13th International Workshop on
Computational Physics and Materials Science:
Total Energy and Force Methods

11 - 13 January 2007

Exploring chemical reactivity in biological systems
with hybrid QM-MM methods

Dario A. ESTRIN
University of Buenos Aires
School of Sciences
Department of Chemistry
Pab. 2, 1424 Buenos Aires
ARGENTINA

These are preliminary lecture notes, intended only for distribution to participants
Exploring chemical reactivity in biological systems with hybrid QM-MM methods

Darío A. Estrin
University of Buenos Aires
ARGENTINA
Outline:

- Introduction
- Validation: the chorismate mutase case
- Oxygen binding in globins: distal, proximal and dynamical effects
- NO detoxification by truncated hemoglobin of M. Tuberculosis
- Problems and perspectives
Computer simulation in Chemistry and Biochemistry:

- Tools based on physical-chemistry ideas

1) Models: aimed to evaluate potential energy surfaces (PES)

\[ E = E(r_1, r_2, r_3, ...) \]

Quantum chemical and classical models

2) Schemes to extract relevant information from PES: molecular dynamics (MD), Monte Carlo (MC), optimizations.
QM models:

- Necessary to describe electronic excitations, and chemical reactions (changes in bonding patterns).
- Limitation in system sizes! Biomolecules: DFT method of choice

Classical models:

- Based on classical ideas. Very efficient numerically.
- Cannot (usually) deal with reactive processes, but are extremely useful for non reactive processes (i.e. protein folding)
- Amber, Charmm, Gromos, etc.
QM-MM hybrid schemes:
Strategy to be able to investigate real chemistry in biomolecules

Reactive region: QM
Environment: MM

Our group implementation: SIESTA DFT/Amber

\[ E_{TOT} = E_{QM} + E_{MM} + E_{QM-MM} \]

Key term: QM-MM coupling

\[ E_{QM-MM} = \sum_{i=1}^{C} q_i \int \frac{\rho(r)}{|r - \tau_i|} \, dr + \sum_{i=1}^{C} \sum_{a=1}^{A} \frac{q_i Z_a}{|R_a - \tau_i|} + E_{LJ} \]

*QM nuclei / environment partial charges interaction*

**QM density - partial charges interaction**

**LJ term: short range, dispersion, etc.**
Extracting valuable info from simulations:

- Geometry optimizations, constrained optimizations, energy changes, reaction profiles. NO TEMPERATURE!!
- Molecular dynamics: thermal effects. Time evolution in the nanosecond range.
- Advanced sampling techniques: umbrella sampling, multiple steering molecular dynamics (Jarzinski’s method): Free energy!

Understanding molecular determinants of a given property

Obtaining information not accessible experimentally
Experiment/Theory collaboration essential
Methodology validation: application to chorismate mutase

(Biosynthesis of aromatic amino acids in bacteria, plants, and fungi)

One bond forms and other breaks, simple choice of reaction coordinate

Scheme 1. Chorismate to Prephenate Conversion Reaction

First approach: constrained optimizations

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

Addition of harmonic term

Siesta PBE-DZVP/Amber level

2 possible QM subsystems: reactant, reactant plus nearby aminoacids
QM subsystem choice: little influence on the profiles

The catalytic effect is reproduced

Thermal and entropic effects neglected. Is it OK?

Figure 3. Energy profile for the reaction of chorismate to prephenate in aqueous solution (circles) and in the enzyme with the two different QM subsystems: substrate (squares), substrate plus the charged side chains glu78 and arg90 (rombus).
How to incorporate thermal and entropic effects?

Free energy profiles: much more expensive!

Umbrella sampling or
Multiple Steered Molecular Dynamics MSMD: interesting strategy

\[ e^{-\beta \Delta G(\lambda)} = < e^{-\beta W(\lambda)} > \]  

Jarzynski equation

A set of steered MD are performed. From the irreversible works the free energy change can be estimated

$$e^{-\beta \Delta G(\lambda)} = \langle e^{-\beta W(\lambda)} \rangle$$

$$H(r, \lambda) = H_0(r) + \frac{1}{2}k [\lambda(r) - \lambda_0 - vt]^2$$
MSMD is more efficient than Umbrella Sampling for the user and is easily parallelized.

**Figure 2.** Free energy profile from chorismate ($\xi \approx 1.75$ Å) to prephenate ($\xi \approx -1.75$ Å), calculated using Jarzynski’s equality (both forward and reverse data are used) for set 1 (red), for set 2 (green), and for umbrella sampling scheme (blue).

*Multiple-Steering QM-MM Calculation of the Free Energy Profile in Chorismate Mutase*  
Heme proteins

*Ideal benchmarks for QM-MM*

**Active site: heme**
- Iron coordinated to a porphyrin

**Very different roles:**
- $O_2$ transport
- Electron transfer
- Hormone biosynthesis
- Detoxification
Dioxygen binding in Globins: crucial process involved in transport and chemistry

$O_2$ entry ($k_{on}$): mainly related with ligand migration

Classical MD simulations

$O_2$ exit ($k_{off}$): mainly related with bond breaking:

QM-MM calculations

Distal (direct) effects
Proximal (indirect) effects
Distal Effects:

O₂ acquires a negative charge upon binding (electrostatic) H bonding and steric effects

Free heme: 22.0 kcal/mol

TrHbN: 37.2 kcal/mol
AscHb: 34.3 kcal/mol
Cyt c’: 8.4 kcal/mol
Mb: 27 kcal/mol

Dynamical distal effects: multiple conformations
CerHb case

QM-MM calculations, DFT SIESTA level and classical MD simulations


Other similar cases: LegHb, Hb P. caudatum
Proteins, in press.
More subtle effects through the protein backbone: proximal effects

Try to analyze trends by constructing model systems in the first place:

- Histidine rotational position (staggered vs eclipsed)
- Fe-Histidine bond distance

QM calculations, DFT SIESTA level, PBE, DZVP basis sets

$21.4 \text{kcal/mol}$  $24.5 \text{kcal/mol}$
Charge transfer to histidine effects: intermediate degree enhances Fe-O₂ binding

<table>
<thead>
<tr>
<th></th>
<th>ΔE₀² (kcal/mol)</th>
<th>Δq₀² (e)</th>
<th>Δq_prox (e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>isolated</td>
<td>22.2</td>
<td>-0.202</td>
<td>0.158</td>
</tr>
<tr>
<td>heme–imidazole complex</td>
<td>22.0</td>
<td>-0.197</td>
<td>0.160</td>
</tr>
<tr>
<td>Fix-L⁴⁰</td>
<td>24.7</td>
<td>-0.248</td>
<td>0.200</td>
</tr>
<tr>
<td>Hbβ⁴¹</td>
<td>18.5</td>
<td>-0.259</td>
<td>0.325</td>
</tr>
</tbody>
</table>

Full protein: QM-MM calculation

LegHb: “strange” protein!

NO detoxification by a truncated hemoglobin (trHbN) of *Mycobacterium Tuberculosis*

NO produced by the immune system, is inactivated by the oxygenated Hb:

\[
\text{toxic} \rightarrow \text{innocuous}
\]

\[
\text{NO} + \text{Fe(II)} \text{O}_2 \rightarrow \text{Fe(III)} + [\text{NO}_3^-] -
\]

Relevant process in physiological (Hb,Mb) and pathological processes
Reaction mechanism:

1) $O_2$ migration (classical MD)
2) $O_2$ binding (QM-MM)
3) NO migration (classical MD)
4) Reaction of NO with $O_2$ (QM-MM)

\[
NO + Fe(II)O_2 \rightarrow Fe(III) + [NO_3]^-
\]
2 channel system proposed on the basis of x-ray results (Bolognesi’s group at Milano) But how do $O_2$ and NO migrate?

Long (100 ns) classical MD (Amber 9) of oxy and deoxy proteins

Free energy profiles for ligand migration computed based on classical MD (no QM) (Our Jarzinski implementation available in Amber9)

What about the chemical processes?

QM-MM optimizations
Siesta PBE-DZVP/Amber level

Oxygen affinity
Hydrogen bond with OH of Tyr33 (B10)

<table>
<thead>
<tr>
<th>isolated</th>
<th>Mb</th>
<th>TrHb</th>
<th>Tyr33→Phe33</th>
</tr>
</thead>
<tbody>
<tr>
<td>-21.4</td>
<td>-25.0</td>
<td>-37.2</td>
<td>-34.3 kcal/mol</td>
</tr>
</tbody>
</table>

Effect of Tyr33→Phe33 mutation on $k_{off}$ is reproduced. Affinity is large (consistent with the detoxification role).
Chemical reaction:

\[
\text{Fe(II)-O}_2 + \text{NO} \rightarrow \text{Fe(III)} [-\text{OONO}] \quad (1) \quad \xi_1 = d(O_{O2}-N_{NO})
\]

\[
\text{Fe(III)} [-\text{OONO}] \rightarrow \text{Fe(IV)}=\text{O} + \text{NO}_2 \quad (2) \quad \xi_2 = d(O_{O1}-O_{O2})
\]

\[
\text{Fe(IV)}=\text{O} + \text{NO}_2 \rightarrow \text{Fe(III)}[-\text{NO}_3] \quad (3) \quad \xi_3 = d(O_{O1}-N_{NO})
\]
Step 1

Vacuum

Step 2

Water

Step 3

Protein

Almost barrierless Reaction in protein is similar to that in water
Conclusions:

NO detoxification ability of trHbN is related to:

- Existence of an adequate ligand migration pathway
- High affinity for O$_2$ due to distal stabilization (H bonds)
- No significant protein effects on chemical reaction. Only spatial restriction of intermediates.

Flaws:

DFT at GGA level. Underestimates reaction barriers, problems with spin energetics of transition metals.
Sampling problems. In many cases optimizations are not enough!

Perspectives:

GGA+ U; other functionals; making program more efficient to be able to sample more, PBC to treat solvation, etc…….
In collaboration with:

Buenos Aires’ s Group

D. Elola, D. A. Scherlis, A. Turjanski, M. L. Fernández, A. Crespo,
A. Nadra, M. Martí, L. Capece, M. González Lebrero, L. Perissinotti, D. Bikiel, L. Boechi,

J. Luque (U. Barcelona), P. Ordejón (ICMB, Barcelona), M. Bolognesi (U. Milano), A. Roitberg (U. Florida).
Charge density partial charges interaction is computed self consistently.

\[ E = \begin{cases} 
\sum_{n=1}^{\text{grilla MM}} \sum_{i=1}^{\text{MM}} \frac{\rho(r_n)q_i}{R_{\text{cutQM}}} & \left| \tau_i - r_n \right| \leq R_{\text{cutQM}} \\
\sum_{n=1}^{\text{grilla MM}} \sum_{i=1}^{\text{MM}} \frac{\rho(r_n)q_i}{\left| \tau_i - r_n \right|} & R_{\text{cutQM}} < \left| \tau_i - r_n \right| \leq R_{\text{cutQMMM}} \\
0 & \left| \tau_i - r_n \right| > R_{\text{cutQMMM}} 
\end{cases} \]

Generally: \( R_{\text{cutQM}} \) 0.2-0.3 Å and \( R_{\text{cutQMMM}} \) 8-10 Å.
Link atom: *frontier between QM and MM systems*

- $C_{MM} - C_{QM}$ is broken and an H atom is added.
- Forces over $H_L$ are divided, scaled, and added to $C_{MM}$ and $C_{QM}$.
- Classical terms involving $C_{MM}$ are computed.