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 $O(N^4)$ 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 Modelling Approach

- Couple quantum mechanics and molecular mechanics approaches
- QM treatment of the active site
  - reacting centre
  - excited state processes (e.g. spectroscopy)
  - problem structures (e.g. complex transition metal centre)
- Classical MM treatment of environment
  - enzyme structure
  - zeolite framework
  - explicit solvent molecules
  - bulky organometallic ligands
The QM/MM Idea

Multi-layered method

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 $^b$

---


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

\[
V(\vec{R}) = \sum_{i \in \text{bonds}} k_{i,\text{bond}} \left( r_{i,\text{bond}} - r_{i,\text{bond}}^{eq} \right) + \sum_{i \in \text{angles}} k_{i,\text{angle}} \left( \theta_{i,\text{angle}} - \theta_{i,\text{angle}}^{eq} \right) + \cdots \\
+ \frac{1}{2} \sum_{i \neq j}^{i,j \in \text{atoms}} \frac{q_i q_j}{|r_i - r_j|} + \frac{1}{2} \sum_{i \neq j}^{i,j \in \text{atoms}} V_{LJ}^{ij} \left( |r_i - r_j| \right)
\]

Not clear which force fields to use – much experience with expected accuracy of \textit{ab initio} methods alone and MM methods alone, but not much with QM/MM
No direct map from wavefunction to parameters
Subtractive QM/MM Coupling

Subtractive QM/MM coupling: The QM/MM energy of the total system (*left hand side of the equation*) is assumed to be equal to the energy of the isolated QM subsystem, evaluated at the QM level, plus the energy of the complete system evaluated at the MM level, minus the energy of the isolated QM subsystem, evaluated at the MM level. The last term is subtracted to correct for double counting of the contribution of the QM subsystem to the total energy. A prerequisite for the calculation is that a force field for the QM subsystem is available.
Additive QM/MM Coupling

Coupling between the QM and MM subsystems in the additive QM/MM schemes. The top panels (a) (c) show bonded interactions between QM and MM atoms. These interactions are handled at the force field level (MM). Panel d shows the Van der Waals interactions between an atom in the QM region and three MM atoms. These interactions are modelled by the Lennard-Jones potential. Panel e illustrates the link atom concept. This atom caps the QM subsystem and is present only in the QM calculation. Panel f demonstrates how the electrostatic QM/MM interactions are handled. In the electrostatic embedding approach, the charged MM atoms enter the electronic Hamiltonian of the QM subsystem. In the mechanical embedding, partial MM charges are assigned to the QM atoms and the electrostatic interactions are computed by the pairwise Coulomb potential.
General QM/MM Methodology

Two main strategies:

- **Additive Method**

  \[ H_{tot} = H_{QM}(QM) + H_{MM}(MM) + H_{QM/MM}(Inter.) \]

- **Subtractive Method**

  \[ E_{tot} = E_{QM}(QM) - E_{MM}(QM) + E_{MM}(All) \]
Hybrid Computational Schemes

Termination Scheme
- Link Atoms, or Boundary zone
- Chemical type (hydrogen atoms, pseudopotentials adjusted connection atoms, localised orbitals)
- Charge perturbations (none, charge deletion, charge shift, selection of 1e integrals, double link atoms)

QM/MM Couplings
- Unpolarised or “mechanical embedding”
- Polarisation of QM region
  - “electrostatic embedding”
- MM polarisation
  - shell model or dipole polarisabilities

Total Energy Expression
- Additive, Uncorrected
  \[ E(M,MM) + E(QL,QM) + E(QM/MM) \]
- Additive, Boundary corrected
  \[ E(M,MM) + E(QL,QM) - E(L,MM) \ldots + \]
- Subtractive
  \[ E(MQ,MM) + E(QL,QM) - E(QL,MM) \]
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...

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

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

\[ E_{QM} : HF \]
\[ E_{MM} : AMBER \]
\[ E_{QM/MM} : \text{Link atom} \]
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

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)

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

Link Atoms

- QM/MM with extra Link Atoms (L) to terminate broken covalent bonds
Boundary Regions

- Boundary region approaches introduce no new atoms, e.g. Solid State Systems with ionic character – Link atoms are inappropriate

  - Range of representations within QM code include modified \textit{ab-initio} atom with model potential, Semi-empirical parameterisation, Frozen orbitals, Design atoms
  - Often associated with re-parameterised MM potentials
Subtractive Schemes

- **Energy Expression**
  \[ E(OI, MM) + E(IL, QM) - E(IL, MM) \]
  - includes link atom correction
  - can treat polarisation of both the MM and QM regions at the force-field level

- **Termination**
  - Any (provided a force field model for IL is available)

- **Advantages**
  - Potentially highly accurate and free from artefacts
  - Can also be used for QM/QM schemes (e.g. IMOMO, Morokuma *et al*)

- **Disadvantages**
  - Need for accurate forcefields (mismatch of QM and MM models can generate catastrophes on potential energy surface)
  - Usually no electrostatic influence on QM wavefunction included (e.g. QMPot), (but can be extended to electrostatic embedding: ONIOM-EE)
## Additive Schemes

### Energy Expressions
- **Without link atom correction**
  \[ E(O,MM) + E(I,QM) + E(IO,QM/MM) \]
- **Link atom correction**
  \[ E(O,MM) + E(IL,QM) + E(IO,QM/MM) - E(L,MM) \]
- **Boundary methods**
  \[ E(OB,MM) + E(IB,QM) + E(IFO,QM/MM) \]

### Advantages
- No requirement for forcefield for reacting centre
- Can naturally build in electrostatic polarisation of QM region - effects of environment of excitations etc

### Disadvantages
- Electrostatic coupling of the two regions, \( E(IO,QM/MM) \) is problematic with link atoms
- Need for boundary atom parameterisation

### Highly variable in implementation
- QM/MM couplings,
- QM termination etc
Other Approaches

ONIOM\textsuperscript{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

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

which relies on the approximation

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

• The ‘model’ system still has to be properly terminated

• Extension to three level systems is relatively straightforward (e.g. \textit{ab initio} core, semi-empirical boundary, MM surroundings)

\textsuperscript{a}Svensson, M. \textit{et. al.} \textit{J. Phys. Chem.} \textbf{1996}, \textit{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 \textit{ab initio} calculations

- Treatment of bonds between the ‘true’ QM region and the fragments is still problematic


Adaptive QM/MM Schemes (i)

- Change of the QM region during the simulation
  - Potential for discontinuity in energy and forces
  - Generally based on principle that forces on atoms in a buffer region are *interpolated* between QM values and MM values, depending on the distance from QM zone

- Rode’s “Hot Spot” method

- ONIOM-XS

- LOTF schemes (MM with on-the-fly parameterisation)
Adaptive QM/MM Schemes (ii)

- Truhlar’s schemes define a conserved potential energy by performing multiple QM/MM calculations (permuting boundary molecules between QM and MM zones) with geometry dependent weights to apply interpolation.
- Number of possible contributions is $2^N$ (where $N$ molecules in the boundary zone).
- Schemes linear in $N$ are also possible, weighting functions are quite complex.


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

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.
Coarser than QM/MM?

Continuum solvation models – treat solvent as a dielectric continuum (PCM=Polarizable Continuum Model; SCRF=Self-Consistent Reaction Field)
Continuum Solvation

Algorithm:
- Compute “reaction field” – polarization of dielectric continuum which generates electric field acting on solute
- Compute electronic wavefunction in presence of new solvent-generated field
- Loop until reaction field does not change

Issues:
- Shape of cavity (spherical and ellipsoidal are rarely acceptable at present)
- Dielectric of solvent – zero vs infinite frequency?
- H-bonding between solvent and solute will not be properly represented
- Atomic radii used to generate cavity
Supermolecule Approach

- Explicit solvent molecules in first solvation shell
- Surround with dielectric continuum
- Expensive, but can be very accurate
- Not feasible if solute is very large

Related approach – QM/MM/PCM
Chorismate Mutase

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

Chorismate

Prephenate

Prephenate dehydratase

Prephenate dehydrogenase

Phenylpyruvate

4-Hydroxyphenylpyruvate

Aminotransferase

Phe

Tyr

Chorismate mutase

pheA

AF0227

NAD+

NADH+

H+

CO₂, H₂O
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
Define \textbf{X} number of steps to describe the pathway of interest
\[ \Delta H^\ddagger = 6.1 \]

\[ \Delta H = -18.5 \]
$\Delta H^\ddagger = 6.0$

$\Delta H = -18.9$
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 Method for Computing Free Energy Barriers
Off-Path Simulation Results: Butane at 300K

<table>
<thead>
<tr>
<th>Method</th>
<th>ΔG‡</th>
<th>ΔG</th>
<th>ΔG³</th>
<th>ΔG</th>
<th>ΔG‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binning</td>
<td>3.43</td>
<td>0.96</td>
<td>5.74</td>
<td>0.99</td>
<td>3.44</td>
</tr>
<tr>
<td>O.P.S.</td>
<td>3.49</td>
<td>1.02</td>
<td>5.72</td>
<td>1.02</td>
<td>3.48</td>
</tr>
<tr>
<td>Normal Modes</td>
<td>3.48</td>
<td>1.04</td>
<td>5.64</td>
<td>1.04</td>
<td>3.48</td>
</tr>
</tbody>
</table>
Off-Path Simulation Results: Maltose at 300K

Potential of Mean Force (kcal/mol)

Replica Number (i.e. Torsional Angle)

- Black line: 300K 50ps
- Red line: 300K 100ps
- Green line: 300K 200ps
- Blue line: 300K 500ps
- Cyan line: 300K 1000ps

Inset graph shows curves labeled X, Y, and C.
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...**
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 Method for Computing Free Energy Barriers
Off-Path Simulation Results: Butane at 300K

<table>
<thead>
<tr>
<th>Method</th>
<th>$\Delta G^\ddagger$</th>
<th>$\Delta G$</th>
<th>$\Delta G^\ddagger$</th>
<th>$\Delta G$</th>
<th>$\Delta G^\ddagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binning</td>
<td>3.43</td>
<td>0.96</td>
<td>5.74</td>
<td>0.99</td>
<td>3.44</td>
</tr>
<tr>
<td>O.P.S.</td>
<td>3.49</td>
<td>1.02</td>
<td>5.72</td>
<td>1.02</td>
<td>3.48</td>
</tr>
<tr>
<td>Normal Modes</td>
<td>3.48</td>
<td>1.04</td>
<td>5.64</td>
<td>1.04</td>
<td>3.48</td>
</tr>
</tbody>
</table>
Off-Path Simulation Results: Maltose at 300K
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...