2145-20

Spring College on Computational Nanoscience

17 - 28 May 2010

QM and Hybrid QM-MM Simulation of Biomolecules
Part II

Dario A. ESTRIN

Univ. of Buenos Aires
Argentina

Hybrid QM-MM Simulations of Biomolecules

Dario A. Estrin
University of Buenos Aires, Argentina

ICTP

May 2010

Outline

1. Motivation
2. The quantum mechanical energy
3. The coupling term
4. Treatment of the QM-MM boundary
5. Possible implementations
6. A representative example
    - Exploring enzymatic catalysis
7. Concluding remarks

Molecular Simulation

Quantum Mechanical Methods

1 - 1000 atoms, 1 - 100 ps

Solve the Schrödinger equation for a system of nuclei and electrons, based on some approximation. Obtain the electronic structure and, thereafter, any molecular property

Molecular Mechanics Methods

10^3 - 10^5 particles, 1 - 100 ns

Use pairwise or many-body potentials to describe the interactions between particles. The conectivity of atoms within molecules is fixed: unable to represent bonds dissociation, reactive processes, charge transfer, or other “quantum” effects

Quantum Mechanical Models

- Based on the Born-Oppenheimer approximation

$$E_{tot} = E_{elec} + \sum_{i \neq j} \frac{Z_i Z_j}{R_{ij}}$$

- E_{elec} is obtained by solving the time independent Schrödinger equation for the electrons in the field of fixed nuclei

$$H\Psi(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}) = E_{elec}\Psi(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}) \Rightarrow h_i\psi_i(\mathbf{x_1}) = \varepsilon_i\psi_i(\mathbf{x_1})$$

- Different approximations and levels of accuracy:
  - ab-initio
    Hartree-Fock, DFT, MP2, MP4, CI, CASSCF, etc
  - semiempirical
    AM1, CNDO, Tight-Binding, etc

Molecular Mechanics or Classical Models

- The potential energy is modeled through pairwise (or many-body) interactions, normally including electrostatics plus dispersive-repulsive terms

$$E_{tot} = E_{electrostatic} + E_{vDW} + E_{bonded}$$

- Parameters based on experimental or ab-initio data.
  Some standard force fields: SPC, TIP4P (water), CHARMM, AMBER (organic and biological molecules)
- Very fast, but no chemistry. Appropriate to sample configurations, fluctuations, phase transitions, transport, folding, etc.

QM-MM: a multiscale approach

In large molecular complexes, it is often the case that quantum-mechanical transformations take place on a reduced portion of the system. In such cases it could be tempting to restrain the electronic description to the “reactive region”, while modelling the rest of the particles with a classical mechanics approach. Both fragments are then allowed to interact somehow.

Original idea by Warshel (1976), in the context of biological chemistry.

First applied using the Empirical Valence Bond method, in which reaction states are described in terms of pre-established valence bond configurations. Warshel proposed to solve this problem in the presence of a classical environment of point-charges.

QM-MM: a multiscale approach

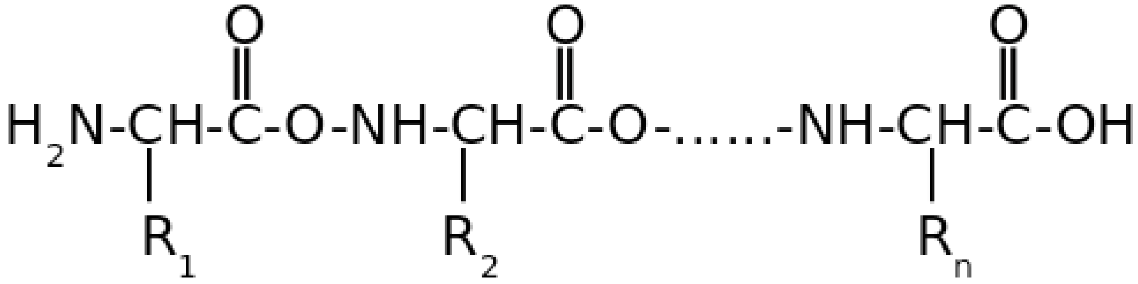
QM-MM methods are meant to describe reactive or quantum mechanical processes in complex environments, such as biomolecules

Two regions

- Reactive part (or solute): treated quantum-mechanically, determines the cost of the simulation
- Environment (or solvent): treated classically, in principle adds little computational load

The QM-MM method in the biological sciences

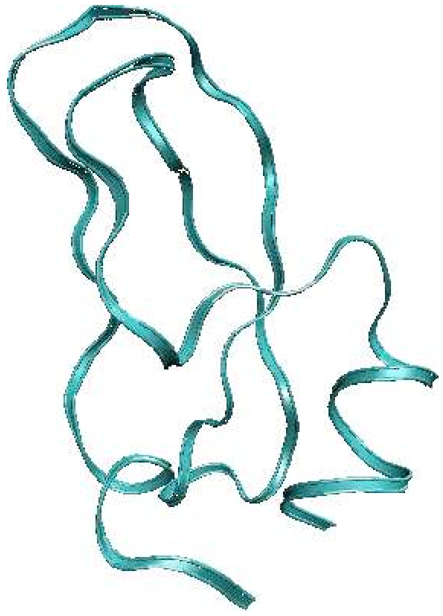
Proteins: catalysis, transport, electron transfer, structure



$$H_2N\text{-}\underset{\underset{R_1}{|}}{CH}\text{-}\overset{\overset{O}{\|}}{C}\text{-}O\text{-}NH\text{-}\underset{\underset{R_2}{|}}{CH}\text{-}\overset{\overset{O}{\|}}{C}\text{-}O\text{-}\ldots\ldots\text{-}NH\text{-}\underset{\underset{R_n}{|}}{CH}\text{-}\overset{\overset{O}{\|}}{C}\text{-}OH$$

The sequence of aminoacids R determines the structure of the protein

- α-helices
- β-sheets
- prostetic groups



The QM-MM method in the biological sciences

What can QM-MM methods do?

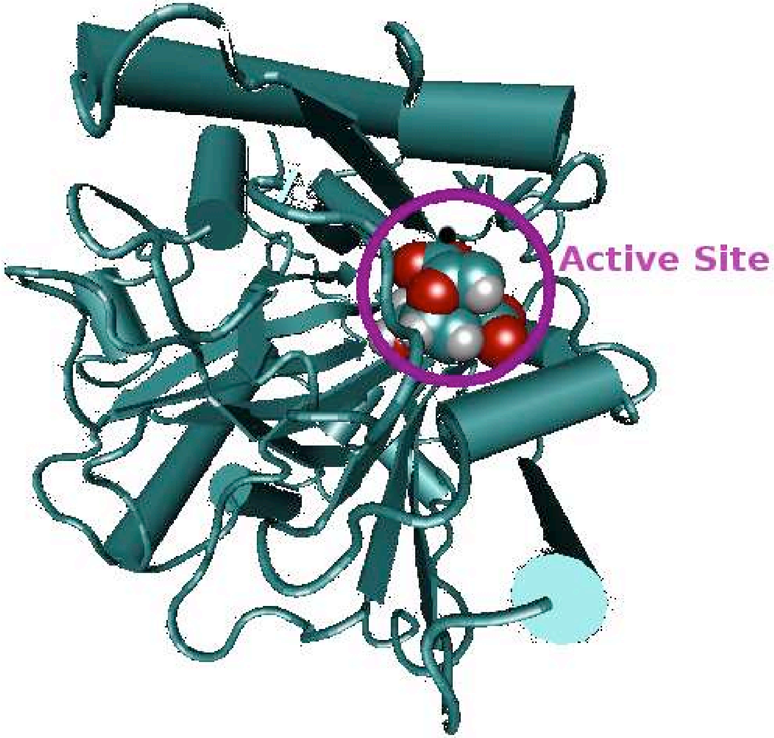
Catalysis in enzymes takes place at the “active site”, which size is usually within reach of QM methods. The structure and reactivity of the active site is modulated by the polypeptide, through mechanical or long range electrostatic interactions

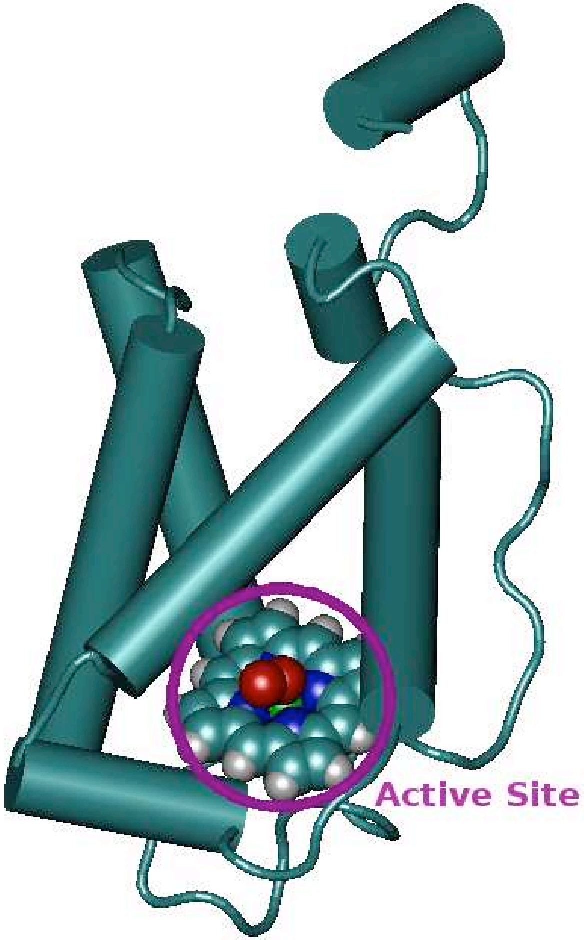
QM-MM methods can be used to:

- Obtain affinities for ligands (O_2, NO, CO, etc) and organic substrates
- Compute activation barriers of enzymatic reactions
- Get structural information and stability on the active site
- Estimate vibrational, electronic, and/or NMR spectra

Some situations in which QM-MM could be helpful

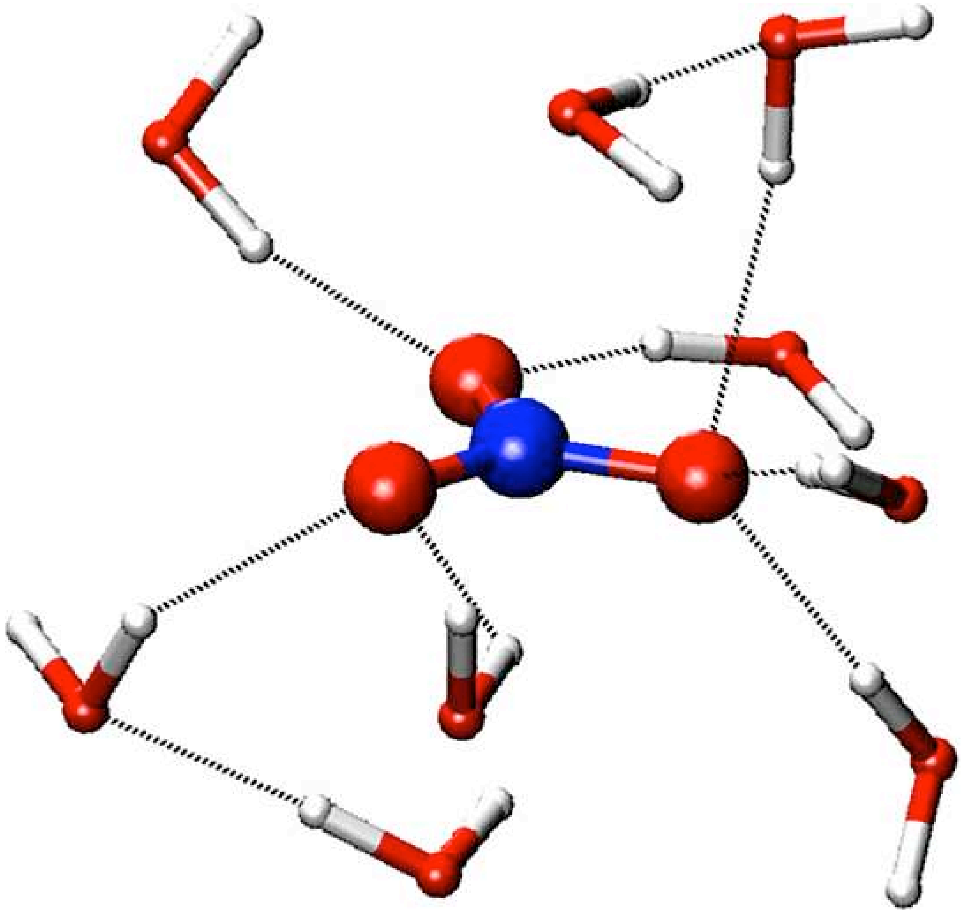
Active sites in proteins





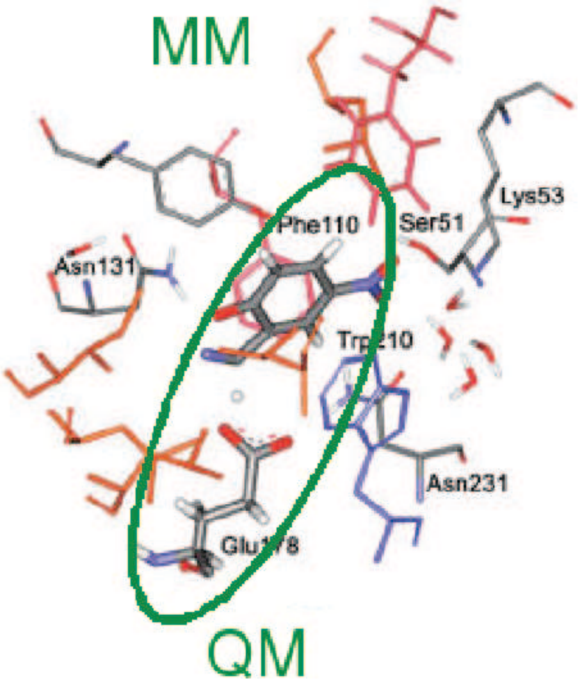
Some situations in which QM-MM could be helpful

Chemistry in solution



QM-MM Hamiltonian

The additive scheme



$$\hat{H}_{TOT}\,\Psi(\mathbf{r},\mathbf{R},\tau) = E_{TOT}(\mathbf{R},\tau)\Psi(\mathbf{r},\mathbf{R},\tau)$$

$$\hat{H}_{TOT} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM-MM}$$

$$E_{TOT} = \langle\Psi|\hat{H}|\Psi\rangle = E_{QM} + E_{MM} + E_{QM-MM}$$

The quantum mechanical energy

First-principles or ab-initio approaches

Born-Oppenheimer: quantum-mechanical electrons and classical nuclei

$$\hat{H}_{QM} = \hat{H}_{elec} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{r_{AB}}$$

$$E_{QM} = \langle \Psi | \hat{H}_{QM} | \Psi \rangle = E_{elec} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{r_{AB}}$$

The coupling term

$$\hat{H}_{QM-MM} = \hat{H}^{vdW}_{QM-MM} + \hat{H}^{electrostatic}_{QM-MM} \quad (+\hat{H}_{link})$$

$$\hat{H}^{vdW}_{QM-MM} = \sum_{\alpha \in QM} \sum_{i \in MM} 4\epsilon_{\alpha i} \left[\left(\frac{\sigma_{\alpha i}}{|\mathbf{R}_\alpha - \mathbf{R}_i|} \right)^{12} - \left(\frac{\sigma_{\alpha i}}{|\mathbf{R}_\alpha - \mathbf{R}_i|} \right)^{6} \right]$$

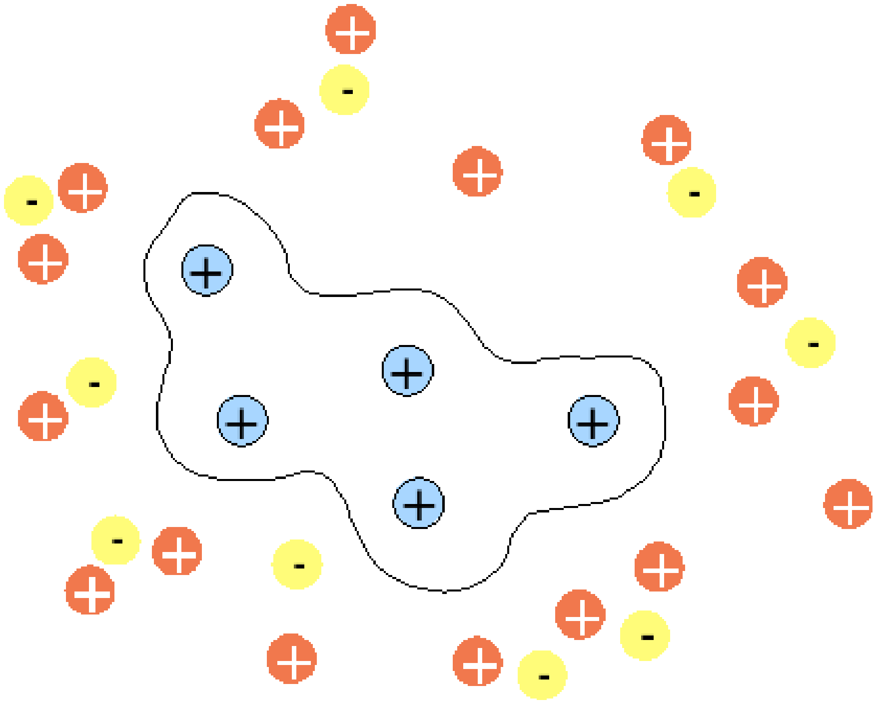
The parameters $\epsilon_{\alpha i}$ and $\sigma_{\alpha i}$ are typically taken from MM force fields. However, they may not always be transferable. Consistency with the other interactions must be checked.

$$E^{vdW}_{QM-MM} = \langle \Psi | \hat{H}^{vdW}_{QM-MM} | \Psi \rangle = \hat{H}^{vdW}_{QM-MM}$$

The coupling

Electrostatic interaction

$$\hat{H}^{electrostatic}_{QM-MM} = \sum_{\alpha \in QM} \sum_{i \in MM} \frac{Z_\alpha q_i}{|\mathbf{R}_\alpha - \mathbf{R}_j|} - \sum_{electrons} \sum_{i \in MM} \frac{q_i}{|\mathbf{R}_e - \mathbf{R_i}|}$$



The coupling

Electrostatic interaction in Hartree-Fock

$$E_{elec} = \langle \Psi | \hat{H}_{el} | \Psi \rangle = \sum_{i=1}^{N} H_{ii}^{core} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (J_{ij} - K_{ij})$$

The classical charge distribution only affects H_{ii}^{core}

$$H_{ii}^{core} = -\frac{1}{2} \langle \chi_i(1) | \nabla_1^2 | \chi_i(1) \rangle - \langle \chi_i(1) | \sum_{\alpha \in QM} \frac{Z_\alpha}{r_{1\alpha}} | \chi_i(1) \rangle$$

$$- \langle \chi_i(1) | \sum_{i \in MM} \frac{q_i}{r_{1i}} | \chi_i(1) \rangle$$

The coupling

Electrostatic interaction in DFT (and in ab-initio methods)

$$E_{QM-MM}^{electrostatic} = -\sum_{i\in MM} q_i \int \frac{\rho(\mathbf{r})}{|\,\mathbf{r}-\mathbf{R}_i\,|}\, d\mathbf{r} + \sum_{i\in MM}\sum_{\alpha\in QM} \frac{q_i Z_\alpha}{|\,\mathbf{R}_\alpha - \mathbf{R}_i\,|}$$

$$E_{QM-MM} = \sum_{i\in MM} q_i \int \frac{\rho(\mathbf{r})}{|\,\mathbf{r}-\mathbf{R}_i\,|}\, d\mathbf{r} + \sum_{i\in MM}\sum_{\alpha\in QM} \frac{q_i Z_\alpha}{|\,\mathbf{R}_\alpha - \mathbf{R}_i\,|} + E_{vdW}^{QM-MM}$$

E_{vdW}^{QM-MM} prevents MM charges of negative sign to approach QM nuclei too closely

The coupling

Embedding

- **Mechanical embedding**
  The QM wave-function is not perturbed by the external MM charges.
  Interaction is described by a force field

$$E_{QM-MM}^{electrostatic} \equiv \sum_{i \in MM} \sum_{j \in QM} \frac{q_i q_j}{r_{ij}}$$

- **Electrostatic embedding**
  QM and MM atoms respond self-consistently to each other

$$\hat{H}_{QM-MM}^{electrostatic} = \sum_{\alpha \in QM} \sum_{i \in MM} \frac{Z_\alpha q_i}{|\mathbf{R}_\alpha - \mathbf{R}_j|} - \sum_{electrons} \sum_{i \in MM} \frac{q_i}{|\mathbf{R}_e - \mathbf{R_i}|}$$

- **Polarizable embedding**
  Requires a polarizable force field: MM charges affected by the electron distribution of the QM region

Polarizable force fields

Fluctuating charge models

$$E_{MM}^{electrostatic} = \sum_i \left[E_i^o + \chi_i^o q_i + \frac{1}{2} J_i^o q_i^2 \right] + \sum_{i,j} \frac{q_i q_j}{r_{ij}}$$

E^o, χ^o, J^o parameters related to electronegativity and polarizability

$\{q_i\}$ are computed to minimize $E_{MM}^{electrostatic}$ under the constraint of electroneutrality within molecules

For implementation in QM-MM molecular dynamics it is convenient to define a fictitious Langrangian:

$$L = \sum_{\alpha \in QM} \frac{1}{2} m_\alpha \dot{R}_\alpha^2 + \sum_{i \in MM} \frac{1}{2} m_i \dot{R}_i^2 - E_{pot}[R_\alpha, R_i] + \sum_{i \in MM} \frac{1}{2} m_q \dot{q}_i^2 - \sum_{I}^{molecules} \lambda_I \sum_{i \in I} q_i$$

Polarizable force fields

Fluctuating charge models

A fictitious mass m_q is defined to allow for the charges q_i to evolve following a Car-Parinello molecular dynamics,

$$m_\alpha \ddot{R}_\alpha = -\frac{\partial E[R_\alpha, R_i]}{\partial R_\alpha} \qquad m_q \ddot{q}_i = -\frac{\partial E[R_\alpha, R_i]}{\partial q_i} - \lambda_i$$

- m_q must be small enough to avoid termal coupling between real and fictitious dynamics.
- QM and MM polarization is determined self-consistenly. Time consuming.
- Requires many parameters, not so easy to fit.

The QM-MM boundary

What if we want the QM and the MM atoms to be chemically bound?

Covalent bonds: enzymes, organic molecules and polymers

Ionic bonds: solid state

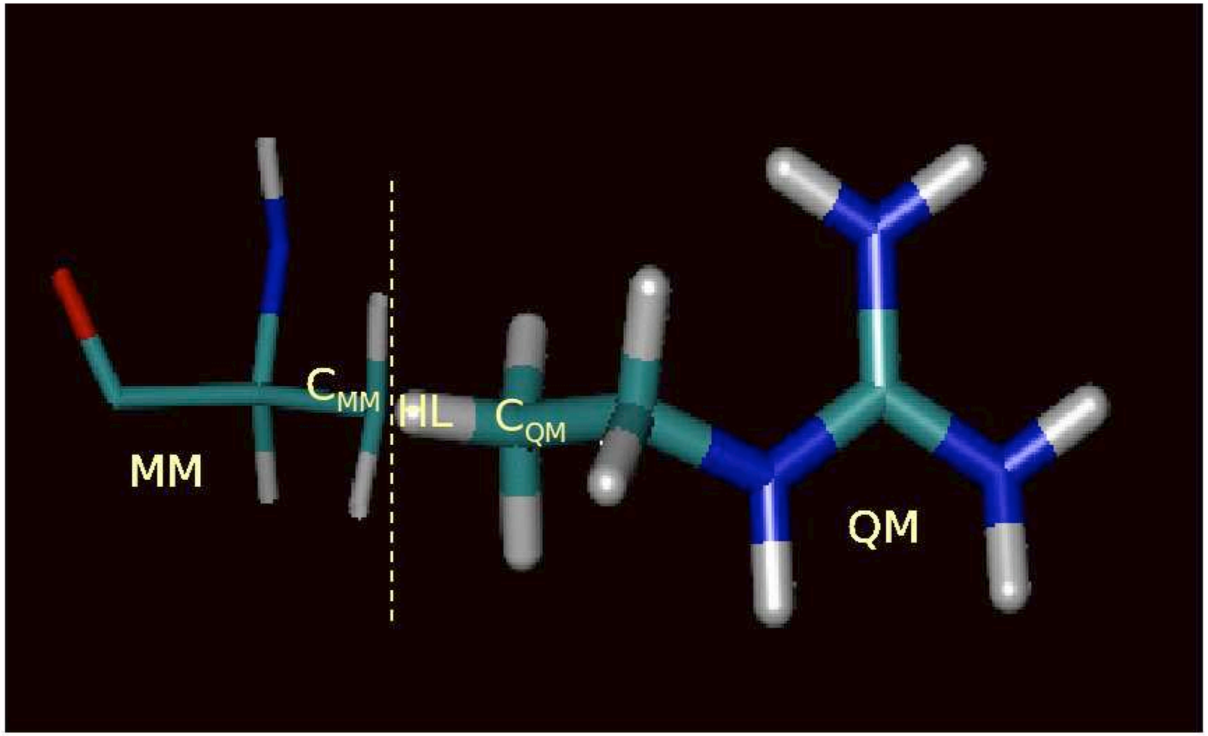
Two major approaches:

- **Link atom**
- **Local orbitals**

The QM-MM boundary

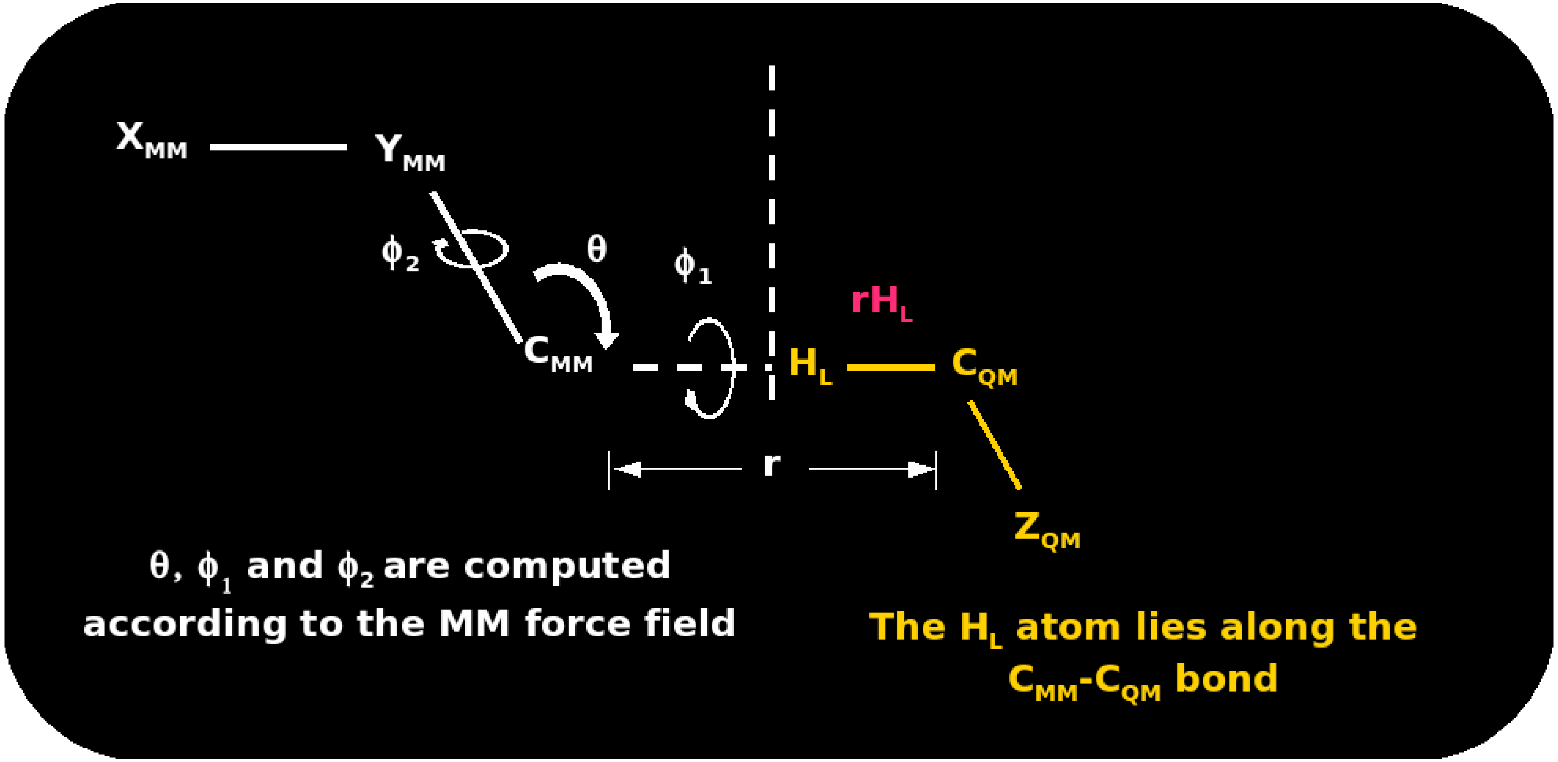
Link atom

In biomolecules the link is usually set at a C-C bond. A QM hydrogen atom H_L is added to saturate coordination.



The QM-MM boundary

Link atom



θ, ϕ_1 and ϕ_2 are computed according to the MM force field

The H_L atom lies along the C_{MM}-C_{QM} bond

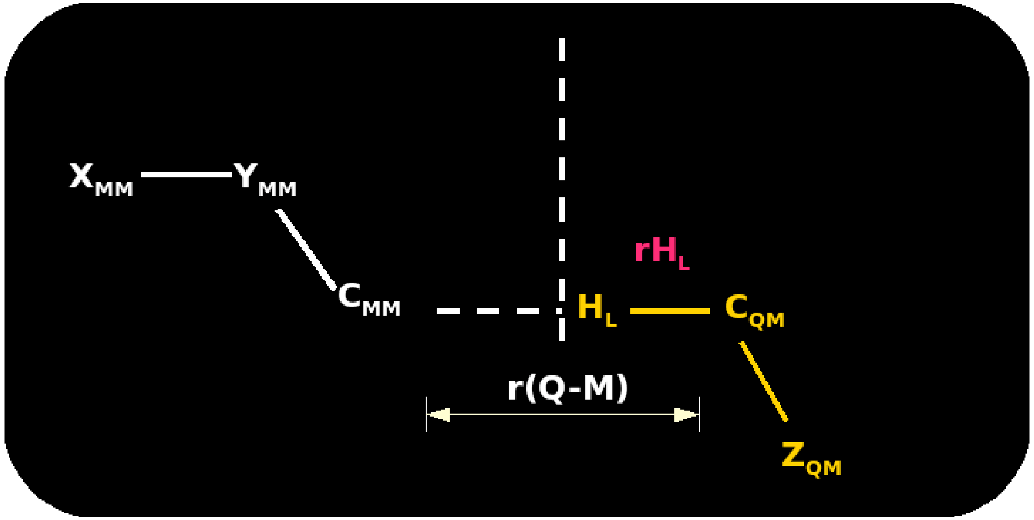
- The charge on C_{MM} (and maybe also on Y_{MM} and X_{MM}) is set to zero, to avoid unrealistic polarization of the H_L-C_{QM} bond.
- Redistribution of charges on neighboring MM atoms is required to preserve neutrality.

The QM-MM boundary

Link atom

What about the C_{QM}-H_L and C_{QM}-C_{MM} bond lengths?

r(Q-M) = r_{C-C} (MM)



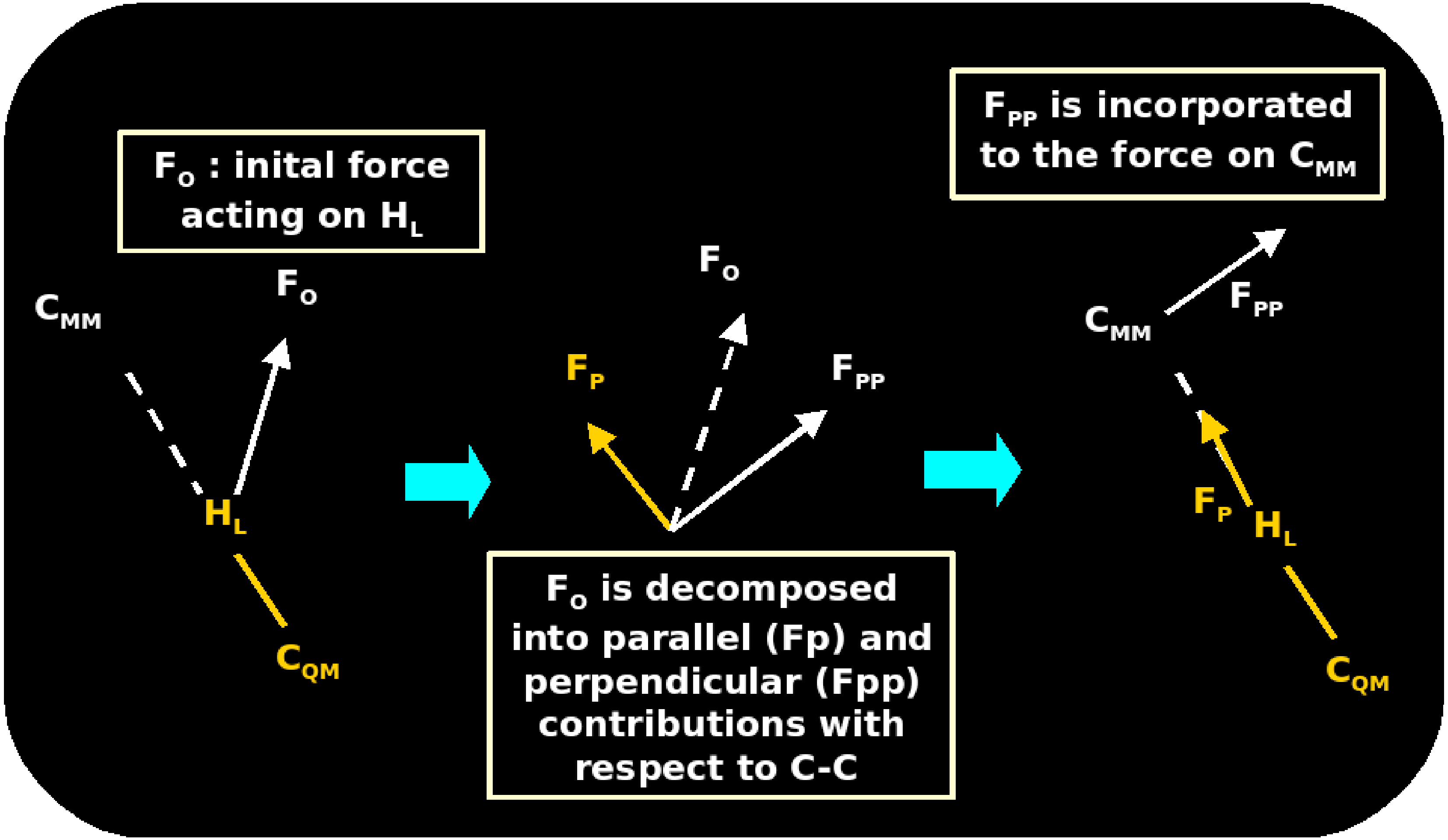
Different “scaling” procedures for rH_L:

- $\mathrm{r}H_L = \frac{r^\circ_{C-H}(MM)}{r^\circ_{C-C}(MM)} r(Q-M)$ (Morokuma et al. **1999**)
- $\mathrm{r}H_L = r^\circ_{C-H} + \frac{k_{CC}}{k_{CH}} \left[r(Q-M) - r^\circ_{C-C}(MM)\right]$
  $\Delta E_{stretch} = k_{CC}\left(1 - \frac{k_{CH}}{k_{CC}}\right)\left[r(Q-M) - r^\circ_{C-C}(MM)\right]^2$ (Eichinger et al. **1999**)

The QM-MM boundary

Link atom

Treatment of the forces on H_L



F_O : inital force acting on H_L

F_O is decomposed into parallel (Fp) and perpendicular (Fpp) contributions with respect to C-C

F_{PP} is incorporated to the force on C_{MM}

The QM-MM boundary

More on link atoms

Alternatives to the use of hydrogen in the Link Atom scheme

- Within semiempirical methods, the link atom does not need to be a real atom but may involve a parameterized Hamiltonian to mimic the original bond
- In ab-initio and semiempirical methods H atoms can be replaced by “dummy groups” or “pseudohalogens”
    - Pseudobond approach: a broken C-C bond is mimicked by a fluor-like carbon bond with a parameterized effective core potential on the pseudo-fluor atom (Zhang, Yang et al. **1999**/**2000**)
    - Effective group potentials: molecular groups of the environment approximated by pseudopotentials (Poteau et al. **2001**)

QM-MM implementations

Some DFT-based schemes

- Using Gaussian basis sets
  - Standard additive schemes:
    GAMESS-UK, ADF, CHEMSHELL (modular), and many others
  - ONIOM
- Using real space grids
  - SIESTA
  - Plane waves methods
- Adaptive QM-MM

QM-MM implementations

Schemes based on Gaussian basis functions

$$\rho(\mathbf{r}) = \sum_i |\chi_i(\mathbf{r})|^2 \qquad \chi_i(\mathbf{r}) = \sum_{\mu=1}^{M} C_{\mu i}\phi_\mu(\mathbf{r})$$

$$\Downarrow$$

$$\rho(\mathbf{r}) = \sum_i \sum_\mu \sum_\nu C_{\mu i} C_{\nu i} \phi_\mu(\mathbf{r}) \phi_\nu(\mathbf{r})$$

$$\Downarrow$$

$$E_{QM-MM}^{electrons} = \sum_{m \in MM} q_m \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_m|} d\mathbf{r} = \sum_{m \in MM} \sum_{i,\mu,\nu} q_m C_{\mu i} C_{\nu i} \int \frac{\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_m|} d\mathbf{r}$$

Integrals of Gaussians have analytic solutions:

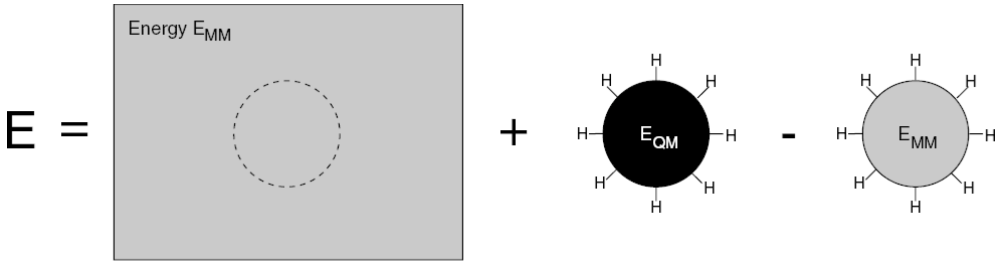
E_{QM-MM} fast and easy to compute

QM-MM implementations

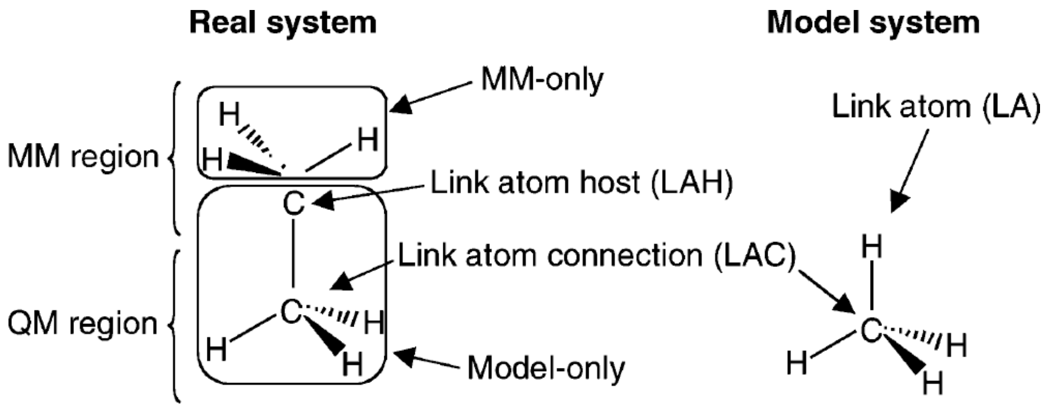
ONIOM

(Morokuma, Vreven, Frisch, and others, **1999—2006**)

- Substractive scheme



$$E_{ONIOM} = E_{MM}^{real} + E_{QM}^{model} - E_{MM}^{model}$$



QM-MM implementations

ONIOM

- Originally formulated for mechanical embedding, with atoms of the MM region represented by point charges
- More recently extended to electrostatic embedding

  (*J. Chem. Theory Comput.* **2006**, 2, 815)
- Advantage of the substractive scheme:
  allows for combination of multiple QM levels, e.g. DFT/AM1/MM
- If QM and MM regions are not chemically connected, the ONIOM energy is equal to the energy of an additive scheme

QM-MM implementations

Real space methods in periodic boundary conditions

In most DFT schemes in PBC the charge density $\rho(\mathbf{r})$ and the electrostatic potential $V_e(\mathbf{r})$ are expressed in terms of Fourier series:

$$\rho(\mathbf{r}) = \sum_G \tilde{\rho}(\mathbf{G}) e^{i\mathbf{G}\mathbf{r}} \qquad V_e(\mathbf{r}) = \sum_G \tilde{V}_e(\mathbf{G}) e^{i\mathbf{G}\mathbf{r}}$$

In this framework $V_e(\mathbf{r})$ can be efficiently calculated from $\rho(\mathbf{r})$ by inverting the Poisson equation with a Fourier transform technique,

$$\nabla^2 V_e(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \Longrightarrow \tilde{V}_e(\mathbf{G}) = -4\pi \frac{\tilde{\rho}(\mathbf{G})}{|\mathbf{G}|^2}$$

Therefore the potential and the electron density are discrete functions in a grid, $V_e(n)$ and $\rho(n)$.

QM-MM implementations

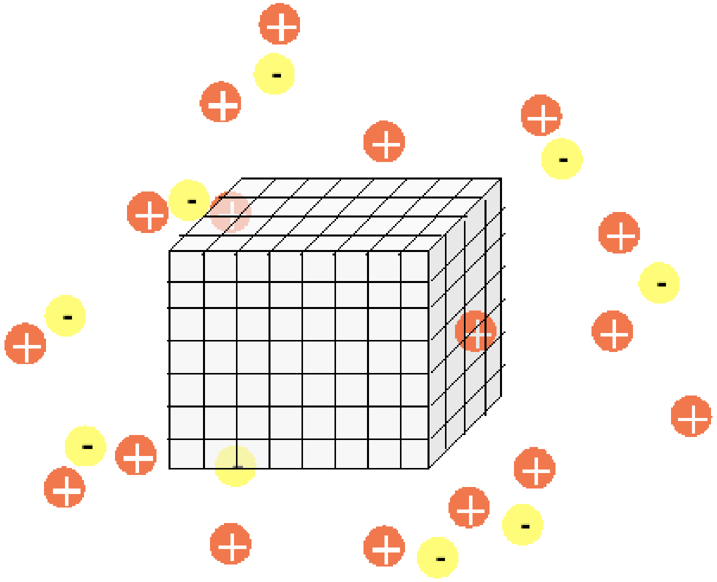
Real space methods in periodic boundary conditions

$\rho(\mathbf{r}) \rightarrow \rho(n)$ defined in a grid of $M = N_x \times N_y \times N_z$ mesh points

$$E^{electrons}_{QM-MM} = \sum_{m \in MM}^{A} q_m \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_m|} d\mathbf{r} = \sum_{m \in MM}^{A} \sum_{n=1}^{M} q_m \frac{\rho(n)}{|\mathbf{r_n} - \mathbf{R}_m|}$$

Analytic integrals
replaced by sums on the mesh,
scales up with $M \times A$
(the number of grid points
times the number of MM atoms)

Grid contains QM atoms only.
No periodicity in the MM system



QM-MM implementations

Plane waves methods

QM-MM schemes based on DFT and plane waves have been coded in the Car-Parrinello molecular dynamics framework

Eichinger, Tavan, Hutter & Parrinello **1999** Laio, VandeVondele & Rothlisberger **2002**

Car-Parrinello Lagrangian:

$$L = \sum_i \frac{1}{2}\mu \int |\psi_i(r)|^2 dr + \sum_I \frac{1}{2} M_I \dot{R}_I^2 - E[\psi, R] + \sum_{i,j} \Lambda_{ij} \Big[\int \psi_i^*(r)\psi_j(r) dr - \delta_{ij} \Big]$$

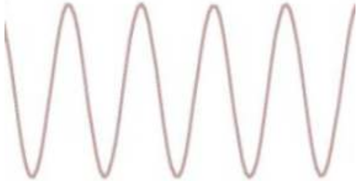
- **Born-Oppenheimer:** nuclei move according to the electronic forces. QM wave function needs to be relaxed at every time step.
- **Car-Parrinello:** nuclei and electronic wave functions evolve simultaneously. The electrons are coupled to a pseudo-classical dynamics through a fictitious mass μ.

QM-MM implementations

Plane waves methods

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{G}} C_i(\mathbf{G}) e^{i\mathbf{Gr}} \qquad \Gamma-point$$

The basis $\{e^{i}\mathbf{Gr}\}$ is delocalized.
Electron spill-out: charge density builds on MM atoms.
Effect not present with localized basis sets.



Very smooth pseudopotentials required to prevent spill-out

$$\lim_{r\to\infty} v(r) = \frac{1}{r}$$

$$v(r) = \frac{r_c^n - r^n}{r_c^{n+1} - r^{n+1}}$$

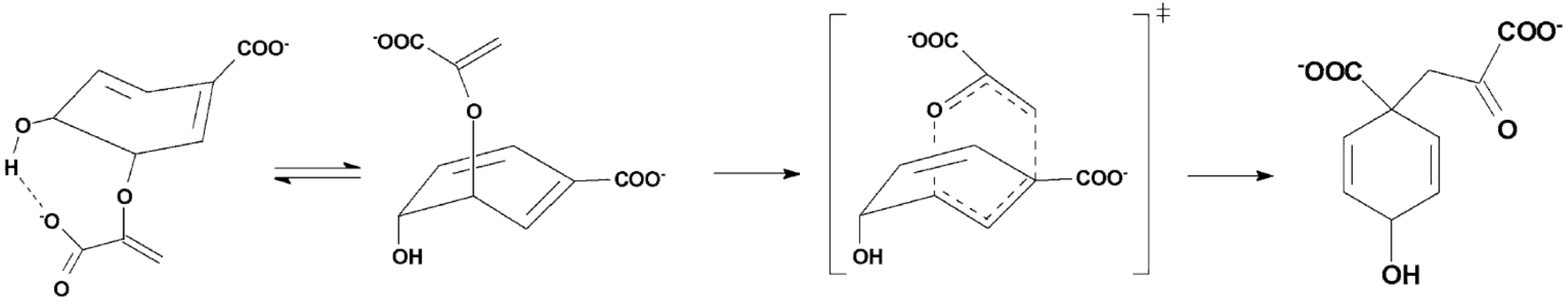
$$\lim_{r\to 0} v(r) = \frac{1}{r_c}$$

J. Chem. Phys. **2002** 116, 6941

Enzymatic catalysis

Role of the protein on the activation barrier

Chorismate Mutase: catalyzes an essential step in the biosynthesis of aromatic amino acids in bacteria, fungi and plants.



diequatorial chorismate

diaxial chorismate

transition state

prephenate

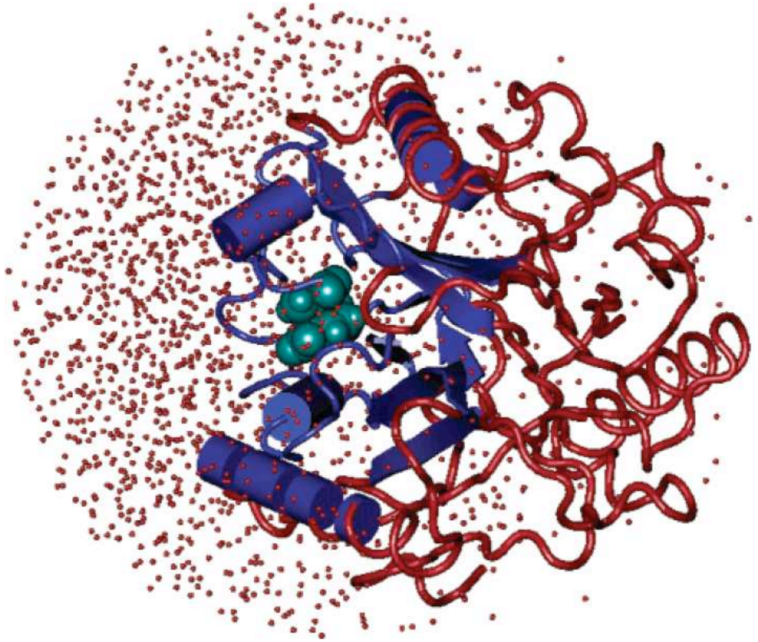
Experimental data for the conversion

Aqueous solution: $\Delta H^{\ddagger}$ = 20.7 kcal/mol $\Delta S^{\ddagger}$ = -12.9 J/K mol

Chorismate mutase: $\Delta H^{\ddagger}$ = 12.7 kcal/mol $\Delta S^{\ddagger}$ = -9.1 J/K mol

Enzymatic catalysis

Constructing the model



Cyan: QM substrate (24-45 QM atoms)

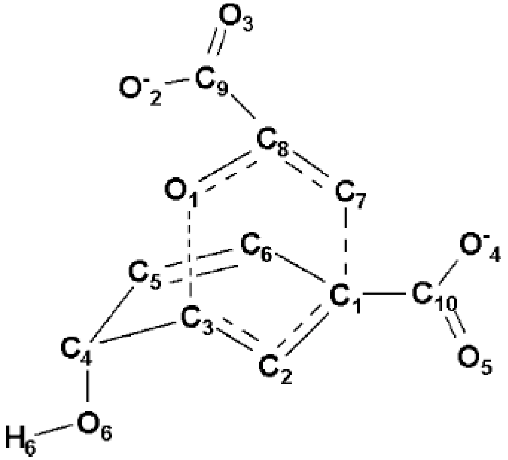
Blue: MM flexible atoms
Red: MM frozen atoms

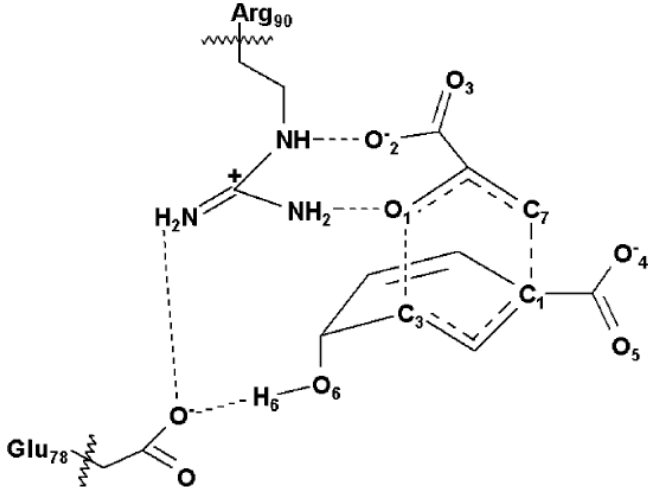
MM: 5627 protein atoms + 1534 TIP3P waters

J. Phys. Chem. B **2003**, 107, 13728

J. Am. Chem. Soc. **2005**, 127, 6940

Two QM models proposed



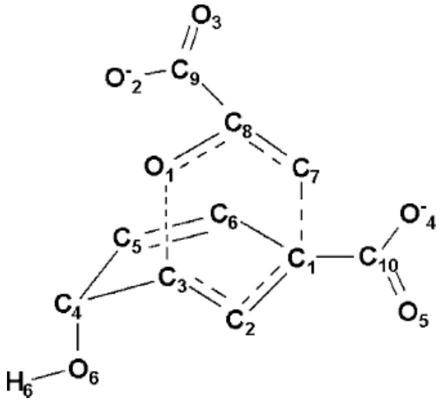


Enzymatic catalysis

Reaction path search

Reaction coordinate

$$\xi = d_{C1-C7} - d_{C3-O1}$$

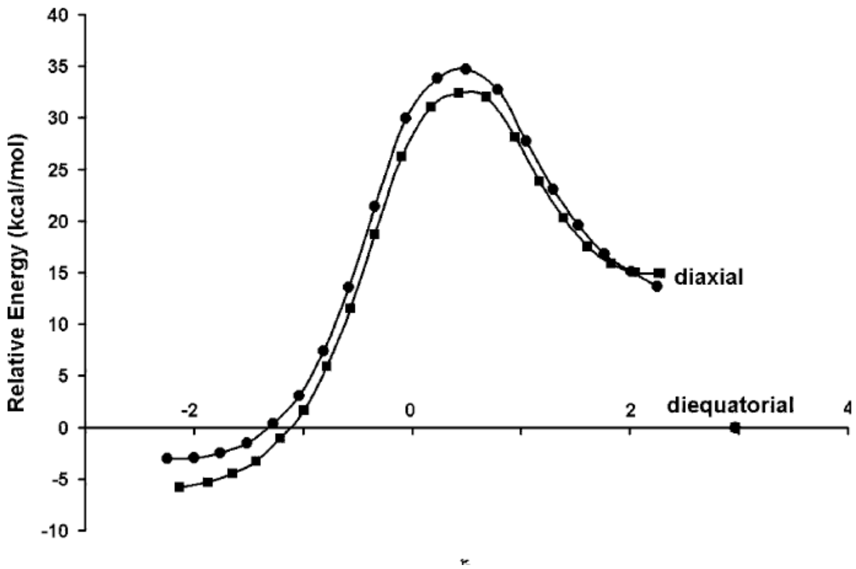


Restrained minimizations are performed with a penalty V_R added to the potential energy.

$$V_R = \frac{1}{2}k(\xi - \xi_0)^2$$

To build the curve, this term is substracted from the total energy.

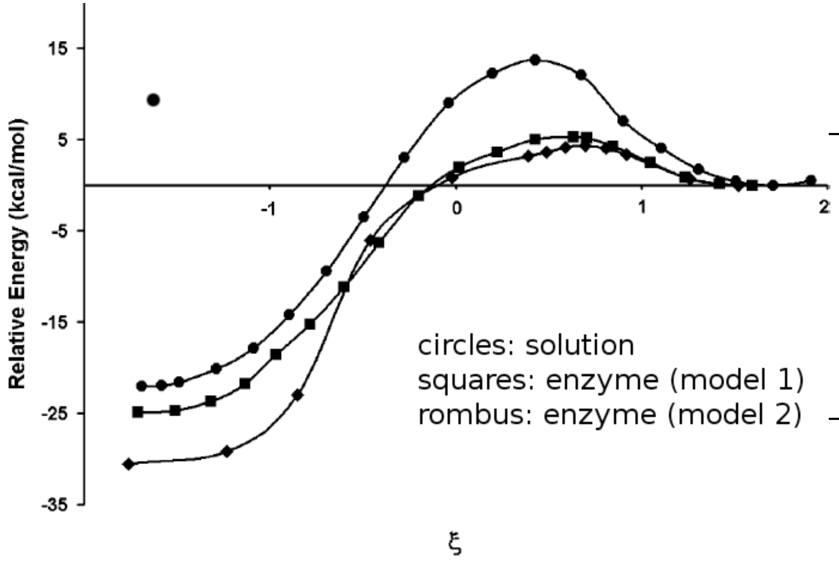
Energy profile in vacuum



Enzymatic catalysis

Solution versus protein

Energy profile in solution and in the enzyme



	ΔE kcal/mol	$\Delta E^{\ddagger}$ kcal/mol	$\Delta E^{\ddagger}_{exp}$ kcal/mol
Vacuum	-5.8	32.4	
Solution	-22.0	13.8	20.7
Enzyme (1)	-24.9	5.3	12.7
Enzyme (2)	-30.6	4.3	12.7

Net stabilization with respect to water, $\Delta(\Delta E^{\ddagger})$:

Experimental: 8.0 kcal/mol

QM-MM: 8.5 kcal/mol (model 1) or 9.5 kcal/mol (model 2)

Enzymatic catalysis

How does the enzyme help?

$$\Delta E^{\ddagger} = \Delta E_{QM} + \Delta E_{MM} + \Delta E_{QM-MM}$$

	Solution	**Enzyme**
ΔE_{QM}	13.5	11.5
E_{QM} rel	0	2.1
ΔE_{QM-MM}	5.9	-4.7
ΔE_{MM}	-5.6	-1.5
ξ (Å)	1.71	1.60
C1-C7 (Å)	3.62	3.13

Catalytic activity mainly due to both a major electrostatic stabilization of the transition state and a minor steric compression/destabilization of the substrate

Enzymatic catalysis

Including thermal and entropic effects

$$\Delta A = W, \text{ reversible process}$$

Jarzynski equality: $e^{-\beta \Delta A(\xi)} \leq \langle e^{-\beta W(\xi)} \rangle$

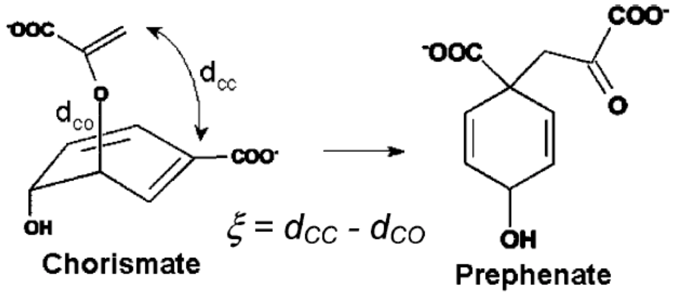
$$H(\mathbf{r}, \xi) = H_0(\mathbf{r}) + \frac{1}{2} k [\xi(\mathbf{r}) - \xi_0 - v \cdot t]^2 \Rightarrow W = \int \dot{\xi} \frac{\partial H}{\partial \xi}(\mathbf{r}(t)) dt$$

The free energy of a process along a reaction coordinate can be computed by performing a number of finite time transformations, and then averaging the work done at each time step

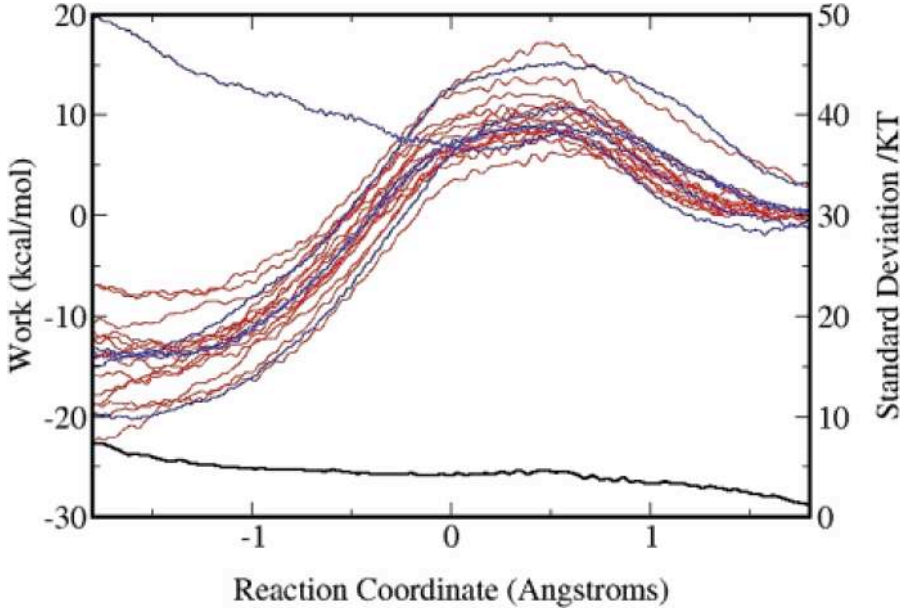
Jarzynski, *Phys. Rev. Lett.* **1997**, 78, 2690

Enzymatic catalysis

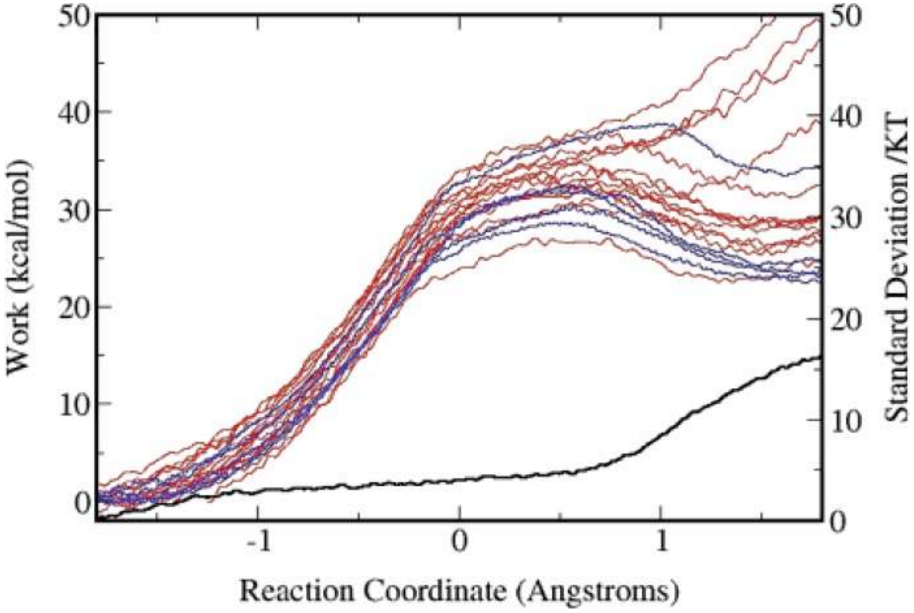
Free energy calculations



$\xi = d_{CC} - d_{CO}$



chorismate to prephenate



prephenate to chorismate

Enzymatic catalysis

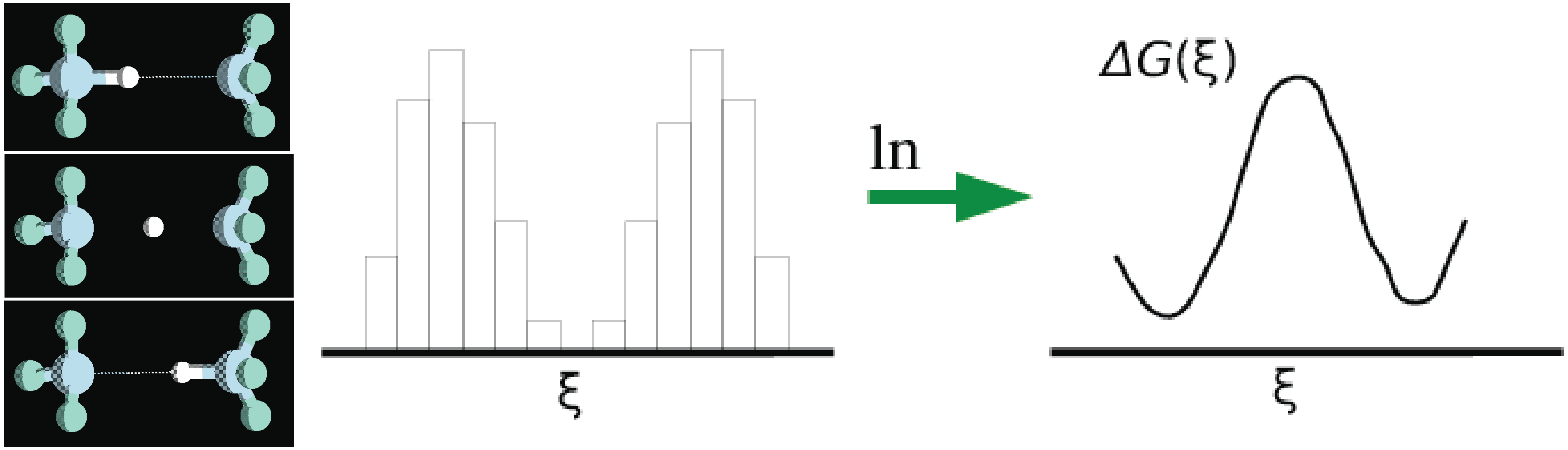
A dynamic picture: **Getting the free energy**

$$A = -kT \ln Q$$

$$\text{con } Q = \int d\mathbf{p}^N d\mathbf{r}^N e^{-E(\mathbf{p}^N,\mathbf{r}^N)/kT}$$

For a system constrained to a reaction coordinate ξ:

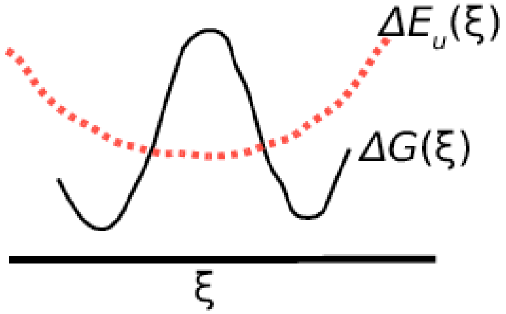
$$A(\xi) = -kT \ln P(\xi) + cte$$



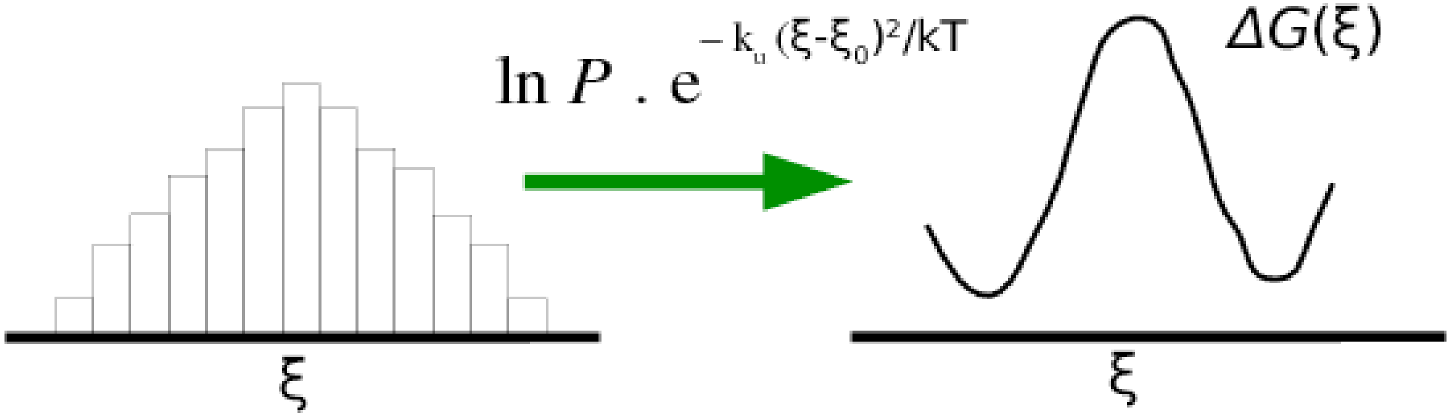
Enzymatic catalysis

A dynamic picture: **Umbrella Sampling**

$$E(\mathbf{r}^N) = E(\mathbf{r}^N) + k_u(\xi - \xi_0)^2$$

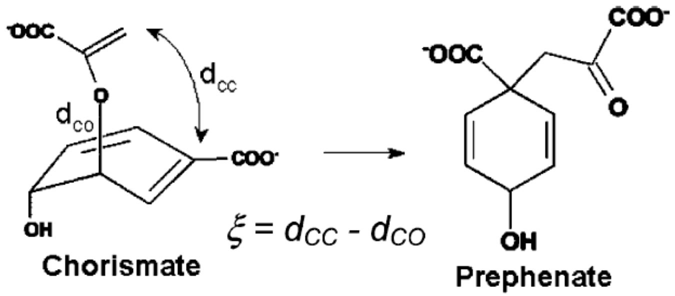


$$\langle A \rangle = \frac{\langle A(\xi) e^{k_u(\xi-\xi_0)^2/kT} \rangle_u}{\langle e^{k_u(\xi-\xi_0)^2/kT} \rangle_u} \Rightarrow A(\xi) = -kT \ln P(\xi) - k_u(\xi - \xi_0)^2 + cte$$

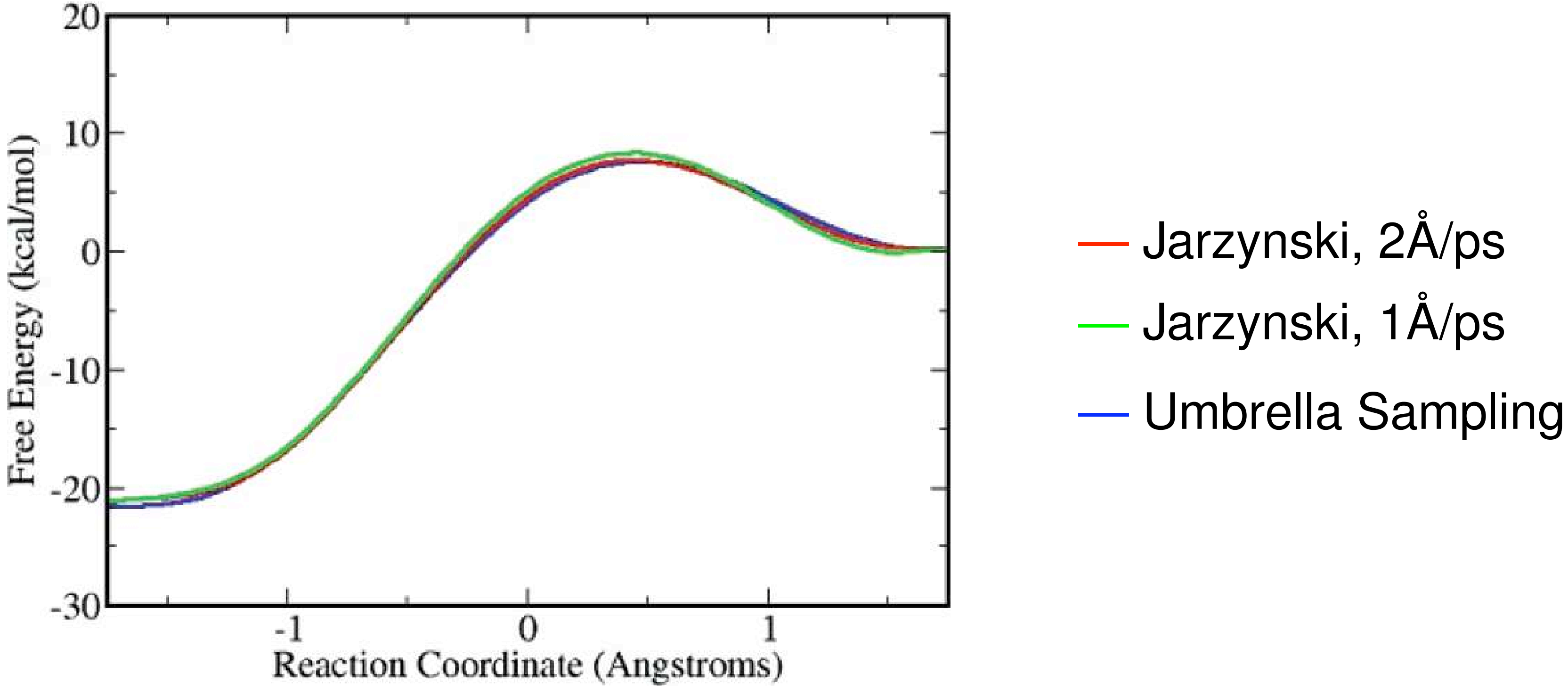


Enzymatic catalysis

The free energy



$$\Delta G^{\ddagger}_{exp} = 15.4 \text{ kcal/mol}$$
$$\Delta G^{\ddagger}_{QM-MM} = 8 \text{ kcal/mol}$$



Concluding remarks

General considerations

- Applications in solution chemistry, and biomolecules: large complex system with a singular active site
- Geometry minimizations and molecular dynamics. Important to discriminate massive contribution from the MM atoms
- No black boxes (yet)

Some challenges

- Description of the QM-MM boundary
- Development of schemes allowing net charge transfer between QM and MM domains: if this happen QM-MM fails!