Modeling membrane proteins

James C. (JC) Gumbart
Georgia Institute of Technology, Atlanta
Why do living cells need membrane proteins?

- Living cells 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 interactions are the driving force
- Self-assembly in water
- Tendency to close on themselves
- Self-sealing (a hole is unfavorable)
- Extensive: up to millimeters
Self-assembly visualized in simulation

Coarse-grained simulation of lipids randomly placed in water
Lipid Diffusion in a Membrane

- Lateral diffusion
  - $D_{\text{lip}} = 10^{-8} \text{ cm}^2/\text{s}$
  - (50 Å in ~ 5 µs)
- Tranverse diffusion (flip-flop)
  - $D_{\text{wat}} = 2.5 \times 10^{-5} \text{ cm}^2/\text{s}$
  - Once in several hours!
    - (~ 50 Å in ~ $10^4$ s)

~9 orders of magnitude slower ensuring bilayer asymmetry can be maintained.
fluid mosaic model

refined version (much more dense, varied)
Membrane protein basics

- one of the most **abundant** classes of proteins
- up to **30%** of the human genome encodes membrane proteins
- over **550** distinct membrane transporters discovered in *E. coli*

Many different ways to associate with membrane

**α-helical** (most membranes)

**β-barrel** (outer membrane)
Types of membrane proteins

- channels and transporters
- receptors
- enzymes
- cell adhesion
- peripheral (not technically membrane proteins)
membrane receptors
permit communication between outside and inside of the cell

three classes:
1) enzyme linked, typically single TM
2) ligand-gated ion channels common example: neurotransmitter receptors (right)
3) G-protein coupled, examples include rhodopsin, beta-2 adrenergic receptor (left)

2012 Nobel Prize in Chemistry (R.J. Lefkowitz, B.K. Kobilka)
Cell adhesion molecules

CAMs are on the cell surface, involved in binding to other cells

Example: integrin

Extracellular domain interacts with other CAMs or EC Matrix

Conformational change initiated by signal from inside or outside the cell

Communicate chemical, mechanical states

Transmembrane domain

Intracellular domain interacts with the cytoskeleton
enzymes

typically only **membrane anchored** by a single TM

typical examples include oxidoreductases, transferases and hydrolases

**Ex: cytochrome P450**
- catalyze oxidation of organic substances
- exist in all domains of life, 18,000 forms known
- in humans, primarily membrane-associated
- responsible for 75% of reactions in drug metabolism

channels

passive transport, solutes flow down (electro)chemical gradient

most common solutes are ions

open to both sides of membrane simultaneously

gramicidin, an unusual antibiotic ion channel

KcsA, a bacterial K+ channel

aquaporin, a water channel

membrane transporters

open to only one side of membrane at a time
substrate binds from one side and releases to other
primary active transporters
couple the hydrolysis of ATP to drive transport

Examples include ion pumps, ATP synthase, ABC (ATP-binding cassette) transporters

structure of the Na+/K+ pump
Ex: ABC transporters

transport cycle for importer (exporter slightly different)

Ex: homology model of Cystic Fibrosis Transmembrane Regulator (CFTR) evolved to be more channel-like (not strongly coupled) for Cl-$\Delta 508$ mutation found in 1/30 people, prevents expression in respiratory epithelial cells

secondary active transporters
transport energy comes from co-transport of an ion

In the plasma membrane of animal cells, Na\(^+\) is the usual co-transported ion in bacteria/yeast (and organelles!) often H\(^+\)

- Example: sodium-glucose linked transporter (SGLT) in the kidneys
- Example: lactose permease in bacteria (left)
alternating access model of transport

transporter cycles through a number of distinct states

three primary states:
1) outward open
2) occluded
3) inward open

no transporter has structures in ALL states
channel structures

Peripheral membrane proteins
only temporarily associate with the membrane

A few examples:

enzymes
phospholipase A2 - involved in lipid metabolism, also in many venoms (promotes cell lysis)

structural
GLA domain - involved in blood coagulation cascade


hydrophobic molecule transporters
glycolipid transfer protein
Binding of a GLA domain

Energetics and the potential of mean force

Potential of mean force (PMF) projects full free-energy space onto one (or more) selected reaction coordinates

\[
\int d\mathbf{x}_1 \cdots d\mathbf{x}_N \ e^{-U(\mathbf{x}_1, \ldots, \mathbf{x}_N)/kT} = \int dz \ e^{-W(z)/kT}
\]

also can be expressed in terms of probability:

\[ W(z) = -kT \ln\left(\frac{P(z)}{P_0}\right) \]

Knowledge of PMF permits determination of many properties, e.g., conductance, average times, binding sites, etc.

2D PMF for ion transport through gramicidin A

Energetics: Glycophorin A

GypA-E expressed at surface of red blood cells, acts as a receptor, prevents aggregation, etc.

NMR structure of TM domain only

PMF for helix-helix association in membrane as function of separation distance
dimer is favored by 10 kcal/mol over separate monomers, mediated by GxxG motif

**Energetics: AmtB**

AmtB - an ammonia ($\text{NH}_3$)/ammonium ($\text{NH}_4^+$) channel

homologous to RhxG ($x=A,B,C$) proteins in mammalian blood cells

channel is **hydrophobic** - $\text{NH}_4^+$ likely changes protonation states at entrance/exit

**PMF** for $\text{NH}_3$ moving through channel shows minima at crystallographically resolved binding sites

determined using **adaptive biasing forces (ABF)** in **NAMD**
Building a membrane-protein system
Building a membrane-protein system (steps)

Step 1: Get the protein PDB from the PDB databank

Step 2: Build a PSF, including repeated subunits if necessary

Step 3: Build the membrane, using VMD (POPE, POPC only) or CHARMM-GUI

Step 4: Orient the protein in the membrane and combine them, removing overlapping lipids - write a new PSF/PDB

Step 5: Add water above and below using VMD Solvate, removing any that accidentally get placed inside the membrane

Step 6: Add ions; prepare inputs for minimization and equilibration

These are the steps in the Membrane Protein Tutorial
Building a membrane-protein system (easier)

Go to the Orientations of Proteins in Membranes (OPM) database

Look up your protein to see the details of its multimeric state, orientation in the membrane, and the membrane that it’s found in

http://opm.phar.umich.edu/
Building a membrane-protein system (easier)

CHARMM-GUI can read the aligned, multimeric protein directly from OPM and build the membrane, water, and ions around it.

http://charmm-gui.org/

Membrane Builder

Membrane Builder helps the user generate a series of CHARMM inputs necessary to build a protein/membrane complex for molecular dynamics simulations. A brief description of each step is given below. Among various other building schemes, either the "insertion" or the "replacement" method can be chosen by the user in step 3. (user can choose one of them in step 3, see below).

- Insertion method
  A protein is inserted into a pre-equilibrated lipid bilayer with a hole whose size is comparable to the protein size (the libraries of lipid bilayers are available in archive)

- Replacement method
  A protein is first packed by lipid-like spheres whose positions are subsequently used to place randomly chosen lipid molecules from the library (the libraries of lipid molecules are available in archive)

Please note that

- If you are not familiar with Membrane Builder, please first watch these video demos and also read the relevant references below.
- NAMD inputs (v2.7b3 or after) are provided for equilibration and production (see STEP6). Input files can be found in "namd" directory when you download tar archive ("charmm-gui.tgz") after all the input file generation.
- GROMACS inputs (v5.0 or after) are provided for minimization, equilibration and production (see STEP6). Input files can be found in "gromacs" directory when you download tar archive ("charmm-gui.tgz") after all the input file generation.
- AMBER inputs (v16 or after) are provided for minimization, equilibration and production (see STEP6). Input files can be found in "amber" directory when you download tar archive ("charmm-gui.tgz") after all the input file generation. See amber/README.
- GENESIS inputs (v1.1.0 or after) are provided for minimization, equilibration and production (see STEP6). Input files can be found in "genesis" directory when you download tar archive ("charmm-gui.tgz") after all the input file generation.
Building a membrane-protein system (easier)

Ex: AmtB (PDB 1U7G) an NH$_3$/NH$_4^+$ channel

OPM shows that it is a trimer

CHARMM-GUI can take output from OPM directly
Building a membrane-protein system (easier)

Think carefully about what to include!

Three copies of the protein

BOG: β-octylglucoside (detergent for crystallization)

NH₃/NH₄⁺ (substrates of the channel)

Crystallographic water

Model/Chain Selection Option:
Click on the chains you want to select.

Select Model # 1 □ □ Read all models?

<table>
<thead>
<tr>
<th>Type</th>
<th>SEGID</th>
<th>PDB ID</th>
<th>Residue ID</th>
<th>Engineered Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>PROA</td>
<td>A</td>
<td>3 385</td>
<td>None</td>
</tr>
<tr>
<td>Protein</td>
<td>PROB</td>
<td>B</td>
<td>3 385</td>
<td>None</td>
</tr>
<tr>
<td>Protein</td>
<td>PROC</td>
<td>C</td>
<td>3 385</td>
<td>None</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETA</td>
<td>D</td>
<td>3 385</td>
<td>BOG</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETB</td>
<td>D</td>
<td>3 385</td>
<td>NH₄⁺</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETC</td>
<td>D</td>
<td>3 385</td>
<td>NH₃</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETD</td>
<td>E</td>
<td>3 385</td>
<td>BOG</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETE</td>
<td>E</td>
<td>3 385</td>
<td>NH₄⁺</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETF</td>
<td>E</td>
<td>3 385</td>
<td>NH₃</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETG</td>
<td>F</td>
<td>3 385</td>
<td>BOG</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETH</td>
<td>F</td>
<td>3 385</td>
<td>NH₄⁺</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETI</td>
<td>F</td>
<td>3 385</td>
<td>NH₃</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETJ</td>
<td>F</td>
<td>3 385</td>
<td>DUM</td>
</tr>
<tr>
<td>Water</td>
<td>WATA</td>
<td>D</td>
<td>3 385</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>WATB</td>
<td>E</td>
<td>3 385</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>WATC</td>
<td>F</td>
<td>3 385</td>
<td></td>
</tr>
</tbody>
</table>
Building a membrane-protein system (easier)

There are a number of other choices to make along the way.

For example, how to patch the termini of the proteins.

NTER and CTER usually appropriate.
Building a membrane-protein system (easier)

Which lipids to use for the membrane?

1u7g » Ammonia Channel

- **Type**: 1. Transmembrane (3 classes)
- **Class**: 1.1. Alpha-helical polytopic (119 superfamilies)
- **Superfamily**: 1.1.017. Ammonia and urea transporters (2 families) 1.A.11 (TCDB)
- **Family**: 1.1.17.01. Ammonia transporter Amt (9 proteins) 1.A.11 (TCDB)
- **Species**: *Escherichia coli* (273 proteins)
- **Localization**: Bacterial Gram-negative inner membrane (548 proteins)

| Hydrophobic Thickness | 29.8 ± 1.3 Å |

Ideally, want to select lipids to match the native membrane composition!

Search textbooks, papers, etc. for estimates of the lipids and ratios for the membrane of interest
Building a membrane-protein system (easier)

Gram-negative inner membrane

mitochondrial membrane

Gram-negative outer membrane

mammalian plasma membrane
Building a membrane-protein system (easier)

Which lipids to use for the membrane?

For simplicity, used single-component POPE here
Building a membrane-protein system (easier)

The system looks reasonable overall, but there are a few potential problems. After a few more choices, a complete system is output (step 5).

The system looks reasonable overall, but there are a few potential problems.
Size of the membrane?

Initial guess of membrane size (100 Å x 100 Å) is too small

want to have more than 2-3 layers of lipids between periodic protein images

only 20 Å between images - will shrink after equilibration! (need 30 Å at least after eq.)
Water inside the membrane

These are not the fault of CHARMM-GUI! Instead they are co-crystallized waters

Model/Chain Selection Option:
Click on the chains you want to select.

Select Model # 1 or Read all models?

<table>
<thead>
<tr>
<th>Type</th>
<th>SEGID</th>
<th>PDB ID</th>
<th>Residue ID</th>
<th>Engineered Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>PROA</td>
<td>A</td>
<td>3 385</td>
<td>None</td>
</tr>
<tr>
<td>Protein</td>
<td>PROB</td>
<td>B</td>
<td>3 385</td>
<td>None</td>
</tr>
<tr>
<td>Protein</td>
<td>PROC</td>
<td>C</td>
<td>3 385</td>
<td>None</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETA</td>
<td>D</td>
<td></td>
<td>BOG</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETB</td>
<td>D</td>
<td></td>
<td>NH4</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETC</td>
<td>D</td>
<td></td>
<td>NH3</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETD</td>
<td>E</td>
<td></td>
<td>BOG</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETE</td>
<td>E</td>
<td></td>
<td>NH4</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETF</td>
<td>E</td>
<td></td>
<td>NH3</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETH</td>
<td>F</td>
<td></td>
<td>BOG</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETI</td>
<td>F</td>
<td></td>
<td>NH4</td>
</tr>
<tr>
<td>Hetero</td>
<td>HETJ</td>
<td>F</td>
<td></td>
<td>NH3</td>
</tr>
<tr>
<td>Water</td>
<td>WATA</td>
<td>D</td>
<td></td>
<td>DUM</td>
</tr>
<tr>
<td>Water</td>
<td>WATB</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>WATC</td>
<td>F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Option 1: rebuild but leave these boxes un-checked

Option 2: use a script to delete intra-membrane waters (see tutorial)
Building a membrane-protein system (easier)

If you look closely at the waters, they look very strange! Why is there a third bond???

When adding water with VMD Solvate, that extra bond is commented out in the topology file.

NAMD doesn’t care about it (it is harmless, just ugly!)

Warning: Ignored 19521 bonds with zero force constants.
Equilibration stages

System has to be relaxed carefully to avoid distortions

First, relax lipid tails for (water/prot/lipid heads restrained) for ~0.5 ns

Second, relax lipids and water (protein restrained) for 3-5 ns to ensure a good packing of lipids around the protein

Finally, can run with everything released in **NpT** ensemble

**NOTE:** CHARMM27 lipids do not maintain correct area/lipid but CHARMM36 lipids do! *Always use the latest force field!*