Modeling and Molecular Dynamics of Membrane Proteins

Emad Tajkhorshid
Department of Biochemistry, Center for Biophysics and Computational Biology, and 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

\[ D_{\text{lip}} = 10^{-8} \text{ cm}^2\text{.s}^{-1} \]
\[ (50 \text{ Å in } \sim 5 \times 10^{-6} \text{ s}) \]
\[ D_{\text{wat}} = 2.5 \times 10^{-5} \text{ cm}^2\text{.s}^{-1} \]

Modeling mixed lipid bilayers!

Once in several hours!
\[ (\sim 50 \text{ Å in } \sim 10^4 \text{ s}) \]

\sim9 orders of magnitude slower ensuring bilayer asymmetry
Fluid Mosaic Model of Membrane

Lateral Diffusion Allowed

Flip-flop Forbidden

Ensuring the conservation of membrane asymmetric structure
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.
Protein/Lipid ratio

- Pure lipid: insulation (neuronal cells)
- Other membranes: on average 50%
- Energy transduction membranes (75%)
  Membranes of mitochondria and chloroplast
  Purple membrane of halobacteria

- Different functions = different protein composition
Protein / Lipid Composition

The purple membrane of halobacteria
Gramicidin A

Might be very sensitive to the lipid head group electrostatic and membrane potential
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 AQPO (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
A Recipe for Membrane Protein Simulations

• Align the protein along the z-axis (membrane normal): OPM, Orient.

• Decide on the lipid type and generate a large enough patch (MEMBRANE plugin in VMD, other sources). Size, area/lipid, shrinking.

• Overlay the protein with a hydrated lipid bilayer. Adjust the depth/height to maximize hydrophobic overlap and matching of aromatic side chains (Trp/Tyr) with the interfacial region

• Remove lipids/water that overlap with the protein. Better to keep as many lipids as you can, so try to remove clashes if they are not too many by playing with the lipids. Add more water and ions to the two sides of the membrane (SOLVATE / AUTOIONIZE in VMD)

• Constrain (not FIX) the protein (we are still modeling, let’s preserve the crystal structure; fix the lipid head groups and water/ion and minimize/simulate the lipid tails using a short simulation.)
A Recipe for Membrane Protein Simulations

• Continue to constrain the protein (heavy atoms), but release everything else; minimize/simulate using a short “constant-pressure” MD (NPT) to “pack” lipids and water against the protein and fill the gaps introduced after removal of protein-overlapping lipids.

• Watch water molecules; They normally stay out of the hydrophobic cleft. If necessary apply constraints to prevent them from penetrating into the open cleft between the lipids and the protein.

• Monitor the volume of your simulation box until the steep phase of the volume change is complete (.xst and .xsc files). Do not run the system for too long during this phase (over-shrinking; sometimes difficult to judge).

• Now release the protein, minimize the whole system, and start another short NPT simulation of the whole system.

• Switch to an NP$_n$AT or an NVT simulation, when the system reaches a stable volume. Using the new CHARMM force field, you can stay with NPT.
Lipid-Protein Packing During the Initial NpT Simulation
Adjustment of Membrane Thickness to the Protein Hydrophobic Surface
Glycerol-Saturated GlpF
Description of full conduction pathway
Complete description of the conduction pathway
Channel Hydrogen Bonding Sites

...{set frame 0}{frame < 100}{incr frame}{{
    animate goto $frame
    set donor [atomselect top
        "name O N and within 2 of
        (rename GCL and name HO)"
    ]
lappend [donor get index] list1
    set acceptor [atomselect top
        "rename GCL and name O and
        within 2 of (protein and name HN HO)"
    ]
lappend [acceptor get index] list2
}}

...
Channel Hydrogen Bonding Sites

<table>
<thead>
<tr>
<th>Residue</th>
<th>Number</th>
<th>Hydrogen Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLN</td>
<td>41</td>
<td>OE1 NE2</td>
</tr>
<tr>
<td>TRP</td>
<td>48</td>
<td>O NE1</td>
</tr>
<tr>
<td>GLY</td>
<td>64</td>
<td>O</td>
</tr>
<tr>
<td>ALA</td>
<td>65</td>
<td>O</td>
</tr>
<tr>
<td>HIS</td>
<td>66</td>
<td>O ND1</td>
</tr>
<tr>
<td>LEU</td>
<td>67</td>
<td>O</td>
</tr>
<tr>
<td>ASN</td>
<td>68</td>
<td>ND2</td>
</tr>
<tr>
<td>ASP</td>
<td>130</td>
<td>OD1</td>
</tr>
<tr>
<td>GLY</td>
<td>133</td>
<td>O</td>
</tr>
<tr>
<td>SER</td>
<td>136</td>
<td>O</td>
</tr>
<tr>
<td>TYR</td>
<td>138</td>
<td>O</td>
</tr>
<tr>
<td>PRO</td>
<td>139</td>
<td>O N</td>
</tr>
<tr>
<td>ASN</td>
<td>140</td>
<td>OD1 ND2</td>
</tr>
<tr>
<td>HIS</td>
<td>142</td>
<td>ND1</td>
</tr>
<tr>
<td>THR</td>
<td>167</td>
<td>OG1</td>
</tr>
<tr>
<td>GLY</td>
<td>195</td>
<td>O</td>
</tr>
<tr>
<td>PRO</td>
<td>196</td>
<td>O</td>
</tr>
</tbody>
</table>
Channel Hydrogen Bonding Sites

<table>
<thead>
<tr>
<th>Residue</th>
<th>Number</th>
<th>Hydrogen Bonding Sites</th>
<th>Residue</th>
<th>Number</th>
<th>Hydrogen Bonding Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLN</td>
<td>41</td>
<td>OE1 NE2</td>
<td>LEU</td>
<td>197</td>
<td>O</td>
</tr>
<tr>
<td>TRP</td>
<td>48</td>
<td>NE1</td>
<td>THR</td>
<td>198</td>
<td>O</td>
</tr>
<tr>
<td>GLY</td>
<td>64</td>
<td></td>
<td>GLY</td>
<td>199</td>
<td>O</td>
</tr>
<tr>
<td>ALA</td>
<td>65</td>
<td></td>
<td>PHE</td>
<td>200</td>
<td>O</td>
</tr>
<tr>
<td>HIS</td>
<td>66</td>
<td>ND1</td>
<td>ALA</td>
<td>201</td>
<td>O</td>
</tr>
<tr>
<td>LEU</td>
<td>67</td>
<td>O</td>
<td>ASN</td>
<td>203</td>
<td>ND2</td>
</tr>
<tr>
<td>ASN</td>
<td>68</td>
<td>ND2</td>
<td>LYS</td>
<td>33</td>
<td>HZ1 HZ3</td>
</tr>
<tr>
<td>ASP</td>
<td>130</td>
<td>OD1</td>
<td>GLN</td>
<td>41</td>
<td>HE21</td>
</tr>
<tr>
<td>GLY</td>
<td>133</td>
<td>O</td>
<td>TRP</td>
<td>48</td>
<td>HE1</td>
</tr>
<tr>
<td>SER</td>
<td>136</td>
<td></td>
<td>HIS</td>
<td>66</td>
<td>HD1</td>
</tr>
<tr>
<td>TYR</td>
<td>138</td>
<td>O</td>
<td>ASN</td>
<td>68</td>
<td>HD22</td>
</tr>
<tr>
<td>PRO</td>
<td>139</td>
<td>N</td>
<td>ASN</td>
<td>203</td>
<td>HD22</td>
</tr>
<tr>
<td>ASN</td>
<td>140</td>
<td>OD1 ND2</td>
<td>TYR</td>
<td>138</td>
<td>HN</td>
</tr>
<tr>
<td>HIS</td>
<td>142</td>
<td>ND1</td>
<td>ASN</td>
<td>140</td>
<td>HN HD21 HD22</td>
</tr>
<tr>
<td>THR</td>
<td>167</td>
<td>OG1</td>
<td>HIS</td>
<td>142</td>
<td>HD1</td>
</tr>
<tr>
<td>GLY</td>
<td>195</td>
<td>O</td>
<td>GLY</td>
<td>199</td>
<td>HN</td>
</tr>
<tr>
<td>PRO</td>
<td>196</td>
<td>O</td>
<td>ASN</td>
<td>203</td>
<td>HN HD21 HD22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ARG</td>
<td>206</td>
<td>HE HH21 HH22</td>
</tr>
</tbody>
</table>
The Substrate Pathway is formed by $C=O$ groups.
Non-helical motifs are stabilized by two glutamate residues.

The Substrate Pathway is formed by $C=O$ groups.
# Conservation of Glutamate Residue in Human Aquaporins

| AQP0_HUMAN | LNTLHPAVSVGQAVTIVFVFTLCLFVLCLIFATYDE-RRNGQLG |
| AQP1_HUMAN | RNDLADVNSGQGLGEIAGLQLVLCVLATDDR-RRDLGG |
| AQP2_HUMAN | VNALSNSTTQGGAVTIVFVFTLCLFVLCLIFASTDE-RRGENPG |
| AQP3_HUMAN | GIFATYPSGHLDINFGDPGFTASLIVCVLAIVDPYPNNFPRG |
| AQP4_HUMAN | VTVHGNLTHGGLIVELITFQFLVTIFASCDSD-KRTDVGT |
| AQP5_HUMAN | VNALNNNNQAMTVELITFOLALCIFASTDS-RRTSPVG |
| AQP6_HUMAN | INVVRNSVSTGQAVAVELILLLQVLCVFASTDS-RQTS-G |
| AQP7_HUMAN | GIFATYLPDHMTLWRGF1NEALNIGMLQI-CLFAITDQENNPAIQ |
| AQP8_HUMAN | -AASFVQEQVAGAVAEIILILLALAVCNAGIN--EKTGKP |
| AQP9_HUMAN | HIFATYPAPYLSLANAFDVYATMILLIVFAIFLSPNLGAPRP |
| GLPF_ECOLI | GTFHYPNPHINFYQAFAVEMVTAILMLILALTDDNGYPRGP |

**ruler** ...180.......190.......200.......210.......22
Glycerol - water competition for hydrogen bonding sites
Revealing the Functional Role of Reentrant Loops
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 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.

\[ W \geq \Delta G \]

Second law of thermodynamics
In principle, it is possible to obtain free energy surfaces from repeated non-equilibrium experiments.
Steered Molecular Dynamics

- **constant force** (250 pN)
  - Graph showing a downward trend in z/Å over time (ns).

- **constant velocity** (30 Å/ns)
  - Graph showing a downward trend in z/Å over time (ns).
Trajectory of glycerol pulled by constant force

SMD Simulation of Glycerol Passage
4 trajectories

\( v = 0.03, \ 0.015 \ \text{Å/ps} \)

\( k = 150 \ \text{pN/Å} \)

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

\[
W(t) = \int_{0}^{t} dt' \ v f(t')
\]
• Captures major features of the channel
• The largest barrier ≈ 7.3 kcal/mol; exp.: 9.6±1.5 kcal/mol

Features of the Potential of Mean Force

Asymmetric Profile in the Vestibules

Artificial induction of glycerol conduction through AqpZ

Three fold higher barriers

AqpZ 22.8 kcal/mol
GlpF  7.3 kcal/mol

Could it be simply the size?

It is probably just the size that matters!

Water permeation

18 water conducted
In 4 monomers in 4 ns
1.125 water/monomer/ns
Exp. = ~ 1-2 /ns

7-8 water molecules in each channel
Correlated Motion of Water in the Channel

The single file of water molecules is maintained.
Experimental value for AQP1: $0.4-0.8 \times 10^{-5}$

One dimensional diffusion:

$$2Dt = \left\langle (z_t - z_0)^2 \right\rangle$$

Experimental value for AQP1: $0.4-0.8 \times 10^{-5}$
Diffusion of Water in the channel

\[ 2Dt = \left\langle (z_t - z_0)^2 \right\rangle \]

Improvement of statistics
Water Bipolar Configuration in Aquaporins
Water Bipolar Configuration in Aquaporins

[Graph and diagrams illustrating water molecule configurations and amino acid interactions within the aquaporin structure.]
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
Proton transfer through water
Cl\(^{-}\) channel

K\(^{+}\) channel

Aquaporins
A Complex Electrostatic Interaction

“Surprising and clearly not a hydrophobic channel”

A Repulsive Electrostatic Force at the Center of the Channel

QM/MM MD of the behavior of an excessive proton