Conformational change in biology: from amino acids to enzymes and molecular motors.

Computational Facilities: NERSC
Collaborators: Martin Karplus, Eric Vanden-Eijnden, Kwangho Nam, Anne Houdusse, Robert Sauer
Financial support: NIH
Introduction

- Conformational motions in biomolecules define all living things
  - Transport across membranes
  - Enzyme reactions (from proton transfer to DNA replication and repair)
  - Linear (Myosin, Kinesin) and Rotary (F$_1$ATPase, ClpX Dynein) motors

- We would like to understand how chemical energy is used to generate force and motion at the molecular level
  - Biological processes are
    - Inherently “renewable”
    - Efficient (light harvesting by photosynthetic bacteria)
    - Robust w.r.t. environmental perturbations (e.g. from temperature changes to antibiotics)
Enzymes

DNA

hOGG1 repair enzyme

Mutated base

Human DNA repair enzyme
hOGG1 bound to site of DNA mutation

(With Profs. Kwangho Nam, Xray structures : Greg Verdine)

ATP $\rightleftharpoons$ ADP + Pi + H2O

Linear motors

Lever arm

M V Monomer

“Motor” head

Actin

Rotary motors

F1-ATPase

Myosin V dimer taking variable steps

36 nm

40 nm
Example: ClpX protein unfolding machine

- Part of protein “recycling” complex
- Six subunits (identical in sequence) adopt different nucleotide-dependent (ATP vs. ATP) conformations
- Energy of ATP hydrolysis coupled to “threading” motion

\[ \text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{Pi} + \text{work} \]

Collaboration with Prof. Robert Sauer, MIT
Example: ClpX protein unfolding machine

Collaboration with Prof. Robert Sauer, MIT
Example: ClpX protein unfolding machine

- How are force generation and motion coupled to hydrolysis?

Collaboration with Prof. Robert Sauer, MIT
Challenges

- Experiments can be difficult
  - Small spatial scales: \( \sim 10^{-9}\) m, variable temporal scales: \(10^{-12} - 10^{-3}\) s

- Thermal fluctuations can complicate measurements

- Carefully parametrized computer simulations can help elucidate basic thermodynamics and kinetics, e.g.:
  1. Stability of intermediates (\(\Delta F = \Delta H - T\Delta S\))
  2. Pathways of transition
  3. Rate of transition, e.g. Markov state models or TST: \(k_{TST} = \kappa e^{-\frac{\Delta F_{TST}}{k_BT}}\)
• Endpoints are provided from crystallography (or NMR)

• Intermediates must be found (computed)
String methods (2D landscape example)

- Transition pathways between endpoint coordinates can be obtained using equilibration in “path space”

- Note: a medium-sized protein, e.g., calcium-binding calmodulin has $3N-6 \sim 7000$ dof
String methods (2D landscape example)

- Each string point is an (almost) independent MD simulation

- Note: a medium-sized protein, e.g., calcium-binding calmodulin has $3N-6 \sim 7000$ dof
Theoretical/Computational Challenges (II)

- “States” \( \{ B_i \} \) must be defined precisely:
- e.g. reactant, product, transition state

\[
F_{B_i} = -k_B T \log \int_{B_i \in \mathbb{R}^{3N-6}} e^{-\beta E(\vec{r})} d\vec{r}
\]

- Many configurations \( r \) comprise a state (but which ones?)

- Also known as the “reaction coordinate problem”: which \( r \) correspond which stage of reaction (reactants, products, intermediates)?
Transition path theory and string methods

- For overdamped LD: states are locally separated by hyperplanes perpendicular to “average” path
  \[ n(q_i) \parallel \phi'(q_i) \]

- Can also change variables
  \[ r \rightarrow \xi \]
  \[ n_\xi(q_i) \parallel \left( \langle \nabla_r(\xi)^T \nabla_r(\xi) \rangle_{\xi=\phi}^{-1} \phi'(q_i) \right) \]
  \[ M^{-1} \]
  (M: Metric tensor)

- This is the basis for the string method in collective variables

\[ \phi(q) \in \mathbb{R}^{3N-6} \text{ average path (string)} \]
\[ q \in [0, 1] \text{ string parameter} \]
\[ n(q) \text{ normal to hyperplane} \]

\[ q(r) = q_i \]

\[ n(q_i) \cdot [r - \phi(q_i)] = 0 \]

(Hyperplane approximation)

Free energy from tessellations

- The free energy can be computed from a tessellation, e.g.:

\[ B_i = \{ r : \| r - A\phi_i \| < \| r - A\phi_j \|, \forall j \neq i \} \]

by constrained MD in each \( B_i \) and flux matching

\[ 0 = -\sum_{k \neq i} P_{i \rightarrow k} R_{i \rightarrow k} + \sum_{k \neq i} P_{k \rightarrow i} R_{k \rightarrow i} \]

\[ F_i = -kT \ln P_i \]

- (Note: \( P \) is the invariant distribution of a Markov state model with transition matrix \( R \))
- Does not allow sampled surfaces to cross (by construction)
String methods (2D landscape example)

- Free energy profile can be computed from many short quasi-equilibrium simulations
Parallel string method (CHARMM)

- **MULTICOM:**
  - Interactive module to add/modify/assign MPI communicators at runtime

```
| n0 | n1 | n2 | n3 | n4 |
```

- **MPI_COMM_LOCAL** [=COMM_CHARMM] (assigned to different communicators on different nodes)
  - Dynamics

```
| n0 | n1 |
```
```
| n0 | n1 |
```
```
| n0 | n1 | n2 | n3 |
```
```
| n0 | n1 | n2 | n3 | n4 |
```

- **Each _LOCAL group only needs** $\phi_i$, $\phi_{i\pm 1}$
• Transition pathways between endpoint coordinates can be obtained using equilibration in “path space”

• Note: a medium-sized protein, e.g., calcium-binding calmodulin has $3N-6 \sim 7000$ dof
Parallel performance

• TEST SYSTEM: Calmodulin (2272 atoms)
  FACTS implicit solvent model

• No slowdown with increasing resolution (weak scaling)

<table>
<thead>
<tr>
<th>test case</th>
<th>total cores</th>
<th>replicas</th>
<th>cores/rep</th>
<th>total atoms</th>
<th>steps</th>
<th>time (m)</th>
<th>steps/m</th>
<th>speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTSM simulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>128</td>
<td>4</td>
<td>32</td>
<td>9088</td>
<td>20000</td>
<td>10</td>
<td>2000</td>
<td>x 2.9</td>
</tr>
<tr>
<td>2.</td>
<td>4096</td>
<td>128</td>
<td>32</td>
<td>290816</td>
<td>138100</td>
<td>72</td>
<td>1980</td>
<td>x 2.8</td>
</tr>
</tbody>
</table>

• String calculations are 3-4 times slower relative to simple MD

<table>
<thead>
<tr>
<th>test case</th>
<th>total cores</th>
<th>replicas</th>
<th>cores/rep</th>
<th>total atoms</th>
<th>steps</th>
<th>time (m)</th>
<th>steps/m</th>
<th>speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTSM simulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>1024</td>
<td>128</td>
<td>8</td>
<td>290816</td>
<td>7000</td>
<td>10</td>
<td>700</td>
<td>x 1</td>
</tr>
<tr>
<td>MD simulation (non-interacting replicas)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>160</td>
<td>20</td>
<td>8</td>
<td>45440</td>
<td>150000</td>
<td>69</td>
<td>2173</td>
<td>x 3.1</td>
</tr>
</tbody>
</table>

• Force calculations are expensive

• Differentiation of coordinate transformations can be made faster

• Smaller subsets of coarse-grained variables will increase speed

• Use of explicit solvent will mask additional string overhead
Application: $\alpha$-helix $\rightarrow$ $\beta$-sheet transition

- Plays a role in
  - Alzheimer's (\(\beta\)-amyloid aggregation)
  - Diabetes
Barriers are due to the sequential unwinding of $\alpha$-helical turns.

- Highest barrier corresponds to hairpin formation.
- Occurs in the “middle” of the path.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$-helix</th>
<th>$\beta$-sheet</th>
<th>$\Delta_{\beta \rightarrow \alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G$</td>
<td>-75.9±0.3$^\dagger$</td>
<td>-82.6±0.3$^\dagger$</td>
<td>6.7±0.4</td>
</tr>
<tr>
<td>$E$</td>
<td>-332.6±0.5</td>
<td>-346.2±0.5</td>
<td>13.6±0.7</td>
</tr>
<tr>
<td>$TS$</td>
<td>-37.6±0.6$^\dagger$</td>
<td>-44.5±0.6$^\dagger$</td>
<td>6.9±0.8</td>
</tr>
</tbody>
</table>
Application: Myosin VI motor

- ATP-driven “motor”
- Present in all Eukaryotic cells
- MVI dimers “walk” on actin filaments in 36 nm steps
- Transports various cellular cargo (organelles, membranes)
- Walks “backwards”: i.e. In the direction opposite to that of other myosins

Converter/Lever arm in ‘Pre-powerstroke’ (PPS) position

Converter/Lever arm in ‘Rigor’ position

Powerstroke:
- lever arm swing
- Pi release (90 s⁻¹)

[De La Cruz, Ostap & Sweeney J. Biol. Chem. 34, 32373, 2001]
How does MVI walk backwards?

- We would like to understand the powerstroke
- No detailed computational studies have been done
- (1) Small-scale converter rearrangement
  - MVI converter adopts two conformations:
  - Which one is more stable and why?
  - What is the mechanism and rate of isomerization?
    - Is this step rate-limiting in the powerstroke?
    - What mutations could change the rate?
- (2) Large-scale rearrangement
- Future studies
Myosin VI converter rearrangement

Rate: \(2.5 \times 10^6 \text{s}^{-1}\)
Summary and Ongoing work

• String method is a powerful tool to study biomolecular systems:
  • Simulates entire “transition” paths
  • (almost) trivially parallel
  • Independent of force-field (MM, QM/MM)
  • Can be applied Cartesian space or in coarse variable space (e.g. distances between amino acid COMs)

• Ongoing application to other biological/chemical problems:
  • Proton transfer reactions using QM/MM (with Prof. Qiang Cui)
  • DNA remodeling by topoisomerases (with Prof. Ioan Andricioaei)
  • DNA repair enzymes (with Prof. Kwangho Nam)
  • Triose phosphate isomerase (TIM) enzyme (with Dr. Guishan Zheng)