Exploring QM/MM paths for mapping reaction mechanisms

H. Lee Woodcock

Computational Biophysics Section
Laboratory of Computational Biology
National Heart, Lung, and Blood Institute
Modeling Approaches

- **Research Design**
  - *ab initio* Quantum Mechanics
  - Molecular Simulations & Models
  - Macroscopic Properties

- **Mesoscale Simulations**
  - Rice Dwarf Virus Capsid
  - Assembly-level phenomena

- **Molecular Sims (atomic)**
  - Electronic Structure

- **System Size (# atoms)**
  - Time (real)
    - fs, ps, ns, μs, ms, s
Protein

QM Region

Substrate

Chorismate (A)

Chorismate-Prephenate Transition State Model (B)

Prephenate (C)

$\Delta G_1$

$\Delta G^{\ddagger}$

$\Delta G_{RXN}$

$\Delta G_2$
General QM/MM Methodology

Two main strategies:

- **Additive Method**

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

- **Subtractive Method**

\[
E_{tot} = E_{QM}(QM) - E_{MM}(QM) + E_{MM}(All)
\]
Reaction Path Methods
Reaction Path Methods

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

- **Reaction Coordinate Driving:**
  - Predetermined reaction coordinate
  - Usually some linear combination of distances
  - Gradually changed

- **Eigenvector Following Methods:**
  - Typically require transition state to be known a priori
  - Too expensive for high dimensional systems
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
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)
The Replica Path Method
The Replica Path Method
The Replica Path Method

Protein

QM Region

Substrate

Replicated Region
The Replica Path Method
The Replica Path Method

Define $X$ number of steps to describe the pathway of interest
The Replica Path Method

\[ E_{\text{RMS}} = \sum_1^N \frac{1}{2} K_r (r_i - \bar{r})^2 \quad r_i = RMSd_{\text{bestfit}}(i, i+1), \quad \bar{r} = \sum_1^N \frac{r_i}{N} \]

\[ E_{\text{ANGLE}} = \sum_1^N \frac{1}{2} K_\phi (\cos\text{max} - \cos (\phi_i))^2 \quad \text{If } \cos\text{max} > \cos (\phi_i) \]

\[ \text{RMSD}(R_i, R_{\text{ref}}) = \sqrt{\sum_1^N (R_i - R_{\text{ref}})^2 m_i w_i / \sum_1^N m_i w_i} \]
Chorismate Mutase

Plays a key role in the shikimate pathway of bacteria, fungi, and other higher plants.
Chorismate Mutase
<table>
<thead>
<tr>
<th>Level of theory</th>
<th>$\Delta E_{rxn}$</th>
<th>$\Delta E^\ddagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/6-31+G(d)/C22</td>
<td>-24.4</td>
<td>26.2</td>
</tr>
<tr>
<td>B3LYP/6-31+G(d)/C22</td>
<td>-19.5</td>
<td>8.95</td>
</tr>
<tr>
<td>RIMP2/6-31+G(d)/C22</td>
<td>-23.1</td>
<td>8.18</td>
</tr>
<tr>
<td>MP2/6-31+G(d)/C22</td>
<td>-23.1</td>
<td>8.20</td>
</tr>
<tr>
<td>SCC-DFTB</td>
<td>-22.1</td>
<td>5.79</td>
</tr>
</tbody>
</table>
What role does Arg63 play in the reaction?
What role does Arg63 play in the reaction?

- **SEMS: Single Environment, Multiple State**
- Employs Replica Path Method + RESDistance
- Optimize several conformations simultaneously
  - Single environment sees the average of all of the states
- Eliminates noise due to variations in the environment

- Chorismate Mutase: 2 Pathways
  - 2 reactants, 2 transition states, and 2 products
  - What role does Arg63 play in the reaction?
    - Catalytic?
$\Delta H^\ddagger = 6.1$

$\Delta H = -18.5$
\( \Delta H^\ddagger = 6.0 \)

\( \Delta H = -18.9 \)
What Next?

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

Two new methods to explore this...
VSA: Vibrational Subsystem Analysis

- **Goal:** Evaluate free energy of a system in the harmonic limit
  - Separate Hamiltonian
    - **Subsystem:** parts of the molecule that are directly involved in the functionality (for example, catalytic activity or ligand binding)
    - **Environment:** the remaining parts of the complex that move in response to changes in the subsystem
  - **Idea:** Fold environment motion into subsystem as a perturbation

Energy must be divided into two components:

\[ 2E = x^T H x = x_s^T H_{ss} x_s + x_s^T H_{se} x_e + x_e^T H_{es} x_s + x_e^T H_{ee} x_e \]

Subsystem \hspace{2cm} Mixed terms \hspace{2cm} Environment

VSA: Vibrational Subsystem Analysis

• **Potential Uses:**
  – Describe local-global coupling in coarse-grained macromolecular systems
  – Eliminating specific degrees of freedom without the detrimental effects of constraining the motion (i.e. making the system too rigid) or deleting part of the system (i.e. artificially increasing flexibility)
  – Elimination of “noise” when computing the harmonic vibrational free energy large biomolecular systems
  – Combining VSA with simulation approaches
  – Inclusion of very light or mass-less particles into NMA without the need for constraints or inclusion of unwanted high frequency heat capacity
Off-Path Simulation Method for Computing Free Energy Barriers
Off-Path Simulation Method for Computing Free Energy Barriers
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^f)</th>
<th>(\Delta G)</th>
<th>(\Delta G^f)</th>
<th>(\Delta G)</th>
<th>(\Delta G^f)</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>

Potential of Mean Force (kcal/mol) vs Torsional Angle (degrees)
Off-Path Simulation Results: Butane at 300K

<table>
<thead>
<tr>
<th>Method</th>
<th>ΔG‡</th>
<th>ΔG</th>
<th>ΔG‡</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>

Potential of Mean Force (kcal/mol)

Torsional Angle (degrees)
Off-Path Simulation Results: Maltose at 300K
Q-Chem/CHARMM QM/MM Method Development

- **Parallel Reaction Path Methods**
  - Replica Path (RPATH)\(^1\)
  - Nudged Elastic Band (NEB)\(^2,3\)
  - Q-Chem,\(^4\) SCC-DFTB\(^5\)

- **Drude polarization model\(^4\)**

- **Free Energy Perturbation\(^4\)**

- **Delocalized Gaussian MM charge (DGMM) methods\(^6\)**

- **ab initio QM/MM analytic Hessians (i.e. Frequency calculations, Normal Mode Analysis)\(^7\)**

- **General multiscale modelling approach (MSCALE)**

- **QM/MM Micro-iteration scheme\(^8\)**

- **CHARMMing\(^9\): Web portal to CHARMM**

---

Conclusions
Conclusions

- Replica Path Method
  - Chorismate Mutase reaction profile
  - Examined methodological dependence
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)
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)**
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
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...
Acknowledgments

Bernard Brooks – LCB/CBS/NHLBI/NIH
Funding: National Heart, Lung and Blood Institute, NIH
Milan Hodoscek – Nat. Inst. Chem. (NIC)
Yihan Shao – Q-Chem Inc.
Wenjun Zhang – LCB/CBS/NHLBI/NIH
An Ghysels – Ghent University Belgium
Tim Miller – LCB/CBS/NHLBI/NIH
Rishi Singh – Univ. of Virgina
Rich Pastor – LCB/MBS/NHLBI/NIH
Paul Sherwood – Daresbury Laboratory

Thank You