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.
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

\( D_{\text{lip}} = 10^{-8} \text{ cm}^2 \cdot \text{s}^{-1} \)

(50 Å in ~ 5 x 10^{-6} s)

\( D_{\text{wat}} = 2.5 \times 10^{-5} \text{ cm}^2 \cdot \text{s}^{-1} \)

Modeling mixed lipid bilayers!

Transverse diffusion (flip-flop)

Once in several hours!

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

\sim 9 \text{ 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.
by: J. Siewert-Jan Marrink and Alan E. Mark, University of Groningen, The Netherlands
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
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

77 Å
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\textsubscript{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

Constriction region

Selectivity filter
Channel Hydrogen Bonding Sites

...
## Channel Hydrogen Bonding Sites

<table>
<thead>
<tr>
<th>Residue</th>
<th>Number</th>
<th>Hydrogen Bonding Site(s)</th>
<th>Residue</th>
<th>Number</th>
<th>Hydrogen Bonding Site(s)</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>O NE1</td>
<td>THR</td>
<td>198</td>
<td>O</td>
</tr>
<tr>
<td>GLY</td>
<td>64</td>
<td>O</td>
<td>GLY</td>
<td>199</td>
<td>O</td>
</tr>
<tr>
<td>ALA</td>
<td>65</td>
<td>O</td>
<td>PHE</td>
<td>200</td>
<td>O</td>
</tr>
<tr>
<td>HIS</td>
<td>66</td>
<td>O 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>GLY</td>
<td>41</td>
<td>HE21</td>
</tr>
<tr>
<td>GLY</td>
<td>133</td>
<td>O</td>
<td>GLN</td>
<td>41</td>
<td>HE1</td>
</tr>
<tr>
<td>SER</td>
<td>136</td>
<td>O</td>
<td>TRP</td>
<td>48</td>
<td>HE1</td>
</tr>
<tr>
<td>TYR</td>
<td>138</td>
<td>O</td>
<td>HIS</td>
<td>66</td>
<td>HD1</td>
</tr>
<tr>
<td>PRO</td>
<td>139</td>
<td>O N</td>
<td>ASN</td>
<td>68</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>
Channel Hydrogen Bonding Sites

<table>
<thead>
<tr>
<th>Residue</th>
<th>Number</th>
<th>Atom1</th>
<th>Atom2</th>
<th>Residue</th>
<th>Number</th>
<th>Atom1</th>
<th>Atom2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLN</td>
<td>41</td>
<td>OE1</td>
<td>NE2</td>
<td>LEU</td>
<td>197</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>48</td>
<td>O</td>
<td>NE1</td>
<td>THR</td>
<td>198</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>64</td>
<td>O</td>
<td></td>
<td>GLY</td>
<td>199</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>65</td>
<td>O</td>
<td></td>
<td>PHE</td>
<td>200</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>66</td>
<td>O</td>
<td>ND1</td>
<td>ALA</td>
<td>201</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>67</td>
<td>O</td>
<td></td>
<td>ASN</td>
<td>203</td>
<td>ND2</td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>68</td>
<td>ND2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>130</td>
<td>OD1</td>
<td></td>
<td>LYS</td>
<td>33</td>
<td>HZ1</td>
<td>HZ3</td>
</tr>
<tr>
<td>GLY</td>
<td>133</td>
<td>O</td>
<td></td>
<td>GLN</td>
<td>41</td>
<td>HE21</td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>136</td>
<td>O</td>
<td></td>
<td>TRP</td>
<td>48</td>
<td>HE1</td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>138</td>
<td>O</td>
<td></td>
<td>HIS</td>
<td>66</td>
<td>HD1</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>139</td>
<td>O</td>
<td>N</td>
<td>ASN</td>
<td>68</td>
<td>HD22</td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>140</td>
<td>OD1</td>
<td>ND2</td>
<td>TYR</td>
<td>138</td>
<td>HN</td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>142</td>
<td>ND1</td>
<td></td>
<td>ASN</td>
<td>140</td>
<td>HN</td>
<td>HD21</td>
</tr>
<tr>
<td>THR</td>
<td>167</td>
<td>OG1</td>
<td></td>
<td>HIS</td>
<td>142</td>
<td>HD1</td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>195</td>
<td>O</td>
<td></td>
<td>GLY</td>
<td>199</td>
<td>HN</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>196</td>
<td>O</td>
<td></td>
<td>ASN</td>
<td>203</td>
<td>HN</td>
<td>HD21</td>
</tr>
</tbody>
</table>
The Substrate Pathway is formed by $C=O$ groups.
The Substrate Pathway is formed by \( \text{C}=\text{O} \) groups.

Non-helical motifs are stabilized by two glutamate residues.
**Conservation of Glutamate Residue in Human Aquaporins**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQP0_HUMAN</td>
<td>LNTLHPAVSVGCATIVFETLTLCVLCIFATYDE-RRNGQLG</td>
</tr>
<tr>
<td>AQP1_HUMAN</td>
<td>RNLDAGVNSQGGLGIEIQLQLVLCVLATTDY-RRRDLGG</td>
</tr>
<tr>
<td>AQP2_HUMAN</td>
<td>VNALSSTTACAVTVELIYLLQLVLGFASTEDY-PRGENPC</td>
</tr>
<tr>
<td>AQP3_HUMAN</td>
<td>GFAITYPSGHLMDINGFDQFQGATSLIVCVAIVDYPNNFVPRG</td>
</tr>
<tr>
<td>AQP4_HUMAN</td>
<td>VMVHGNLTAGHGLIVEIITPQVLFTIFASCDS-KRTDVTG</td>
</tr>
<tr>
<td>AQP5_HUMAN</td>
<td>VNALNNITGQAMYVELLTFQALCIFASTDS-RRTSPVG</td>
</tr>
<tr>
<td>AQP6_HUMAN</td>
<td>INVVRNVECTGAVVLITLQLYLCLVFASTDS-RQTS--G</td>
</tr>
<tr>
<td>AQP7_HUMAN</td>
<td>GIFATYLPDHMTLWRGFLNEAFLGLMLQ-CIFAITDQENMPALPG</td>
</tr>
<tr>
<td>AQP8_HUMAN</td>
<td>AAFTVQEQQVAGALVEEIIITLLLALAVCNSQENV--EKTKGP</td>
</tr>
<tr>
<td>AQP9_HUMAN</td>
<td>HIFATYPAPYLSLANAFADQVYATMILLIVFAIFDSNLGAPRG</td>
</tr>
</tbody>
</table>

**GLPF_ECOLI**

- CTFSTYPPLNPHICEQAFAVEMVIALMGLILALTDDGNGVPRGP

**ruler**

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

Potassium channel
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.

Second law of thermodynamics $W \geq \Delta G$
Jarzynski’s Equality

Transition between two equilibrium states

\[ \lambda = \lambda_i \rightarrow \lambda = \lambda(t) \rightarrow \lambda = \lambda_f \]

\[ \Delta G = G_f - G_i \]

In principle, it is possible to obtain free energy surfaces from repeated non-equilibrium experiments.

\[ \langle e^{-\beta W} \rangle = e^{-\beta \Delta G} \]

\[ \beta = \frac{1}{k_B T} \]
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
$v = 0.03, \ 0.015 \ \text{Å/ps}$
$k = 150 \ \text{pN/Å}$

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

$$W(t) = \int_0^t dt' \ vf(t')$$
• 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

5 nanosecond Simulation
Correlated Motion of Water in the Channel

The single file of water molecules is maintained.
Diffusion of Water in the channel

One dimensional diffusion: \( 2Dt = \left\langle (z_t - z_0)^2 \right\rangle \)

Experimental value for AQP1: 0.4-0.8 e-5
Diffusion of Water in the channel

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

- Time (ns)
- MSD (A^2)

\[ D = \text{slope}/2 = 0.046 \text{ Å}^2/\text{ps} = 0.46 \times 10^{-5} \text{ cm}^2/\text{s} \]

Improvement of statistics
Water Bipolar Configuration in Aquaporins
REMEMBER:

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
“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
Accelerated Simulation of Membranes and Membrane-Associated Phenomena with a Novel Atomistic Membrane Mimetic Model

Emad Tajkhorshid
Computational Structural Biology and Molecular Biophysics
Beckman Institute, University of Illinois at Urbana-Champaign
www.csbmb.beckman.illinois.edu

GpA insertion in 12 ns
Plasma membranes are asymmetric

Affinity is controlled by lipid content

Ca^{2+} promotes specific lipid conformations

Leaflet asymmetry is vital for coagulation

Courtesy of Jim Morrissey, UIUC
MD Simulation with Full Membrane Representation

Collectively more than 150 ns of biased simulations to capture a single binding event.

Z. Ohkubo and E. Tajkhorshid, Structure 2008
Molecular Determinants of Phospholipid Synergy in Blood Clotting*

Received for publication, April 15, 2011, and in revised form, May 9, 2011. Published, JBC Papers in Press, May 11, 2011, DOI 10.1074/jbc.M111.251769

Narjes Tavoosii, Rebecca L. Davis-Harrison†, Taras V. Pogorelov‡§, Y. Zenmei Ohkubo‡§, Mark J. Arcario*¶, Mary C. Clay¶, Chad M. Rienstra§¶, Emad Tajkhorshid‡§¶, and James H. Morrissey††
HMMM model
Highly Mobile Membrane Mimetic model

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

Optimizing the Tail Length

Optimizing the Tail Length

SERP

\[ n = 0 \]

DRPS

\[ n = 3 \]

DPPS

\[ n = 5 \]

DSPS

\[ n = 7 \]

DOPS

\[ n = 18 \]
Spontaneous and Rapid Formation of a Bilayer

60 x 60 x 120 Å
DVPSs at 3 x 3 x 6 grid points
(22 ns)
Spontaneous and Rapid Formation of a Bilayer

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

Critical for proper description of lipid protein interactions

**HMMM-** Preserving the “Face” of the Lipid Bilayer
HMMM – lipids are much more mobile than full-lipids

HMMM membrane
1 ns

Full membrane
10 ns
HMMM – lipids are more mobile than full-lipids

\[ D = \lim_{t \to \infty} \frac{1}{4t} \left\langle \left| r(t_0 + t) - r(t_0) \right|^2 \right\rangle \]

DVPS DOPS with Na\(^+\) counterions

\[ A_L = 294 / \text{Å}^2 \]
\[ A_L = 144 / \text{Å}^2 \]
\[ A_L = 144 / \text{Å}^2 \]
\[ A_L = 100 / \text{Å}^2 \]
\[ A_L = 68.6 / \text{Å}^2 \]
\[ A_L = 65.3 / \text{Å}^2 \]
\[ A_L = 65.3 / \text{Å}^2 \]

lower density
short tails
Spontaneous Insertion of FVII-GLA
Spontaneous Membrane Binding

\((n = 10)\)
Spontaneous Insertion of Transmembrane Helices

$t = 0$

50 x 50 x 75 Å
Glycophorin A monomers: 2
z-constraint on 2 carbonyl carbons

12 ns

Taras Pogorelov
Highly Mobile Membrane Mimetic Model (HMMM)

Facilitating dynamical studies of membrane-associated phenomena