Modeling and Molecular Dynamics of Membrane Proteins

Emad Tajkhorshid
NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute, University of Illinois at Urbana-Champaign
Why Do Living Cells Need Membrane

- Living cells also need to exchange materials and information with the outside world... however, in a highly selective manner.

Extracellular (outside)

Cytoplasm (inside)
Phospholipid Bilayers Are Excellent Materials For Cell Membranes

- Hydrophobic interaction is the driving force
- Self-assembly in water
- Tendency to close on themselves
- Self-sealing (a hole is unfavorable)
- Extensive: up to millimeters
Lipid Diffusion in a Membrane

Lateral diffusion

\[ D_{\text{lip}} = 10^{-8} \text{ cm}^2\text{s}^{-1} \]
(50 Å in \( \sim 5 \times 10^{-6} \text{ s} \))

\[ D_{\text{wat}} = 2.5 \times 10^{-5} \text{ cm}^2\text{s}^{-1} \]

Modeling mixed lipid bilayers!

Very slow

Tranverse diffusion (flip-flop)

Once in several hours!

(\( \sim 50 \text{ Å in } \sim 10^{4} \text{ s} \))

\(~9 \text{ orders of magnitude slower}\)

\( \text{ensuring bilayer asymmetric} \)
Technical difficulties in Simulations of Biological Membranes

- **Time scale**
- **Heterogeneity of biological membranes 😞**

60 x 60 Å
Pure POPE
5 ns
~100,000 atoms
Coarse-grained modeling of lipids

150 particles

9 particles!

Also, increasing the time step by orders of magnitude.
Analysis of Molecular Dynamics Simulations of Biomolecules

- A very complicated arrangement of hundreds of groups interacting with each other
- Where to start to look at?
- What to analyze?
- How much can we learn from simulations?

It is very important to get acquainted with your system
Aquaporins
Monomeric pores
Water, glycerol, ...

Tetrameric pore
Perhaps ions???

Aquaporins of known structure:

GlpF - E. coli glycerol channel (aquaglycerolporin)
AQP1 - Mammalian aquaporin-1 (pure water channel)
AqpZ and AQP0 (2004)
Functionally Important Features

- Tetrameric architecture
- Amphipatic channel interior
- Water and glycerol transport
- Protons, and other ions are excluded
- Conserved asparagine-proline-alanine residues; NPA motif
- Characteristic half-membrane spanning structure

~100% conserved -NPA- signature sequence
A Semi-hydrophobic channel
Molecular Dynamics Simulations

Protein: ~ 15,000 atoms
Lipids (POPE): ~ 40,000 atoms
Water: ~ 51,000 atoms
Total: ~ 106,000 atoms

NAMD, CHARMM27, PME
NpT ensemble at 310 K
1ns equilibration, 4ns production
10 days /ns - 32-proc Linux cluster
3.5 days/ns - 128 O2000 CPUs
0.35 days/ns - 512 LeMieux CPUs
Protein Embedding in Membrane

Hydrophobic surface of the protein

Ring of Tyr and Trp
Embedding GlpF in Membrane
Animation available at the Nobel web site
One of the most useful advantages of simulations over experiments is that you can modify the system as you wish: You can do modifications that are not even possible at all in reality!

This is a powerful technique to test hypotheses developed during your simulations. Use it!
Electrostatic Stabilization of Water Bipolar Arrangement

Characterizing Protein Forces

QM/MM MD of proton behavior in the channel
Water Bipolar Configuration in Aquaporins
Proton transfer through water

\[ H^+ \rightarrow H^+ \rightarrow H^+ \rightarrow H^+ \rightarrow H^+ \rightarrow H^+ \rightarrow H^+ \rightarrow H^+ \rightarrow H^+ \]
**Steered Molecular Dynamics is a non-equilibrium method by nature**

- A wide variety of events that are inaccessible to conventional molecular dynamics simulations can be probed.

- The system will be driven, however, away from equilibrium, resulting in problems in describing the energy landscape associated with the event of interest.

Second law of thermodynamics: \[ W \geq \Delta G \]
In principle, it is possible to obtain free energy surfaces from repeated non-equilibrium experiments.
AqpZ vs. GlpF

- Both from *E. coli*
- AqpZ is a pure water channel
- GlpF is a glycerol channel
- We have high resolution structures for both channels
Steered Molecular Dynamics

constant force
(250 pN)

constant velocity
(30 Å/ns)
SMD Simulation of Glycerol Passage

Trajectory of glycerol pulled by constant force
Constructing the Potential of Mean Force

4 trajectories
$v = 0.03, \ 0.015 \ \text{Å/ps}$
$k = 150 \ \text{pN/Å}$

\[ f(t) = -k[z(t) - z_0 - vt] \]

\[ W(t) = \int_0^t dt' \ \nu f(t') \]
• Captures major features of the channel
• The largest barrier $\approx 7.3$ kcal/mol; exp.: $9.6 \pm 1.5$ kcal/mol

Features of the Potential of Mean Force

\[ e^{-\frac{\Delta G}{k_B T}} = \langle e^{-\frac{W}{k_B T}} \rangle \]

Asymmetric Profile in the Vestibules

Artificial induction of glycerol conduction through AqpZ

Three fold higher barriers

\[
\text{AqpZ} \quad 22.8 \text{ kcal/mol} \\
\text{GlpF} \quad 7.3 \text{ kcal/mol}
\]

Could it be simply the size?

Battling the Timescale - Case II
Biased (nonequilibrium) simulations


✦ Neurotransmitter Uptake
  » Norepinephrine, serotonin, dopamine, glutamate,…

✦ Gastrointestinal Tract
  » Active absorption of nutrients
  » Secretion of ions

✦ Kidneys
  » Reabsorption
  » Secretion

✦ Pharmacokinetics of all drugs
  » Absorption, distribution, elimination
  » Multi-drug resistance in cancer cells
Alternating Access Mechanism

Outward-facing

Inward-facing


Non-equilibrium methods are required.
Complex Processes Require Complex Treatments

I.1 Defining Practical Collective Variables
Empirical search for practical collective variables for inducing the conformational changes involved in the transition.

I.2 Optimizing the Biasing Protocols
Systematic search for a practical biasing protocol by using different combinations of collective variables.

II. Optimizing the Transition Pathway
Use all of the conformations available to generate the most reliable transition pathway:
1. Bayesian approach for combining the data
2. Post-hoc string method (analysis tool)
3. String method with swarms of trajectories

III.1 Free Energy Calculations
Using the most relevant collective variables (from I.1), biasing protocol (from I.2), and initial conformations (from I.2).

III.2 Assessing the Sampling Efficiency
Detecting the poorly sampled, but potentially important regions, e.g., by using PCA.

Aggressive Search of the Space

Inward-Facing

Outward-Facing

Optimal Path

Refined TMD

TMD
Non-equilibrium Driven Molecular Dynamics:
Applying a time-dependent external force to induce the transition
Along various pathways/mechanisms (collective variables)

\[ U_{dr}(x, t) = \frac{1}{2} k \left( \xi(x) - \xi_A + (\xi_B - \xi_A) \frac{t}{T} \right)^2 \]

Harmonic constant \quad Initial state

Biasing potential \quad Collective variables: RMSD, distance, \( R_g \), angle, ...

Orientation quaternion

Final state \quad Total simulation time

Progressively Optimizing the Biasing Protocol/Collective Variable using non-Equilibrium Work as a Measure of the Path Quality

Example set taken from a subset of 20 ns biased simulations
Mechanistic Insight From Transition Pathways in ABC exporters from Non-Equilibrium Simulations

NBD Doorknob Mechanism

Describing a Complete Cycle (Adding Substrate) Requiring a Combination of Multiple Collective Variables

IF

apo

OF

apo

IF

bound

OF

bound

12 replicas x 40 ns (H1/H7)
50 replicas x 20 ns (10 Hs)

30 r x 20 ns
30 r x 20 ns

12 replicas x 40 ns (H1/H7)
24 replicas x 20 ns (H1/H7)
200 replicas (2D) x 5 ns
50 replicas x 20 ns

150 replicas
## Simulation protocols

<table>
<thead>
<tr>
<th>Transition</th>
<th>Technique</th>
<th>Collective Variables</th>
<th># of Replicas × Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 IF(_a)↔OF(_a)</td>
<td>BEUS</td>
<td>(Q(_1), Q(_7))</td>
<td>12 × 40 ns = 0.5 μs</td>
</tr>
<tr>
<td>2 IF(_a)↔IF(_b)</td>
<td>SMwST</td>
<td>{Q}</td>
<td>1000 × 1 ns = 1 μs</td>
</tr>
<tr>
<td>3</td>
<td>BEUS</td>
<td>{Q}</td>
<td>50 × 20 ns = 1 μs</td>
</tr>
<tr>
<td>4 OF(_a)↔OF(_b)</td>
<td>BEUS</td>
<td>Z(_{Pi})</td>
<td>30 × 40 ns = 1.2 μs</td>
</tr>
<tr>
<td>5</td>
<td>BEUS</td>
<td>({Q}, Z(_{Pi}))</td>
<td>30 × 40 ns = 1.2 μs</td>
</tr>
<tr>
<td>6</td>
<td>BEUS</td>
<td>({Q}, Z(_{Pi}))</td>
<td>30 × 40 ns = 1.2 μs</td>
</tr>
<tr>
<td>7</td>
<td>BEUS</td>
<td>(Q(_1), Q(_7))</td>
<td>24 × 20 ns = 0.5 μs</td>
</tr>
<tr>
<td>8</td>
<td>BEUS</td>
<td>Z(_{Pi})</td>
<td>15 × 30 ns = 0.5 μs</td>
</tr>
<tr>
<td>9 IF(_b)↔OF(_b)</td>
<td>2D BEUS</td>
<td>(ΔRMSD, Z(_{Pi}))</td>
<td>200 × 5 ns = 1 μs</td>
</tr>
<tr>
<td>10</td>
<td>SMwST</td>
<td>({Q}, Z(_{Pi}))</td>
<td>1000 × 1 ns = 1 μs</td>
</tr>
<tr>
<td>11</td>
<td>BEUS</td>
<td>({Q}, Z(_{Pi}))</td>
<td>50 × 20 ns = 1 μs</td>
</tr>
<tr>
<td>12</td>
<td>BEUS</td>
<td>({Q}, Z(_{Pi}))</td>
<td>150 × 50 ns = 7.5 μs</td>
</tr>
</tbody>
</table>

### Total Simulation Time
18.7 μs

---

Battling the Timescale - Case III
Multiscale Simulations

Combining multiple replica simulations and coarse-grained models to describe membrane fusion
Workflow for Multi-Scale Modeling

Parametrically Defined Sine Function

Christopher Mayne, Tajhorshid Lab
Workflow for Multi-Scale Modeling

Christopher Mayne, Tajkhorshid Lab
Workflow for Multi-Scale Modeling

Grid Design and Construction

- Remove plane points within the sphere
- Remove sphere points below plane

Membrane Budding/Fusion

Christopher Mayne, Tajkhorshid Lab
Battling the Timescale - Case IV
Reduced Representations

Highly Mobile Membrane Mimetic model

GpA insertion in 12 ns
Specific lipids regulate various functional aspects of membrane proteins

Integral membrane proteins

Peripheral membrane proteins
Lipid-Dependent Regulation and Activity of Peripheral Membrane Proteins

- Membrane binding is a key regulatory step in the function of diverse proteins:
  - Cytoplasmic enzymes (kinases, Ras, P450, synaptotagmin, ...)
  - Coagulation factors (GLA and C2 domains)
  - Membrane sculpting proteins (BAR domain)
  - Pathogenic systems – viral fusion peptides, synuclein,
  - Immune/apoptotic system (TIM proteins)

- **Lipid-specificity** is a common feature:
  - Mostly at the level of **head groups**: PS, PG, PIP2, PA, ...
  - Requiring all-atom representation of the head groups
  - Slow lateral diffusion of lipids within a bilayer environment makes simulation studies of membrane-associated phenomena even more challenging
Lipid Dependent Binding and Activation

Affinity is controlled by lipid content

Leaflet asymmetry is vital for coagulation

Courtesy of Jim Morrissey, UIUC
Lipid Dependent Binding and Activation

Mode and specificity of lipid-protein interactions constitute one of the main mechanistic aspects
How do we construct an initial model for peripheral proteins in membranes?

- Membrane-Bound State
- Signaling Lipids

- equilibrium MD
- too shallow?
- SMD / Biased
- too deep?
- wrong orientation
- MD
- wrong lipid contacts
- Very long MD relaxation
How do we construct an initial model for peripheral proteins in membranes?

Placement of protein in the membrane

Membrane-Bound State

Membrane-Bound State

Very long MD relaxation

MD relaxation

MD relaxation

too shallow
wrong orientation

wrong lipid contacts

too deep
wrong orientation

too deep
wrong orientation

too shallow
wrong orientation
Simulation of Binding with Full Membrane Representation

Partial list of technical problems:

• Biased simulations
• Unknown depth of insertion
• Single binding event
• Frequently failing
• **Minimal lipid reorganization**

**HMMM model**

**Highly Mobile Membrane Mimetic model**

**Advantages**

- Increased mobility of lipids
- Retain explicit headgroups allowing for atomic details

_Biophys. J., 102: 2130-2139 (2012) (Cover Article)_

[Images of authors: Zenmei Ohkubo, Mark Arcario, Taras Pogorelov, Josh Vermaas, Javier Baylon]
**HMMM-** Preserving the “Face” of the Lipid Bilayer

Perfect match in the membrane profile particularly in the head group region

Critical for proper description of lipid protein interactions
Enhanced Lipid Lateral Diffusion
Without Compromising Atomic Details of the Headgroups

\[ D = \lim_{t \to \infty} \frac{1}{4t} \langle [r(t_0 + t) - r(t_0)]^2 \rangle \]
Enhanced Lipid Lateral Diffusion
Without Compromising Atomic Details of the Headgroups

Conventional membrane (10 ns)  HMMM membrane (1 ns)
HMMM accelerated sampling of lipid-protein interactions
Constructing a superior initial model faster
PS-Dependent Spontaneous Insertion of FVII-GLA

Zenmei Ohkubo
Spontaneous, Unbiased Membrane Binding Accelerated Process Allows for better sampling ($n = 10$)

PS-Dependent Membrane Binding of Talin

Final model converted to full membrane
Stable in 100 ns simulations

Revealing the **Hydrophobic Anchor**

- Snorkeling of lysine acts as a switch which releases a conserved phenylalanine anchor (F261 & F283) into the membrane.
- Reformation of the hydrophobic pocket causes looser binding of talin; suggests a mechanism for unbinding of protein.

Membrane Induced Domain Rearrangement of Talin

Membrane Binding of Influenza Hemagglutinin Fusion Peptide

Membrane Binding of Influenza Hemagglutinin Fusion Peptide

7 different initial orientation each simulated 3 times

Membrane Binding of Influenza Hemagglutinin Fusion Peptide

Spontaneous binding observed in the majority of the simulations:
21 independent simulations starting from 7 different orientations

Remarkable convergence of membrane binding simulations

Remarkable convergence of membrane binding simulations

Robust Tilting of the Anchor Domain in Snare Protein Synaptobrevin
Robust Tilt Observed in Synaptobrevin

Membrane thickness mildly restrained:
one carbon/tail
$k = 0.05$ kcal/mole/A^2

Identifying a Hinge

Cytochrome P450 3A4 (CYP3A4)

- Enzymes essential for the metabolism of xenobiotics and other compounds, found in all domains of life.
- In the human body, CYPs are membrane-bound proteins.
- The interaction with membrane mediates binding of substrates.
- **CYP3A4**: most abundant CYP in the human body, metabolizes about 50%-60% of drugs that are metabolized in the body.

Insertion and Membrane-Induced Conformational Change of Cytochrome P450

Insertion and Membrane-Induced Conformational Change of Cytochrome P450

Within 10 degrees of experimental measurement of the tilt angle (S. Sligar)

Insertion and Membrane-Induced Conformational Change of Cytochrome P450

\[ t = 0 \text{ closed} \]

\[ t = 45 \text{ open} \]