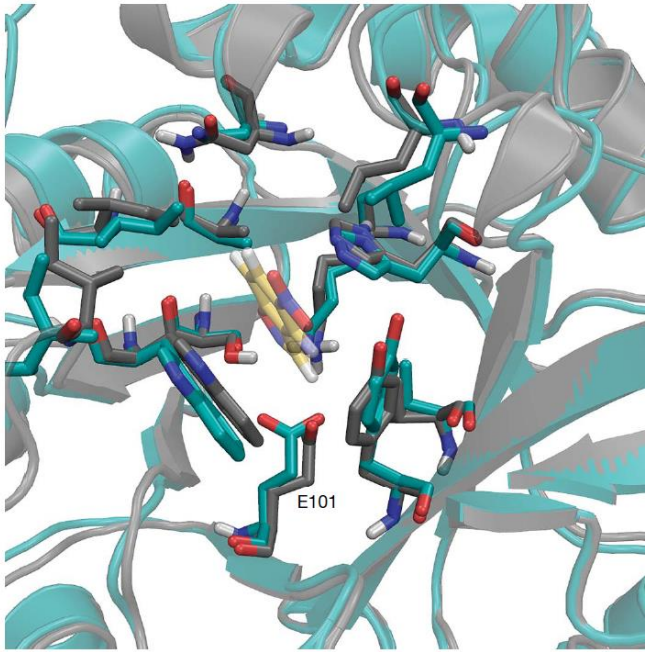


Figure 4 | Comparison of the designed model of KE07 and the crystal structure. The crystal structure (cyan) was solved in the unbound state and shows only modest rearrangement of active site side chains compared to the designed structure (grey) modelled in the presence of the transition state (yellow, transparent). (Backbone r.m.s.d. for the active site is 0.32 Å versus 0.95 Å for the active site including the side chains.) The observed electron

nature Vol 453|8 May 2008|doi:10.1038/nature06879

ARTICLES

Kemp elimination catalysts by computational enzyme design

Daniela Röthlisberger[1]*, Olga Khersonsky[4]*, Andrew M. Wollacott[1]*, Lin Jiang[1,2], Jason DeChancie[6], Jamie Betker[3], Jasmine L. Gallaher[3], Eric A. Althoff[1], Alexandre Zanghellini[1,2], Orly Dym[5], Shira Albeck[5], Kendall N. Houk[6], Dan S. Tawfik[4] & David Baker[1,2,3]

The design of new enzymes for reactions not catalysed by naturally occurring biocatalysts is a challenge for protein engineering and is a critical test of our understanding of enzyme catalysis. Here we describe the computational design of eight enzymes that use two different catalytic motifs to catalyse the Kemp elimination—a model reaction for proton transfer from carbon—with measured rate enhancements of up to 10^5 and multiple turnovers. Mutational analysis confirms that catalysis depends on the computationally designed active sites, and a high-resolution crystal structure suggests that the designs have close to atomic accuracy. Application of *in vitro* evolution to enhance the computational designs produced a >200-fold increase in k_{cat}/K_m (k_{cat}/K_m of 2,600 $M^{-1}s^{-1}$ and k_{cat}/k_{uncat} of $>10^6$). These results demonstrate the power of combining computational protein design with directed evolution for creating new enzymes, and we anticipate the creation of a wide range of useful new catalysts in the future.

Nature **vol 453**, pages190–195 (08 May 2008)

Enzyme engineering

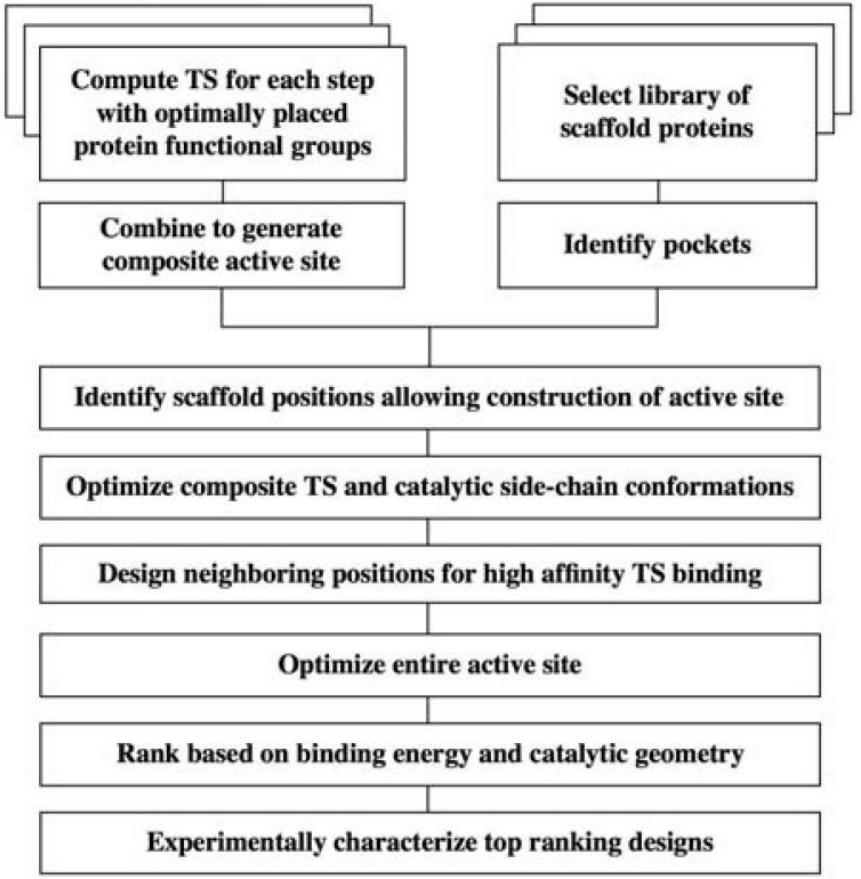
Kendall Houk, UCLA
David Baker, University of Washington

Science 07 Mar 2008:
Vol. 319, Issue 5868, pp. 1387-1391
DOI: 10.1126/science.1152692

De Novo Computational Design of Retro-Aldol Enzymes

Lin Jiang,[1,2]* Eric A. Althoff,[1]* Fernando R. Clemente,[4] Lindsey Doyle,[5] Daniela Röthlisberger,[1] Alexandre Zanghellini,[1,2] Jasmine L. Gallaher,[1] Jamie L. Betker,[1] Fujie Tanaka,[6] Carlos F. Barbas III,[6] Donald Hilvert,[7] Kendall N. Houk,[4] Barry L. Stoddard,[5] David Baker[1,2,3]†

The creation of enzymes capable of catalyzing any desired chemical reaction is a grand challenge for computational protein design. Using new algorithms that rely on hashing techniques to construct active sites for multistep reactions, we designed retro-aldolases that use four different catalytic motifs to catalyze the breaking of a carbon-carbon bond in a nonnatural substrate. Of the 72 designs that were experimentally characterized, 32, spanning a range of protein folds, had detectable retro-aldolase activity. Designs that used an explicit water molecule to mediate proton shuffling were significantly more successful, with rate accelerations of up to four orders of magnitude and multiple turnovers, than those involving charged side-chain networks. The atomic accuracy of the design process was confirmed by the x-ray crystal structure of active designs embedded in two protein scaffolds, both of which were nearly superimposable on the design model.



Enzyme engineering

Kendall Houk, UCLA
David Baker, University of Washington

Science 16 Jul 2010:
Vol. 329, Issue 5989, pp. 309-313
DOI: 10.1126/science.1190239

Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction

Justin B. Siegel,[1,2]* Alexandre Zanghellini,[1,2]*† Helena M. Lovick,[3] Gert Kiss,[4] Abigail R. Lambert,[5] Jennifer L. St.Clair,[1] Jasmine L. Gallaher,[1] Donald Hilvert,[6] Michael H. Gelb,[3] Barry L. Stoddard,[5] Kendall N. Houk,[4] Forrest E. Michael,[3] David Baker[1,2,7]‡

The Diels-Alder reaction is a cornerstone in organic synthesis, forming two carbon-carbon bonds and up to four new stereogenic centers in one step. No naturally occurring enzymes have been shown to catalyze bimolecular Diels-Alder reactions. We describe the de novo computational design and experimental characterization of enzymes catalyzing a bimolecular Diels-Alder reaction with high stereoselectivity and substrate specificity. X-ray crystallography confirms that the structure matches the design for the most active of the enzymes, and binding site substitutions reprogram the substrate specificity. Designed stereoselective catalysts for carbon-carbon bond-forming reactions should be broadly useful in synthetic chemistry.

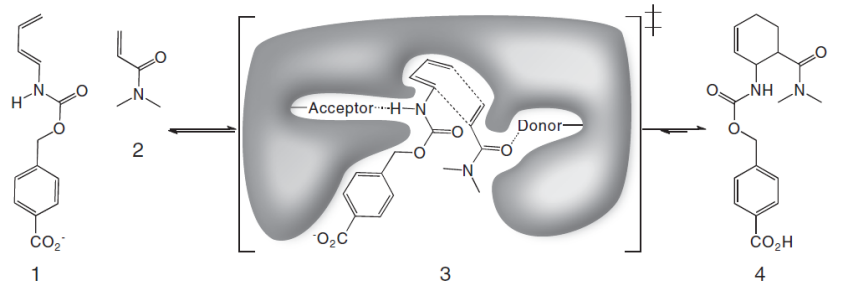
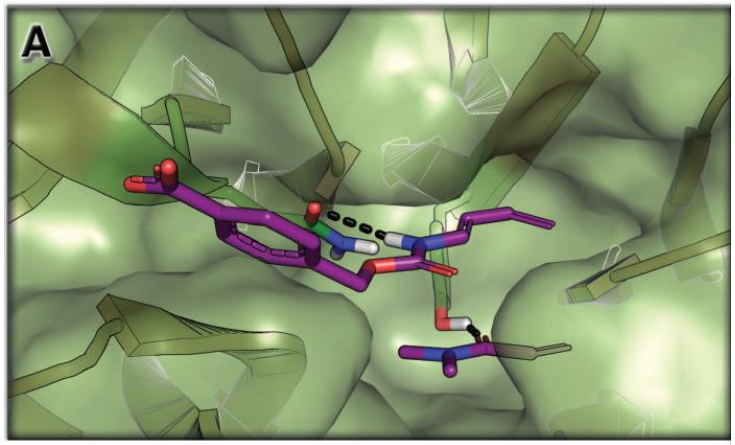


Fig. 1. The Diels-Alder reaction. Diene (1) and dienophile (2) undergo a pericyclic [4 + 2] cycloaddition (3) to form a chiral cyclohexene ring (4). Also shown in (3) is a schematic of the design target active site, with hydrogen bond acceptor and donor groups activating the diene and dienophile and a complementary binding pocket holding the two substrates in an orientation optimal for catalysis.



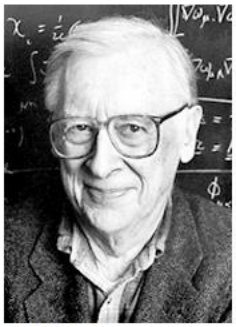
Nobel Prize in Chemistry 1998

Nobel Prizes and Laureates
Chemistry Prizes 1998
▼ About the Nobel Prize in Chemistry 1998
Summary
Press Release
Advanced Information
Illustrated Information
Award Ceremony Video
Award Ceremony Speech
Banquet Video
▸ Walter Kohn
▸ John Pople
All Nobel Prizes in Chemistry
All Nobel Prizes in 1998

The Nobel Prize in Chemistry 1998
Walter Kohn, John Pople

The Nobel Prize in Chemistry 1998

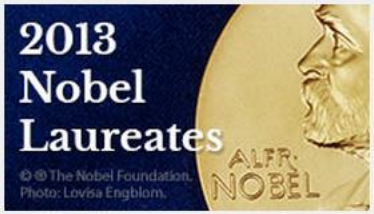
Walter Kohn



John A. Pople

The Nobel Prize in Chemistry 1998 was divided equally between Walter Kohn *"for his development of the density-functional theory"* and John A. Pople *"for his development of computational methods in quantum chemistry"*.

Share | Tell a Friend | Comments



2013 CHEMISTRY PRIZE QUESTION

Did you know that computer models are used to study everything from molecules of life to industrial chemical processes?

Yes
No
Submit

Greetings to the 2013 Nobel

The recipients

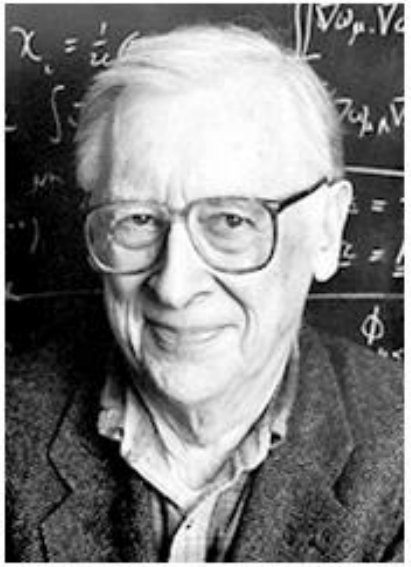
- Walter Kohn

Developed the Density Functional Theory (DFT), which became the most widely used quantum chemistry method due to its efficiency and accuracy.

- John Pople

Developed numerous algorithms for quantum chemistry methods. Main founder behind the currently most widely used quantum chemistry software, Gaussian.

Walter Kohn



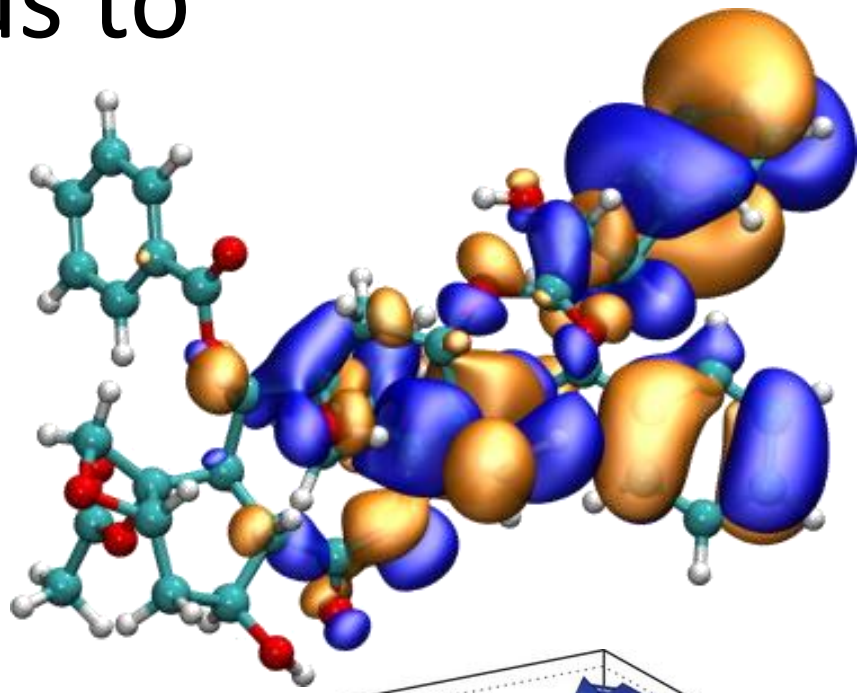
John A. Pople

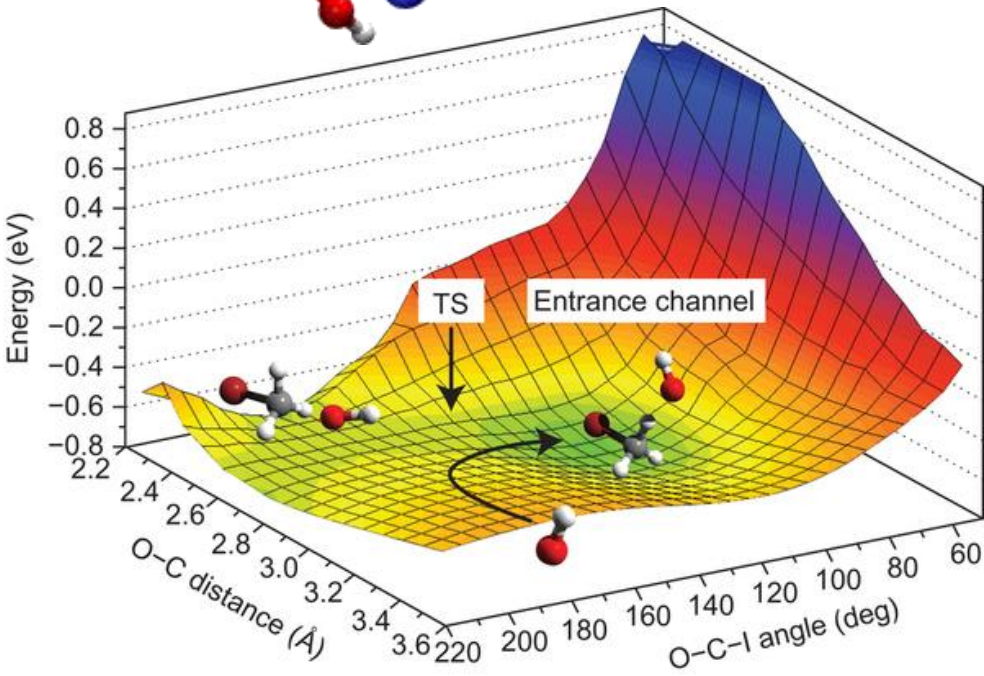
Their work enables us to

- Obtain the solution of the Schrödinger equation using approximate methods

$$H\Psi = E\Psi$$

- Find energies and wave functions of small to medium sized molecules
- Provide accurate models of **chemical structure** and **reactivity**





Nobel Prize in Chemistry 2013

Nobel Prizes and Laureates
Chemistry Prizes 2013
▼ About the Nobel Prize in Chemistry 2013
Summary
Prize Announcement
Press Release
Advanced Information
Popular Information
Greetings
▸ Martin Karplus
▸ Michael Levitt
▸ Arieh Warshel
All Nobel Prizes in Chemistry
All Nobel Prizes in 2013

The Nobel Prize in Chemistry 2013
Martin Karplus, Michael Levitt, Arieh Warshel

The Nobel Prize in Chemistry 2013

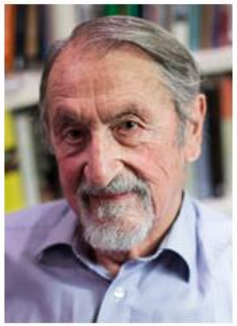
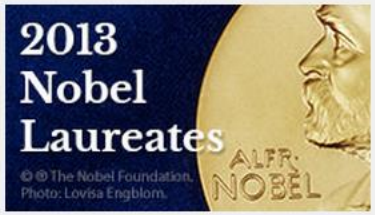

© Harvard University
Martin Karplus


Photo: © S. Fisch
Michael Levitt


Photo: Wikimedia Commons
Arieh Warshel

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

Share | Tell a Friend | Comments



2013 CHEMISTRY PRIZE QUESTION

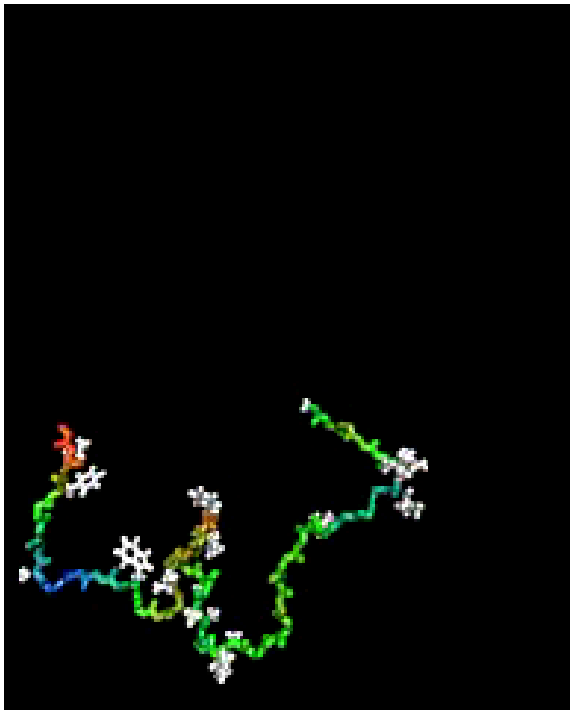
Did you know that computer models are used to study everything from molecules of life to industrial chemical processes?

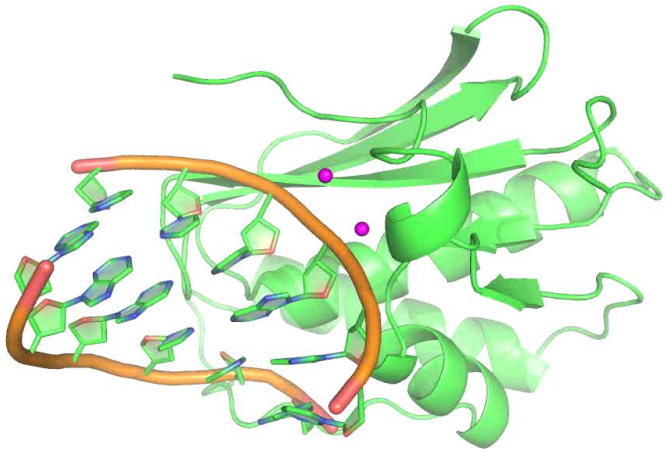
Yes
No
Submit

Greetings to the 2013 Nobel

Their work enables us to

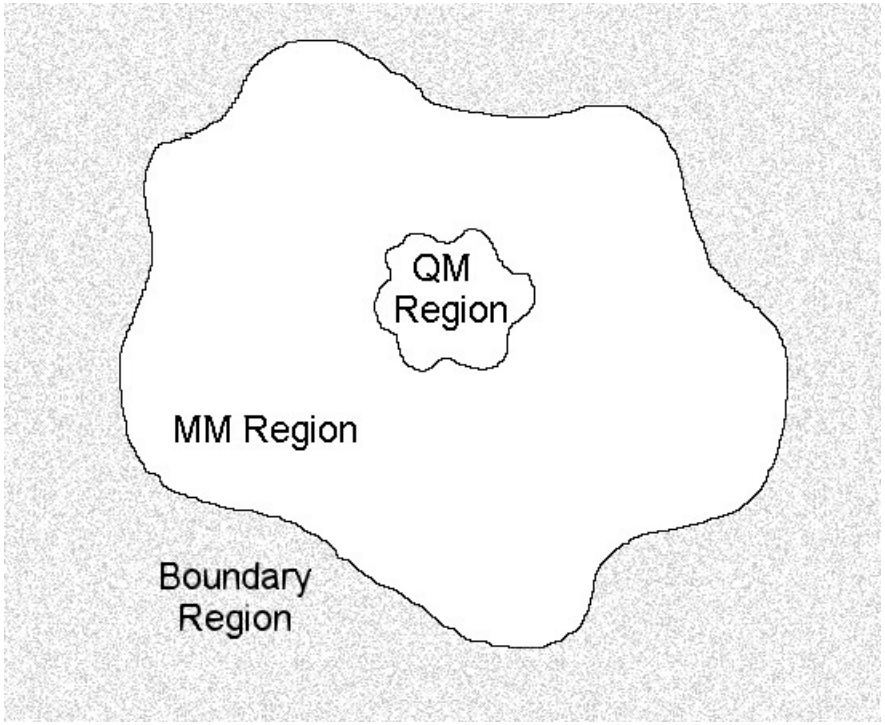
- Find the structures of complex biomolecules by calculating their Newtonian dynamics
- Find reaction mechanisms of enzymes
- Model and predict structure and function of biological systems





A Hybrid QM/MM Approach

The development of hybrid QM/MM approaches is guided by the general idea that large chemical systems may be partitioned into an electronically important region which requires a quantum chemical treatment and a remainder which only acts in a perturbative fashion and thus admits a classical description.

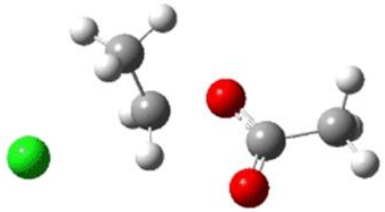


Quantum Mechanics

- **In theory, a very accurate treatment of the system**
- **Largely *ab initio*, i.e. parameter-free**
- **Very expensive — typically scales as $O(N^4)$ 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**

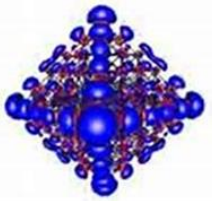
ORCA

Q-CHEM



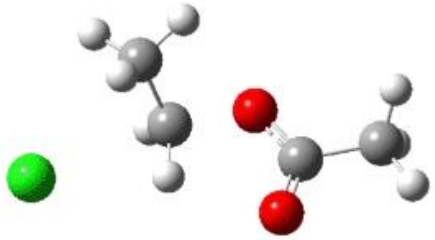
gaussian.com

TURBOMOLE



Software Packages

- **QM software packages:**
  - **ORCA**
  - **Q-Chem**
  - **Gaussian**
  - **Turbomole**
  - **MOPAC**
- **MM software packages with QM/MM:**
  - **NAMD (NEW!)**
  - **CHARMM**
  - **ChemShell**
  - **AMBER**
  - **GROMACS**
  - **Q Site - Schrodinger**



QM/MM implementations

QM/MM Partitioning

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

Interaction between QM and MM

Energy of MM subsystem

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

The Simplest Hybrid QM/MM Model

Hamiltonian for the molecular system in the Born-Oppenheimer approximation:

$$H = -\frac{1}{2}\sum_{i}^{electrons}\nabla^2 - \sum_{i}^{electrons}\sum_{j}^{nuclei}\frac{Z_j}{R_{ij}} + \sum_{i}^{electrons}\sum_{j<i}^{electrons}\frac{1}{r_{ij}} + \sum_{i}^{nuclei}\sum_{j<i}^{nuclei}\frac{Z_i Z_j}{R_{ij}}$$

←"Standard" QM hamiltonian

$$H = -\frac{1}{2}\sum_{i}^{electrons}\nabla^2 - \sum_{i}^{electrons}\sum_{j}^{nuclei}\frac{Z_j}{R_{ij}} + \sum_{i}^{electrons}\sum_{j<i}^{electrons}\frac{1}{r_{ij}} + \sum_{i}^{nuclei}\sum_{j<i}^{nuclei}\frac{Z_i Z_j}{R_{ij}} \underbrace{- \sum_{i}^{electrons}\sum_{k}^{charges}\frac{Q_k}{R_{ik}} + \sum_{i}^{nuclei}\sum_{k}^{charges}\frac{Z_i Q_k}{R_{ik}}}_{Effect\ of\ External\ Charges}$$

The main drawbacks of this simple QM/MM model are:

⇒ it is impossible to optimize the position of the QM part relative to the external charges because QM nuclei will collapse on the negatively charged external charges.

⇒ some MM atoms possess no charge and so would be invisible to the QM atoms

⇒ the van der Waals terms on the MM atoms often provide the only difference in the interactions of one atom type versus another, i.e. chloride and bromide ions both have unit negative charge and only differ in their van der Waals terms.

A Hybrid QM/MM Model

So, it is quite reasonable to attribute the van der Waals parameters (as it is in the MM method) to every QM atom and the Hamiltonian describing the interaction between the QM and MM atoms can have a form:

$$\hat{H}_{QM/MM} = -\sum_{i}^{electrons}\sum_{j}^{MM\ atoms}\frac{Q_j}{r_{ij}} + \sum_{i}^{nuclei}\sum_{j}^{MM\ atoms}\frac{Z_i Q_j}{R_{ij}} + \sum_{i}^{nuclei}\sum_{j}^{MM\ atoms}\left\{\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right\}$$

The van der Waals term models also electronic repulsion and dispersion interactions, which do not exist between QM and MM atoms because MM atoms possess no explicit electrons.

A. Warshel, M. Levitt // Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. // ***J.Mol.Biol. 103(1976), 227-49***

QM/MM implementations

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…

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

Simple QM/QM implementation

- 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

QM/MM implementations

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):

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

QM/MM implementations

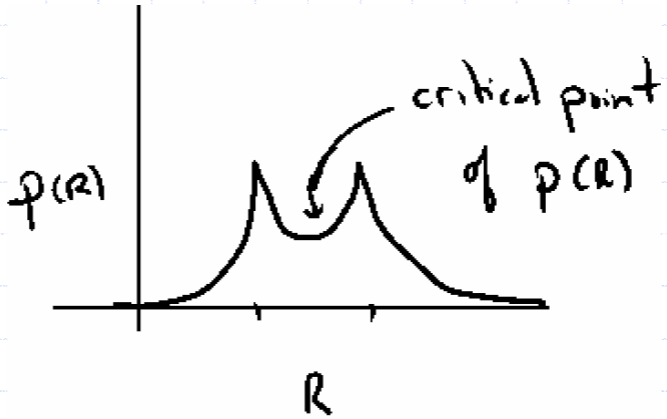
Atomic Charge Schemes

- "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!

QM/MM implementations

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"



QM/MM implementations

Charge Schemes 2

- 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

The Hybrid QM/MM Model

Now we can construct a "real" hybrid QM/MM Hamiltonian:

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

$$\hat{H}_{QM/MM} = -\sum_{i}^{electrons}\sum_{j}^{MM\ atoms}\frac{Q_j}{r_{ij}} + \sum_{i}^{nuclei}\sum_{j}^{MM\ atoms}\frac{Z_i Q_j}{R_{ij}} + \sum_{i}^{nuclei}\sum_{j}^{MM\ atoms}\left\{\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right\}$$

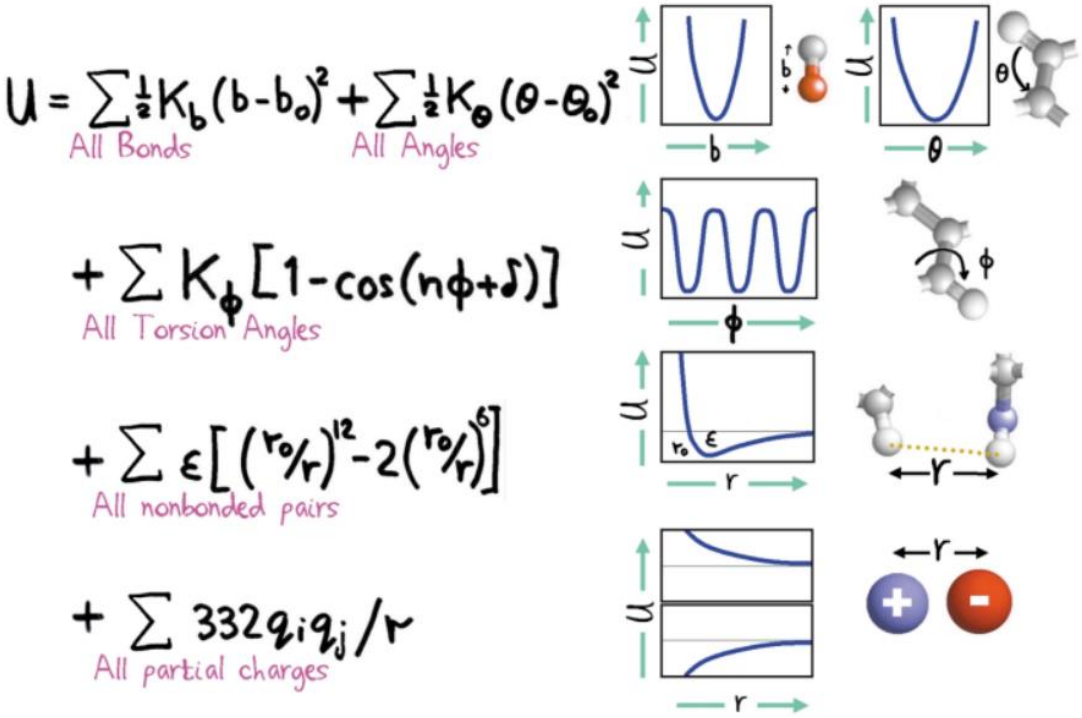
$$\hat{H}^{el}_{QM/MM} = -\sum_{i}^{electrons}\sum_{j}^{MM\ atoms}\frac{Q_j}{r_{ij}} + \sum_{i}^{nuclei}\sum_{j}^{MM\ atoms}\frac{Z_i Q_j}{R_{ij}}$$

$$E = \left\langle \Psi \middle| \hat{H}_{QM} + \hat{H}^{el}_{QM/MM} \middle| \Psi \right\rangle + E^{vdw}_{QM/MM} + E_{MM}$$

A "standard" MM force field can be used to determine the MM energy. For example, CHARMM force field has a form:

$$E_{MM} = \sum_{nonbonded}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{R_{ij}}\right] + \sum_{bonds}\frac{K_b}{2}(R - R_0)^2 + \sum_{angles}\frac{K_\theta}{2}(\theta - \theta_0)^2 + \sum_{dihedrals} V_\phi\left(1 + \cos(n\phi + \delta)\right)$$

The Hybrid QM/MM Model



A “standard” MM force field can be used to determine the MM energy.
For example, CHARMM force field has a form:

$$E_{MM} = \sum_{nonbonded}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{R_{ij}}\right] + \sum_{bonds}\frac{K_b}{2}(R-R_0)^2 + \sum_{angles}\frac{K_\theta}{2}(\theta-\theta_0)^2 + \sum_{dihedrals} V_\phi\left(1+\cos(n\phi+\delta)\right)$$

What does the QM program provide?

We need the QM/MM gradients corresponding to the total energy.

$$E = \left\langle \Psi \middle| \hat{H}_{QM} + \hat{H}^{el}_{QM/MM} \middle| \Psi \right\rangle + E^{vdw}_{QM/MM} + E_{MM}$$

Forces on QM nuclei
(gradient of the energy)

Classical gradient

For derivatives
On MM atoms:

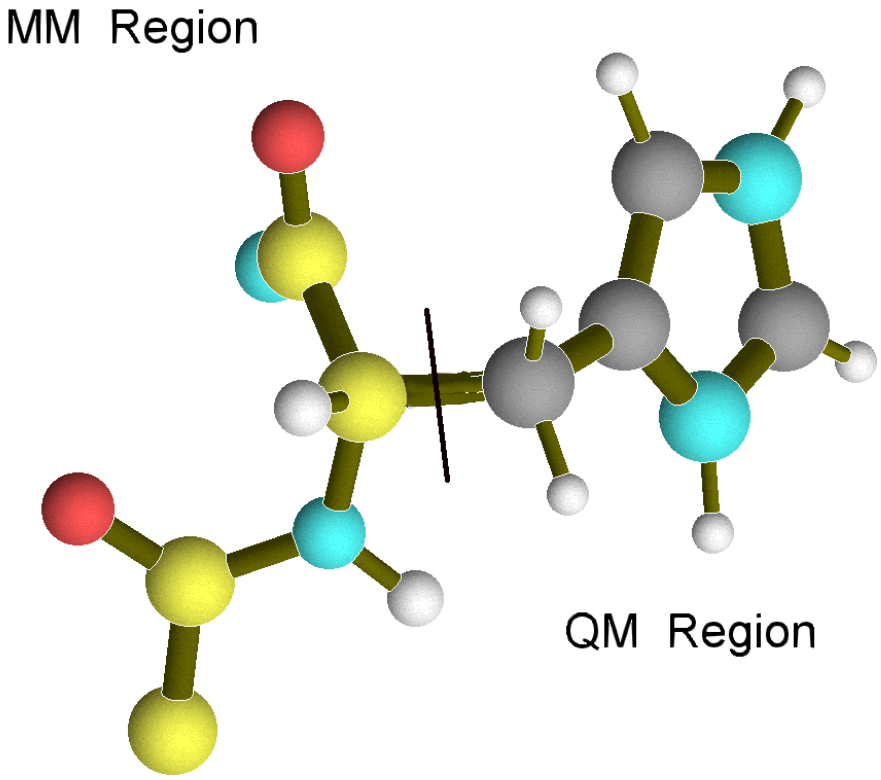
$$\sum_{i \in MM} q_i \int \frac{\rho_{QM}(r)}{|r - r_i|} dr$$

Gradient of the potential (field) at the position of the MM atoms (x q_i)

To evaluate the gradient (and energy) of the QM/MM potential at given nuclei positions, the following practical information is required from the QM calculations:

- Gradient of the wave function in the position of the QM atoms
- Derivative of the electrostatic potential arising from the QM wave function at the position of the MM atoms (Field)
- (Energy of polarized QM system)

Dividing Covalent Bonds across the QM and MM Regions



In many simulations it is necessary to have the QM/MM boundary cut covalent bonds, and a number of additional approximations have to be made.

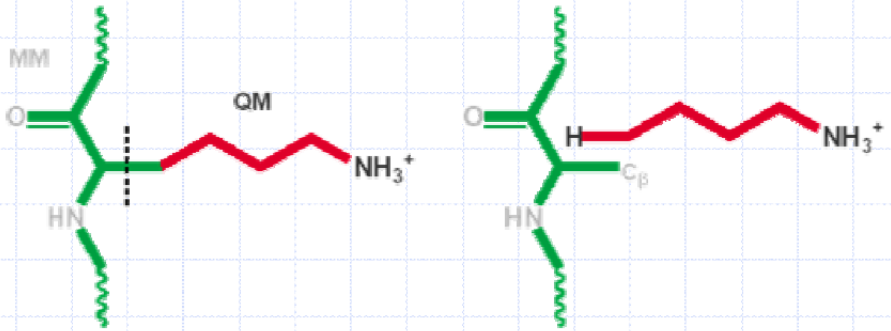
QM/MM implementations

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

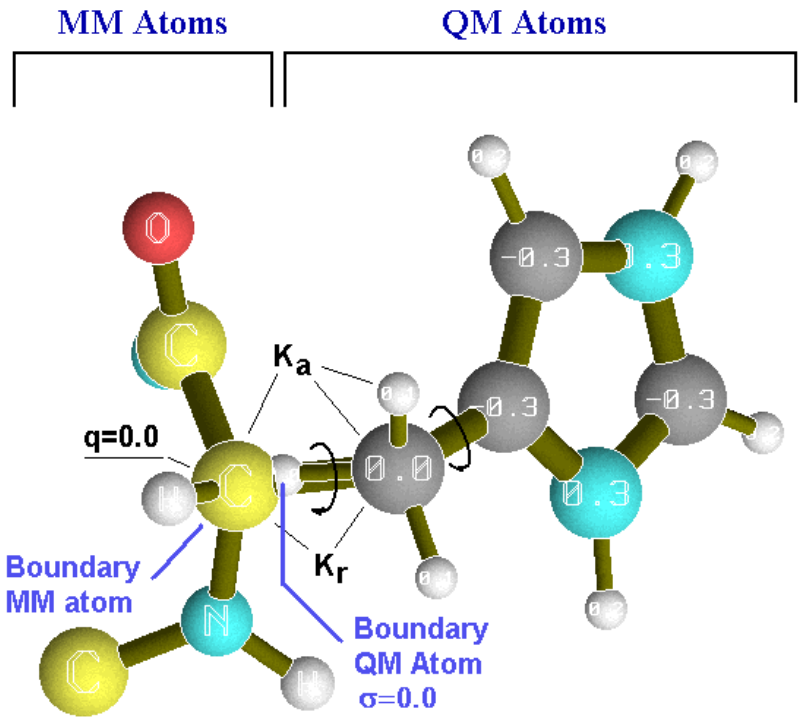


QM/MM implementations

Covalent Embedding 2

- 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

Implementation of "link" atom approach



The link atom is placed along the bond vector joining the QM and MM atom

The default link atom type is hydrogen

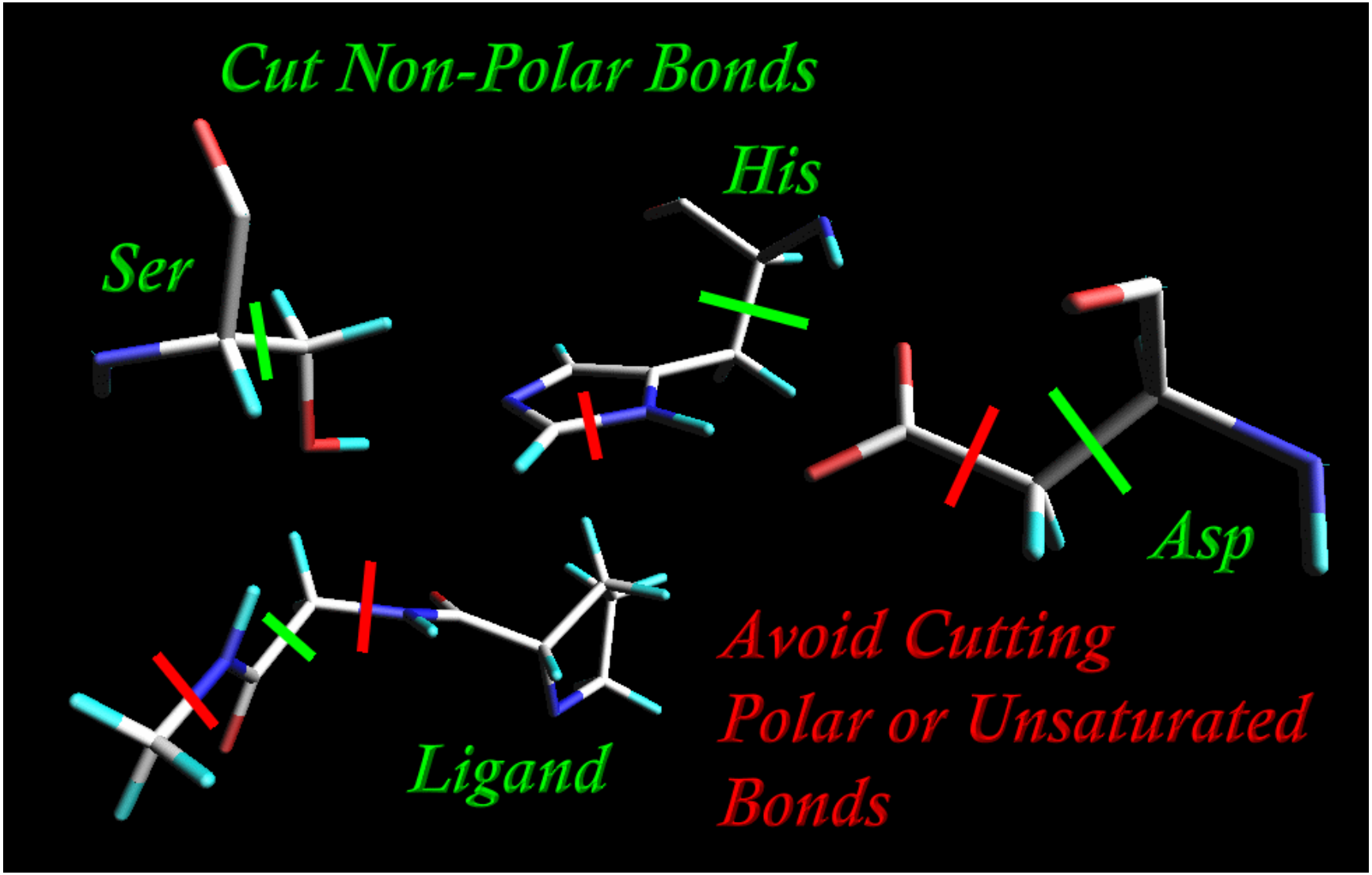
It interacts with MM region only electrostatically (no VDW term).

VdW interaction between QM and MM atoms which form 1-2 and 1-3 "bonded" pairs *is not calculated*.

Bond stretching, angle bending, and torsion interactions between QM and MM regions are calculated as those in MM if 1-2, 1-2-3, or 1-2-3-4 terms contain at least one MM atom

Hints for running QM/MM calculations

Choosing the QM region



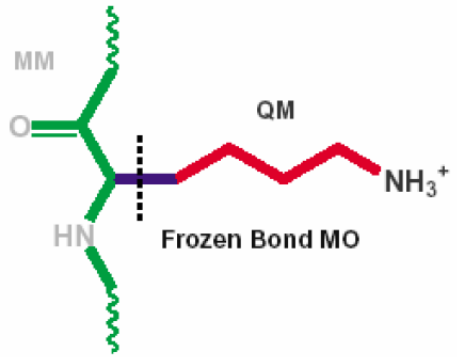
QM/MM implementations

Covalent Embedding 4

Frozen orbital ideas:

Rivail & co-workers (1994)

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



QM/MM implementations

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

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

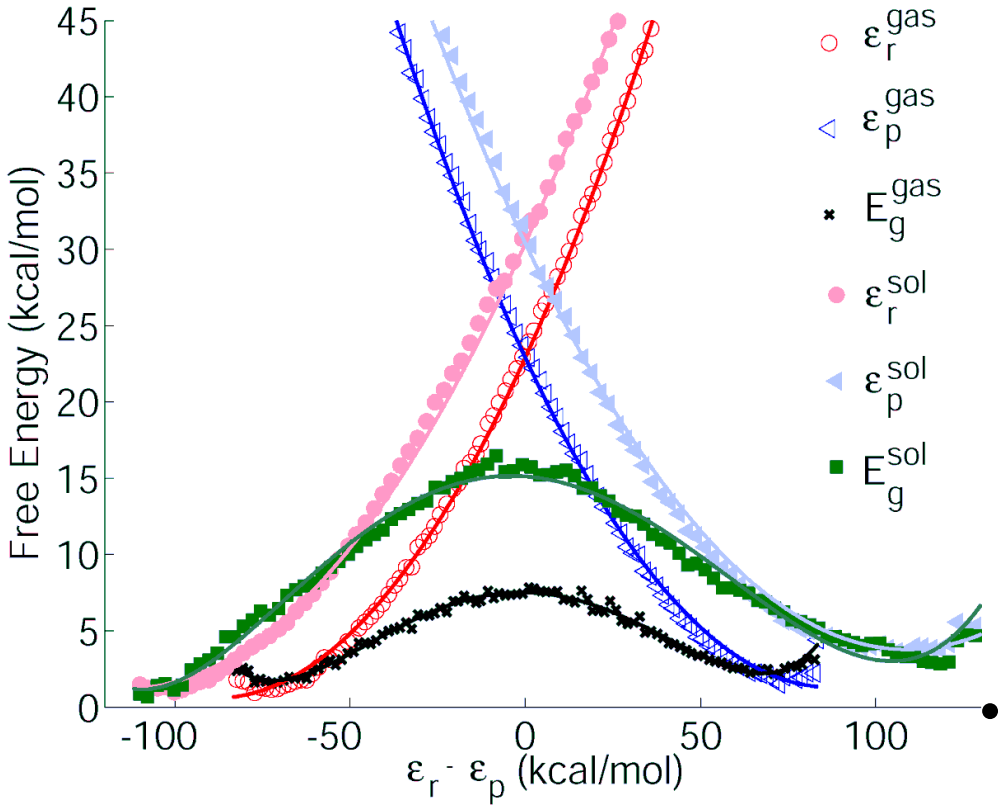
QM/MM implementations

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

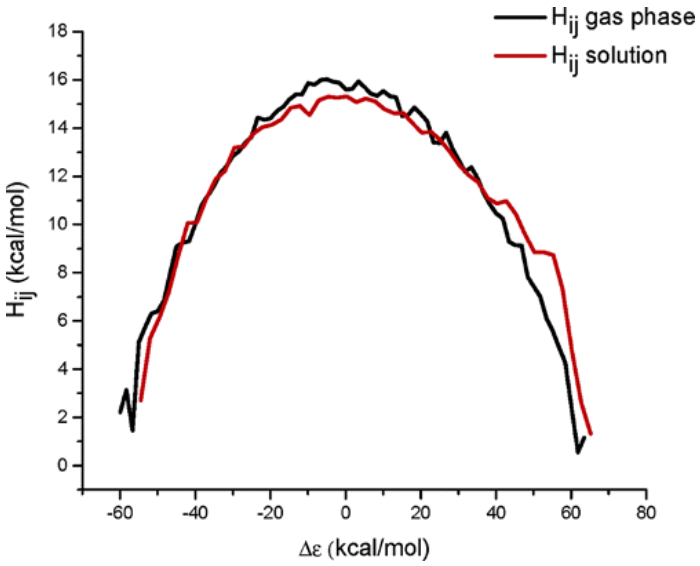
Classical alternative: EVB

$$Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$$



Hong, Rosta, Warshel; J. Chem. Phys., 2006

- H_{rp} is the approximately the same in gas phase and in solution



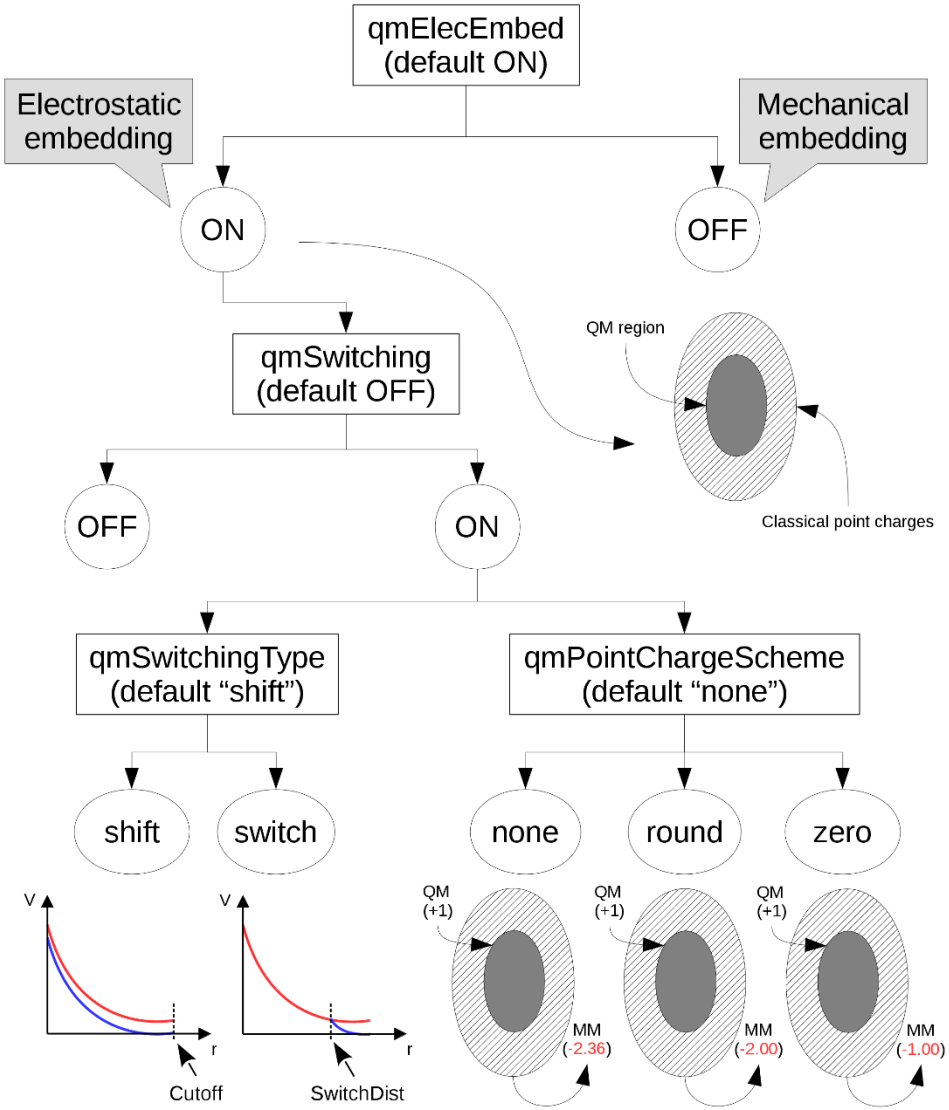
- Calculated free energies: ε_r ε_p with FDFT and E_g with DFT

$$H_{rp} = \sqrt{(\varepsilon_r - E_g)(\varepsilon_p - E_g)}$$

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

QM/MM implementation in NAMD

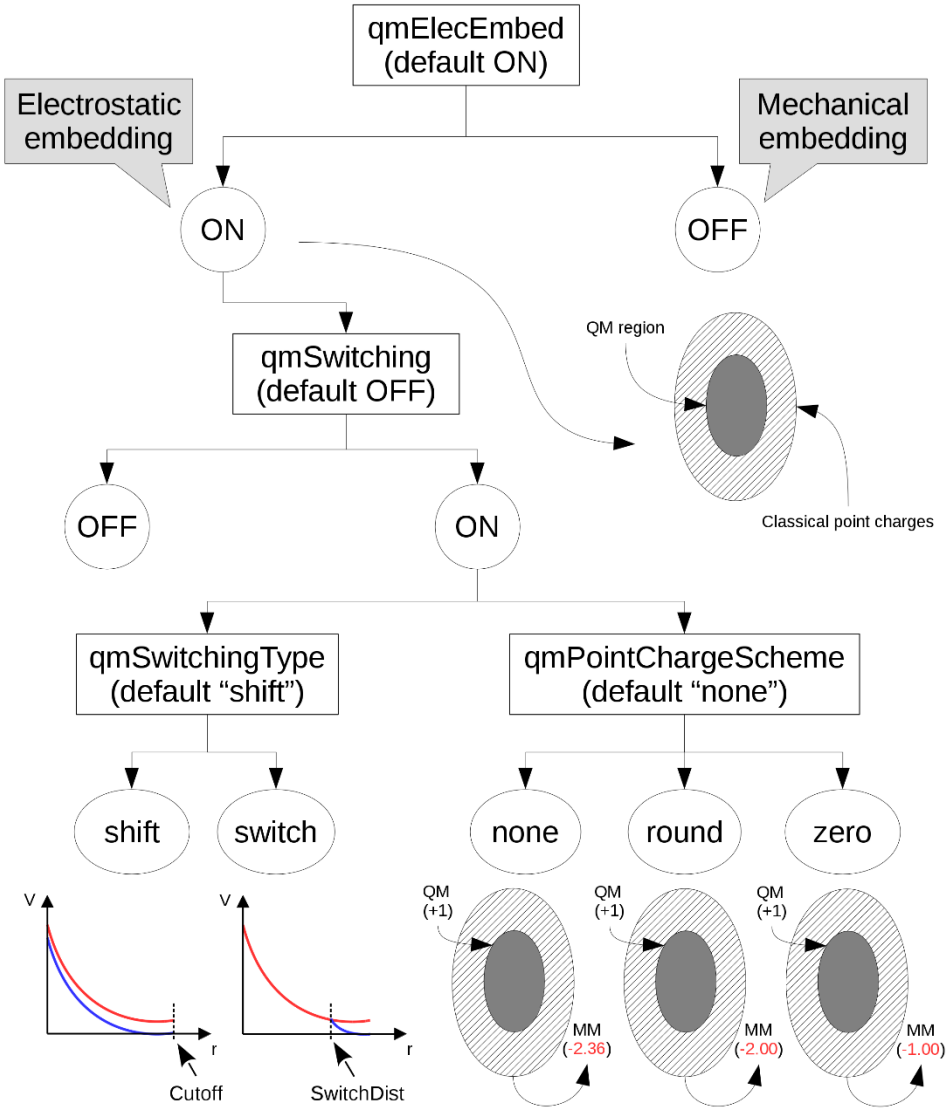
- Electrostatic & Mechanical embedding
- Possible switching function to cut off long range electrostatics



QM/MM implementation in NAMD

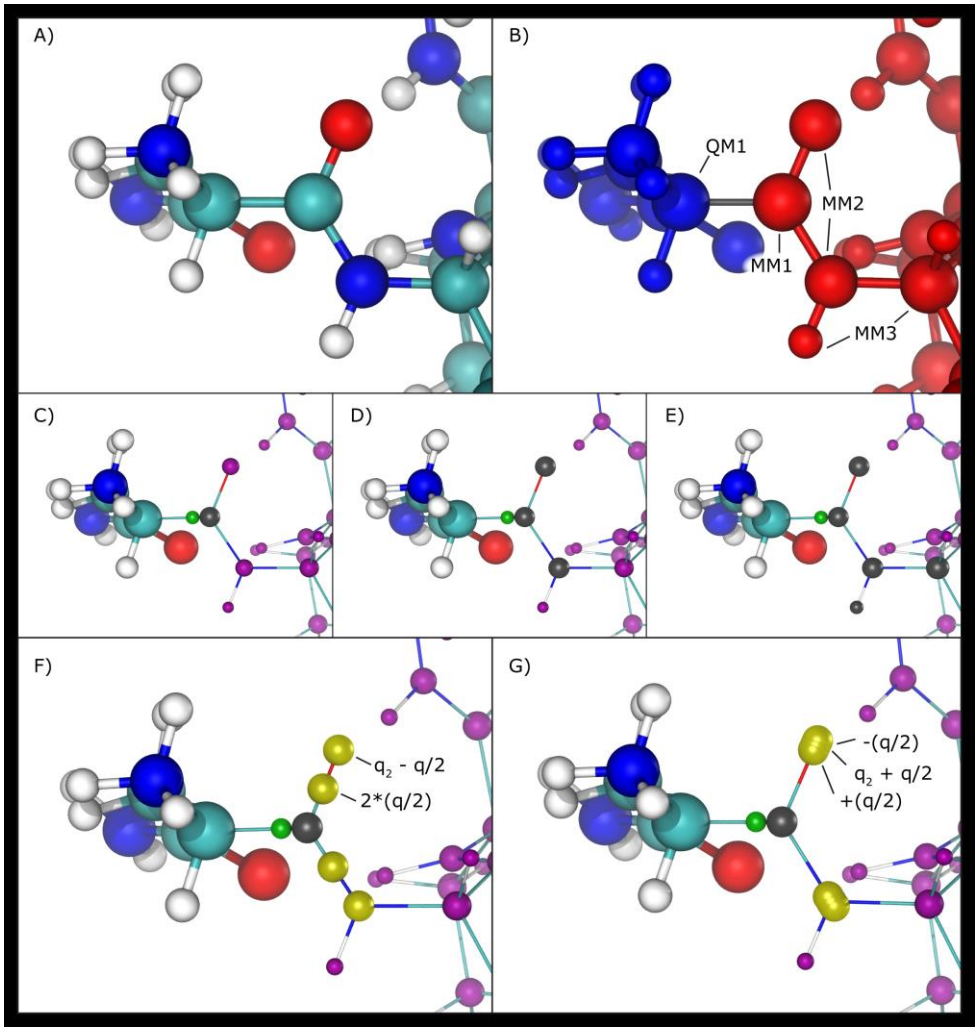
- NAMD uses partial charges for the QM atoms to calculate the electrostatic interaction with classical point charges.
- There are two possibilities for the origin of the QM partial charges:
  - the original partial charges found in the force field parameter files can be used
  - updated partial charges can be gathered at each step form the QM software output (controllable through the "qmChargeMode" keyword).

The continuous update in charge distribution allows for a partial re-parameterization of the targeted molecule at each time step.



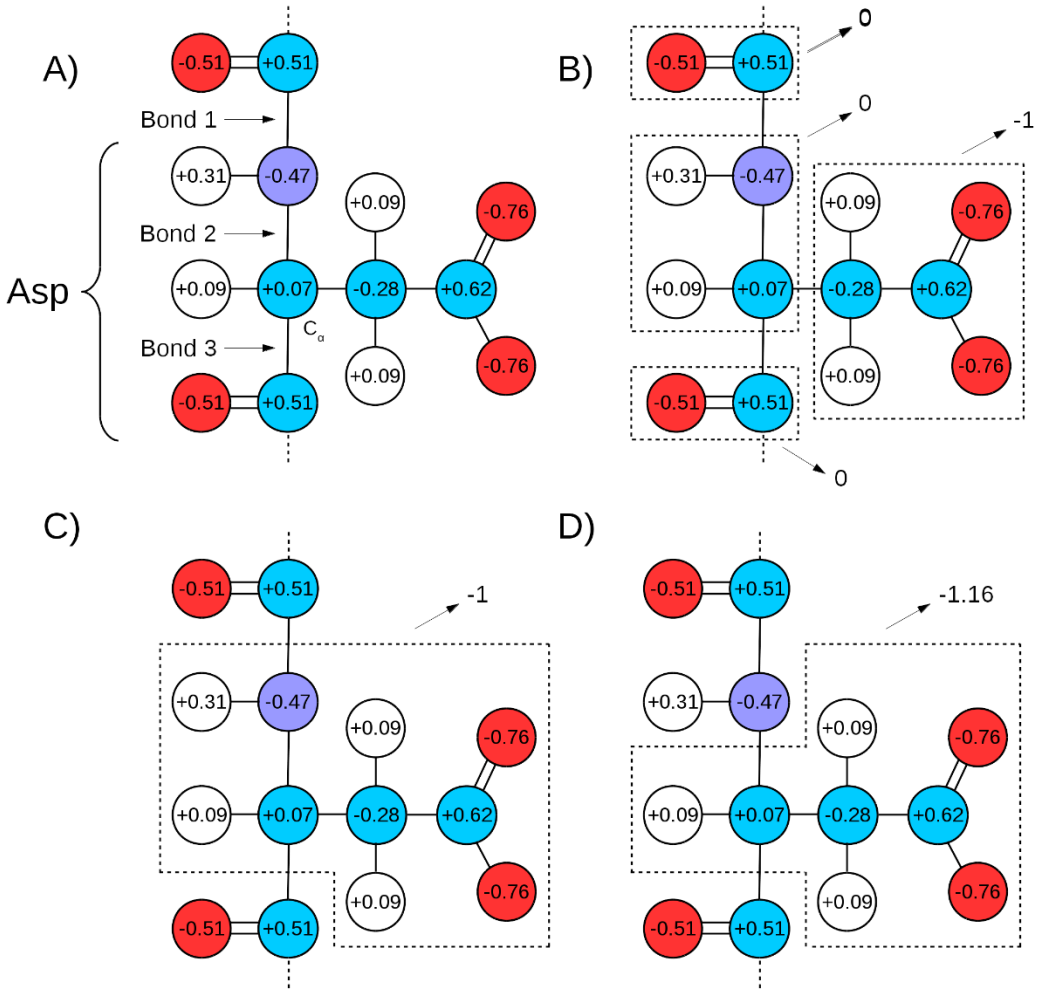
QM/MM implementation in NAMD

- Link atom treatment
- keywords "qmBondDist" to scale the bond length
- "qmLinkElement“ to change the terminating atom type



QM/MM implementation in NAMD

- Several charge redistribution scheme to remove partial charges from nearby QM atoms
- "qmBondScheme" keyword
- No charge or Redistributed Charge and Dipole (RCD) methods

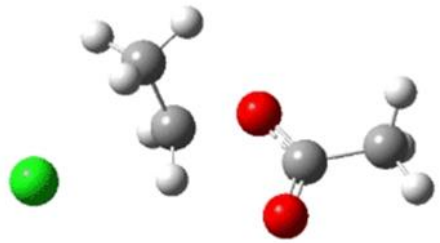


QM/MM implementation in NAMD

- ORCA and MOPAC natively supported
- python wrapper scripts for **GAUSSIAN**, **TeraChem** and **Q-Chem**

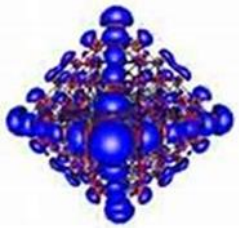
ORCA

Q-CHEM



gaussian.com

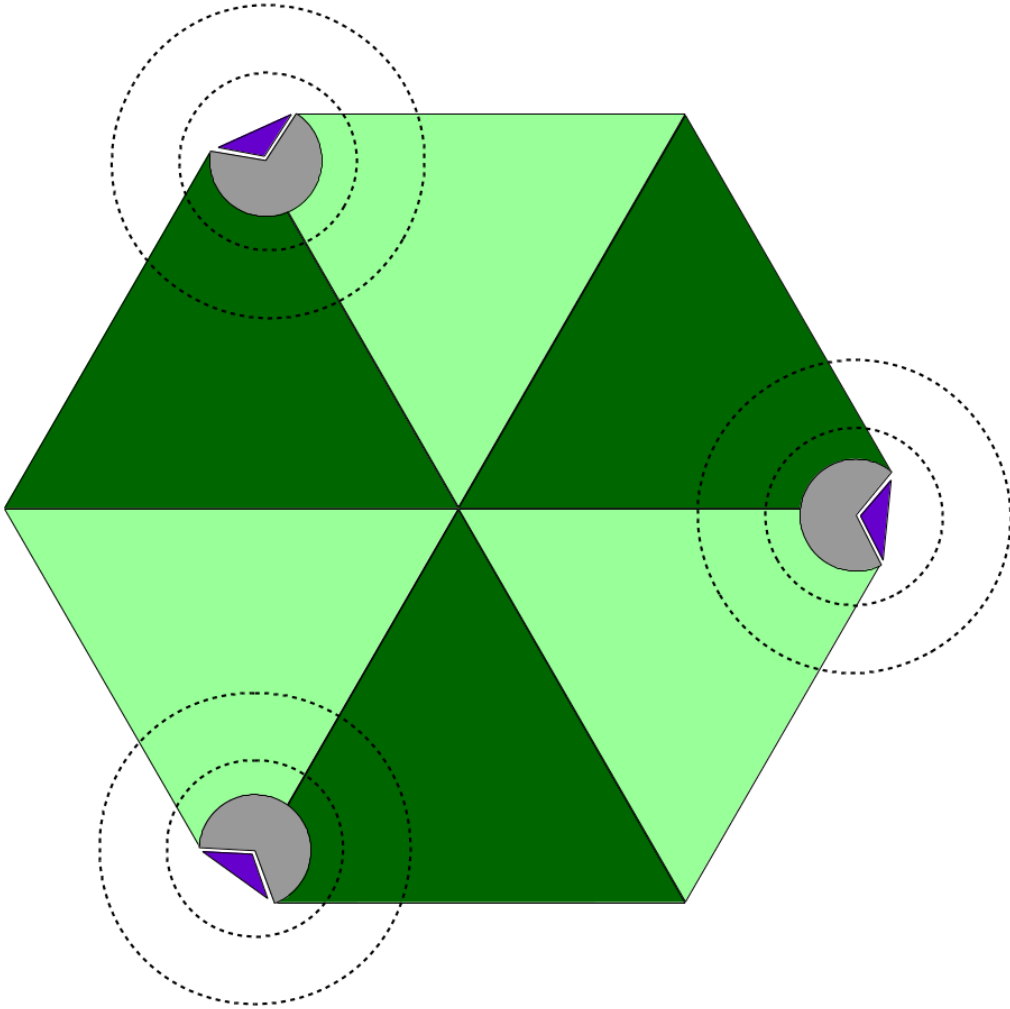
TURBOMOLE



QM/MM implementation in NAMD

- Multiple QM regions are supported
- More on keywords, etc:

http://www.ks.uiuc.edu/Research/qmmm/



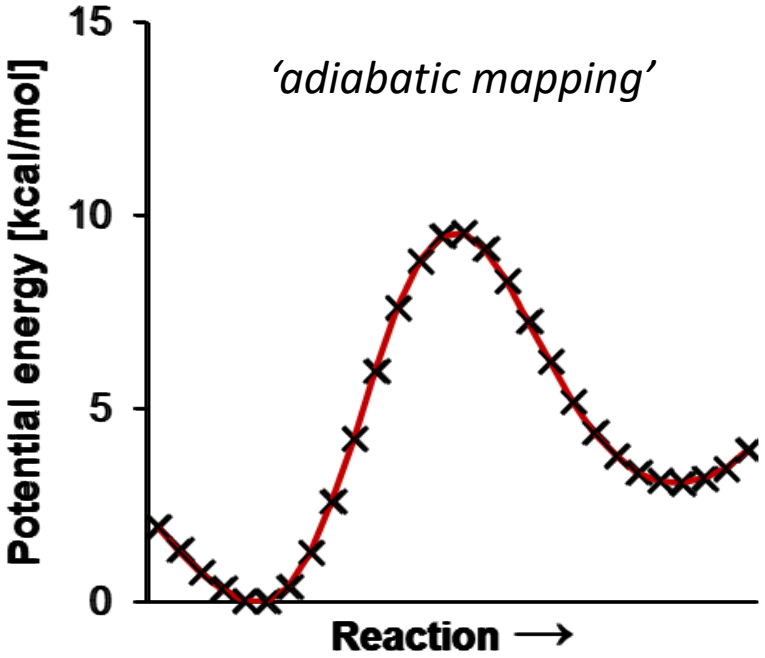
QM/MM reaction modelling

- To overcome activation energy, need to ‘force’ a reaction happening: apply ‘bias’ along a ‘reaction coordinate’

Potential Energy Surface

QM/MM minimization

‘adiabatic mapping’

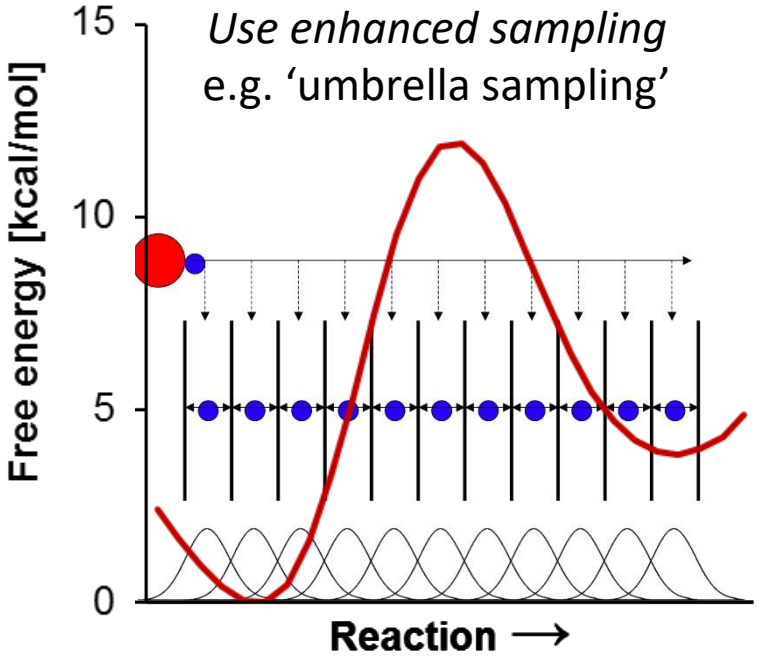


Free Energy Surface

QM/MM MD + statistics

Use enhanced sampling

e.g. ‘umbrella sampling’



Dynamics on QM/MM potential

Minimizations corresponding to 0K is often problematic for larger systems:

- Missing entropy effects are important in e.g., protein folding, ligand binding, etc.
- Even is enthalpic contributions dominate, only local minima are found

There are several methods to model the thermostatted movements of the atoms on a given potential. These should provide a proper thermodynamic ensemble at long time scales.

A popular one is Langevin dynamics to obtain canonical ensemble:

$$M\ddot{X} = -\nabla U(X) - \gamma M\dot{X} + \sqrt{2\gamma k_B T M}R(t)\,,$$

R(t) is a delta-correlated stationary Gaussian process with zero-mean and a constant variance.

γ: damping constant

Others: Andersen, Nose-Hoover, etc…

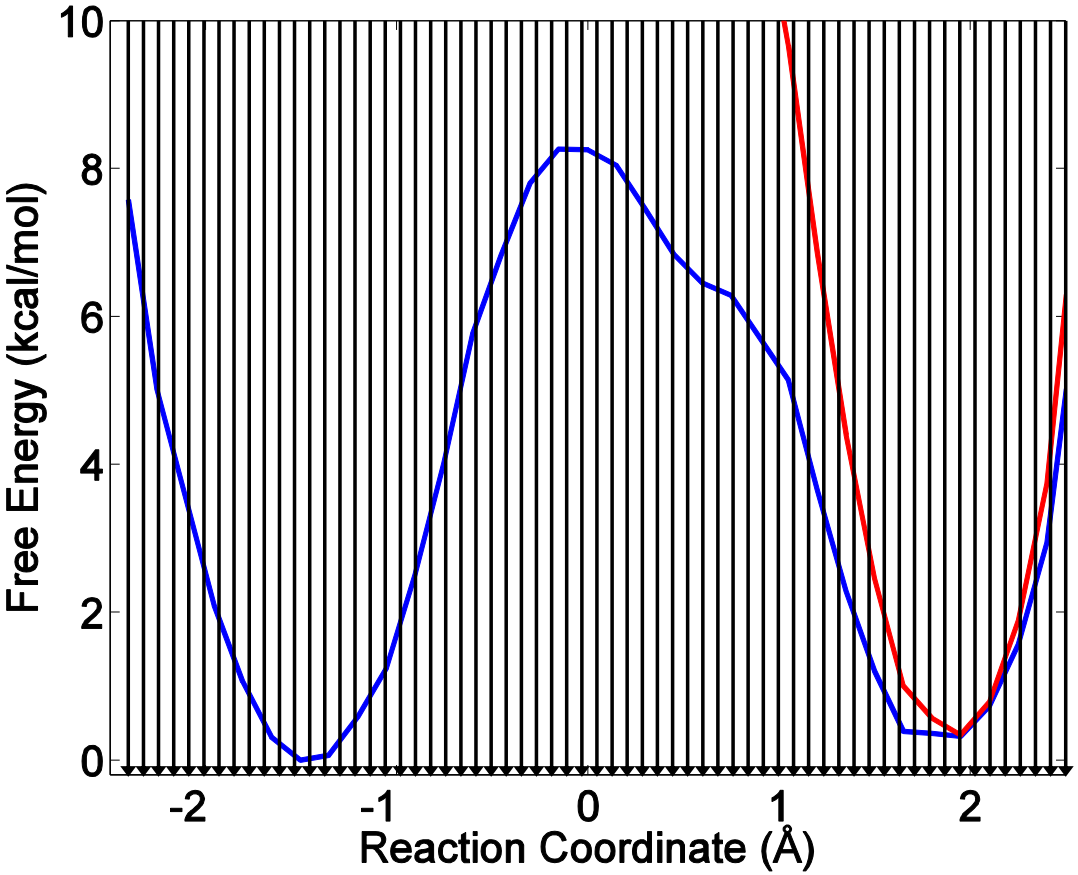
1. Umbrella Sampling

- Run parallel simulations with harmonic constraints moving along the reaction coordinate

Torrie and Valleau; *J. Comp. Phys.* 1977
Cited: ~ 3200 times

- Recover the unbiased free energy surface from combined data using e.g., WHAM

$$E_i(q_A) = U_{pot}(q_A) + \frac{1}{2} k_i (\xi_A - \xi_i)^2$$



Ferrenberg, Swendsen; *Phys. Rev. Lett*. 1989 - Cited: ~ 2800 times
Kumar, Rosenberg, Bouzida, Swendsen, Kollman; *J. Comput. Chem.* 1992 (WHAM) - Cited: ~ **3600** times

1. WHAM

$$\Pr^{(k)} \propto \prod_{i=1}^{Nbin} \left(p_i^{(k)} \right)^{n_i^{(k)}}$$

Probability of observing a trajectory in the *k*-th simulation.

$$\tilde{L} = \ln \prod_{k=1}^{NSim} \prod_{i=1}^{Nbin} \left(p_i^{(k)} \right)^{n_i^{(k)}}$$

$p_i^{(k)}$ Equilibrium probability for bin *i*

$n_i^{(k)}$ Histogram count in bin *i*

$$p_i^{(k)} = f^{(k)} e^{-\beta U_i^{(k)}} p_i = \frac{e^{-\beta U_i^{(k)}} p_i}{\sum_{j=1}^{N_{bin}} e^{-\beta U_j^{(k)}} p_j}$$

$e^{\frac{-U_i^{(k)}}{k_B T}}$ Boltzmann factor for biasing energy

$f^{(k)} = \frac{1}{\sum_{j=1}^{N_{bin}} e^{-\beta U_j^{(k)}} p_j}$ Normalizing factor for the equilibrium probability

$$L = \ln \prod_{k=1}^{NSim} \prod_{i=1}^{Nbin} \left(f^{(k)} e^{-\beta U_i^{(k)}} p_i \right)^{n_i^{(k)}} + \sum_{k=1}^{NSim} \lambda^{(k)} \left(1 - \sum_{i=1}^{Nbin} f^{(k)} e^{-\beta U_i^{(k)}} p_i \right)$$

1. WHAM

$$\Pr^{(k)} \propto \prod_{i=1}^{Nbin} \left(p_i^{(k)} \right)^{n_i^{(k)}}$$

Probability of observing a trajectory in the *k*-th simulation.

$$\tilde{L} = \ln \prod_{k=1}^{NSim} \prod_{i=1}^{Nbin} \left(p_i^{(k)} \right)^{n_i^{(k)}}$$

$p_i^{(k)}$ Equilibrium probability for bin *i*

$n_i^{(k)}$ Histogram count in bin *i*

$$p_i^{(k)} = f^{(k)} e^{-\beta U_i^{(k)}} p_i = \frac{e^{-\beta U_i^{(k)}} p_i}{\sum_{j=1}^{N_{bin}} e^{-\beta U_j^{(k)}} p_j}$$

$$\frac{\partial L}{\partial f^{(k)}} = 0 \qquad \frac{\partial L}{\partial p_i} = 0$$

$$p_i = \frac{\sum_{k=1}^{Msim} n_i^{(k)}}{\sum_{l=1}^{Msim} N^{(l)} f^{(l)} e^{-\beta U_i^{(l)}}}$$

$$L = \ln \prod_{k=1}^{NSim} \prod_{i=1}^{Nbin} \left(f^{(k)} e^{-\beta U_i^{(k)}} p_i \right)^{n_i^{(k)}} + \sum_{k=1}^{NSim} \lambda^{(k)} \left(1 - \sum_{i=1}^{Nbin} f^{(k)} e^{-\beta U_i^{(k)}} p_i \right)$$

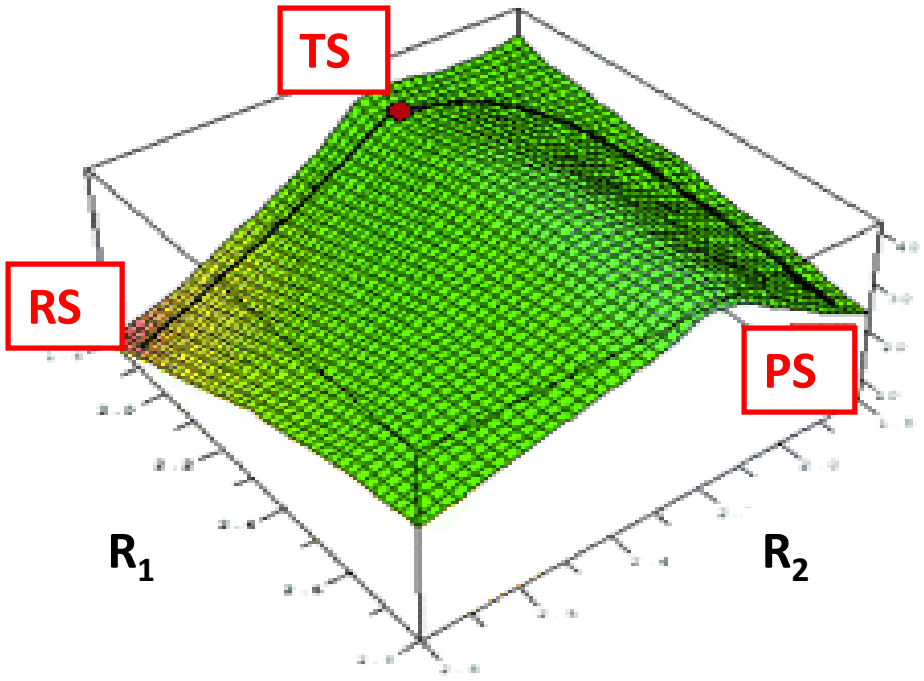
1. Free Energy Simulations Using the String Method

- String-type methods have become highly successful in recent years

  Henkelman, Jonsson, W. Yang, Brooks, …

- Optimized a 1D string in the multidimensional space of the internal reaction coordinates to obtain minimum free energy path

  E, Ren, Vanden-Eijnden, *Phys. Rev. B*, 2002

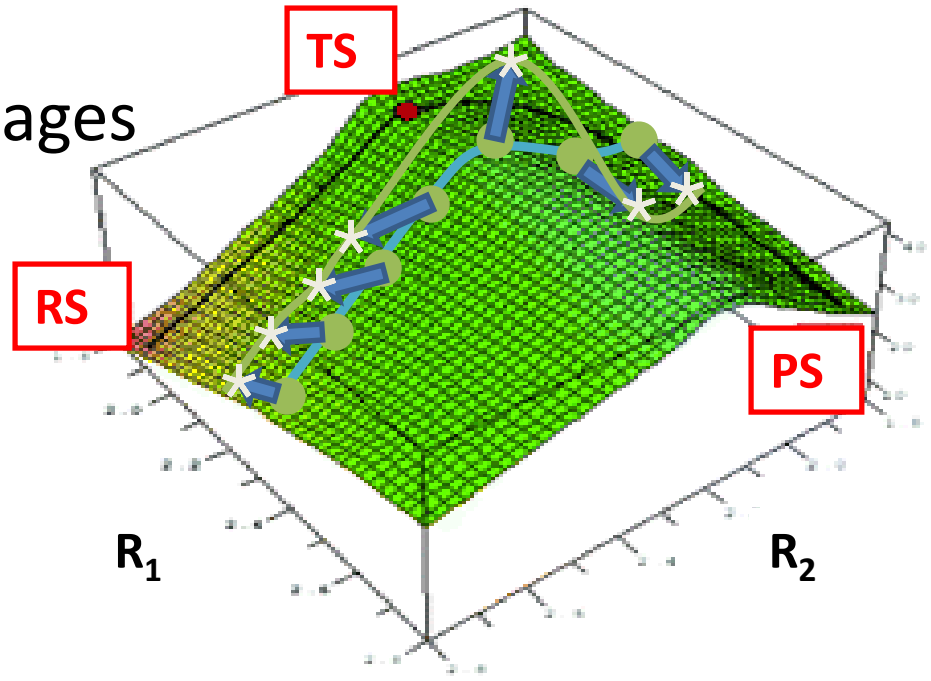


- Hamiltonian replica exchange between string images

  Rosta, Nowotny, Yang, Hummer, *J. Am. Chem. Soc.*, 2011

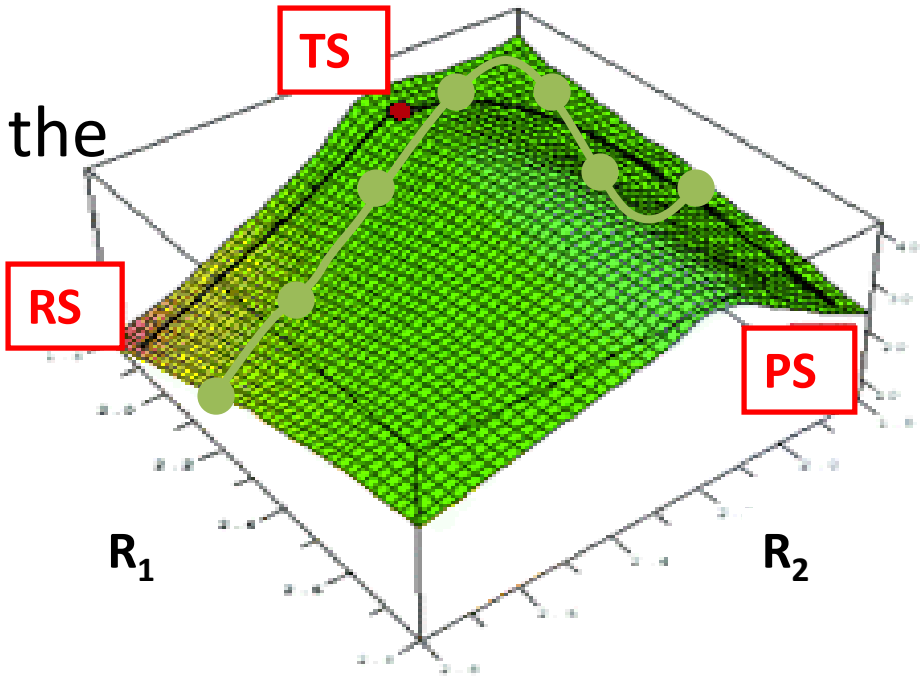
1. Free Energy Simulations Using the String Method

- Start with a guess for the string
- Run Umbrella Sampling simulations
- Determine forces for the images along the string
- Fit new string



1. Free Energy Simulations Using the String Method

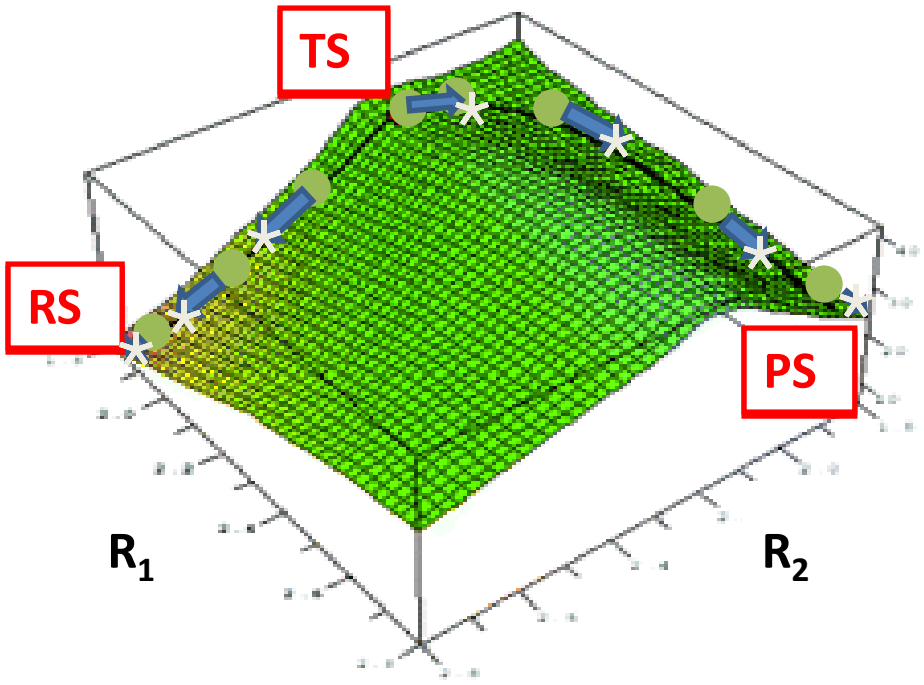
- Start with a guess for the string
- Run Umbrella Sampling simulations
- Determine forces acting on the images along the string
- Fit new string
- Redistribute images
- Run next iteration



1. Free Energy Simulations Using the String Method

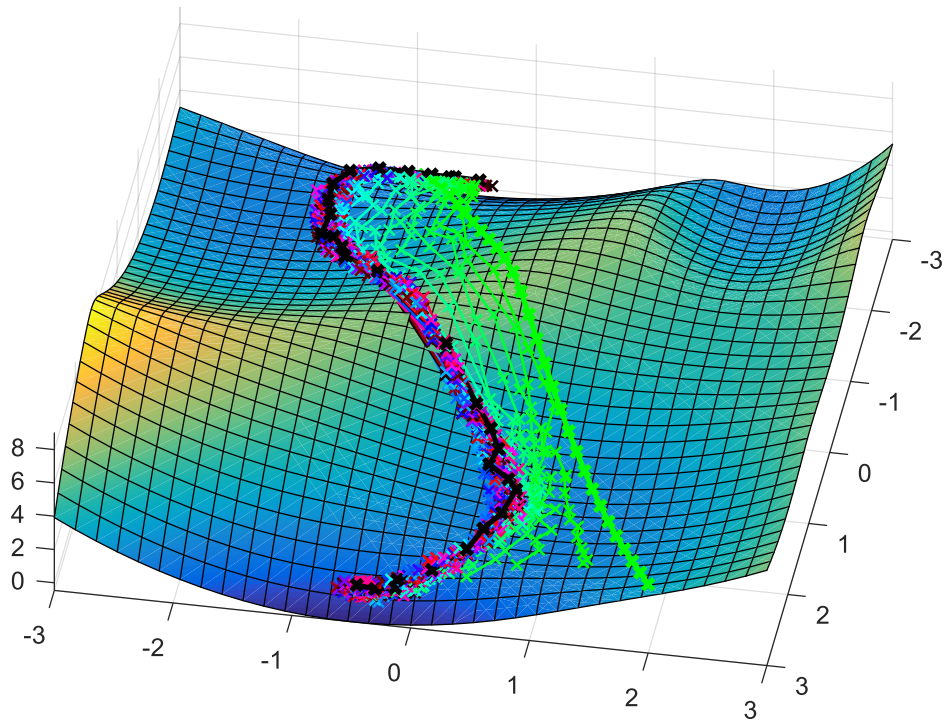
- Converged string:
  - Forces are parallel to string
- We use all data from all string simulations with **Histogram Free** implementation of WHAM: works with **very high dimensionality**

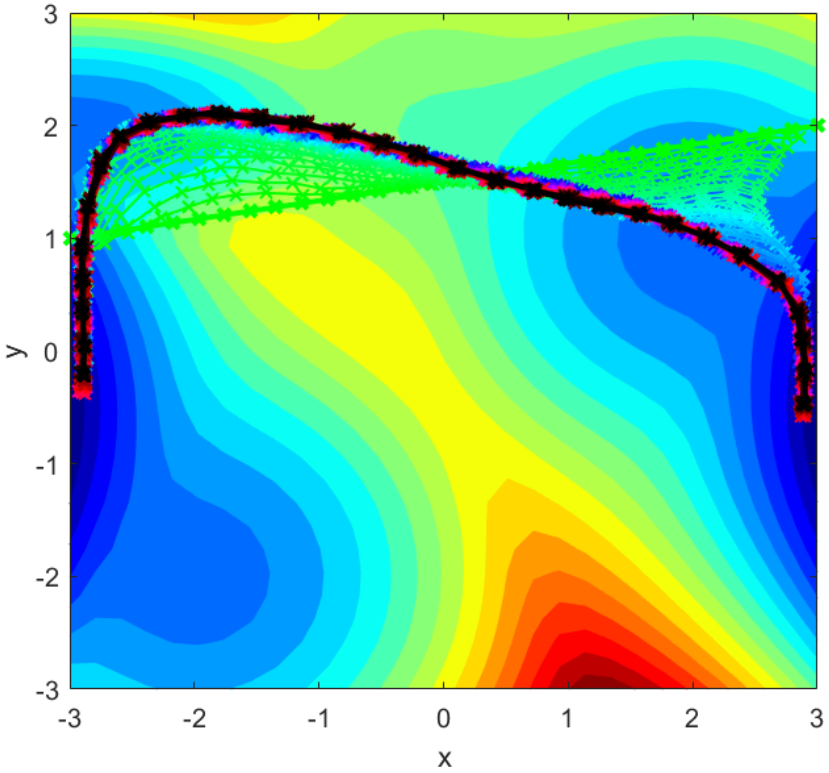
$$\frac{1}{f^{(k)}} = \sum_{k=1}^{NSim} \sum_{A \in NSim} \frac{c^{(k)}\left(\xi_A^k\right)}{\sum_{l=1}^{NSim} N^{(l)} f^{(l)} c^{(l)}\left(\xi_A^k\right)}$$



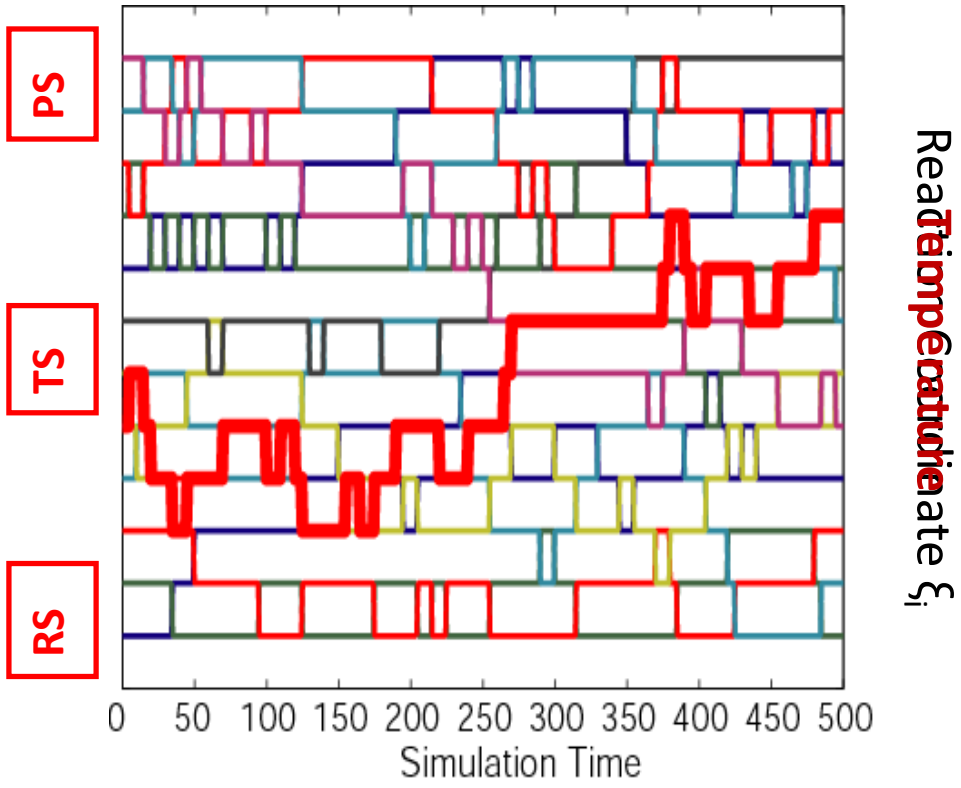
Rosta, Nowotny, Yang, Hummer, *J. Am. Chem. Soc.*, 2011

1. Free Energy Simulations Using the String Method

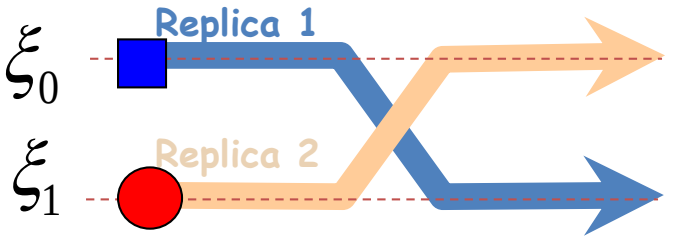




Hamiltonian Replica Exchange

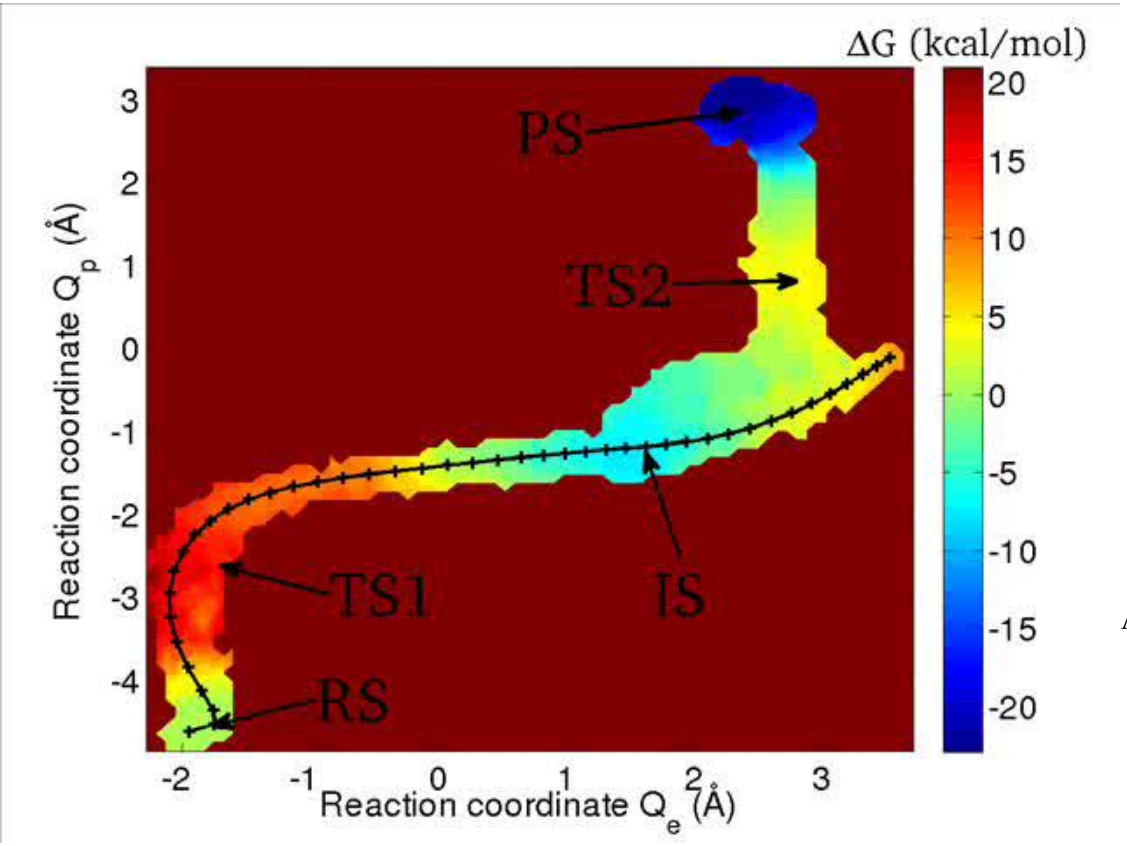


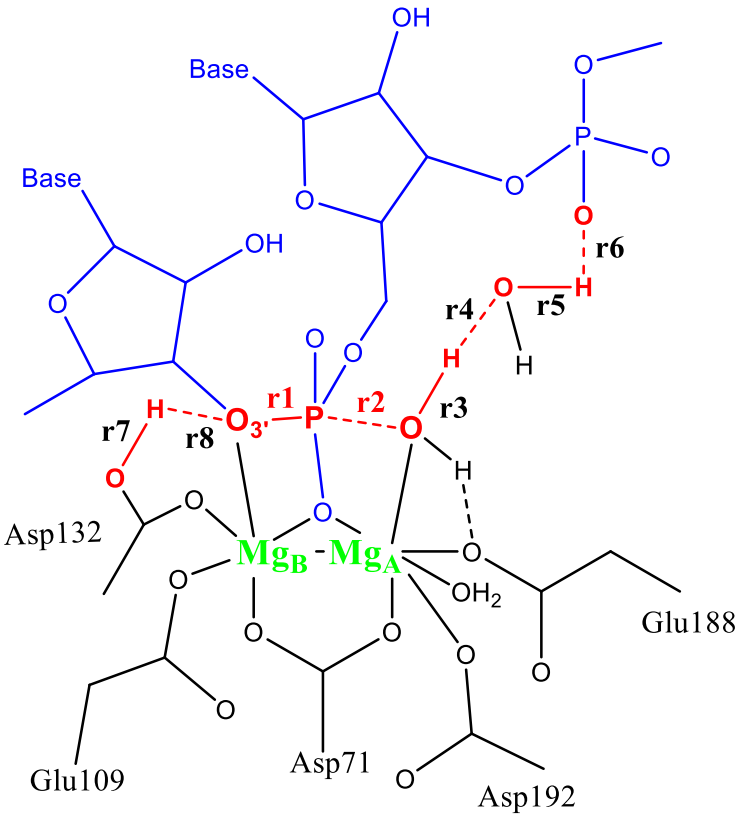
- Running MD at different temperatures in parallel
- Couple the runs in order to speed up lowest temperature's dynamics
- Preserve P_{eq} at each temperature
- Detailed balance condition has to be satisfied



K. Hukushima and K. Nemotto, *J. Phys. Soc. Japan*, 1996
Fukunishi, Watanabe & Takada, *J. Chem. Phys.*, 2002

1. RNase H: Free Energy Surface



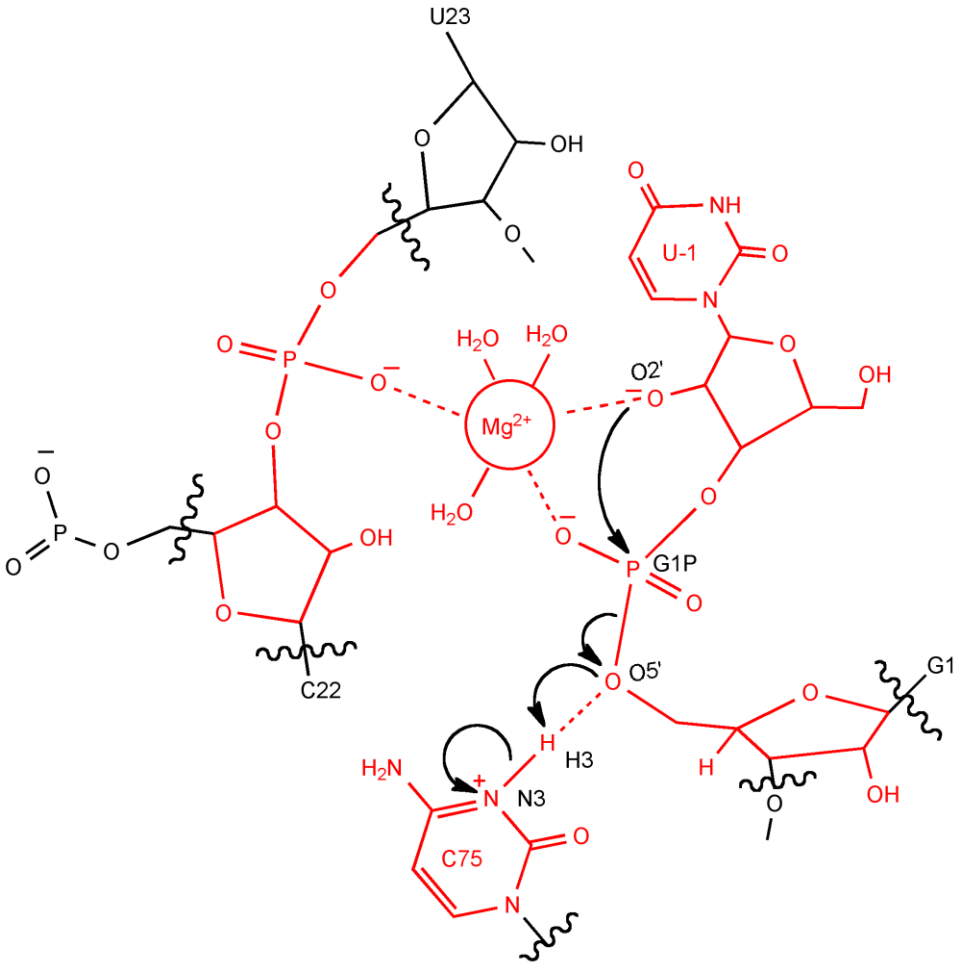


$Q_e=r_1-r_2$

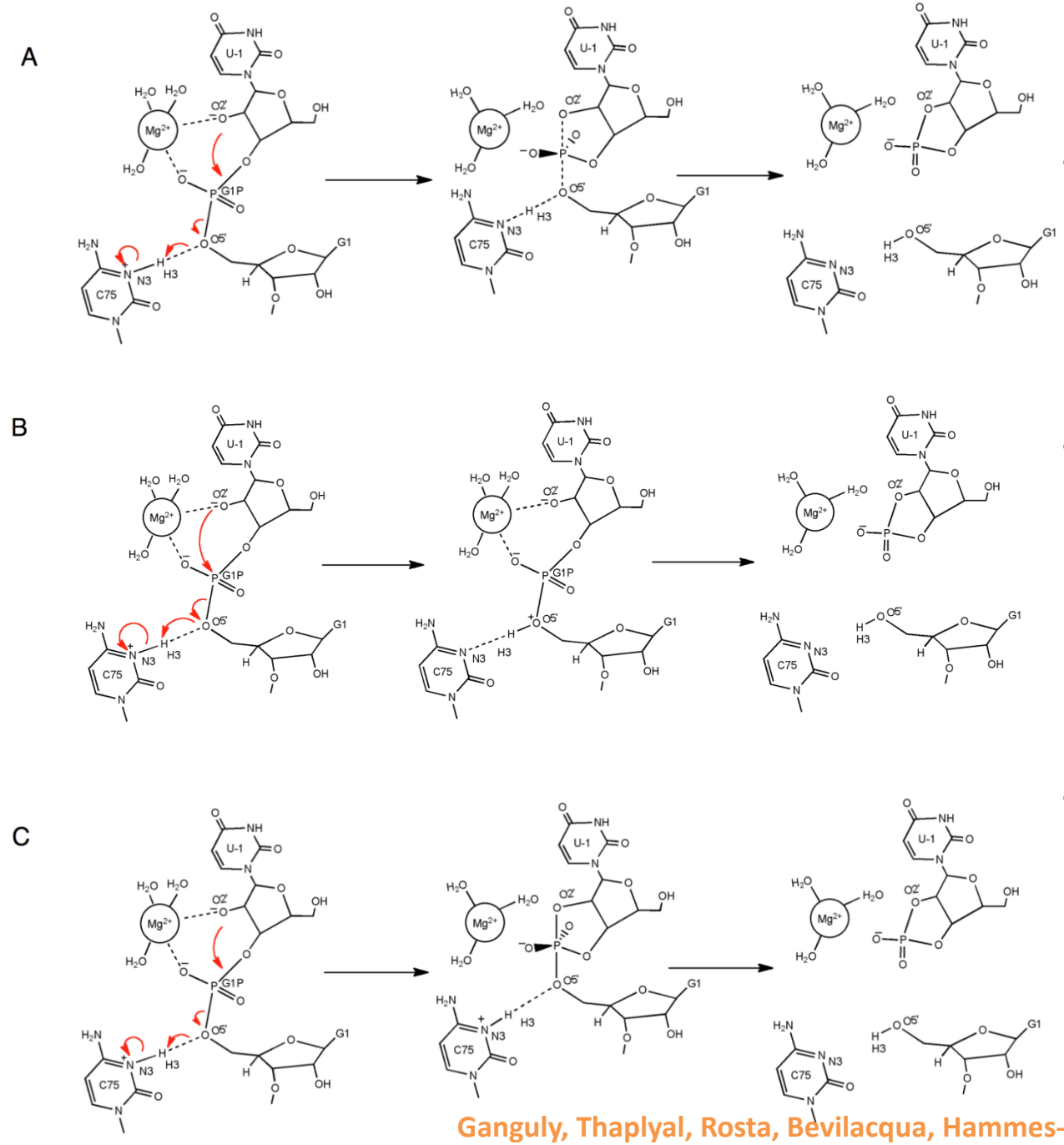
$Q_p=r_3-r_4+r_5-r_6+r_7-r_8$

Rosta, Nowotny, Yang, Hummer, *J. Am. Chem. Soc.*, 2011

HDV ribozyme



Ganguly, Thaplyal, Rosta, Bevilacqua, Hammes-Schiffer, *JACS* 2014

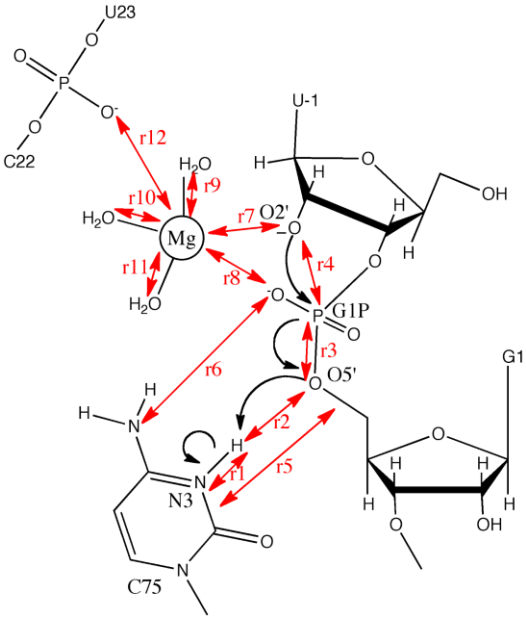


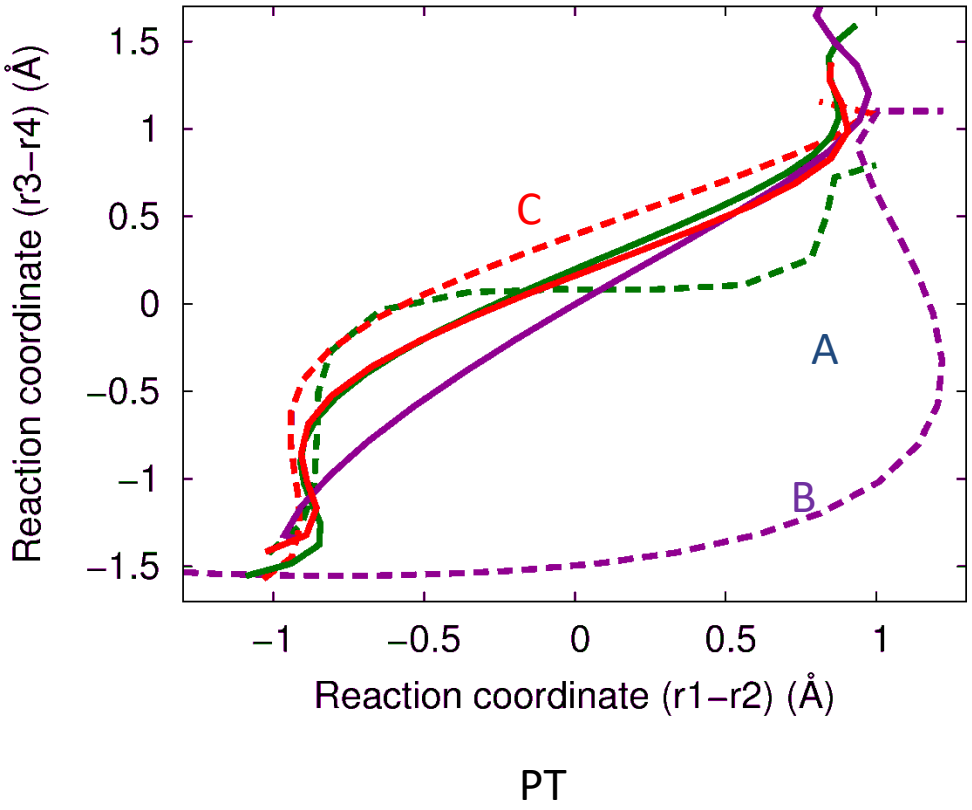
- A: **concerted** mechanism with phosphorane intermediate
- B: **sequential** pathway with a proton-transferred intermediate
- C: **sequential** pathway with a phosphorane intermediate

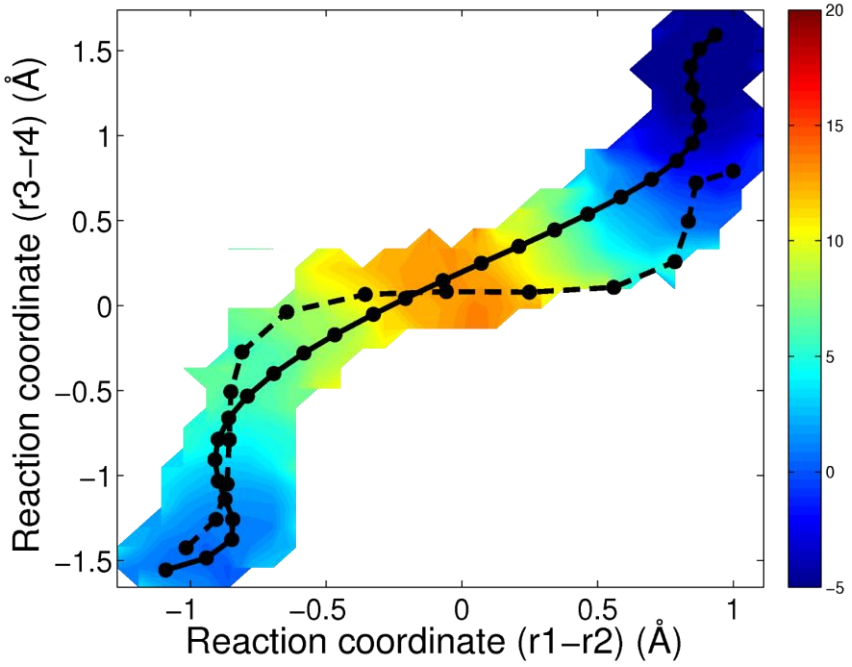
Ganguly, Thaplyal, Rosta, Bevilacqua, Hammes-Schiffer, *JACS* 2014

HDV ribozyme

- Reaction mechanism

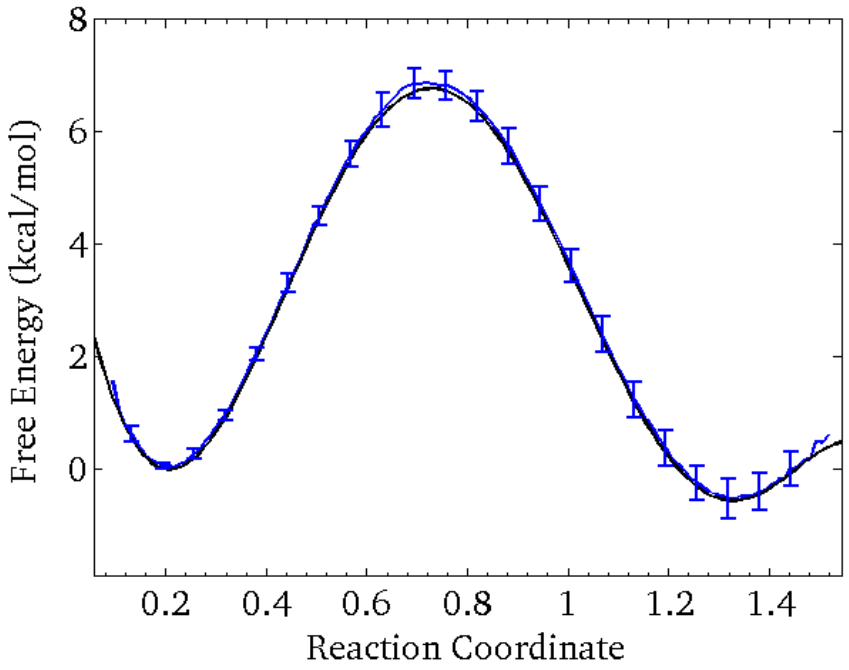


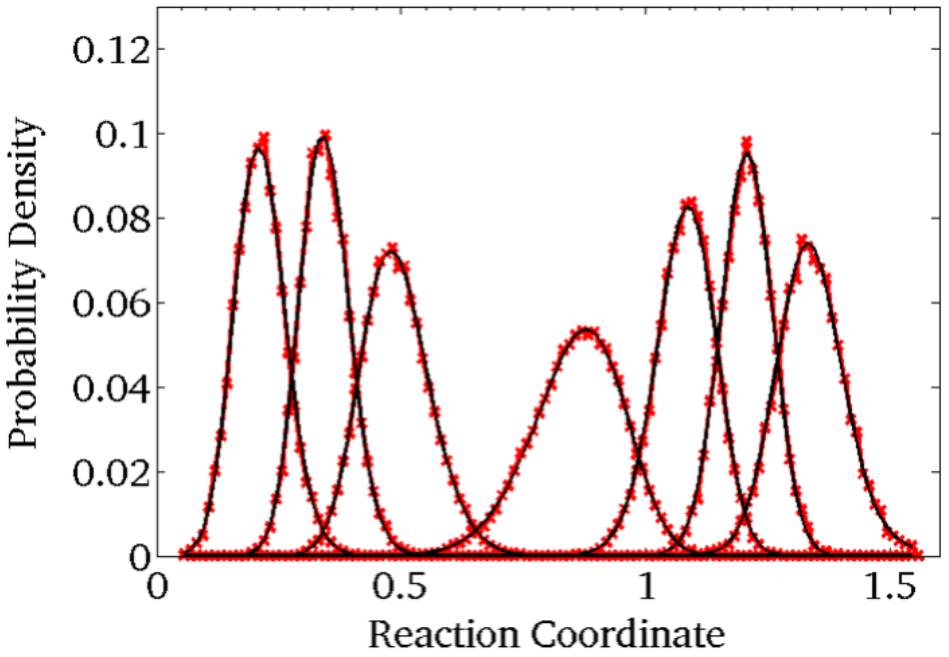




Ganguly, Thaplyal, Rosta, Bevilacqua, Hammes-Schiffer, *JACS* 2014

1. Umbrella Sampling Simulations & WHAM

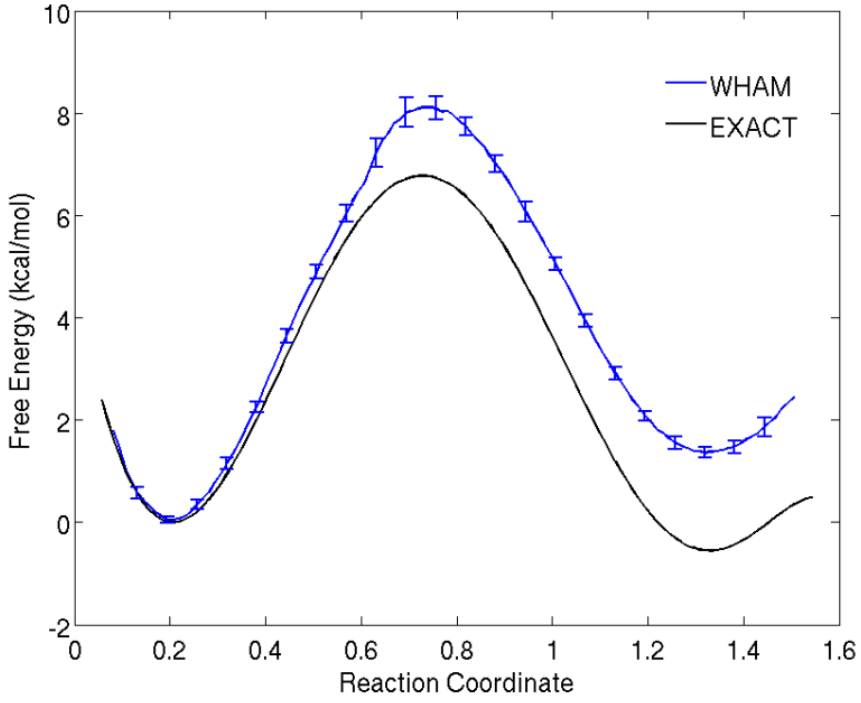


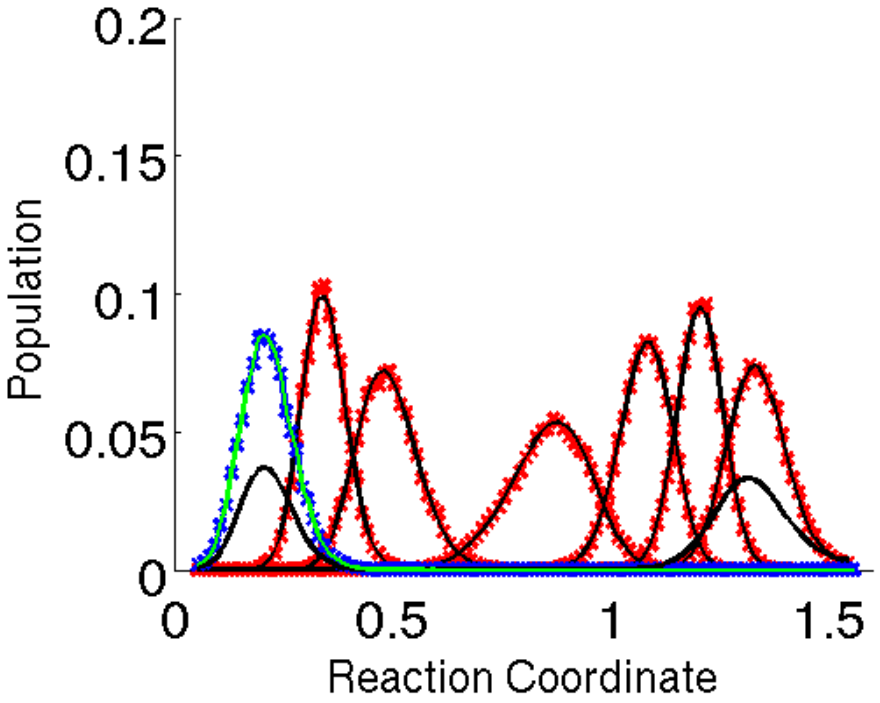


- MC simulations
- 7 Umbrellas with 50 kcal/mol biasing force each

Rosta, Hummer, *JCTC*, 2015

1. Systematic error when using WHAM in conjunction with small biasing force





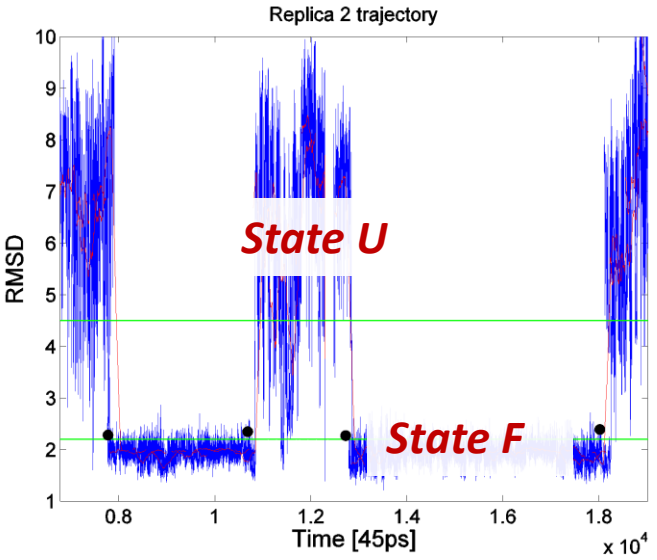
- 6 Umbrellas with 50 kcal/mol biasing force each
- 1 Umbrella with 1 kcal/mol biasing force each

Rosta, Hummer, *JCTC*, 2015

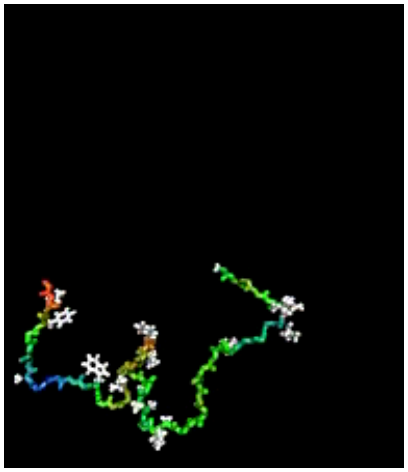
1. Markov-state models of conformational transitions

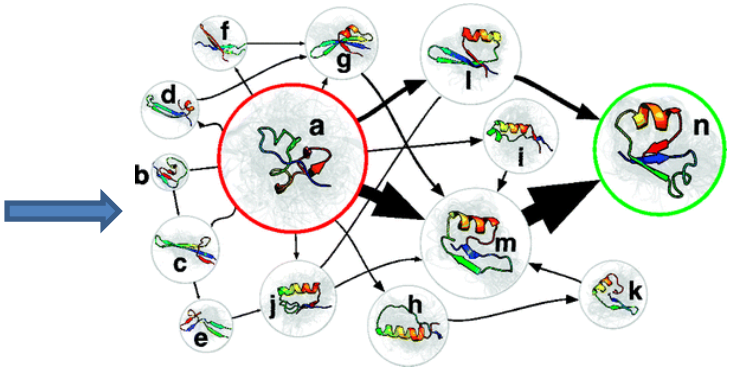
Master equation:

$$\frac{dP_i}{dt}(t) = \sum_{\substack{j=1 \\ (j \neq i)}}^{N} k_{i \leftarrow j} P_j(t) - \sum_{\substack{j=1 \\ (j \neq i)}}^{N} k_{j \leftarrow i} P_i(t)$$



Folding@home, V. Pande, Stanford





$$F \underset{k_F(T)}{\overset{k_U(T)}{\rightleftharpoons}} U$$

Voelz, Bowman, Beauchamp and Pande, *J. Am. Chem. Soc.*, 2010

DHAM: Dynamic Histogram Analysis Method

Our approach:

- Discretized reaction coordinates
- Short lagtimes

- Markov property:

$$P_{1|n-1}(x_n, t_n | x_{n-1}, t_{n-1}; ...; x_1, t_1) = \overbrace{P_{1|1}(x_n, t_n | x_{n-1}, t_{n-1})}^{\text{transition probability}}, \quad t_1 < ... < t_n$$

Most molecular dynamics integrators are Markovian!

1. DHAM: Dynamic Histogram Analysis Method

$$\Pr^{(k)} \propto \prod_{i=1}^{Nbin} \prod_{j=1}^{Nbin} \left(M_{ji}^{(k)} \right)^{T_{ji}^{(k)}}$$

$$\tilde{L} = \ln \tilde{\prod_{k=1}^{NSim}} \tilde{\prod_{i=1}^{Nbin}} \tilde{\prod_{j=1}^{Nbin}} \left(M_{ji}^{(k)} \right)^{T_{ji}^{(k)}}$$

$$M_{ji}^{(k)} = f_i^{(k)} c_{ji}^{(k)} M_{ji} = \frac{c_{ji}^{(k)} M_{ji}}{\sum_{l=1}^{N_{bin}} c_{li}^{(k)} M_{li}}$$

$$\frac{M_{ji}^{(k)}}{M_{ji}^{(0)}} = f_i^{(k)} c_{ji}^{(k)} \approx \exp\left(-\left(u_j^{(k)} - u_i^{(k)} \right) / 2k_B T \right)$$

$$M_{ji} = \frac{\sum_{k=1}^{Msim} T_{ji}^{(k)}}{\sum_{k=1}^{Msim} n_i^{(k)} f_i^{(k)} c_{ji}^{(k)}}$$

$$t_2 = -\tau_{lag} / \ln \lambda_2$$

Biased Dynamical Trajectories

↓

Histogram of Transitions

↓

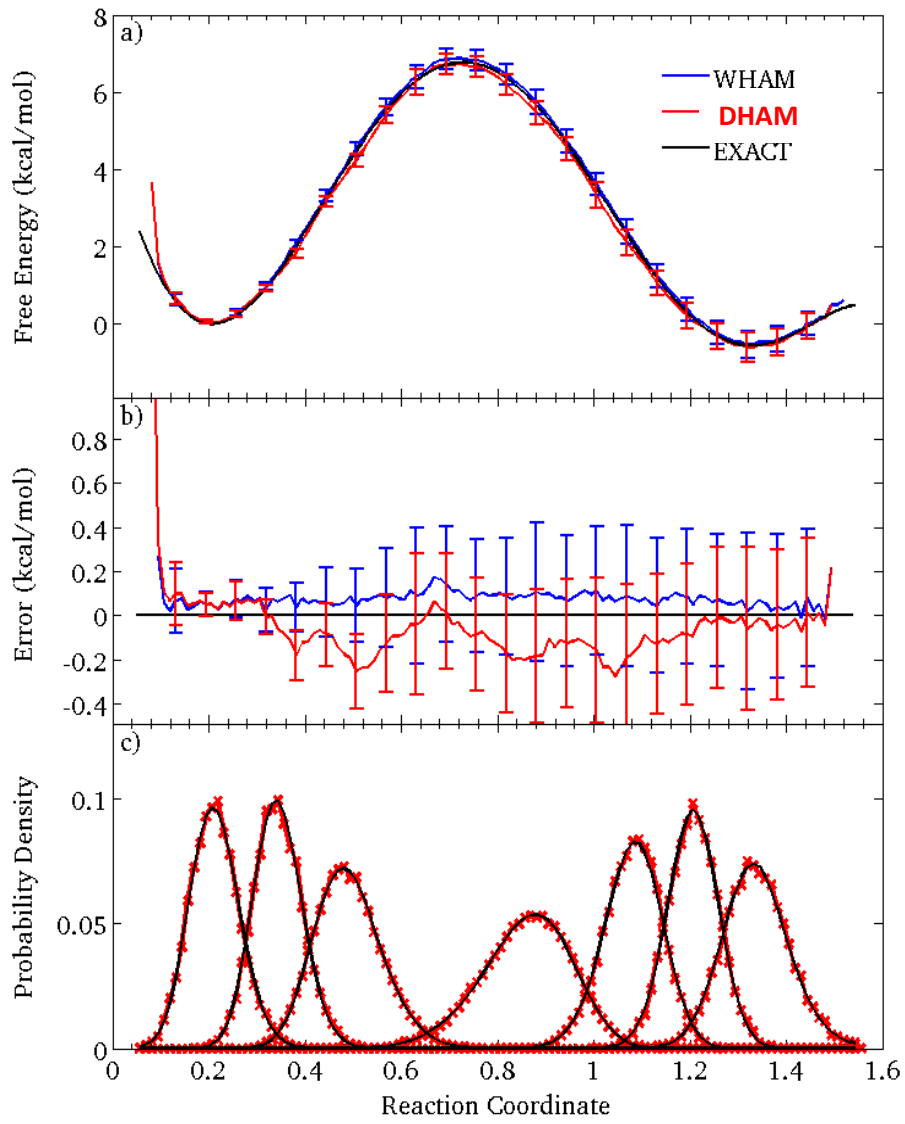
Markov Model

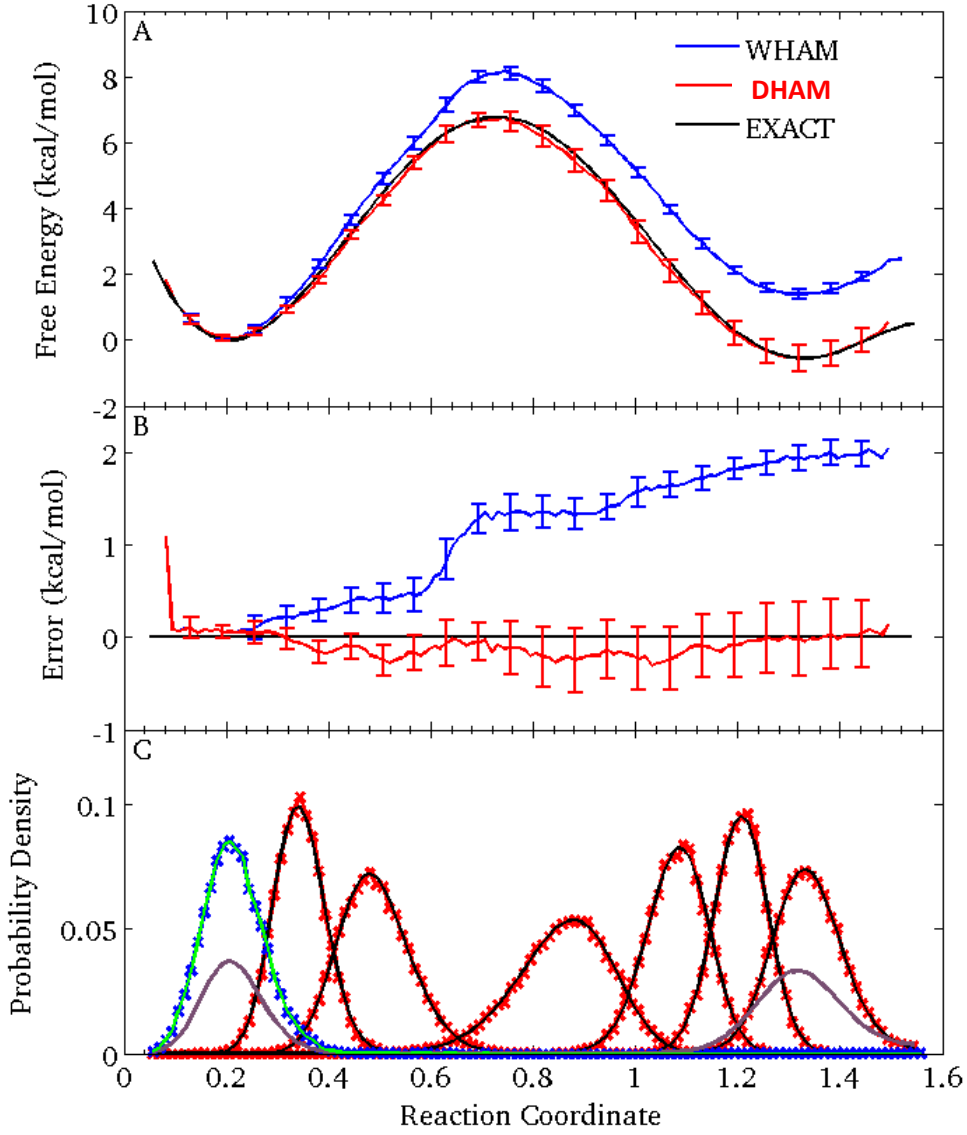
↓

Free Energy & Kinetics

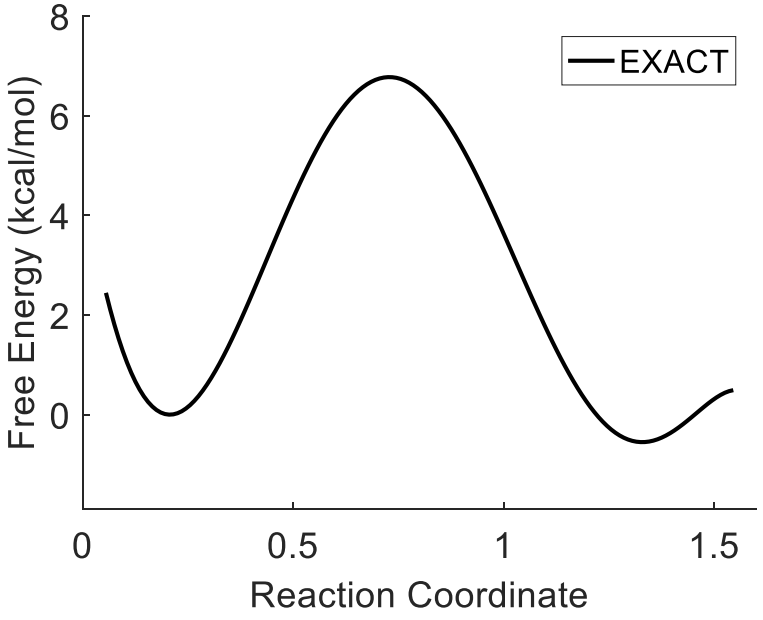
Rosta, Hummer, *JCTC*, 2015

Applications using Umbrella Sampling

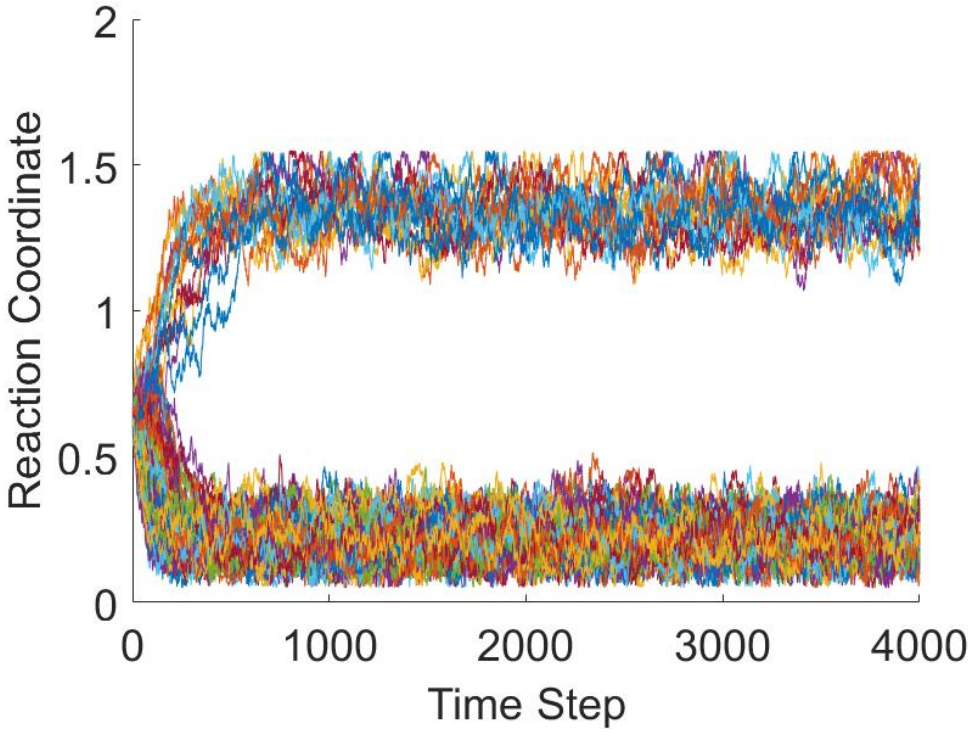




1. Applications for “downhill” unbiased non-equilibrium trajectories



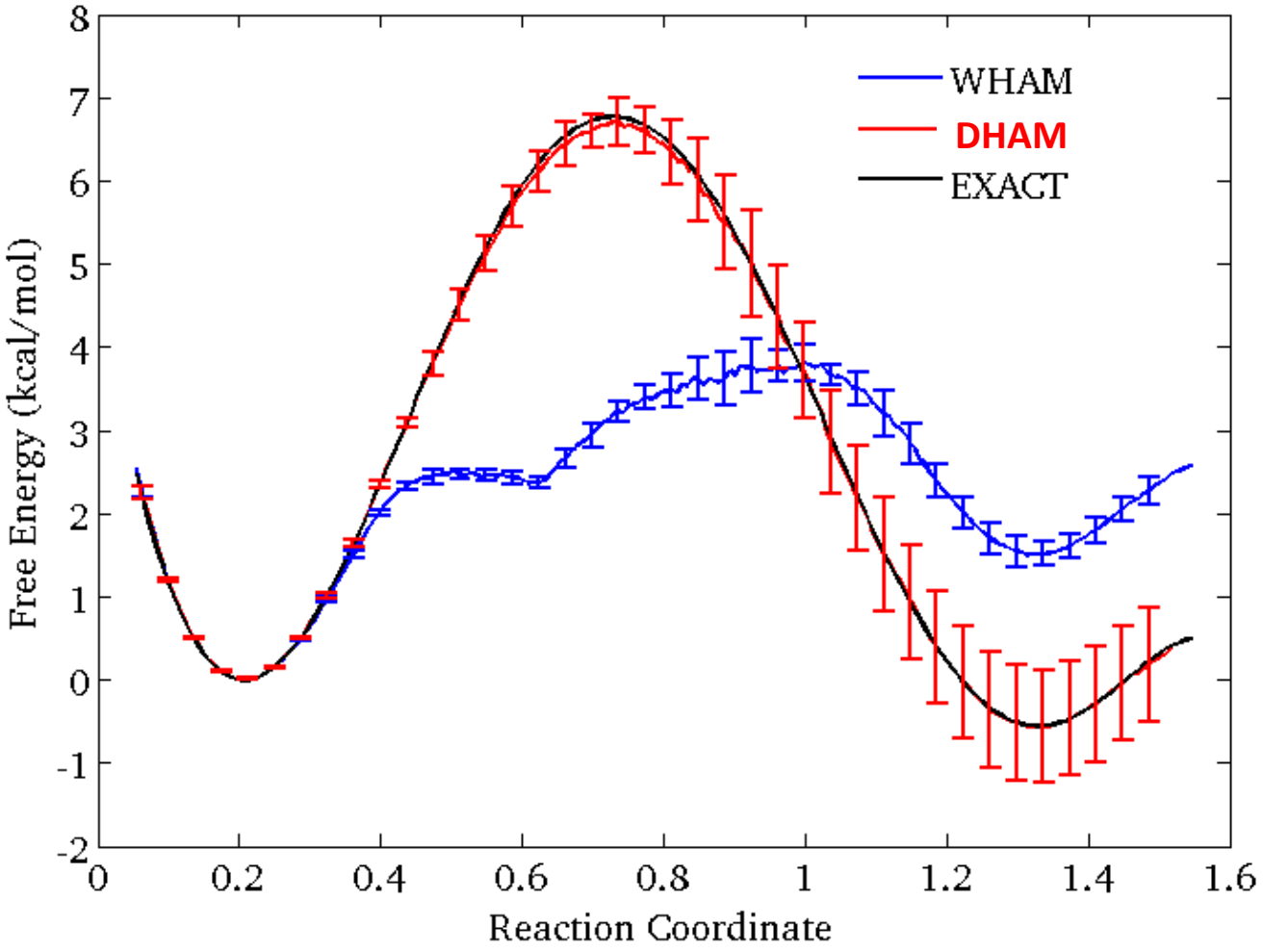
MC trajectories initiated from around x=0.63



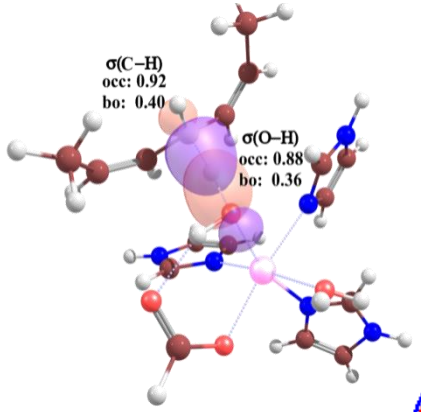
1. Applications for “downhill” unbiased non-equilibrium trajectories

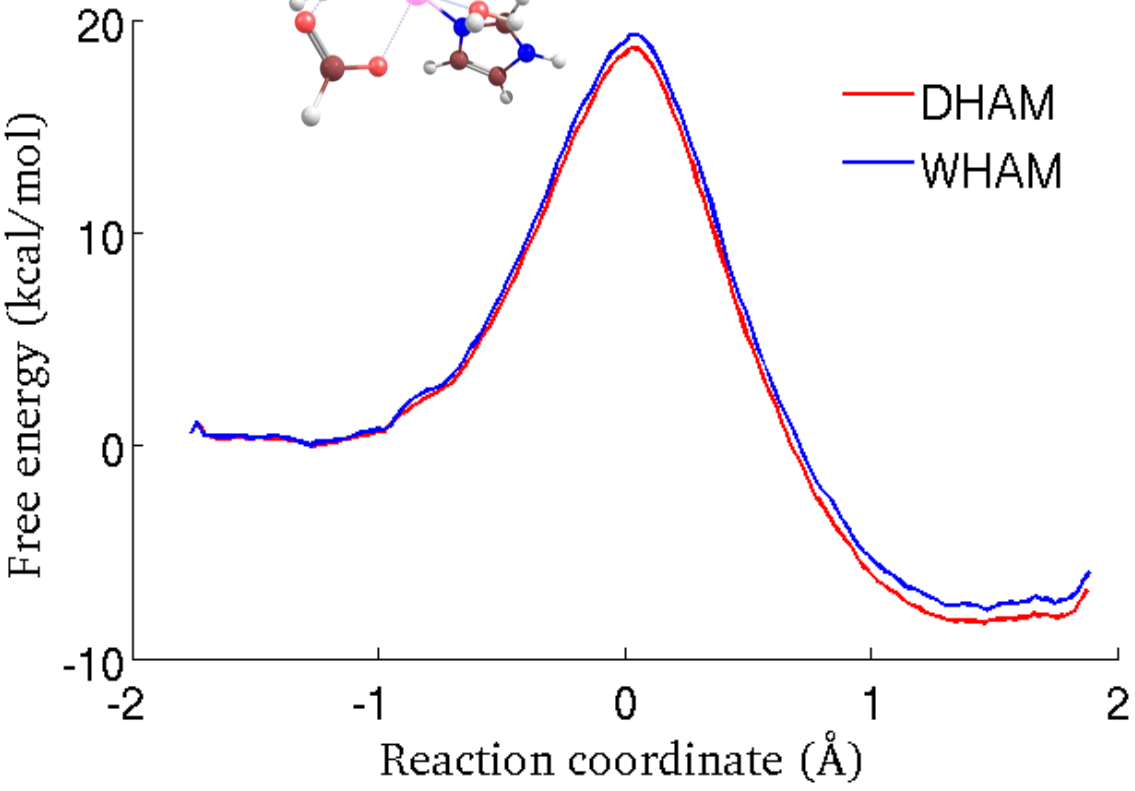
$$M_{ji} = \frac{\sum_{k=1}^{Msim} T_{ji}^{(k)}}{\sum_{k=1}^{Msim} n_i^{(k)}}$$

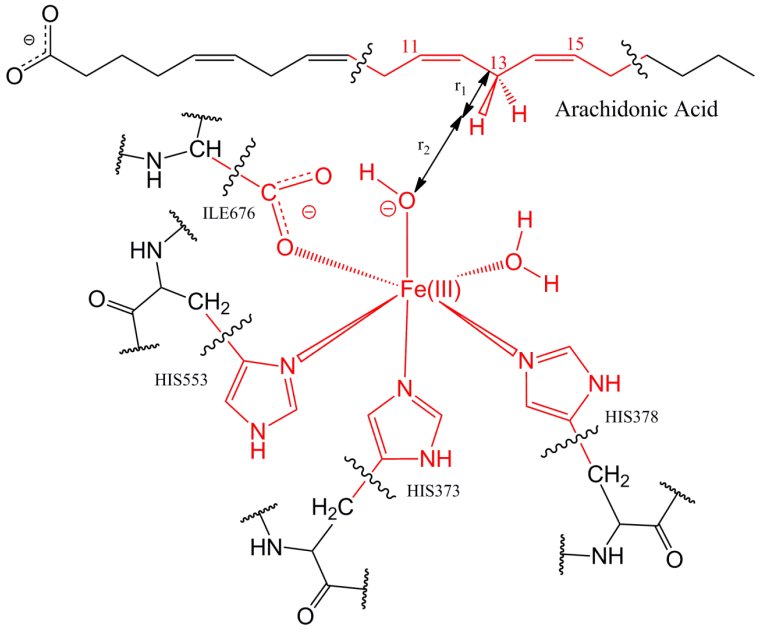
$$\sum_{j=1}^{Nbin} M_{ij} p_j = p_i$$



1. Umbrella sampling QM/MM simulations of catalytic reactions



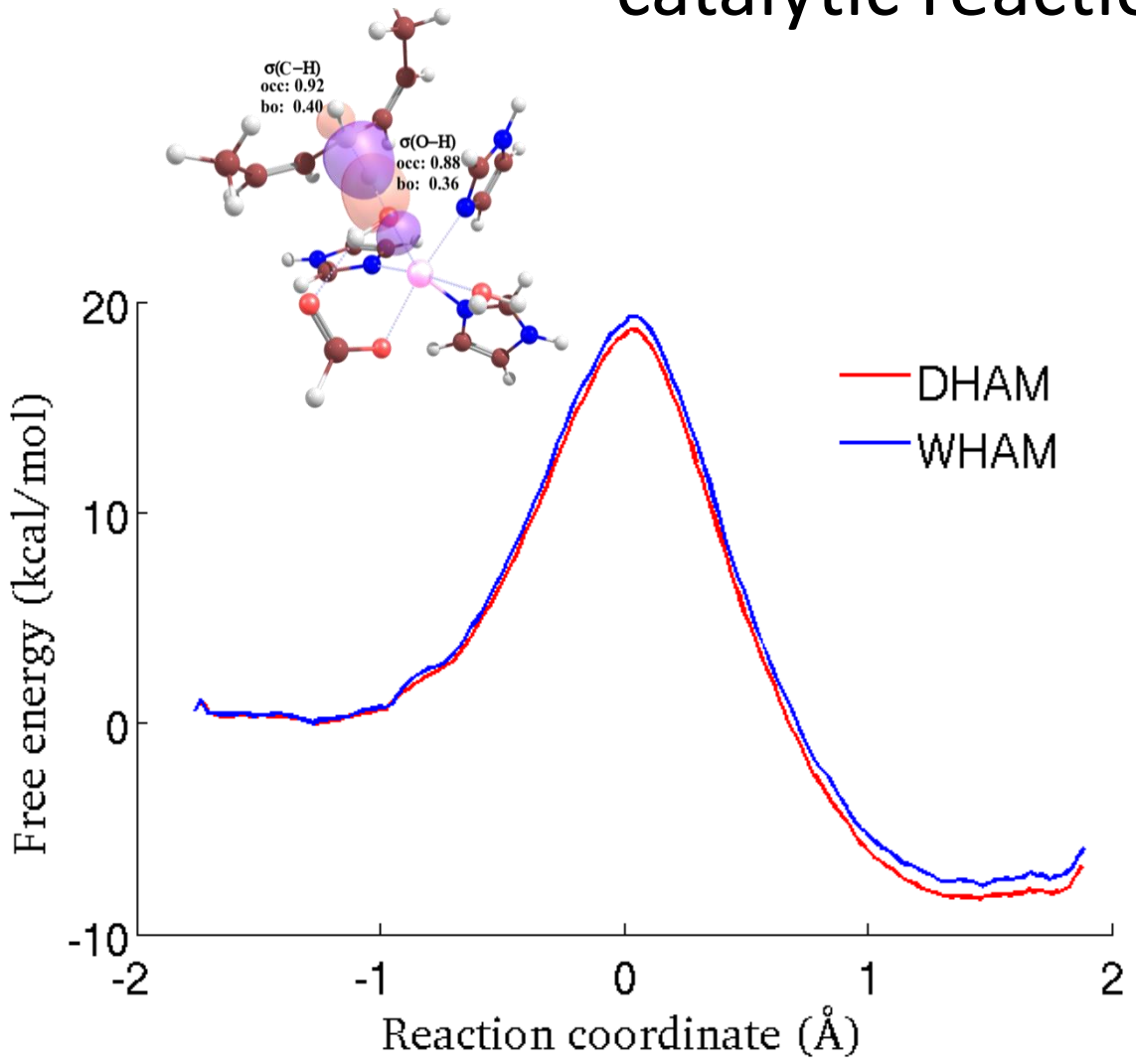




Hydrogen abstraction from arachidonic acid in the catalytic reaction of human 15-LOX-2 lipoxygenase.

R. Suardiaz, P. G. Jambrina, L. Masgrau, A. Gonzalez-Lafont, E. Rosta, J.M. Lluch, *JCTC*, 2016

1. Umbrella sampling QM/MM simulations of catalytic reactions



Lag time	Rate
1 fs	0.24 s^{-1}
10 fs	0.17 s^{-1}
50 fs	0.29 s^{-1}

Experimental k_{cat} = 0.7 s^{-1}

$$k = Ae^{-\Delta G^{\ddagger}/k_B T}$$

k=0.22 s^{-1} @T=300K
Calculated prefactor using a barrier of 18.7 kcal/mol:

A = 8.6x10^{12} s^{-1}

$k_B T/h$ = 6.3x10^{12} s^{-1}

R. Suardiaz, et al, *JCTC*, 2016
P. Saura, R. Suardiaz, et al, *ACS Catalysis*, 2017