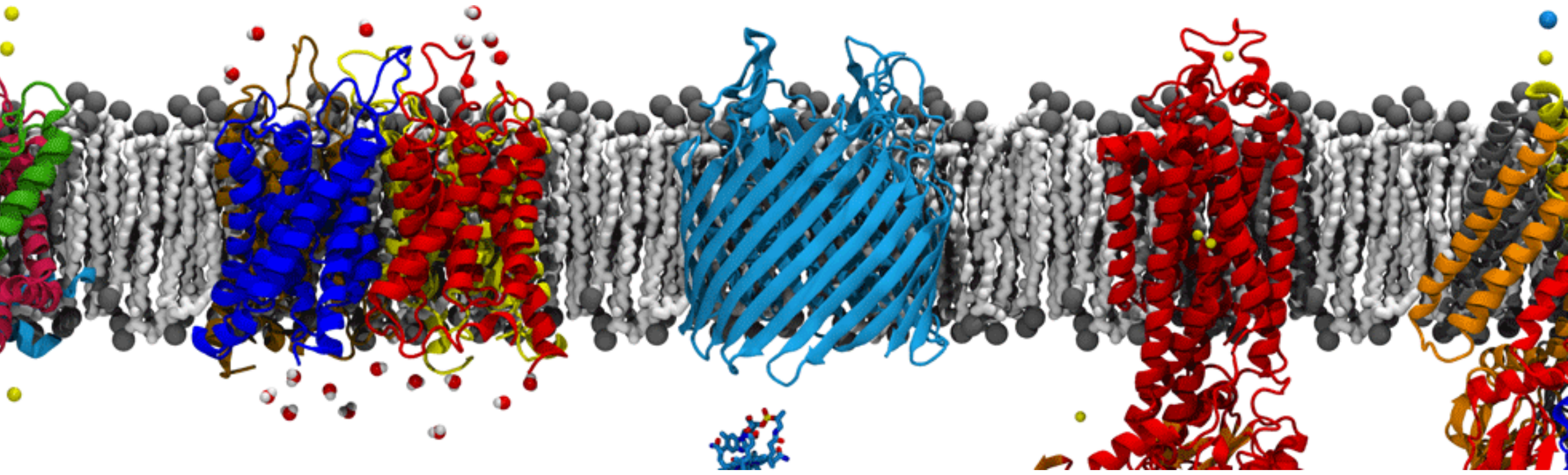
Modeling membrane proteins



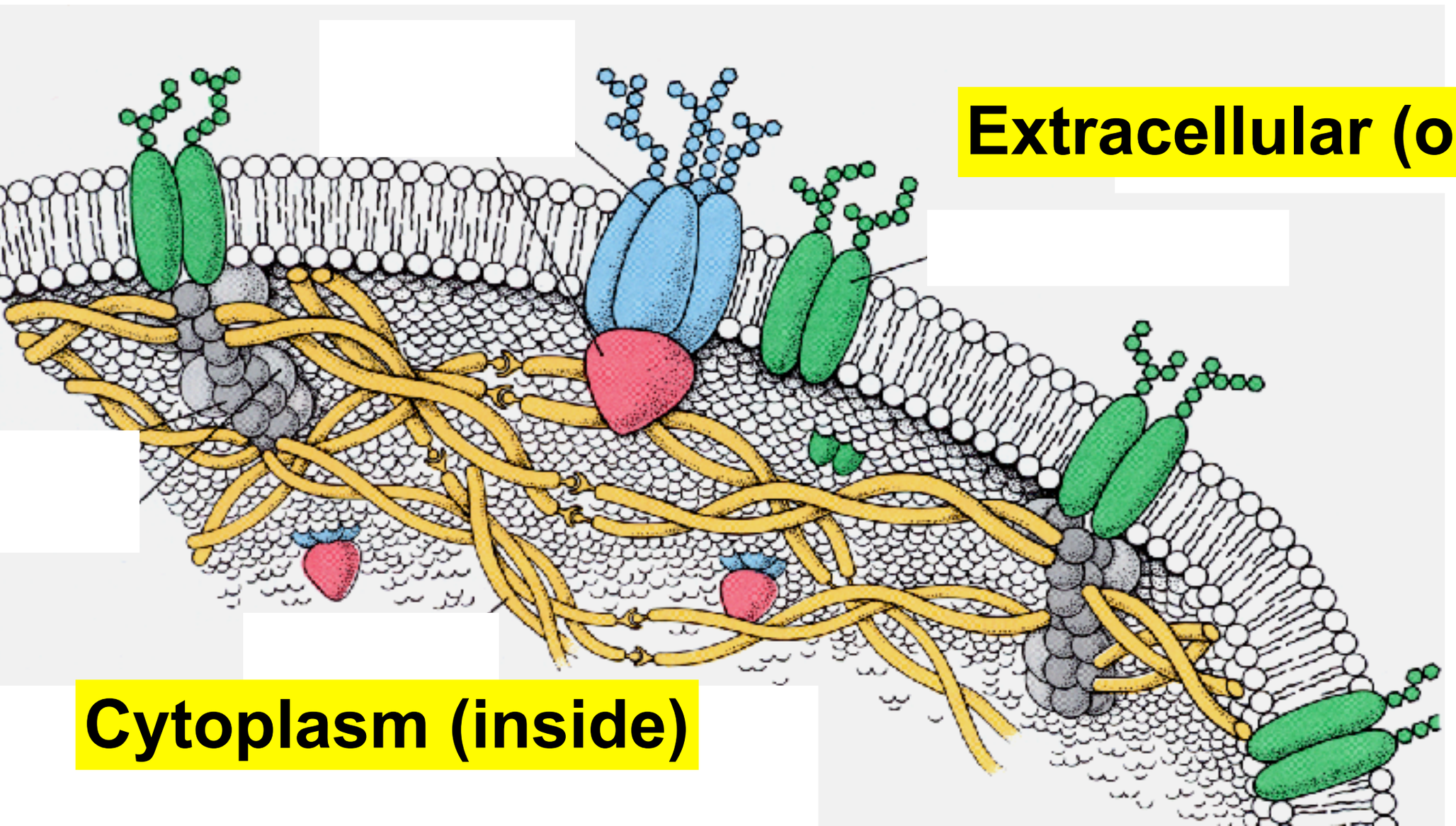
James C. (JC) Gumbart
Georgia Institute of Technology, Atlanta

Computational Biophysics Workshop | DICP | July 12 2018

Why do living cells need membrane proteins?

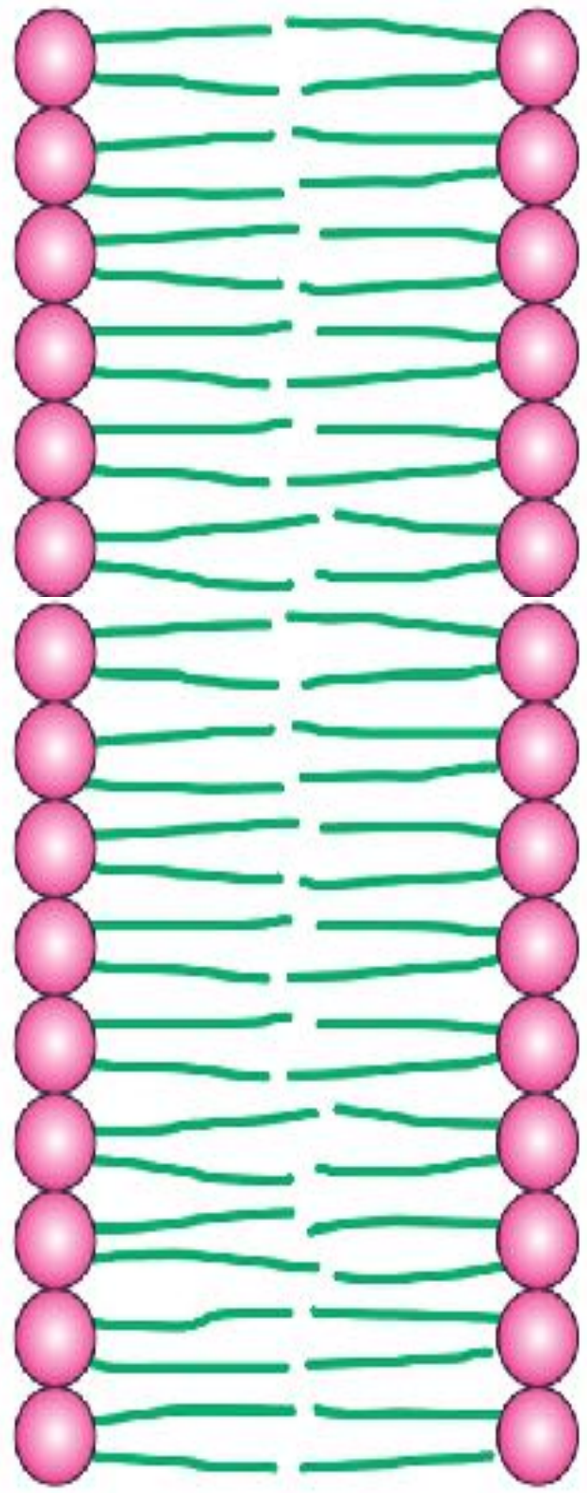
- Living cells need to exchange materials and information with the outside world

… however, in a highly selective manner.

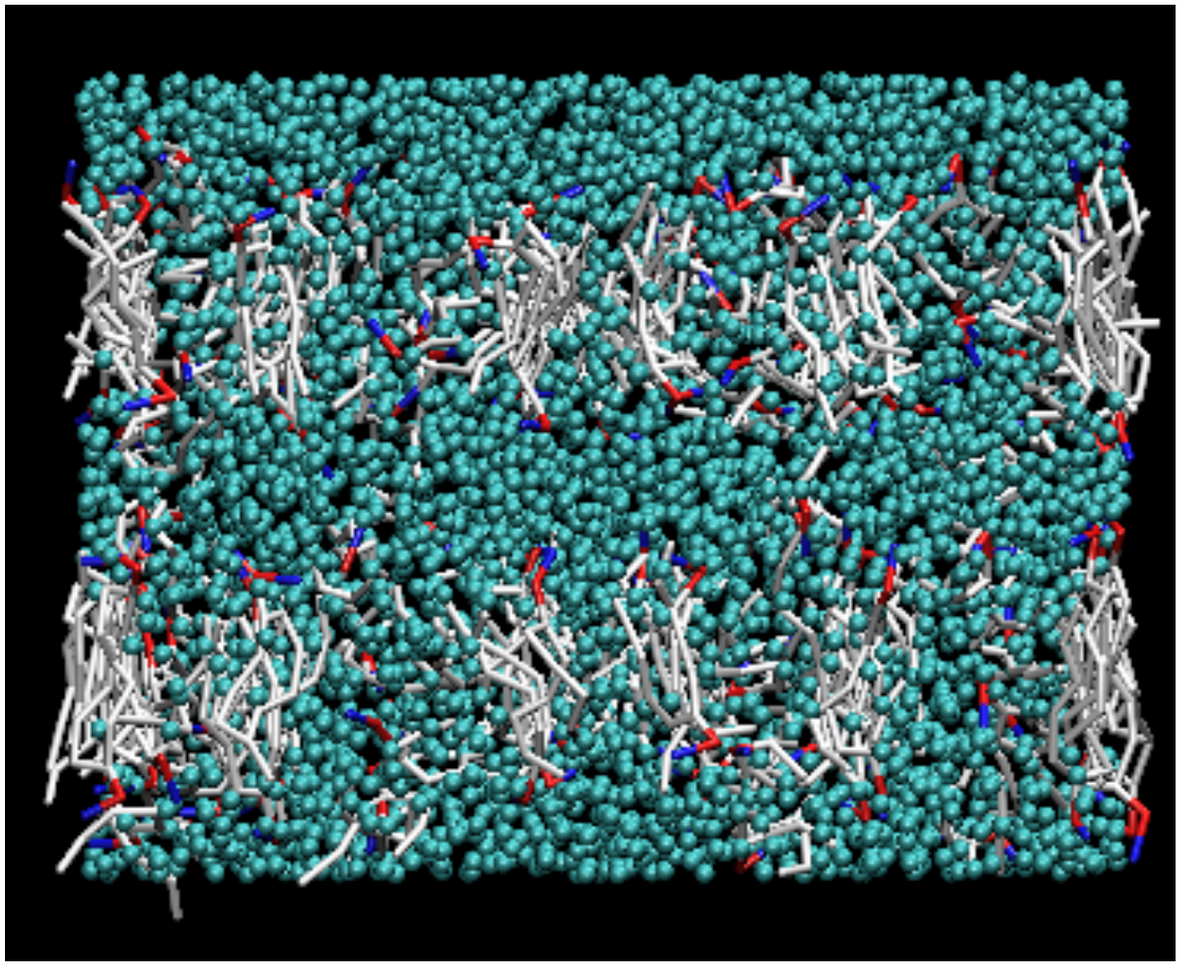


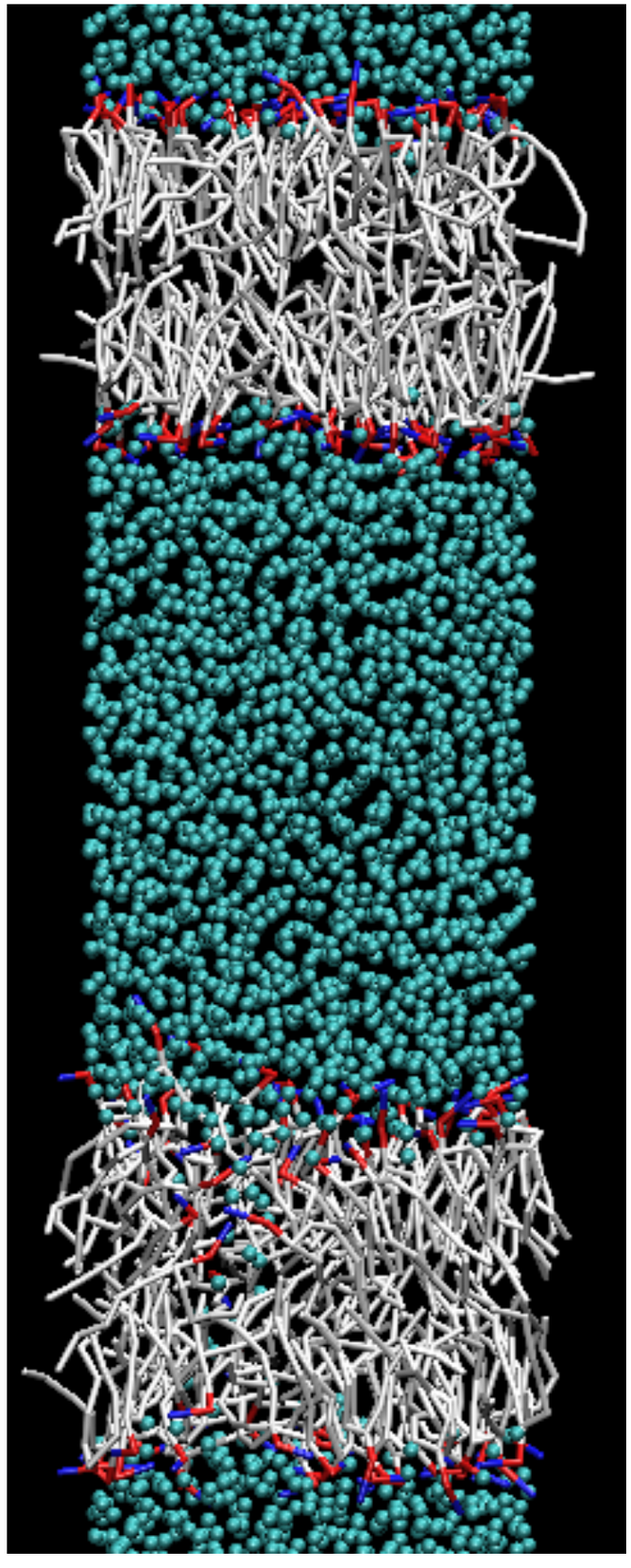
Phospholipid bilayers are excellent materials for cell membranes

- 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



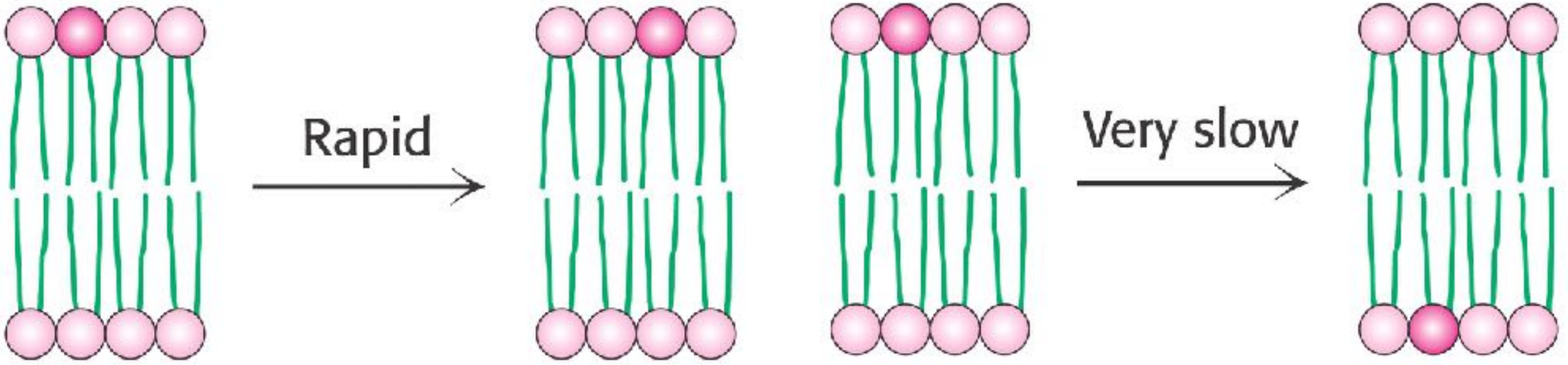
Self-assembly visualized in simulation





Coarse-grained simulation of
lipids randomly placed in water

Lipid Diffusion in a Membrane



$D_{lip} = 10^{-8}$ cm^2/s

(50 Å in ~ 5 μs)

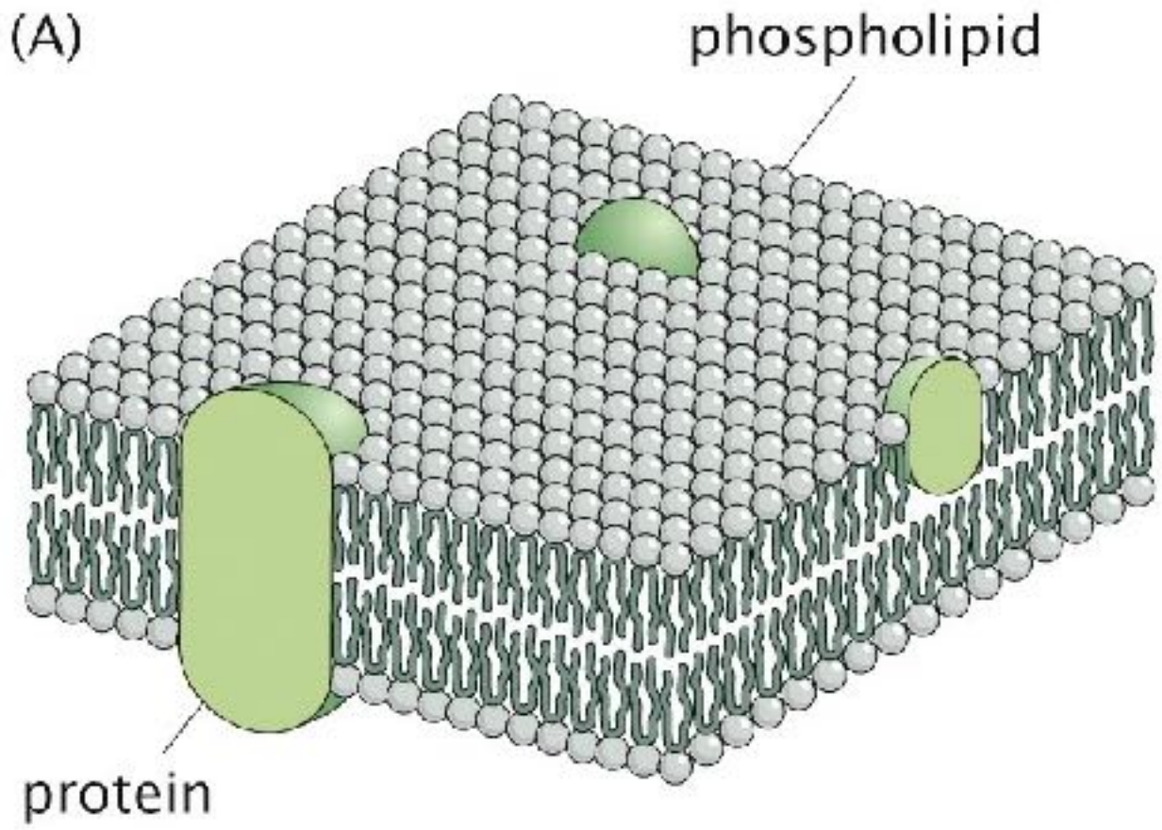
$D_{wat} = 2.5 \times 10^{-5}$ cm^2/s

Once in several hours!

(~ 50 Å in ~ 10^4 s)

~9 orders of magnitude slower
ensuring bilayer asymmetry
can be maintained

Membrane composition



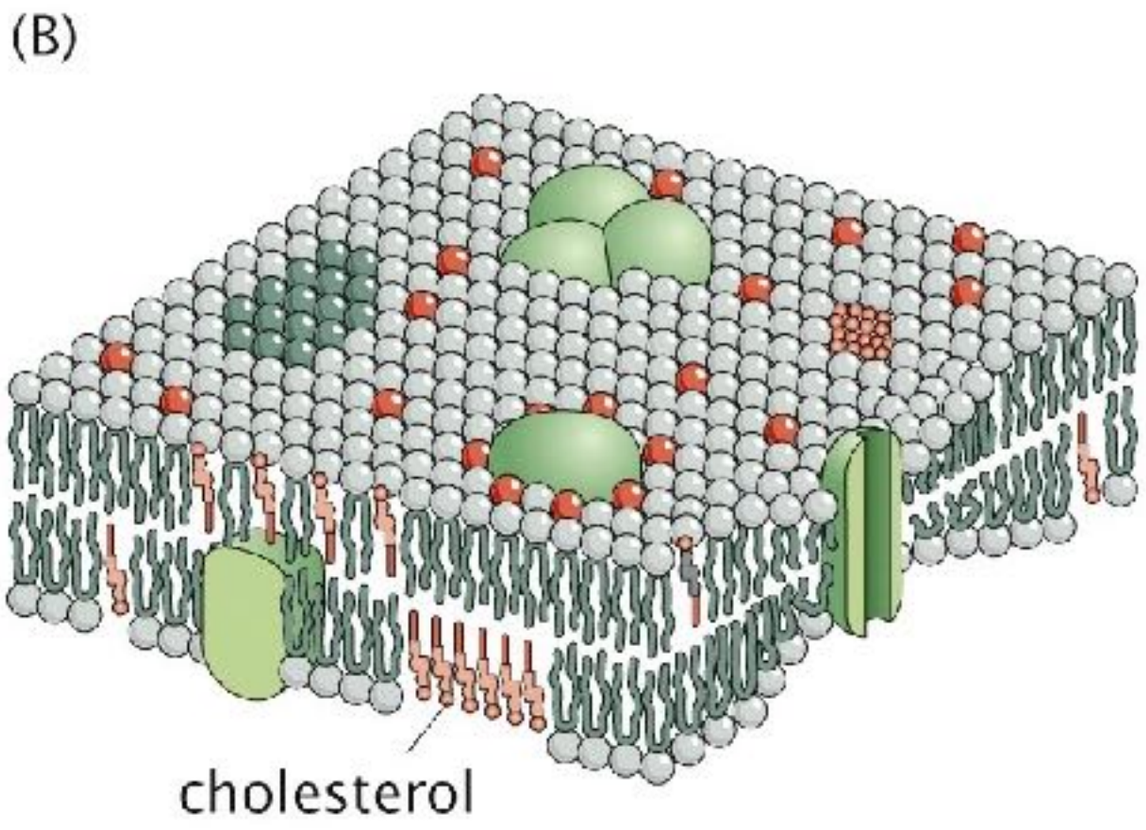


Figure 11.4ab Physical Biology of the Cell, 2ed. (© Garland Science 2013)

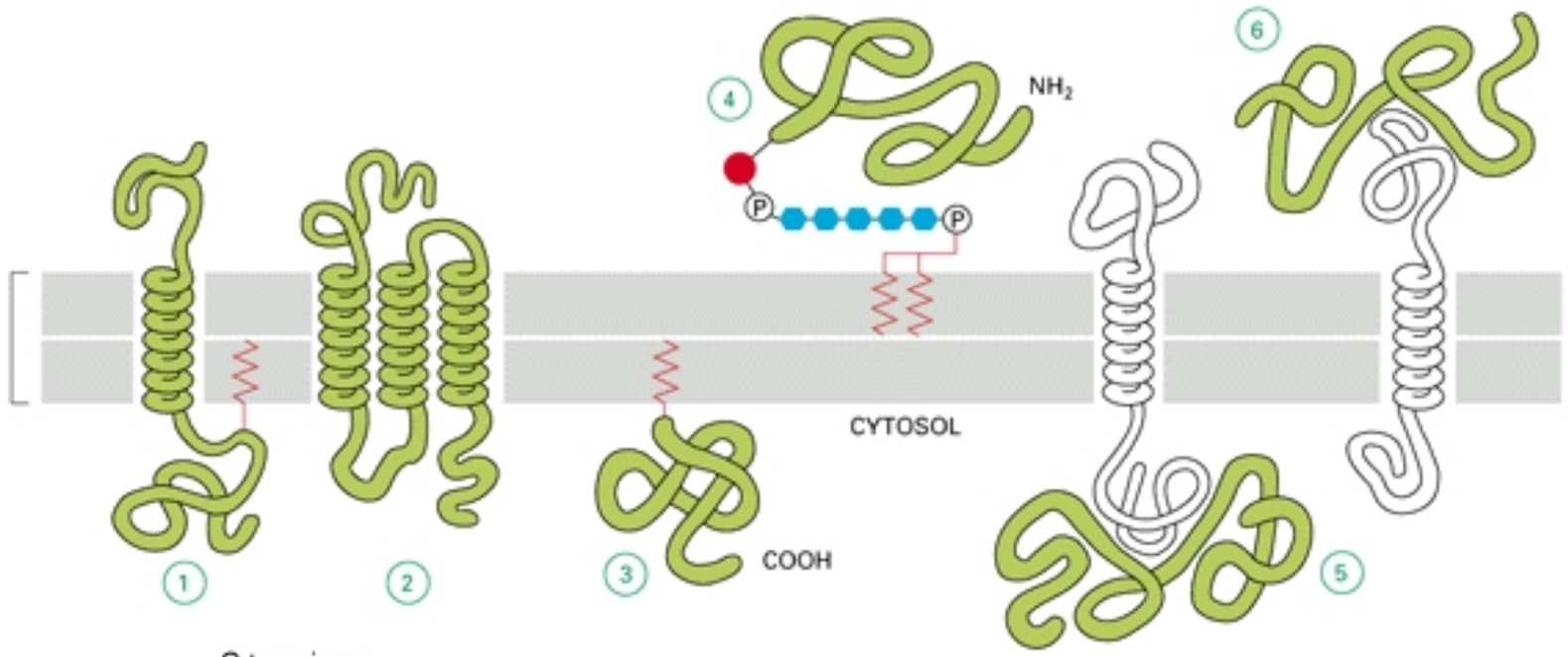
fluid mosaic model

Singer SJ, Nicolson GL (Feb 1972). *Science* **175**: 720–31.

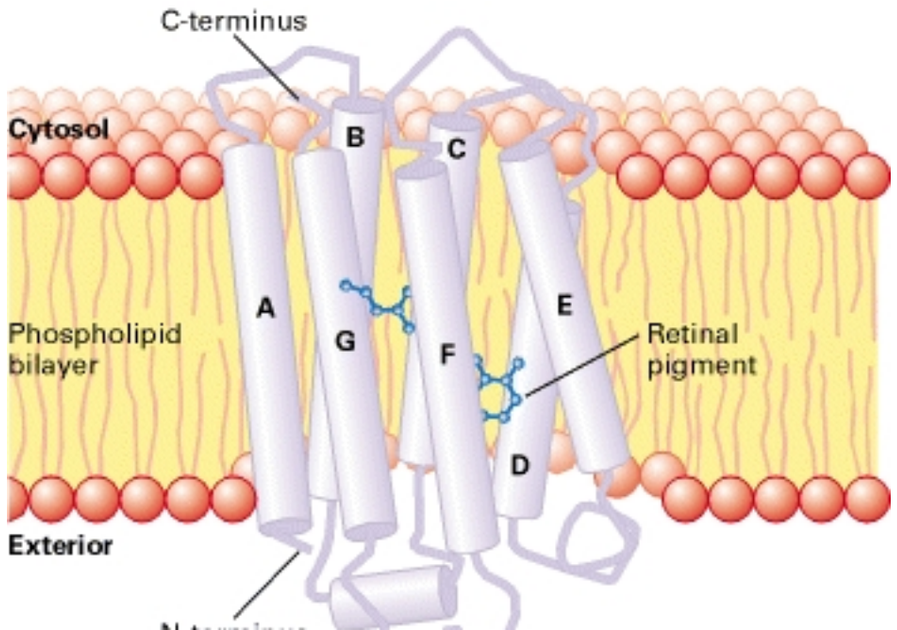
refined version (much more dense, varied)

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*

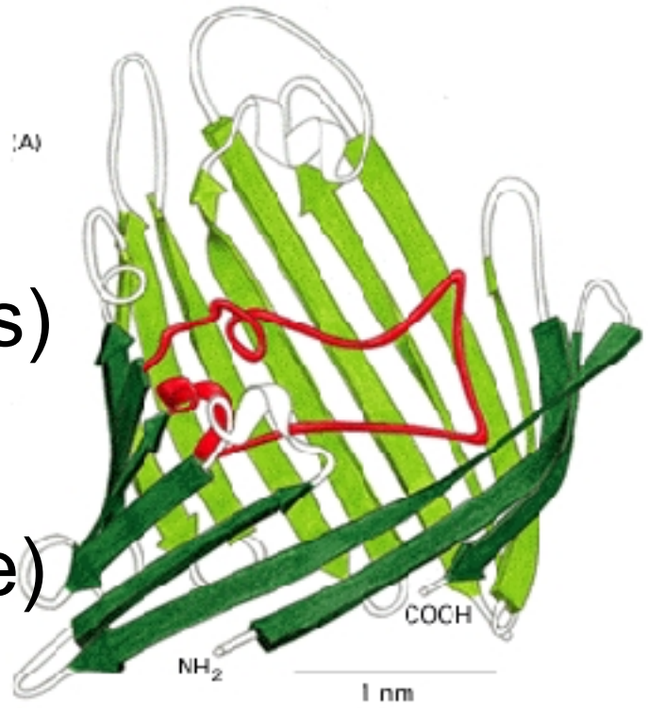


many different ways to associate with membrane



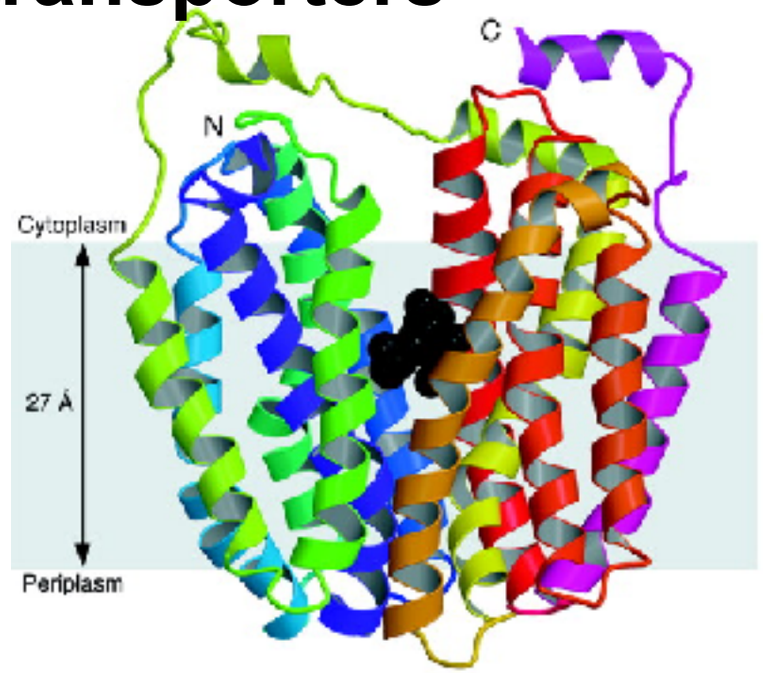
α-helical (most membranes)

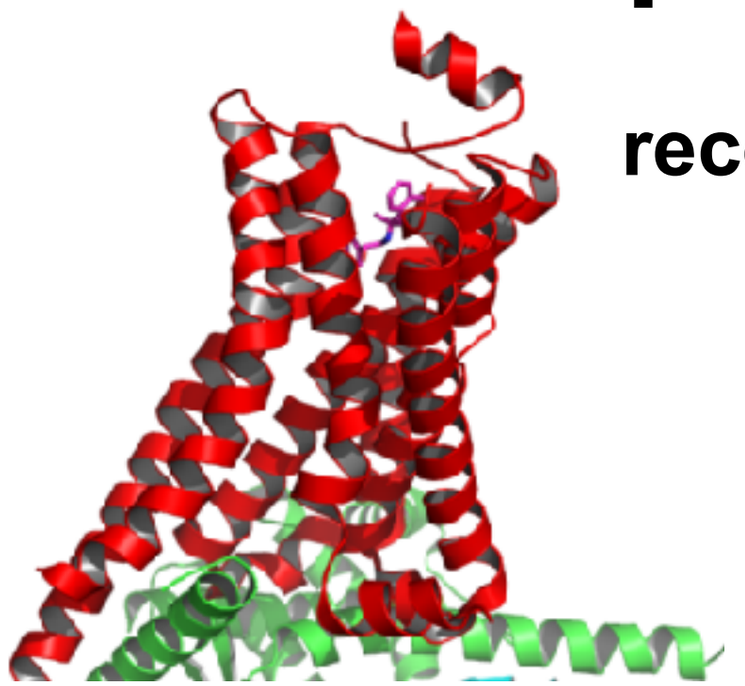
β-barrel (outer membrane)



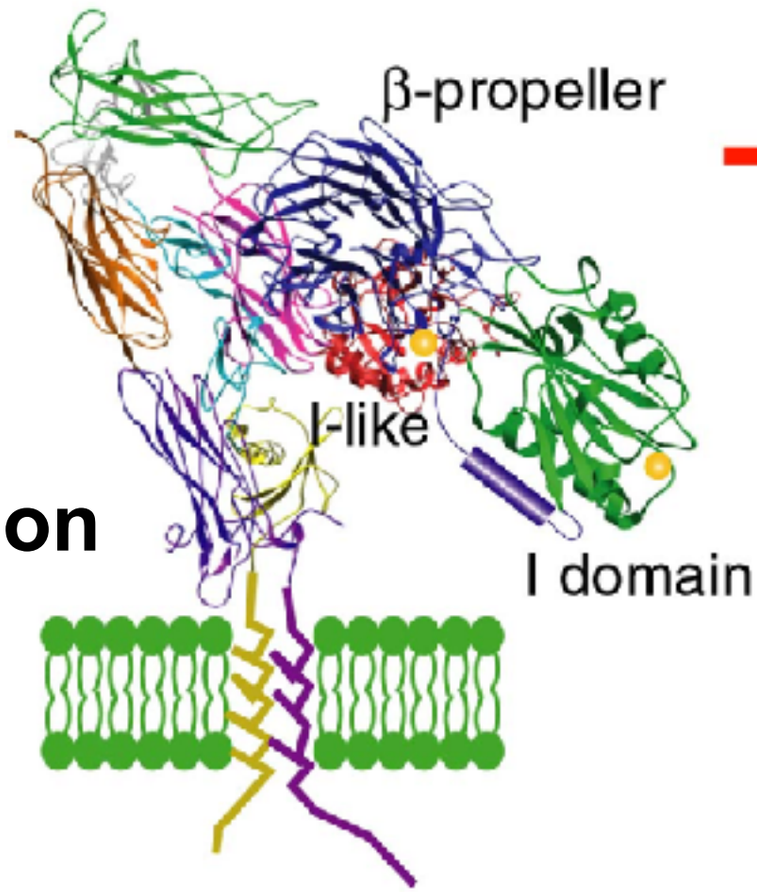
Types of membrane proteins

channels and transporters



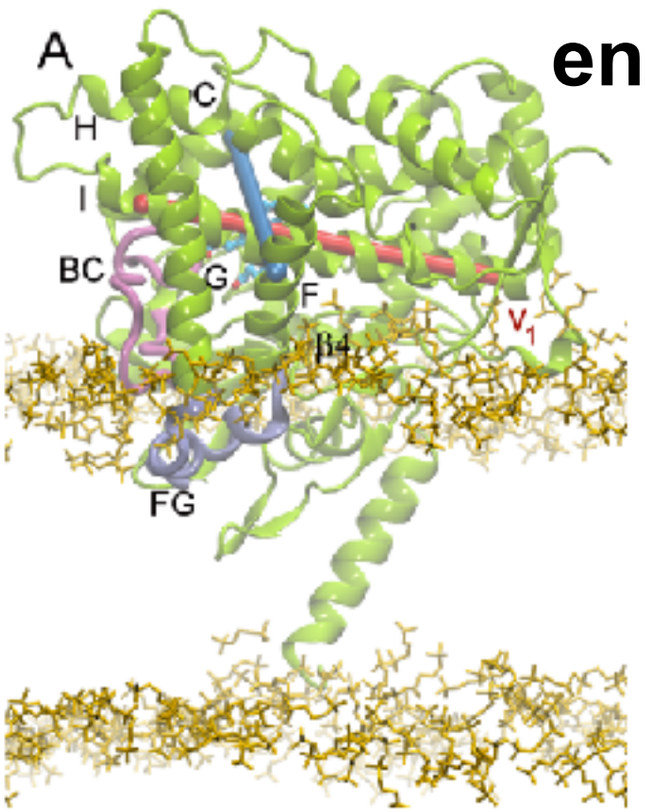


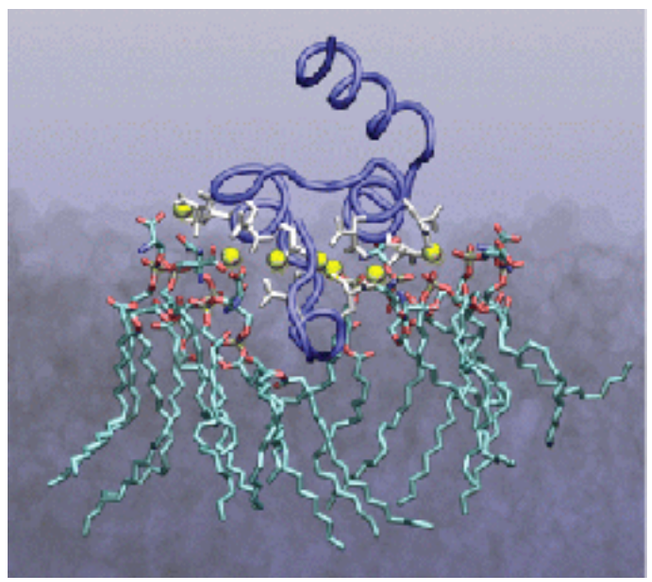
receptors



cell adhesion

enzymes





peripheral (not *technically* membrane proteins)

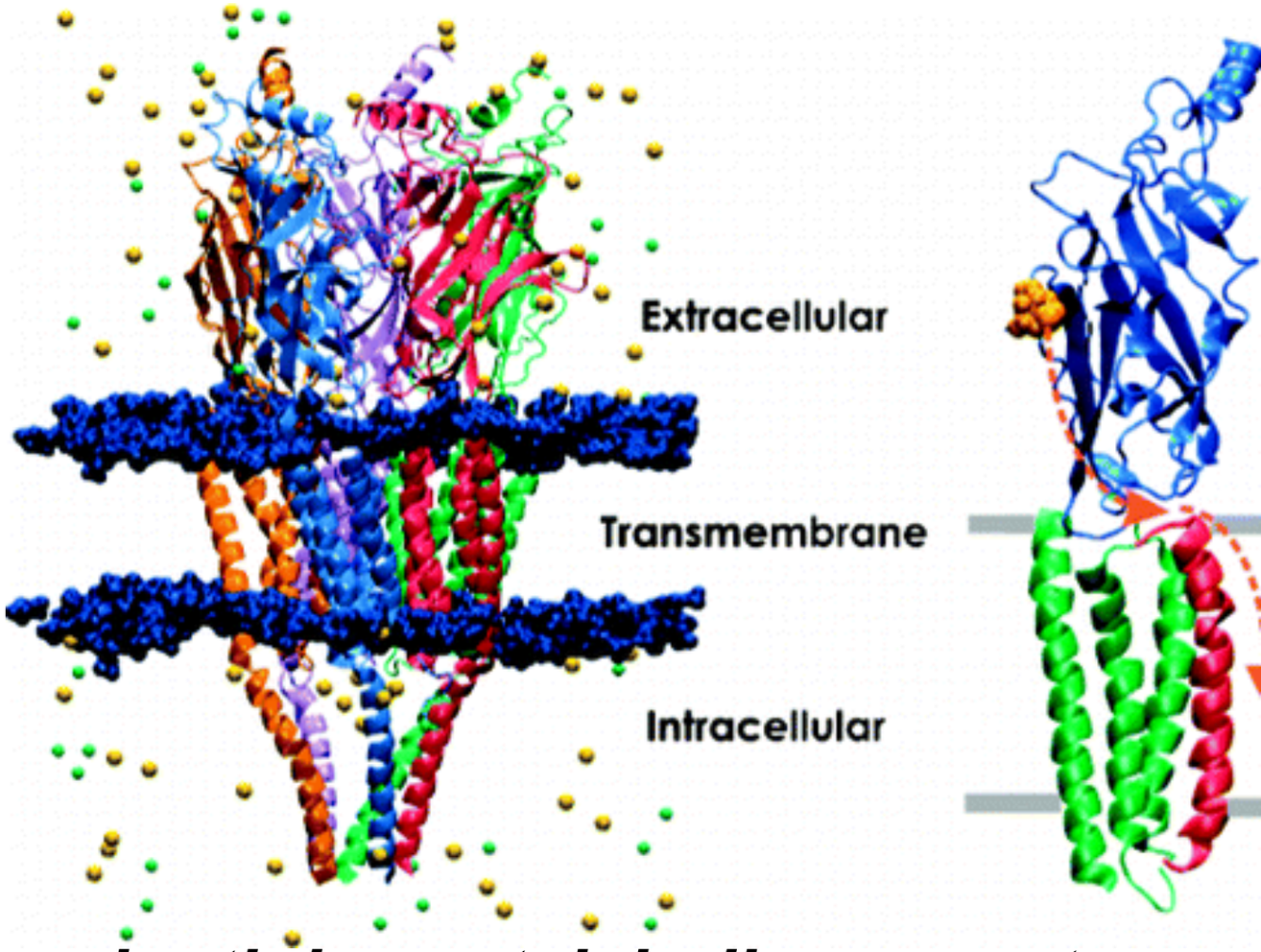
membrane receptors

permit communication between outside and inside of the cell

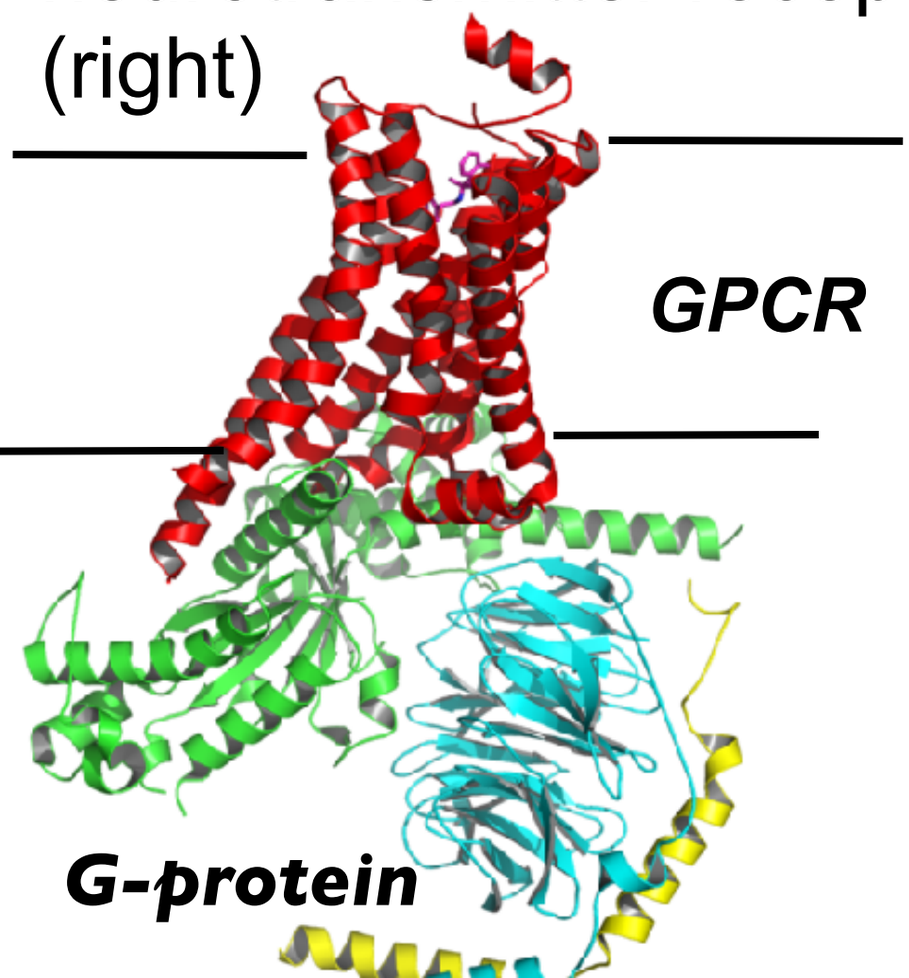
three classes:

1) enzyme linked, typically single TM

2) ligand-gated ion channels common example: neurotransmitter receptors (right)



nicotinic acetylcholine receptor



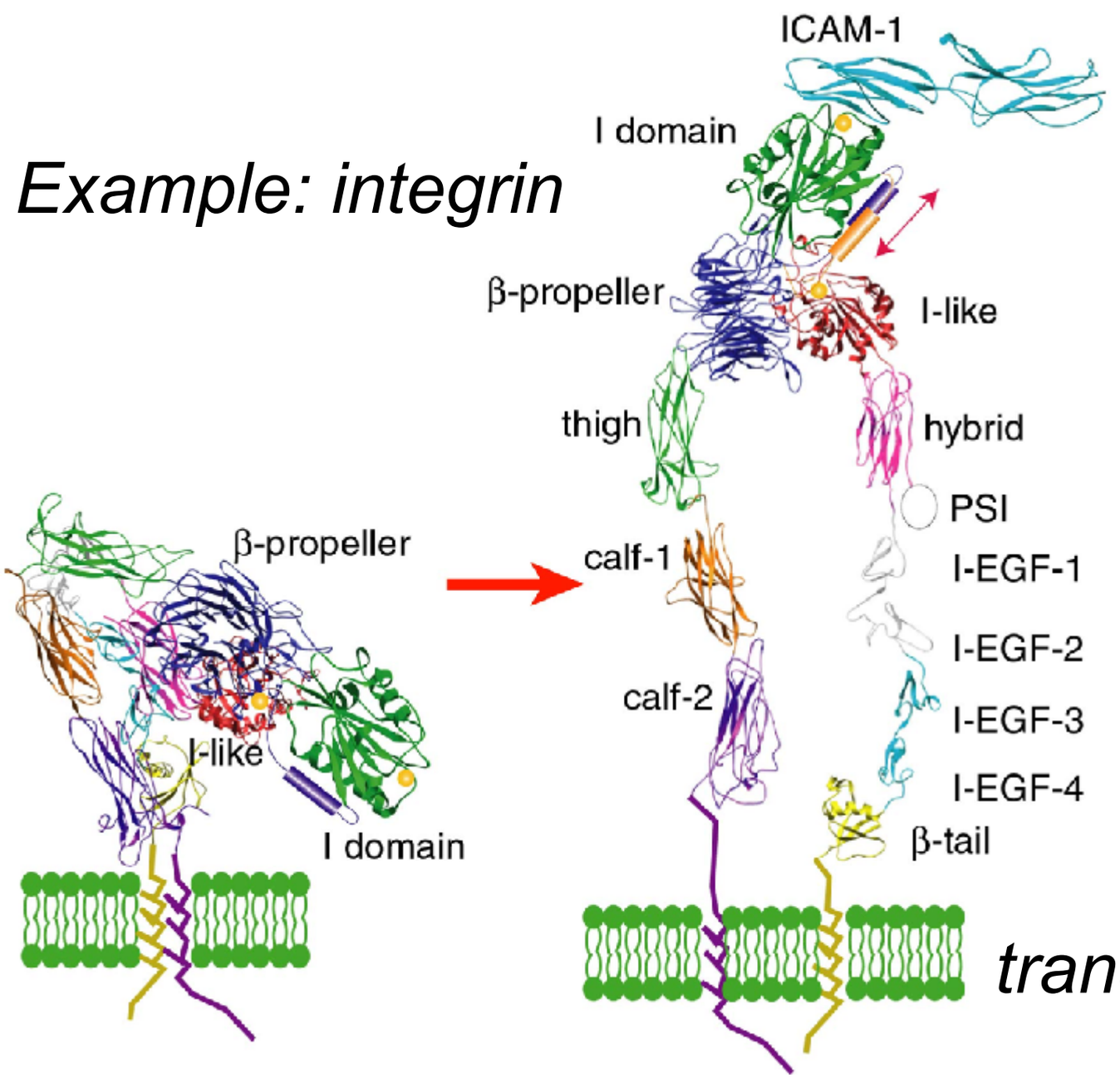
3) G-protein coupled, examples include rhodopsin, beta-2 adrenergic receptor (left)

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

examples include oxidoreductases, transferases and hydrolases

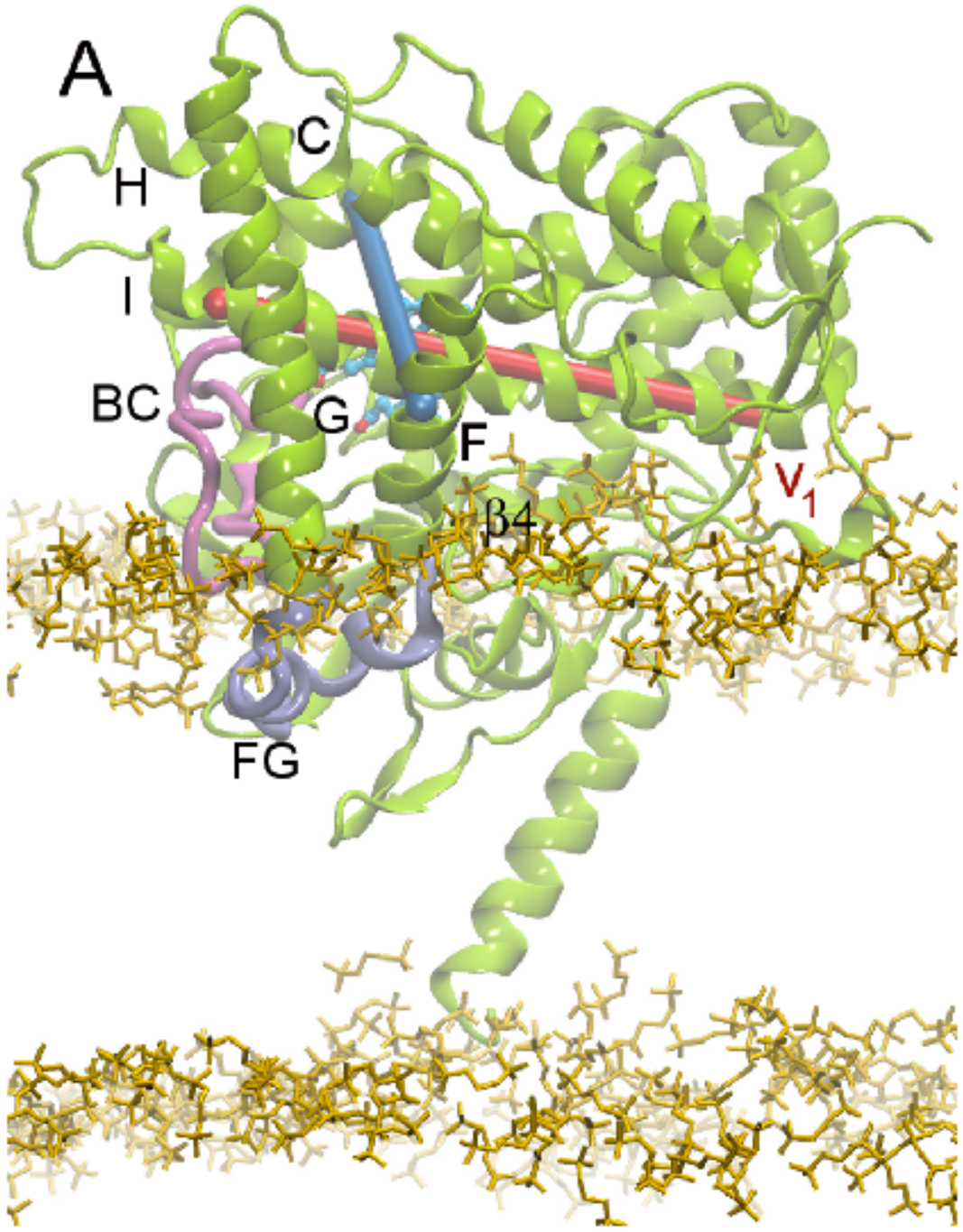
Ex: cytochrome P450

-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



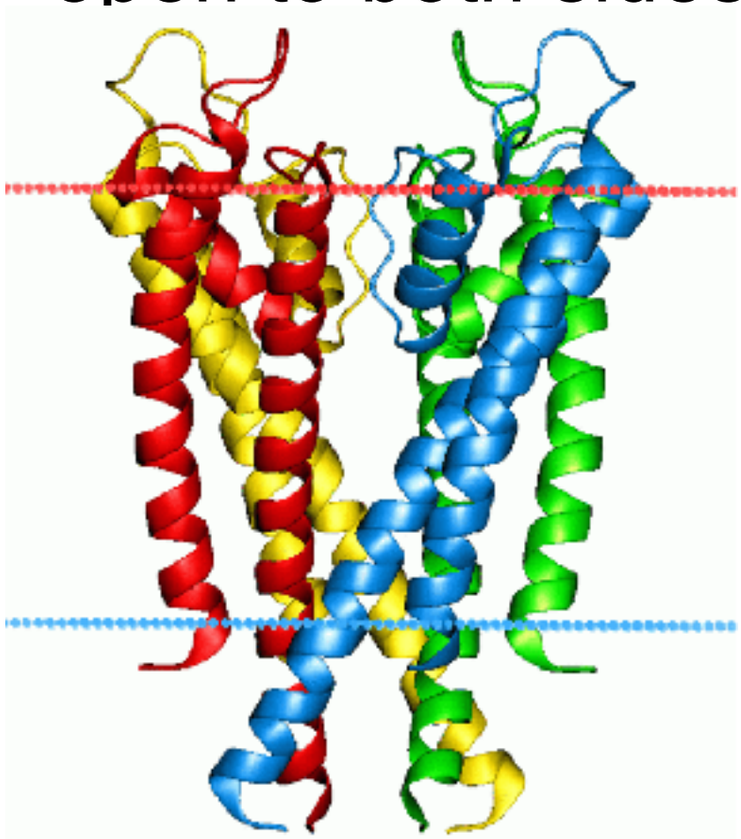
Cojocaru V, Balali-Mood K, Sansom MSP, Wade RC (2011) Structure and Dynamics of the Membrane-Bound Cytochrome P450 2C9. PLoS Comput Biol 7(8): e1002152.

channels

passive transport, solutes flow down (electro)chemical gradient

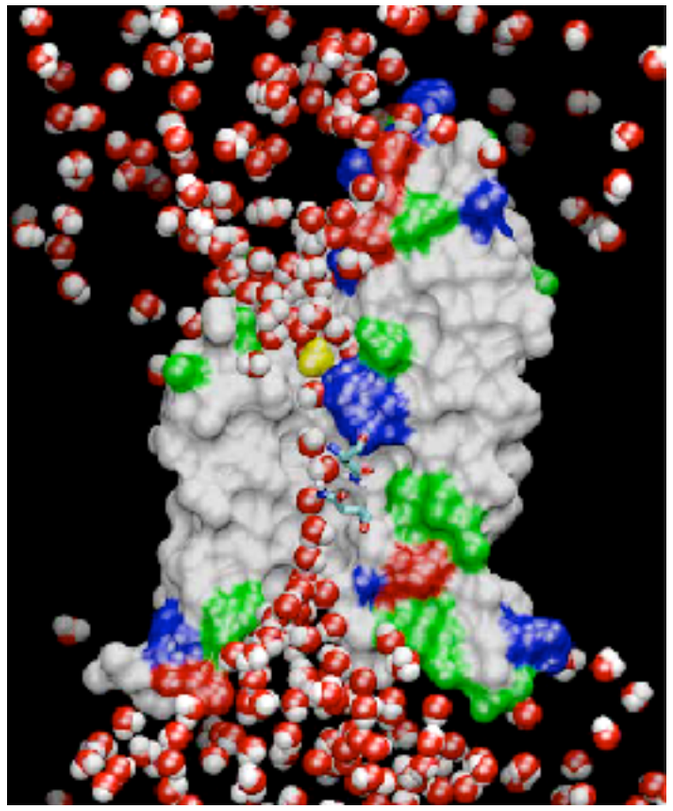
most common solutes are **ions**

open to both sides of membrane simultaneously

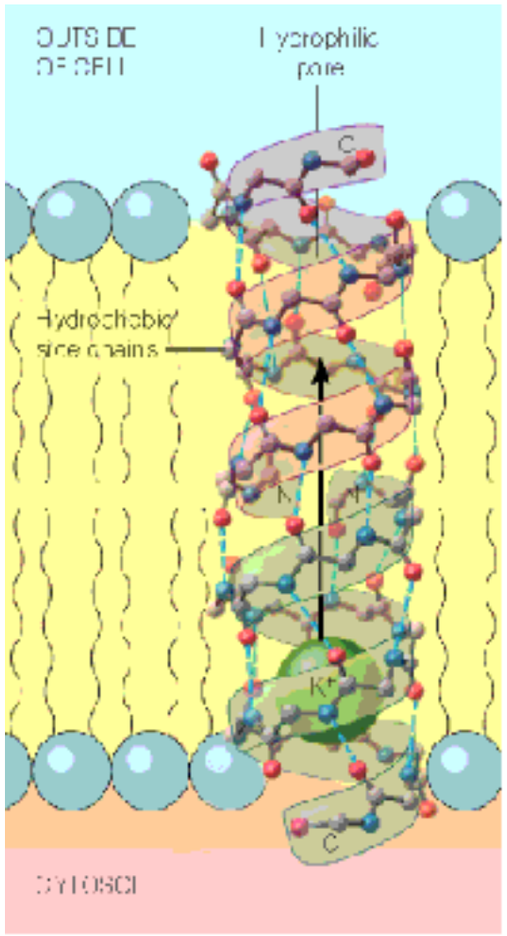


KcsA, a bacterial K+ channel

gramicidin, an unusual antibiotic ion channel

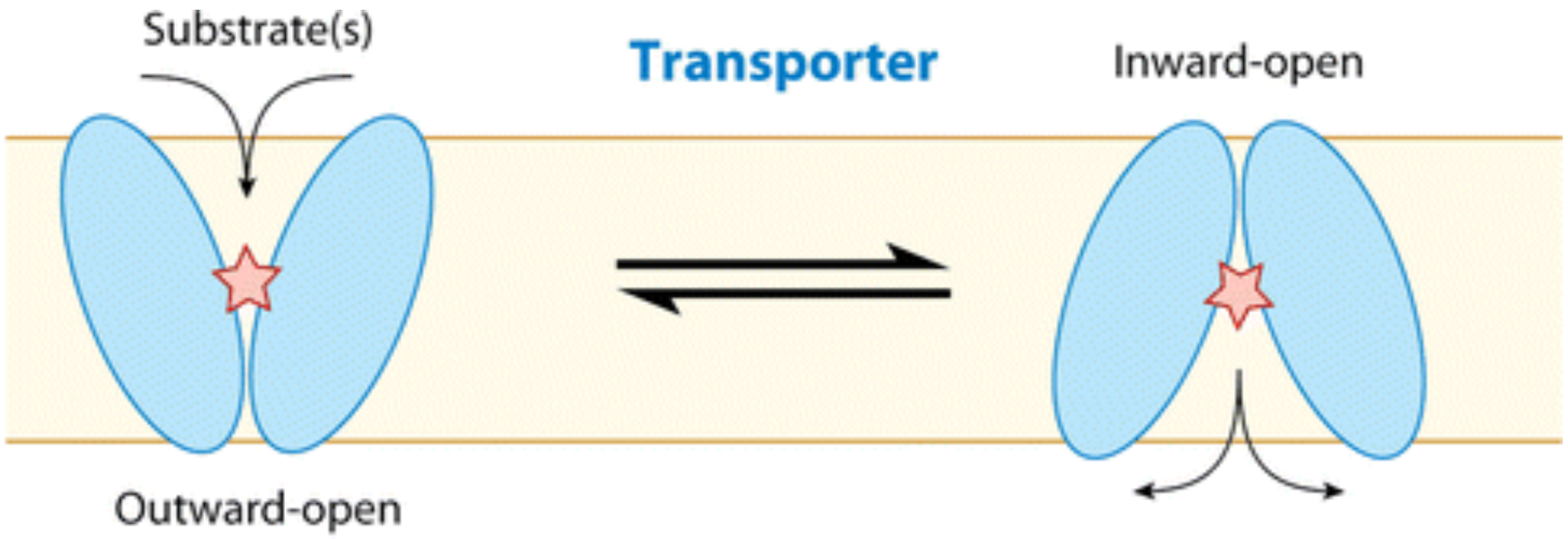


aquaporin, a water channel

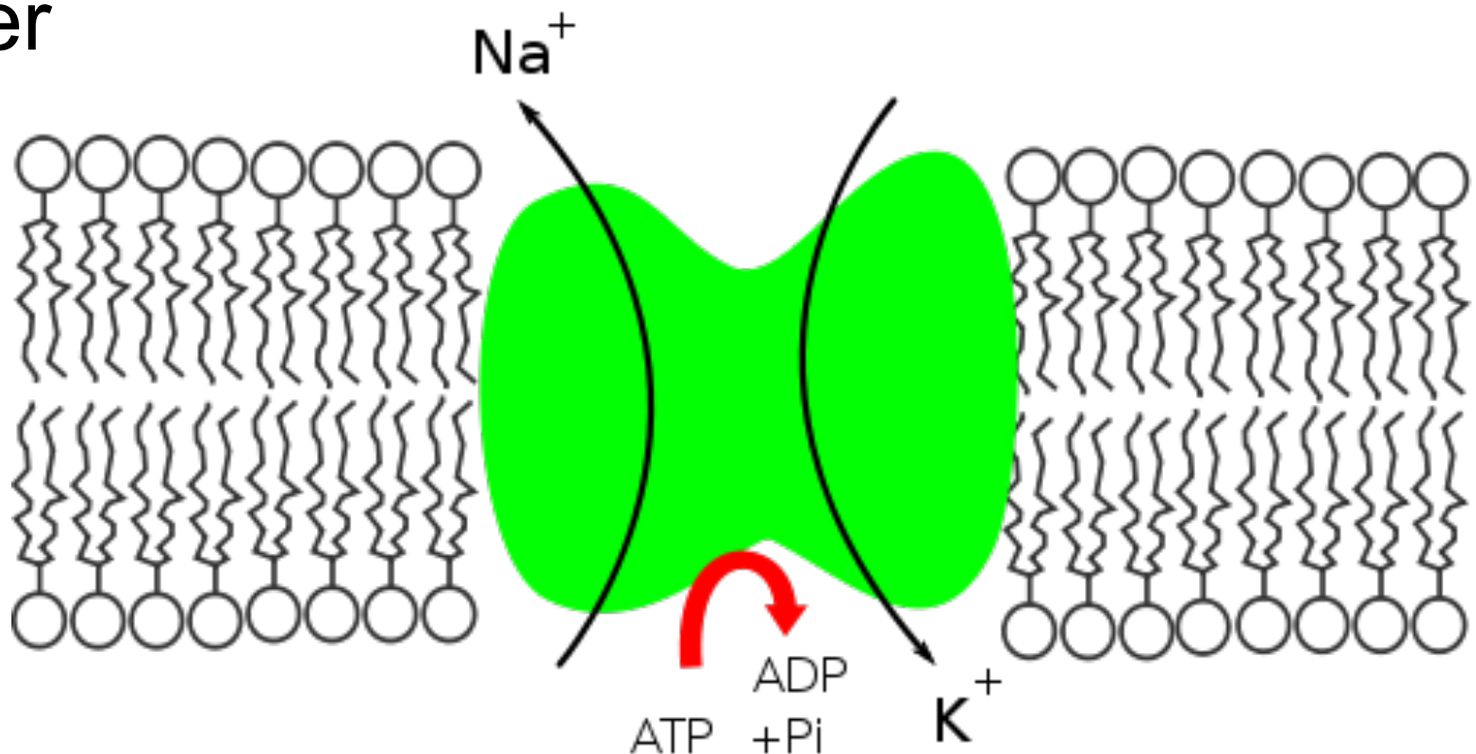


Nobel Prize (2003) for channel structures, **K+**: R. MacKinnon; **aquaporins**: P. Agre

membrane transporters

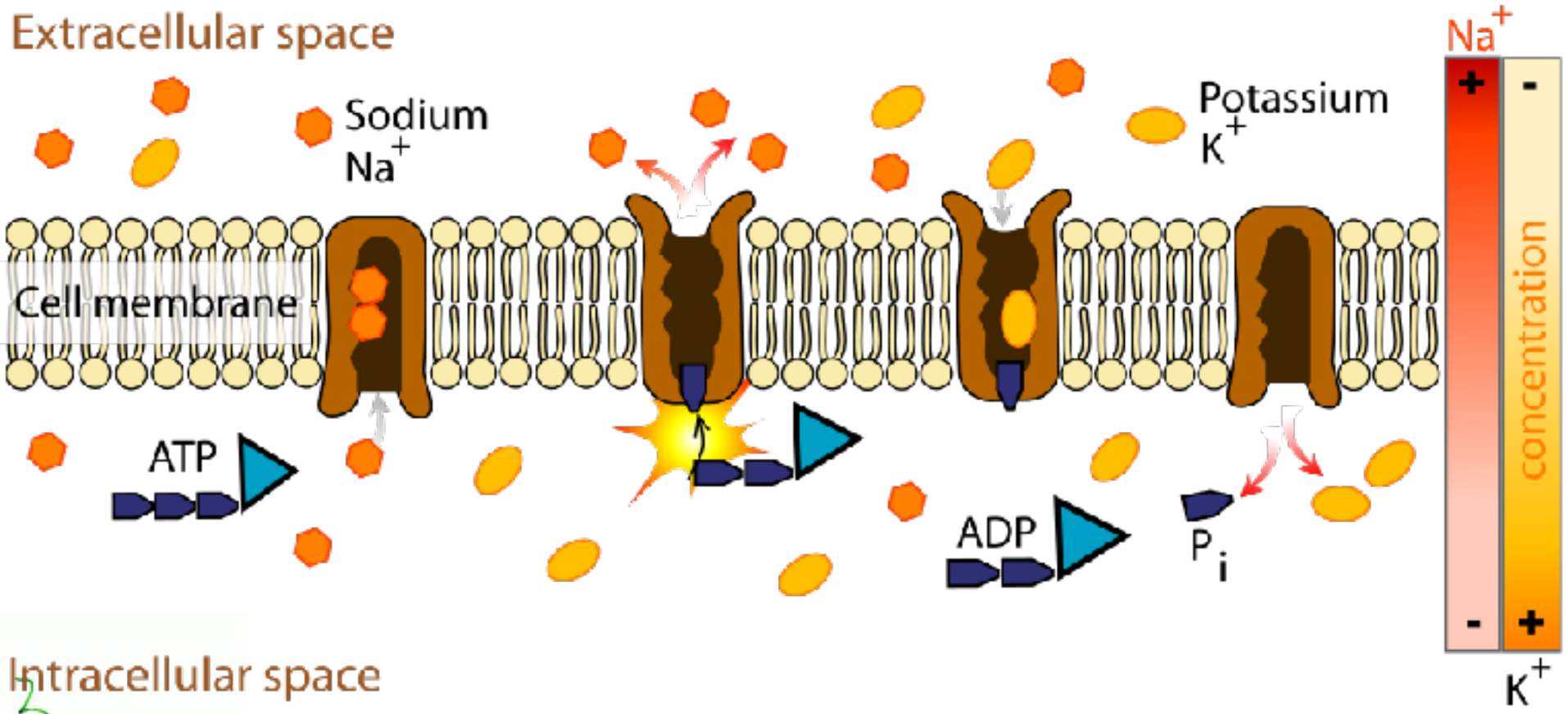


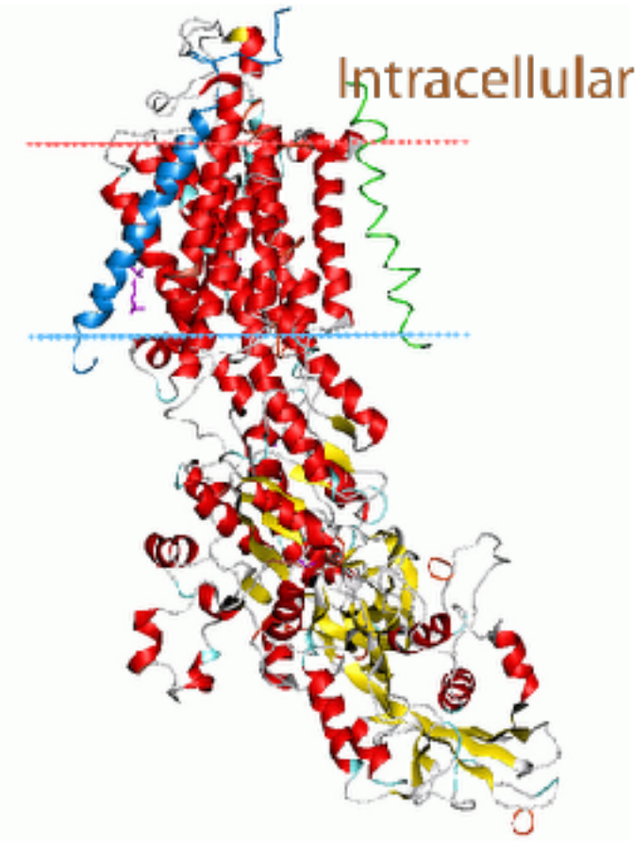
open to only **one** side of membrane at a time
substrate binds from one side and releases to other



primary active transporters

couple the hydrolysis of ATP to drive transport



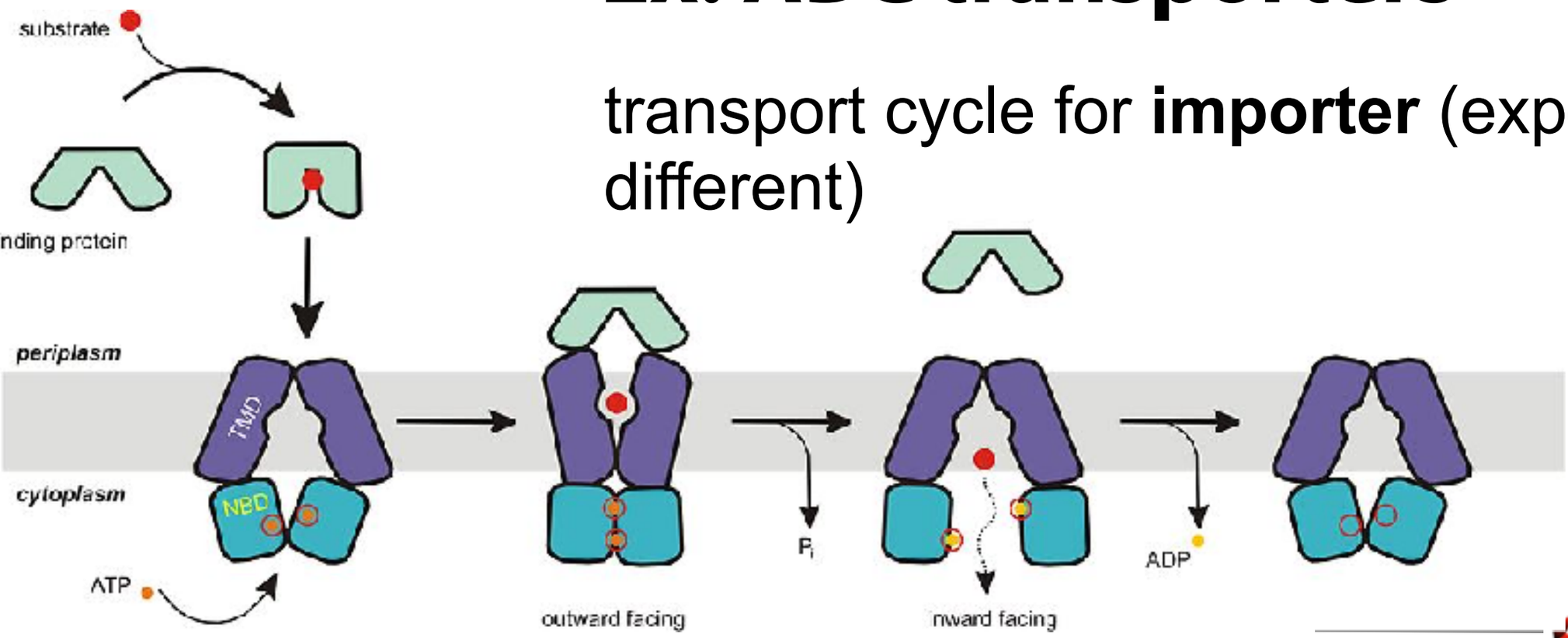


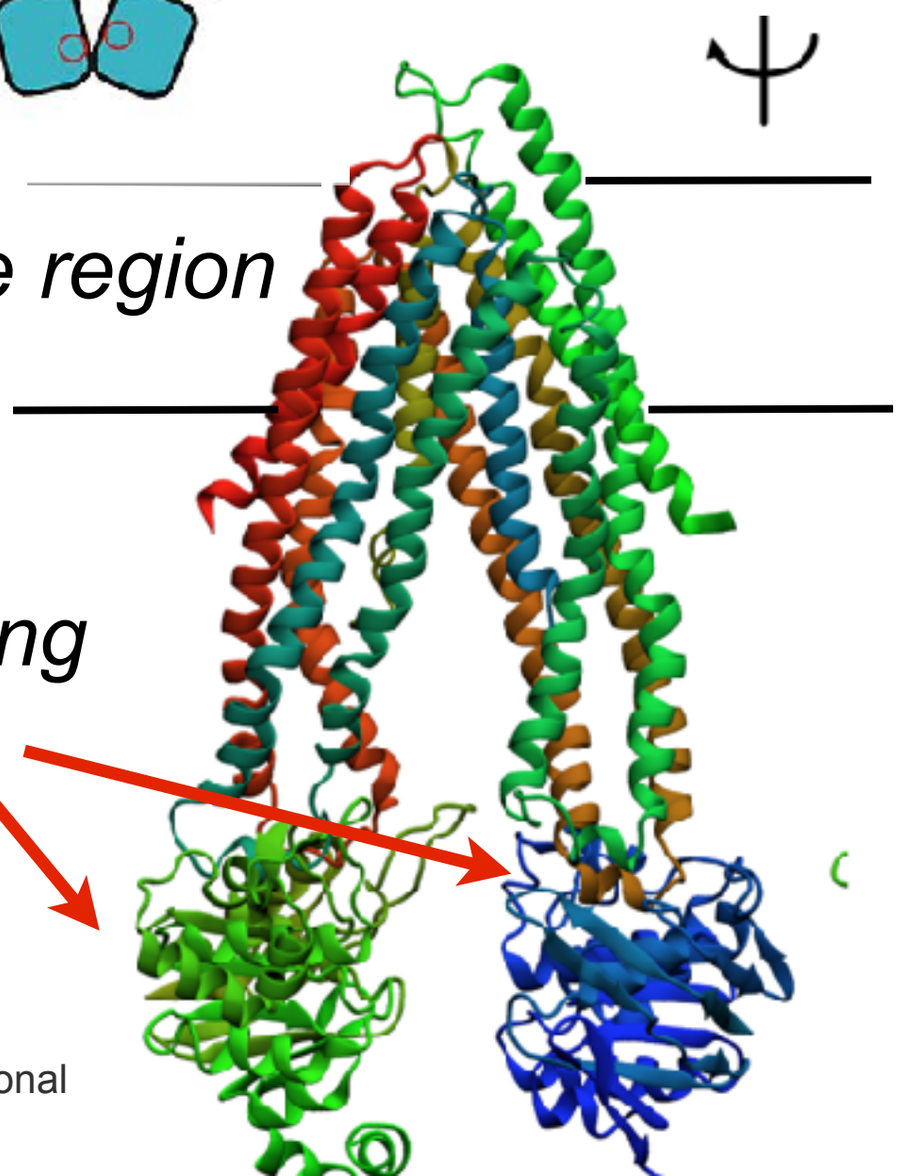
Examples include ion pumps, ATP synthase, ABC (ATP-binding cassette) transporters

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**)

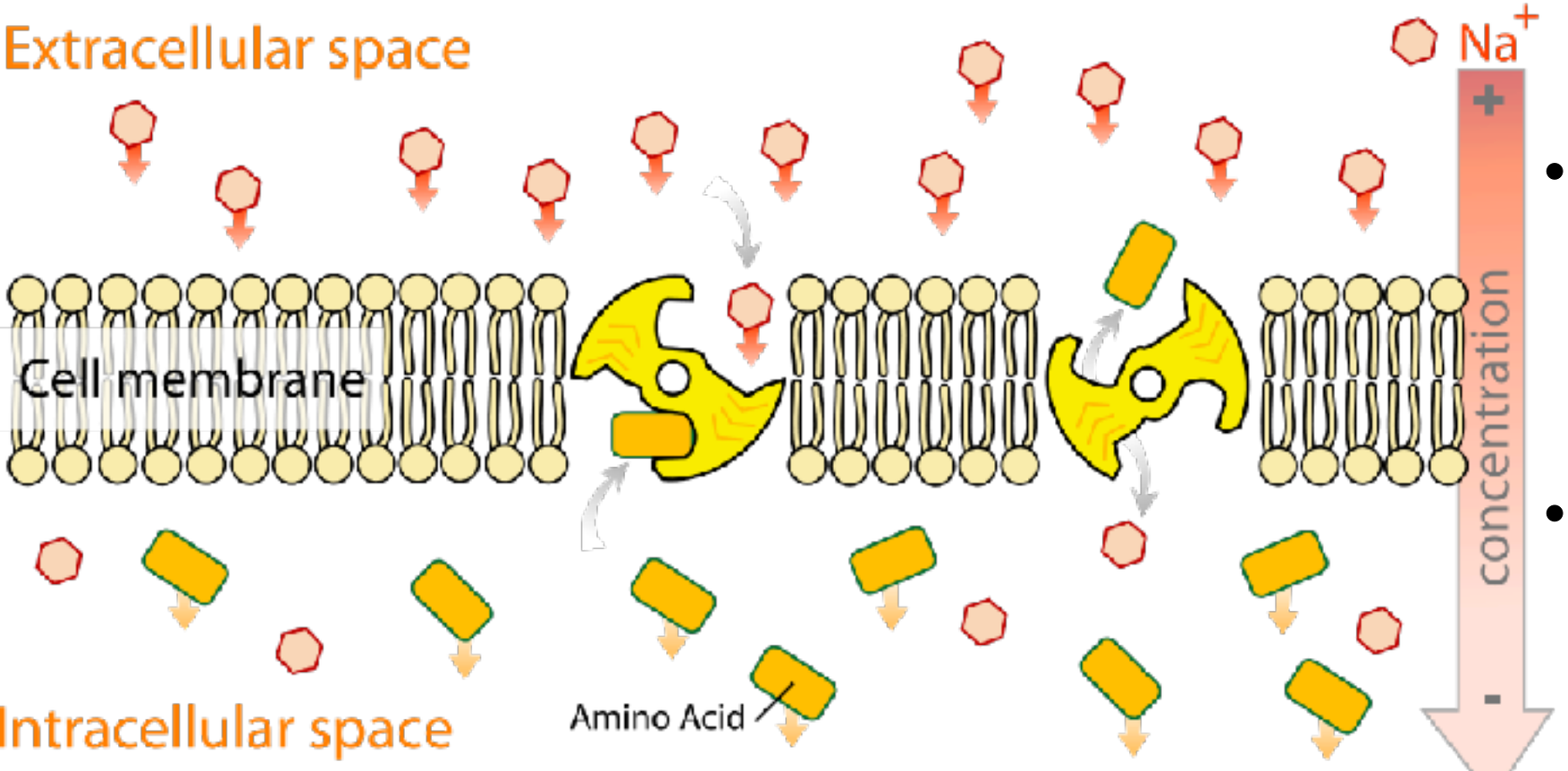
evolved to be more channel-like (not strongly coupled) for Cl-

Δ508 mutation found in 1/30 people, prevents expression in respiratory epithelial cells

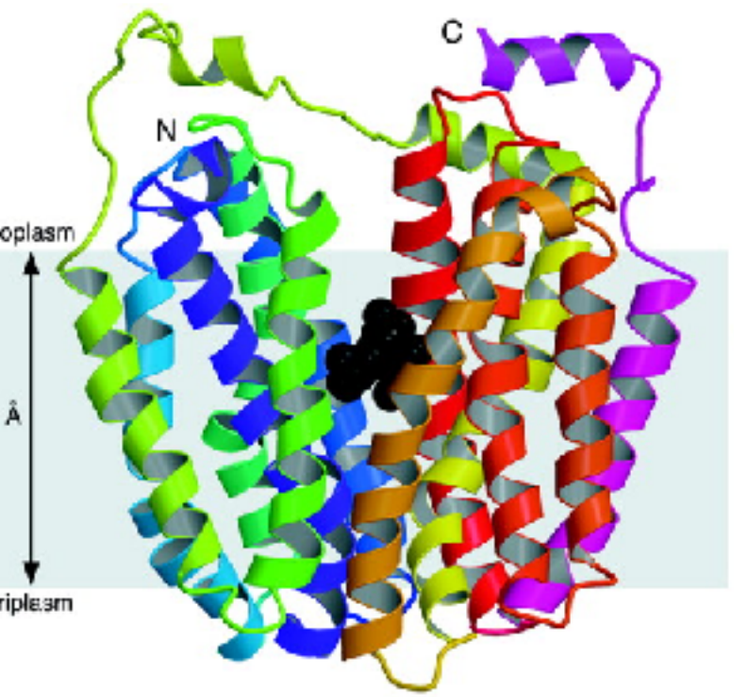
Rahman KS, Cui G, Harvey SC, McCarty NA (2013) Modeling the Conformational Changes Underlying Channel Opening in CFTR. PLoS ONE 8(9): e74574.

secondary active transporters

transport energy comes from co-transport of an ion



- ion goes with substrate → **symporter**
- against substrate → **antiporter**



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*)

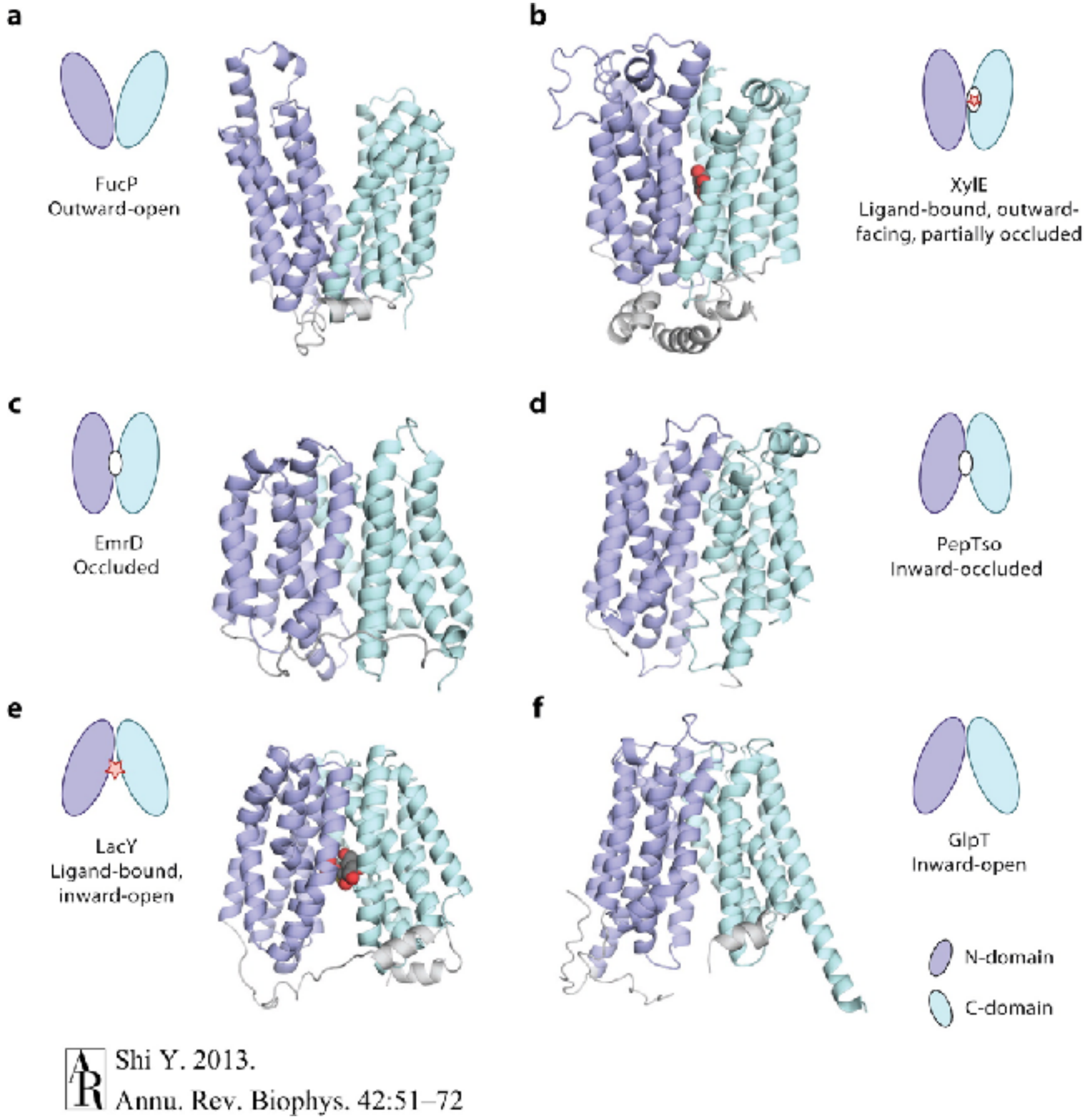
alternating access model of transport

transporter cycles through a number of distinct states

three primary states:

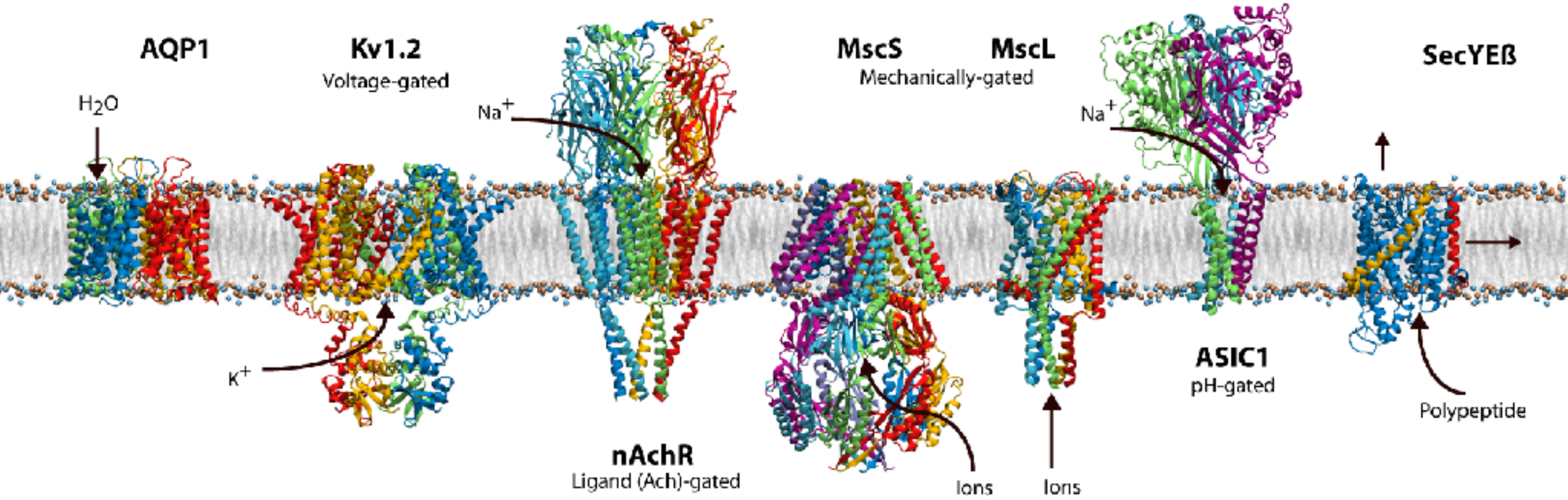
1) outward open
2) occluded
3) inward open

no transporter has structures in **ALL** states

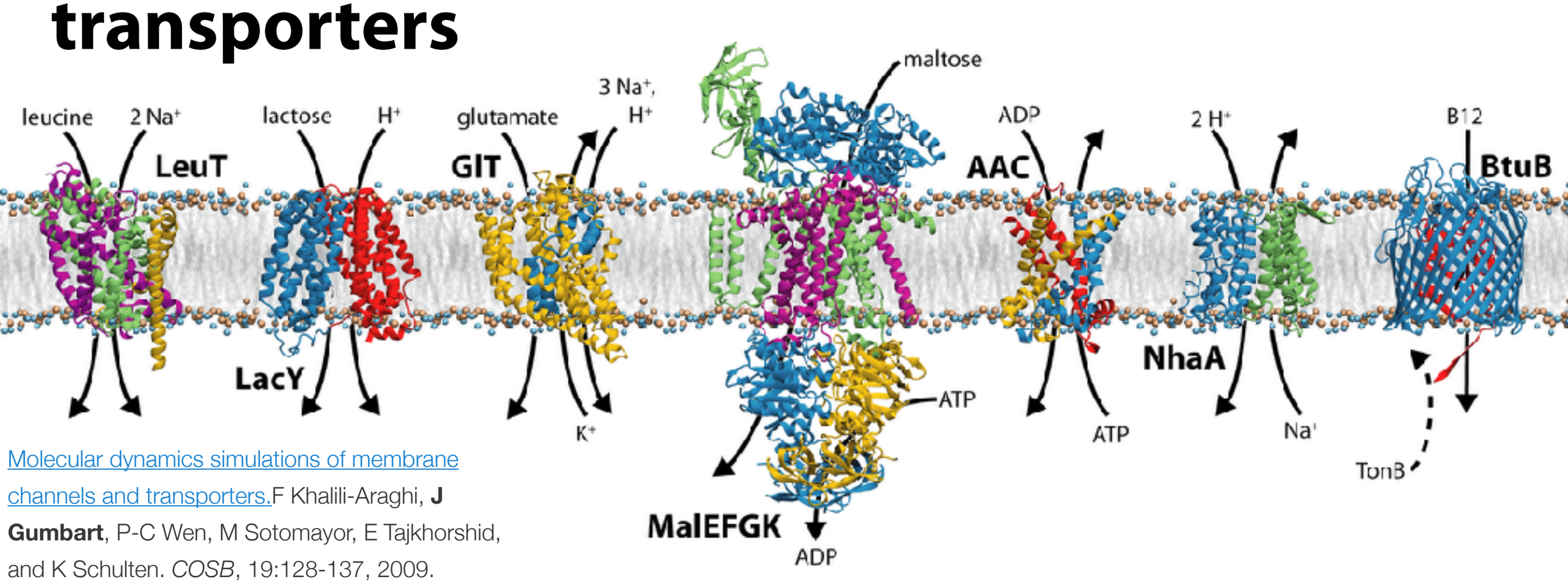


Shi Y. 2013.
Annu. Rev. Biophys. 42:51–72

channel structures



transporters



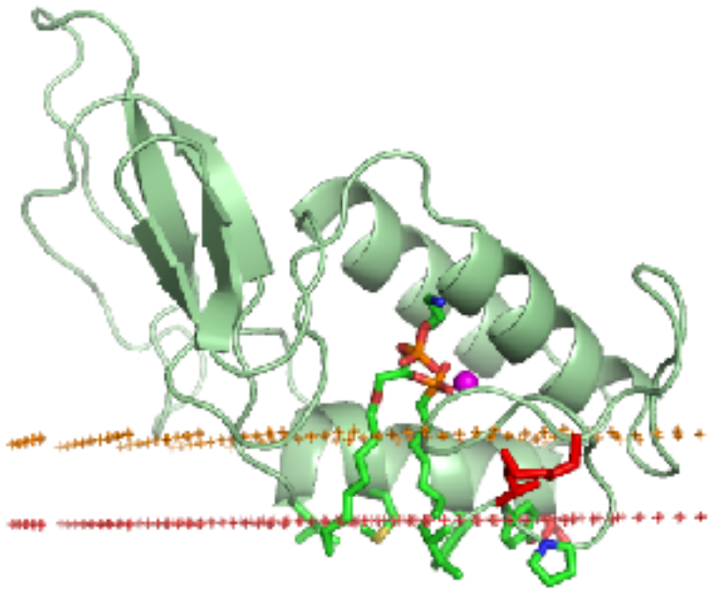
[Molecular dynamics simulations of membrane channels and transporters.](#)F Khalili-Araghi, **J Gumbart**, P-C Wen, M Sotomayor, E Tajkhorshid, and K Schulten. *COSB*, 19:128-137, 2009.

Peripheral membrane proteins

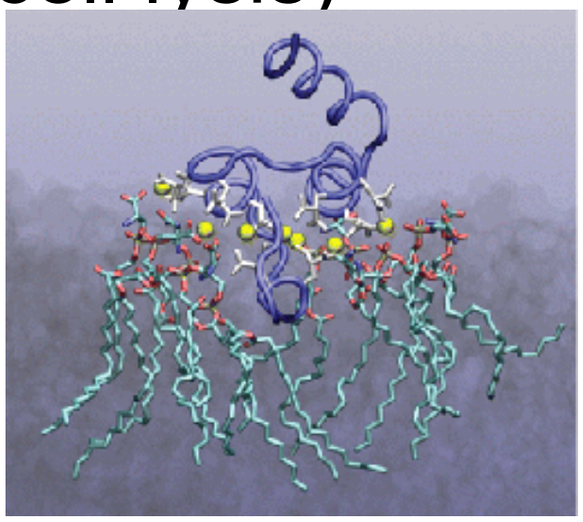
only **temporarily** associate with the membrane

A few examples:

enzymes



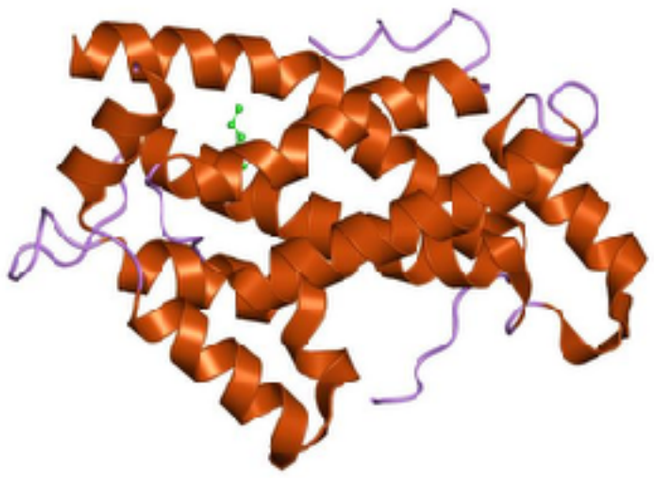
phospholipase A2 - involved in lipid metabolism, also in many venoms (promotes cell lysis)



Tajkhorshid Lab (UIUC): N. Tavoosi,et al. (2011) *JBC*. **286**: 23247.

structural

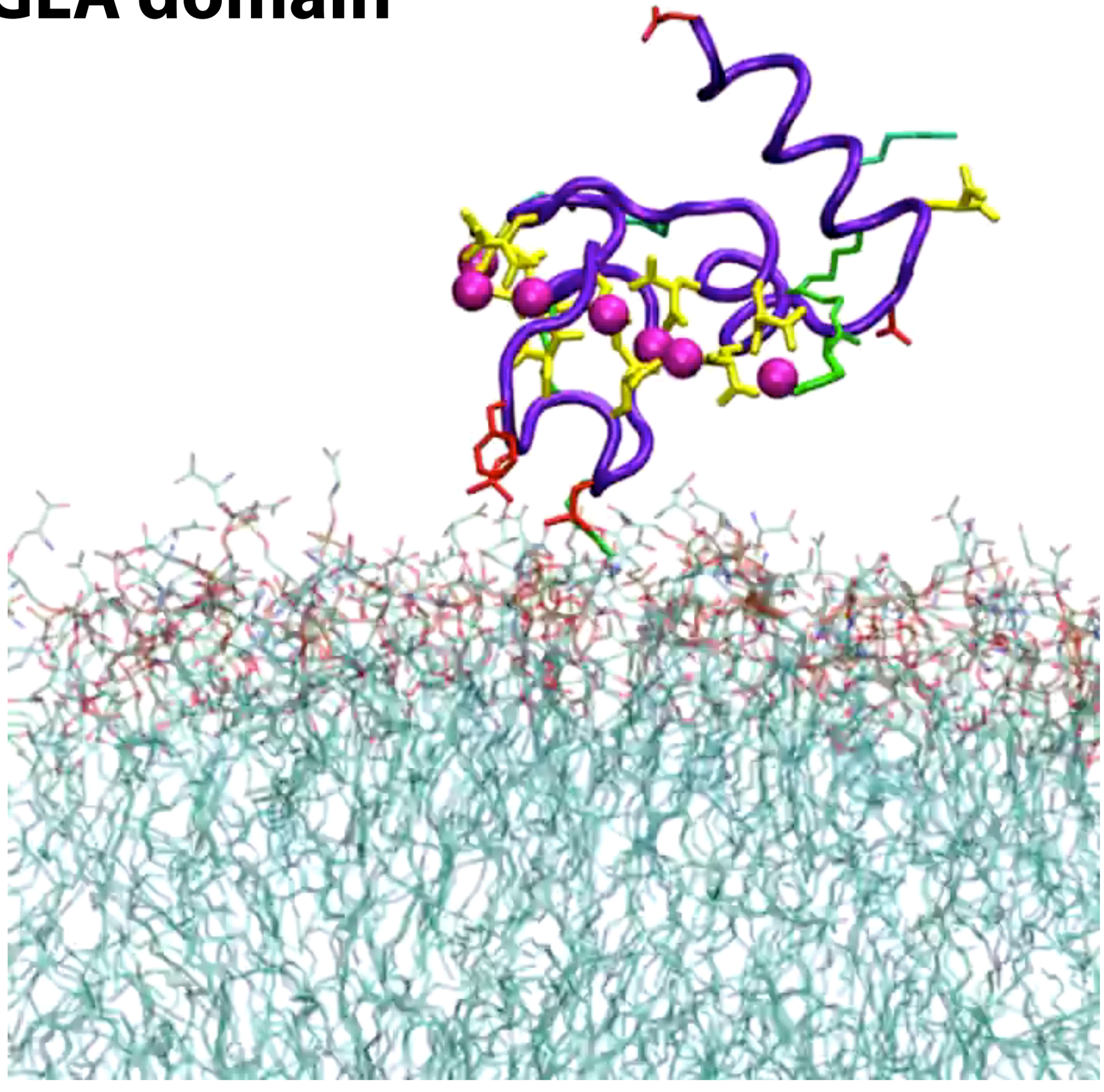
GLA domain - involved in blood coagulation cascade



hydrophobic molecule transporters

glycolipid transfer protein

Binding of a GLA domain



Tajkhorshid Lab (UIUC): N. Tavoosi,et al. (2011) *JBC*. **286**: 23247.

Energetics and the potential of mean force

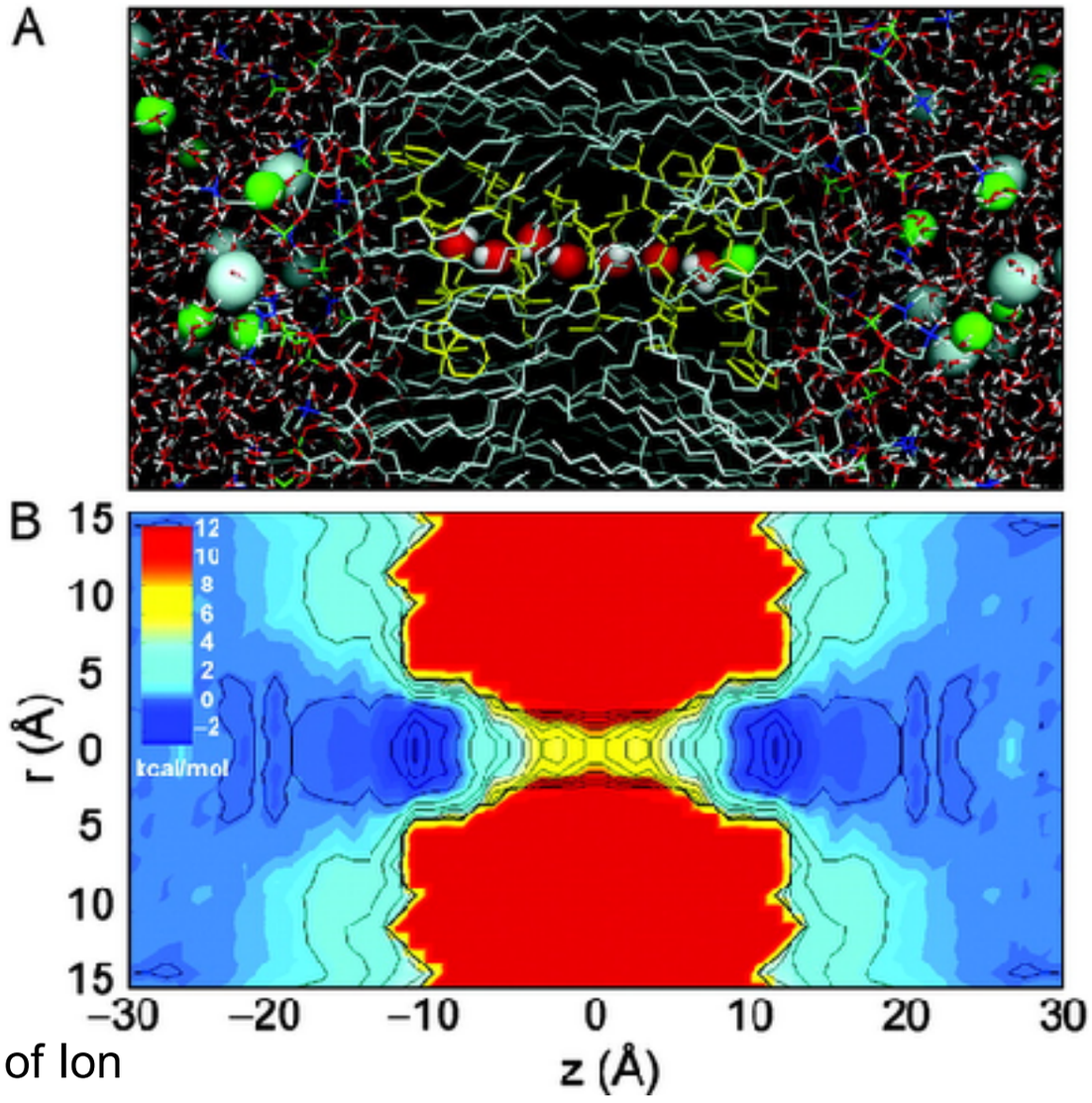
potential of mean force (PMF) projects full free-energy space onto one (or more) selected reaction coordinates

$$\int \mathrm{d}\mathbf{x}_1..\mathrm{d}\mathbf{x}_N \; e^{-U(\mathbf{x}_1,..,\mathbf{x}_N)/kT} = \int \mathrm{d}z \; e^{-W(z)/kT}$$

also can be expressed in terms probability: $W(z) = -kT \ln(\frac{P(z)}{P_0})$

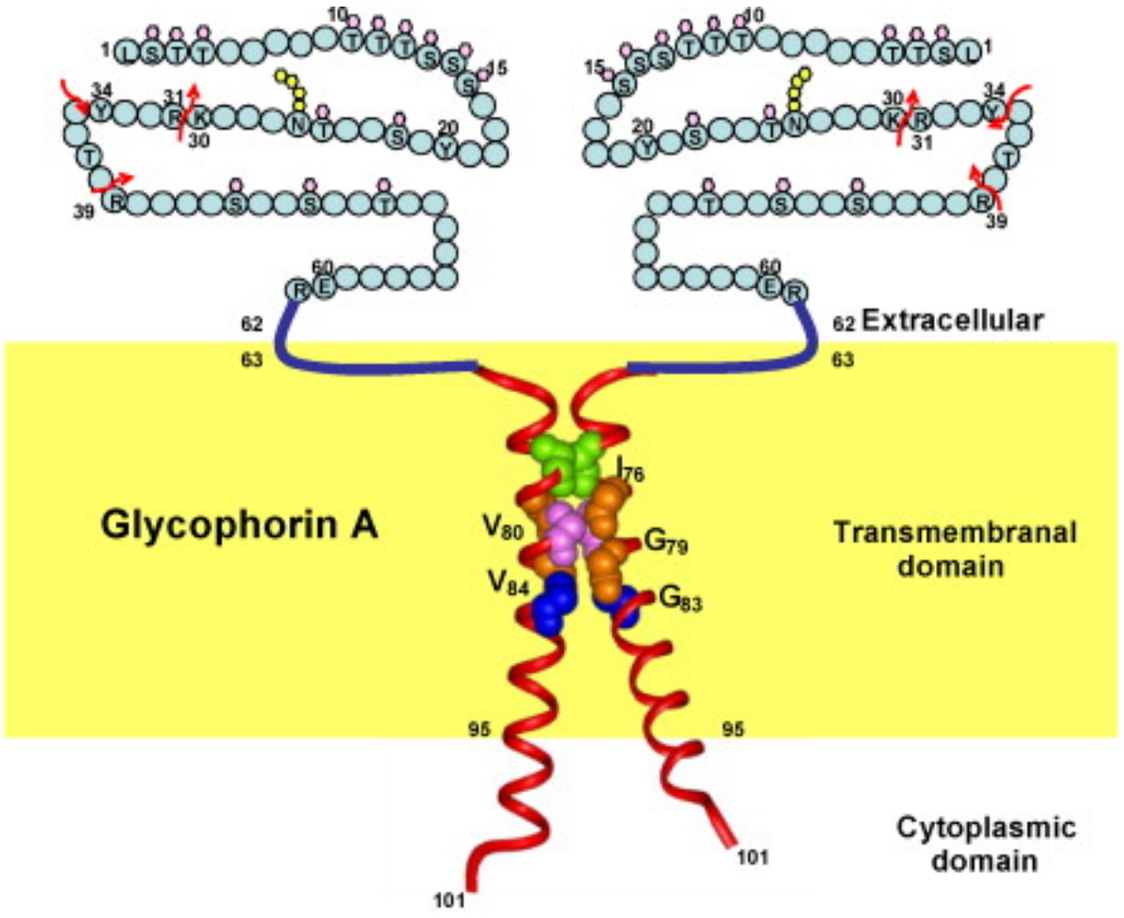
knowledge of PMF permits determination of many properties, e.g., conductance, average times, binding sites, etc.

2D PMF for ion transport through gramicidin A



T. W. Allen, O. S. Andersen and B. Roux. 2004. The Energetics of Ion Conduction in the Gramicidin A Channel.*Proc. Nat. Acad. Sci.* 101, 117-122.

Energetics: Glycophorin A

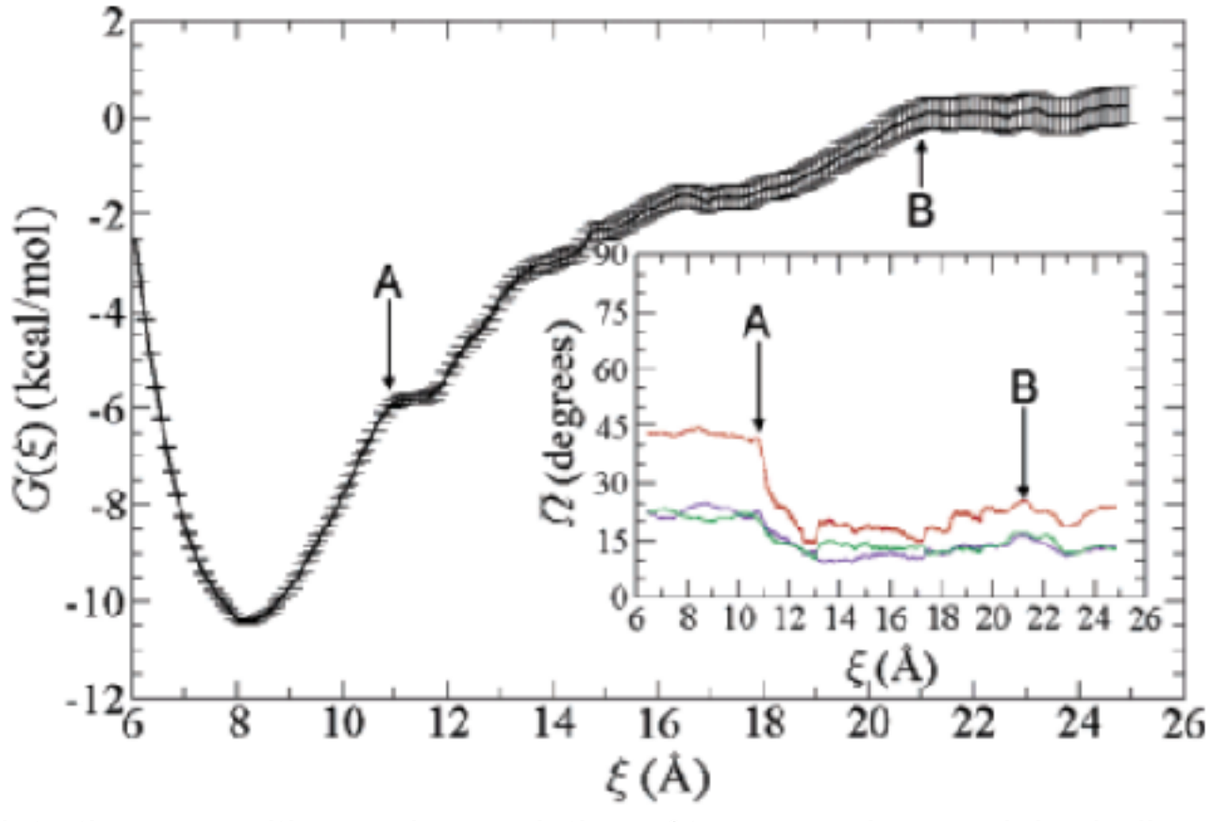


GypA-E expressed at surface of red blood cells, acts as a receptor, prevents aggregation, etc.

NMR structure of TM domain only

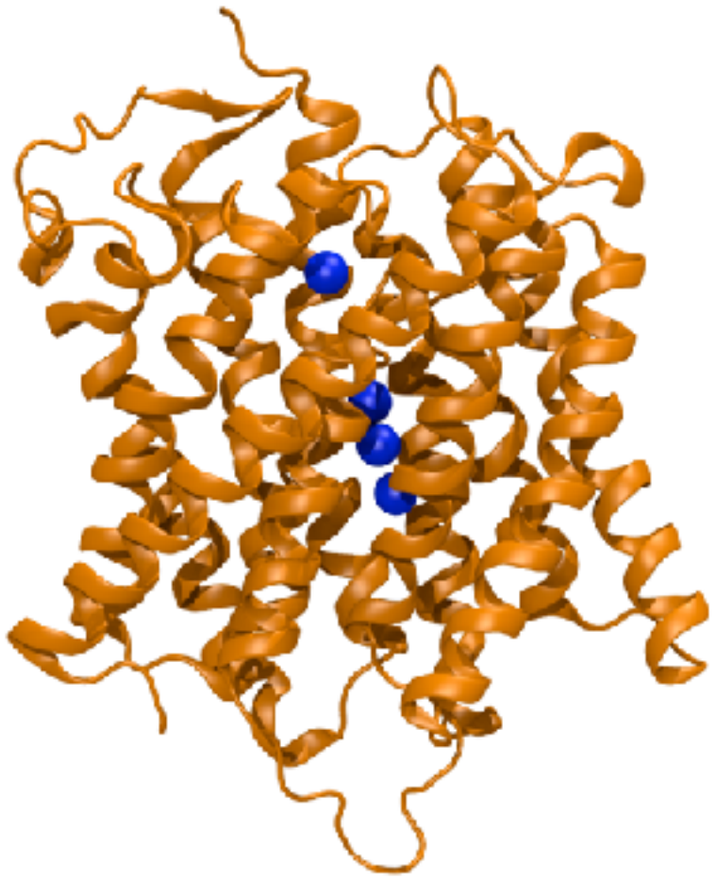
PMF for helix-helix association in membrane as function of separation distance

dimer is favored by **10 kcal/mol** over separate monomers, mediated by GxxG motif



Henin, J.; Pohorille, A.; Chipot, C. Insights into the recognition and association of transmembrane alpha-helices. The free energy of alpha-helix dimerization in glycophorin A *JACS* 2005, 127 (23), 8478-8484.

Energetics: AmtB



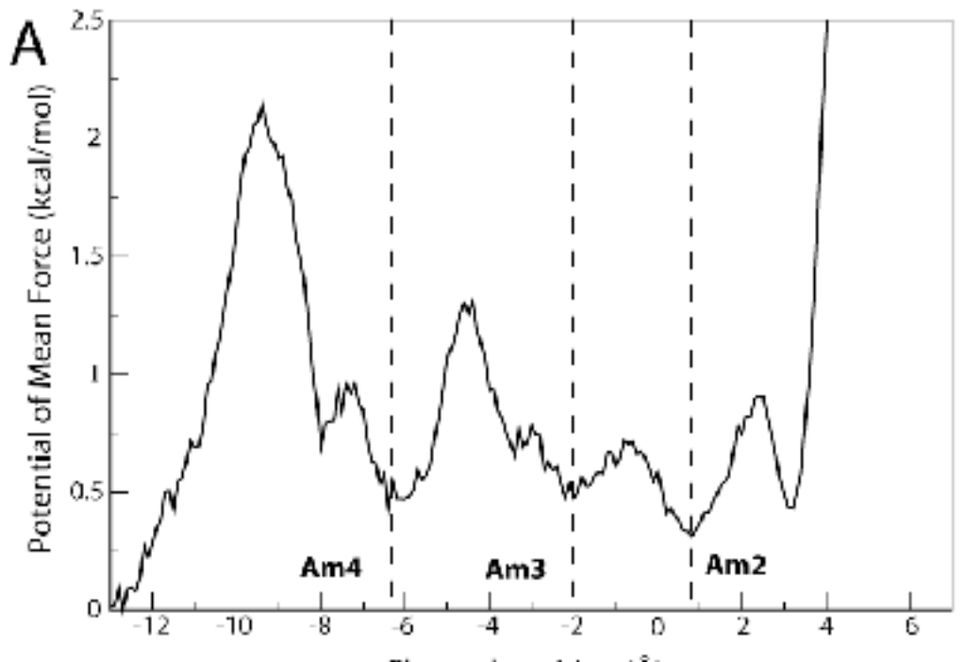
AmtB - an ammonia (NH_3)/ammonium (NH_4^+) channel

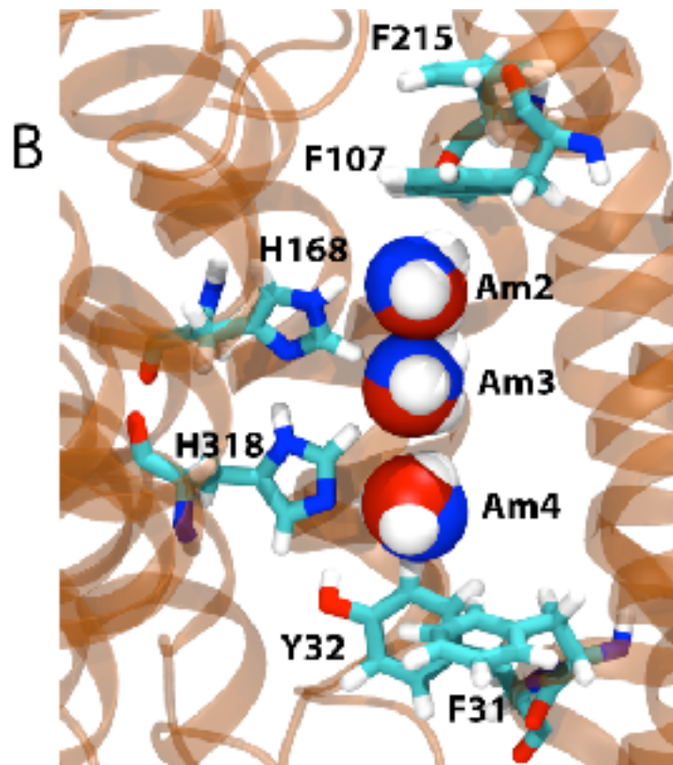
homologous to RhxG (x=A,B,C) proteins in mammalian blood cells

channel is **hydrophobic** - NH_4^+ likely changes protonation states at entrance/ exit

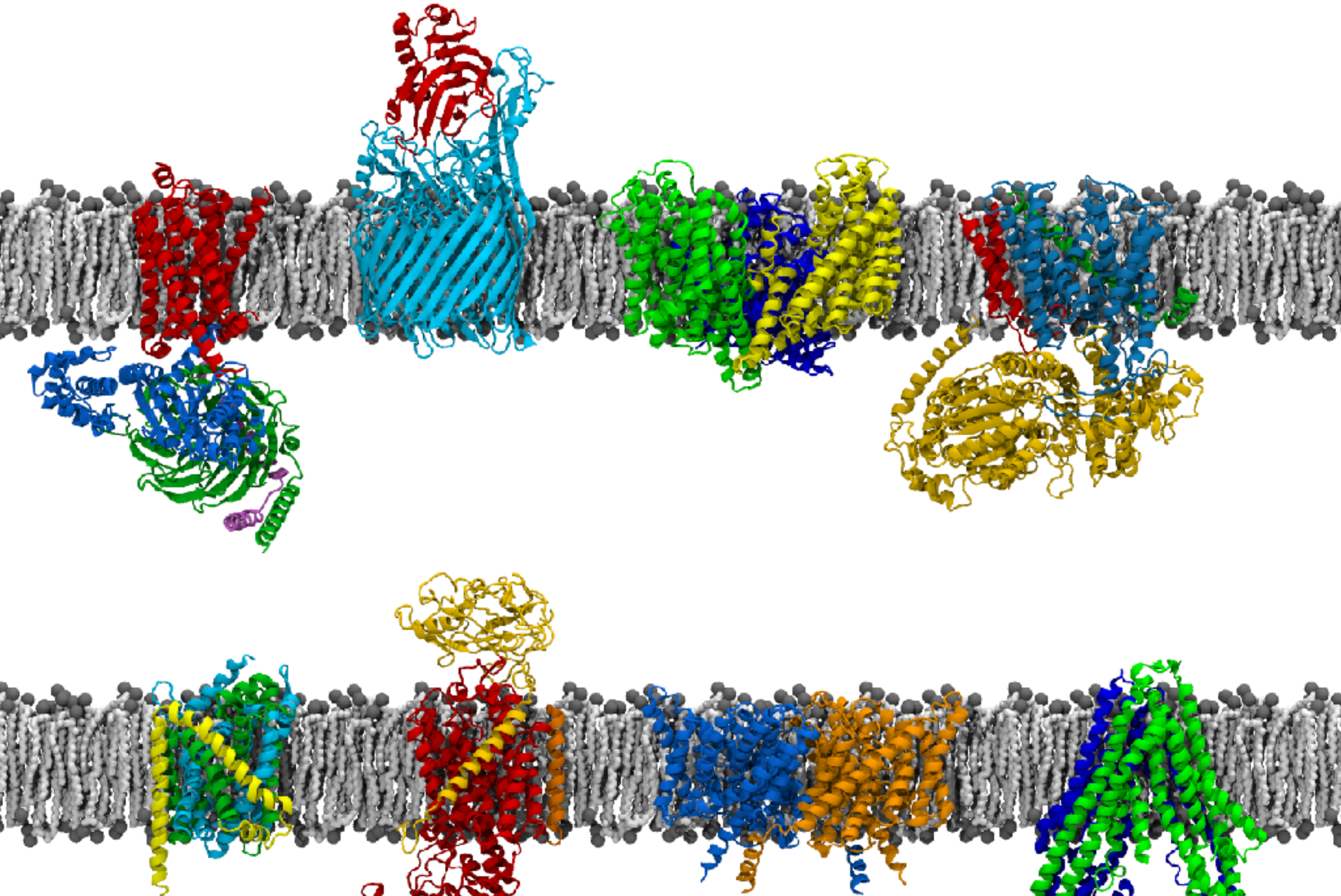
PMF for NH_3 moving through channel shows minima at crystallographically resolved binding sites

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

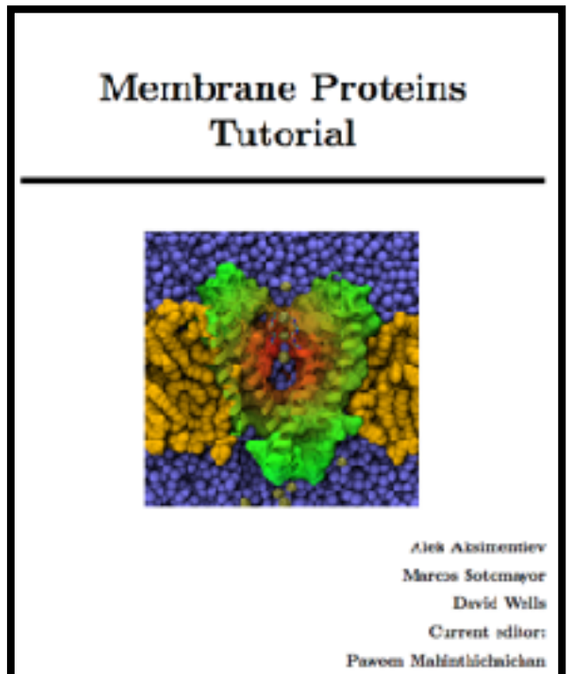
Step 3: Build the membrane, using VMD (POPE, POPC only) or CHARMM-GUI

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/

orientations of proteins in membranes (OPM) database

UNIVERSITY OF MICHIGAN | COLLEGE OF PHARMACY | LOMIZE GROUP

Search OPM

PDB ID or protein name

HOME | ABOUT OPM | DOWNLOAD OPM FILES | CONTACT US | PPM SERVER | LIPID COMPOSITION ATLAS

Protein Classification

Types (3 types)
Classes (11 classes)
Superfamilies (485 superfamilies)
Families (912 families)
Species (788 species)
Localization (24 types)
All proteins in OPM (3913 proteins)

Protein Links

PDB Sum, PDB, OCA, MPKS, PDBTM, MPDB, CGDB

PPM Server

Orientations of Proteins in Membranes (OPM) database

OPM provides spatial arrangements of membrane proteins with respect to the hydrocarbon core of the lipid bilayer.

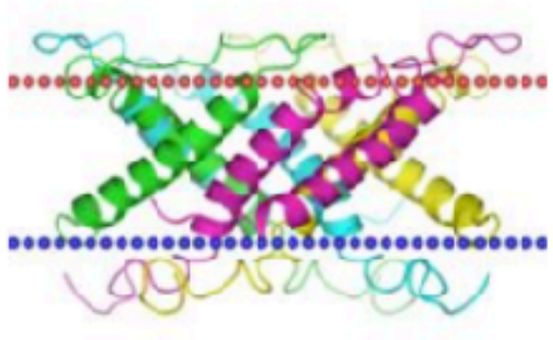
OPM includes all unique experimental structures of transmembrane proteins and some peripheral proteins and membrane-active peptides (Features).

Each protein is positioned in a lipid bilayer of adjustable thickness by minimizing its transfer energy from water to the membrane (Methods).

OPM provides structural classification and sorting according to different criteria (Classification).

Our calculations are in agreement with experimental studies of 24 transmembrane and 39 peripheral peptides and proteins (Comparison with Experiments).

For more information on single-spanning transmembrane proteins please see our Membranome database



3hzq » Mechanosensitive channel MscL, expanded state

See all membrane protein images in OPM

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/

CHARMM-GUI

Effective Simulation Input Generator and More

CHARMM is a versatile program for atomic-level simulation of many-particle systems, particularly macromolecules of biological interest. - M. Karplus

about us :: input generator :: archive :: charmm docs :: lectures :: movie gallery :: video demo :: citations :: update log :: jobs & events :: giving

Some lectures and job postings are now available. See upload log for update history and giving for donation. Contact info is given below.

Input Generator

- PDB Reader
- Glycan Reader & Modeler
- Ligand Reader & Modeler
- Glycolipid Modeler
- LPS Modeler
- Solvator
- Quick MD Simulator
- Drude Prepper
- Membrane Builder
- Martini Maker
- PACE CG Builder
- Boundary Potential Utilizer
- PBEQ Solver
- Implicit Solvent Modeller
- Free Energy Calculator
- NMR Structure Calculator

Membrane Builder

Tutorial

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. See gromacs/README.
- **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.

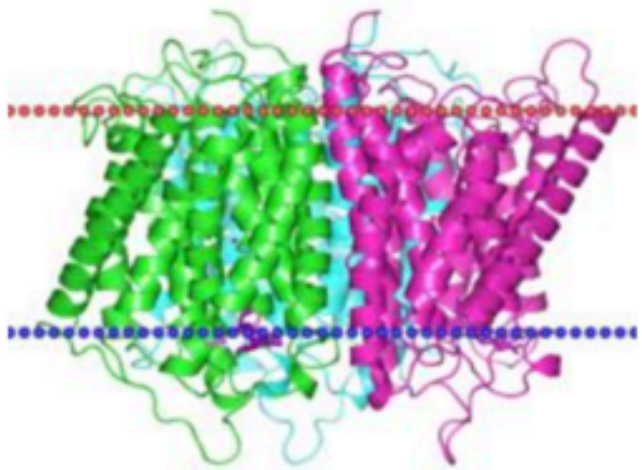
Building a membrane-protein system (easier)

Ex: AmtB (PDB 1U7G) an NH_3/NH_4^+ channel

OPM shows that it is a trimer

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)



1u7g » Ammonia Channel	
Hydrophobic Thickness	29.8 ± 1.3 Å

Protein/Membrane System

Download PDB File: 1u7g Download Source: OPM

Upload PDB File: Choose File No file chosen

PDB Format: PDB PDBx/mmCIF CHARMM

Membrane Only System

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)

NH_3/NH_4^+ (substrates of the channel)

Crystallographic water

Model/Chain Selection Option:

Click on the chains you want to select.

Select Model # 1 ☐ Read all models?

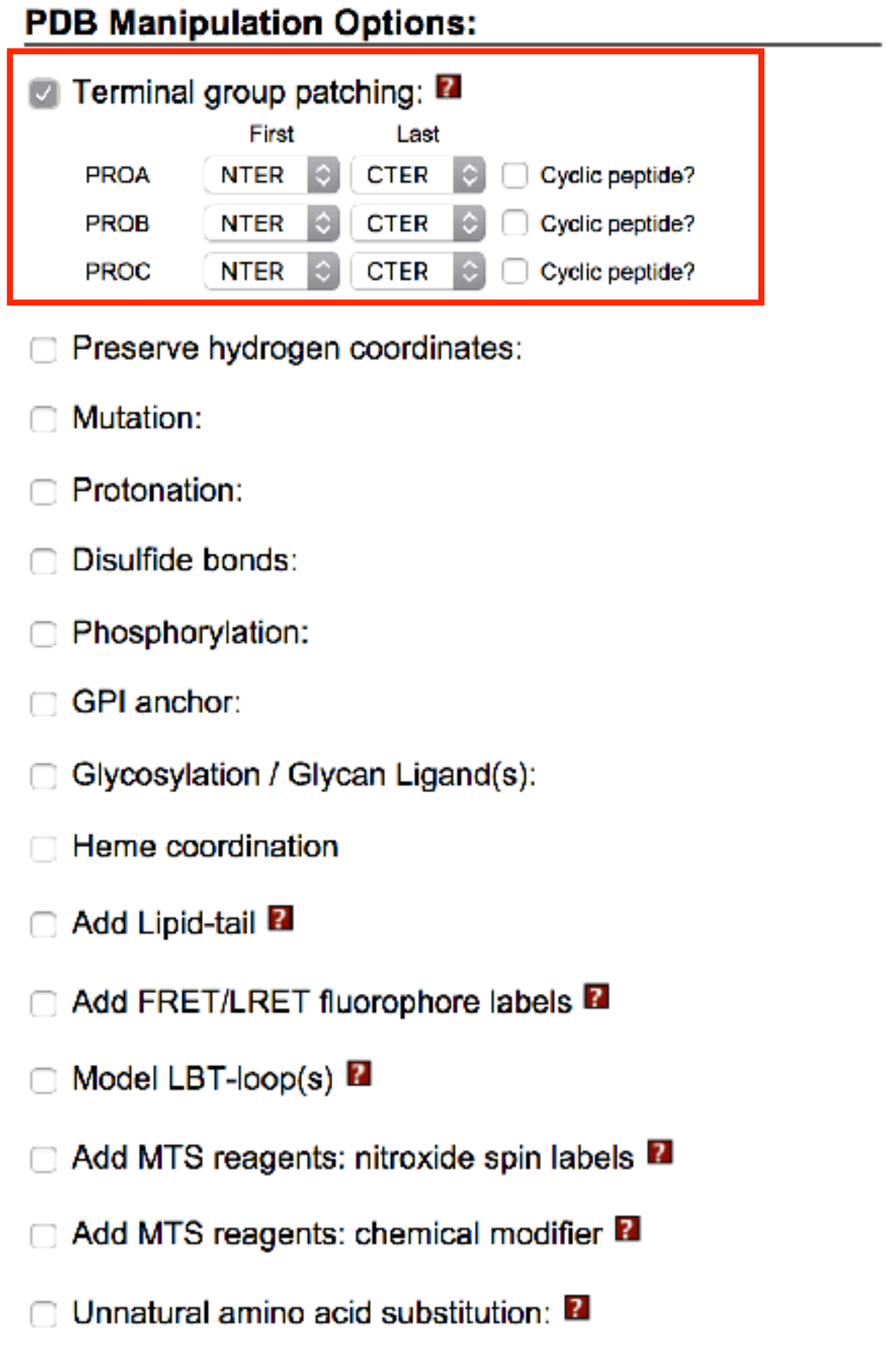
				Residue ID		
	Type	SEGID	PDB ID	First	Last	Engineered Residues
☑	Protein	PROA	A	3	385	None
☑	Protein	PROB	B	3	385	None
☑	Protein	PROC	C	3	385	None
☐	Hetero	HETA	D			BOG
☑	Hetero	HETB	D			NH4
☑	Hetero	HETC	D			NH3
☐	Hetero	HETD	E			BOG
☑	Hetero	HETE	E			NH4
☑	Hetero	HETF	E			NH3
☐	Hetero	HETG	F			BOG
☑	Hetero	HETH	F			NH4
☑	Hetero	HETI	F			NH3
☐	Hetero	HETJ				DUM
☑	Water	WATA	D			
☑	Water	WATB	E			
☑	Water	WATC	F			

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

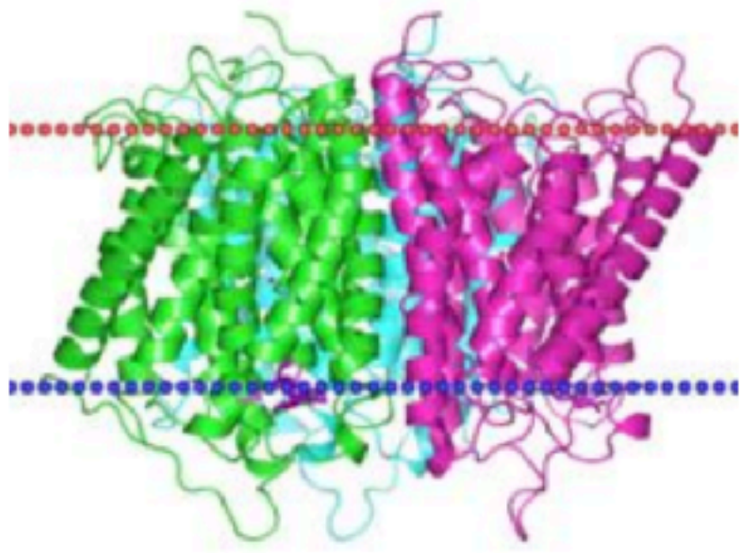


Building a membrane-protein system (easier)

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)

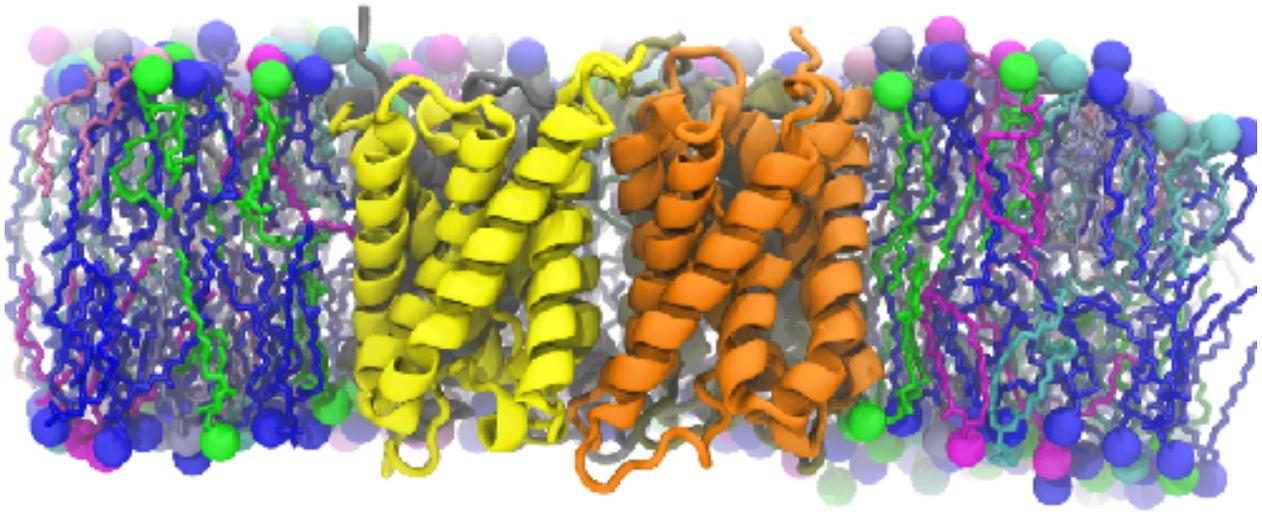


1u7g » Ammonia Channel	
Hydrophobic Thickness	29.8 ± 1.3 Å

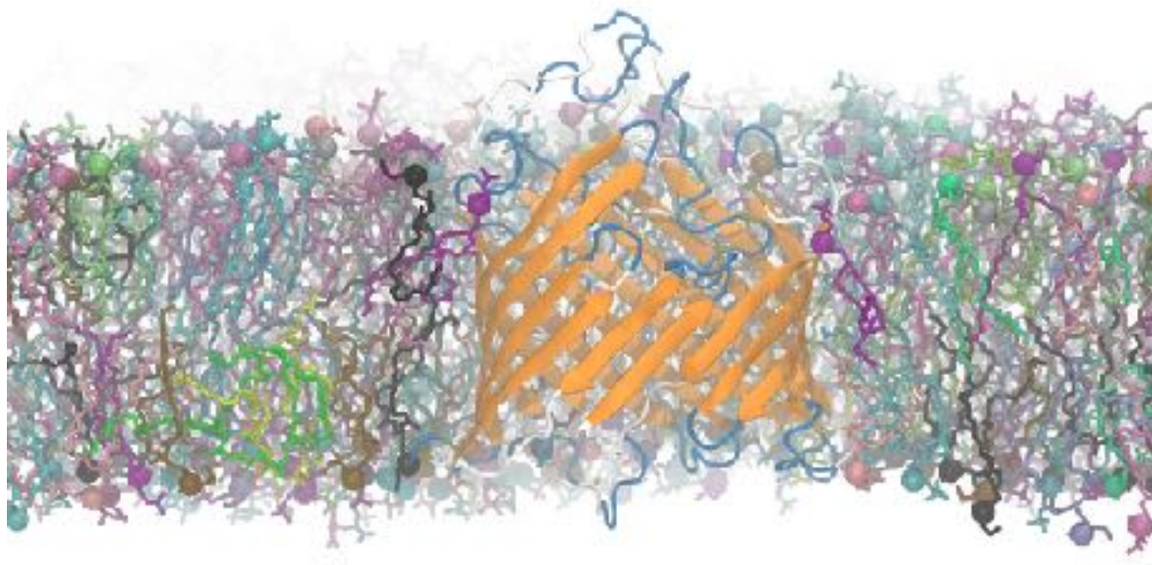
Ideally, want to select lipids to match the native membrane composition!

Search textbooks, papers, etc. for estimates of the lipids and ratios for the membrane of interest

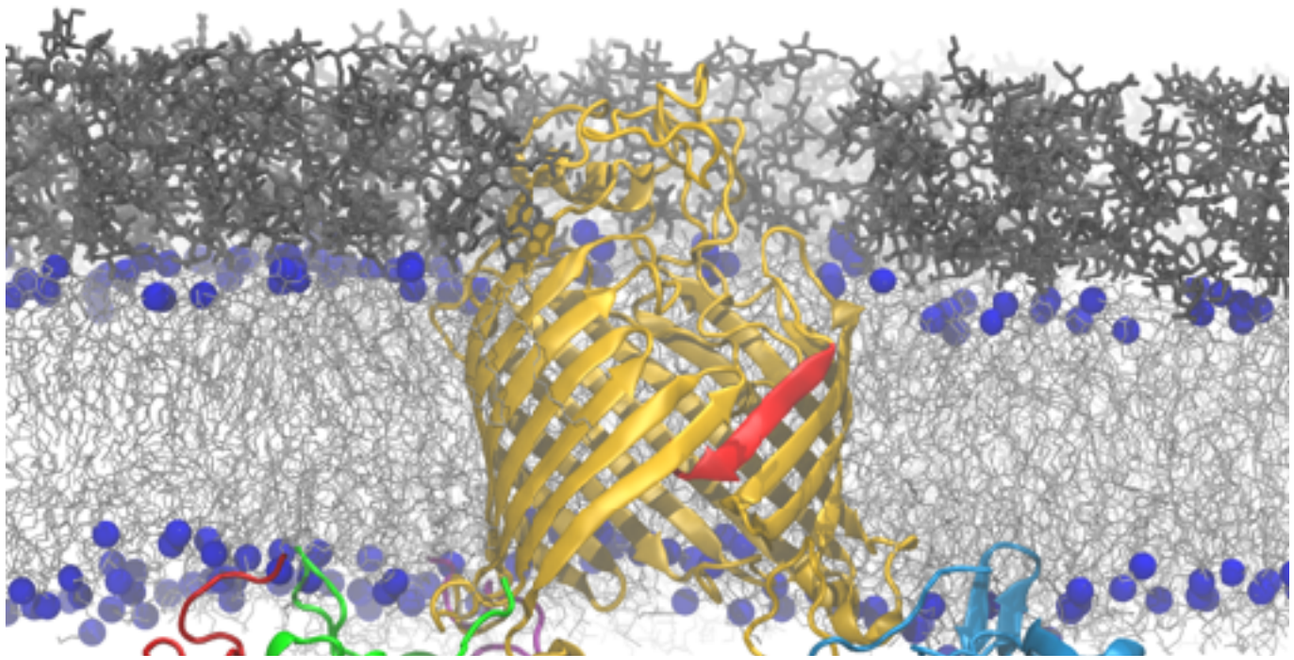
Building a membrane-protein system (easier)



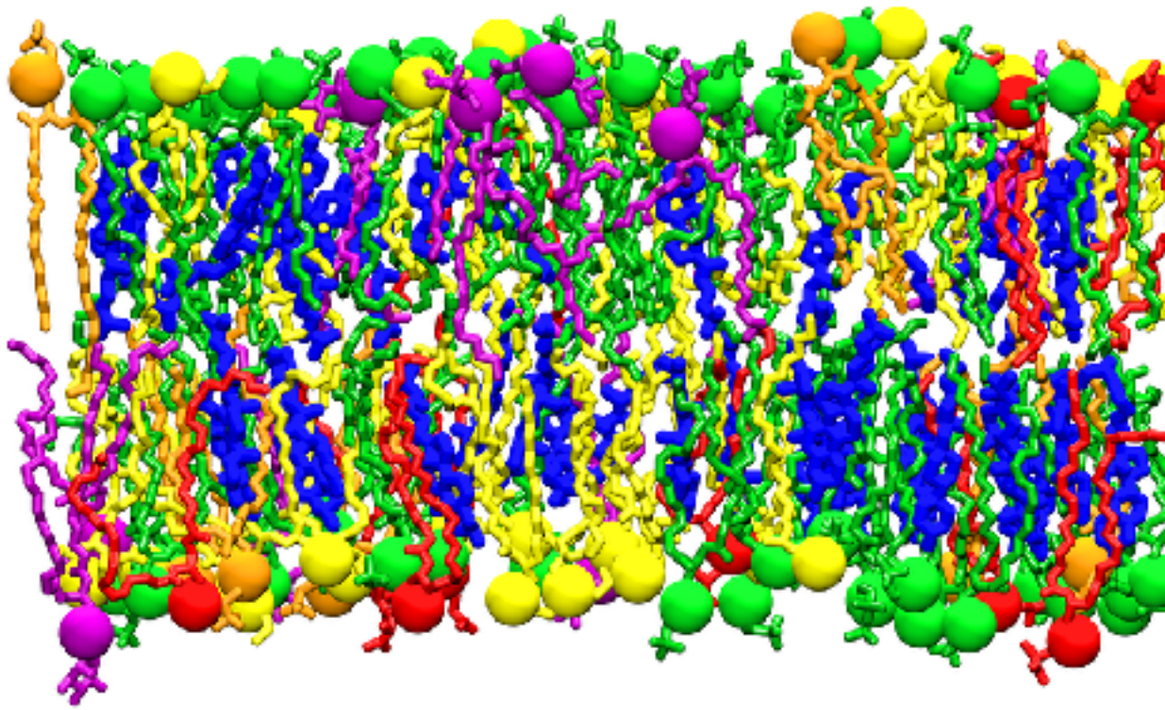
Gram-negative inner membrane



mitochondrial membrane



Gram-negative outer membrane



mammalian plasma membrane

Building a membrane-protein system (easier)

Which lipids to use for the membrane?

System Size Determination Options:

- ○ **Homogeneous Lipid** (we recommend users to use "**Heterogeneous Lipid**" even for homogeneous lipid bilayer building)
- ◉ **Heterogeneous Lipid**

1. Box Type: Rectangular (Currently, only CHARMM, NAMD, and GROMACS support the hexagonal box)

2. Length of Z based on:
   - ◉ Water thickness 17.5 (Minimum water height on top and bottom of the system)
   - ○ Hydration number 50 (Number of water molecules per one lipid molecule)
   - ○ Hydration (w/w) % 50 (Percent ratio of Water/lipid weight)

3. Length of XY based on:
   - ◉ Ratios of lipid components
   - ○ Numbers of lipid components

Length of X and Y: 120 (initial guess)
(The system size along the X and Y must be the same)

Show the system info click this once you fill the following table:

For simplicity, used single-component POPE here

Lipid Type	Charge [e]	Tail Info. [sn1/sn2]	Images	Upperleaflet Ratio (Integer)	Lowerleaflet Ratio (Integer)	Surface Area
▸ **Sterols**						
▸ **PA (phosphatidic acid) Lipids**						
▸ **PC (phosphatidylcholine) Lipids**						
▾ **PE (phosphatidylethanolamine) Lipids**						
DLPE	0	12:0 / 12:0	[Image]	0	0	60.8
DMPE	0	14:0 / 14:0	[Image]	0	0	59.9
DPPE	0	16:0 / 16:0	[Image]	0	0	59.0
DSPE	0	18:0 / 18:0	[Image]	0	0	58.8
PYPE	0	16:0 / 16:1	[Image]	0	0	58.8
POPE	0	16:0 / 18:1	[Image]	1	1	58.8

Calculated Number of Lipids:

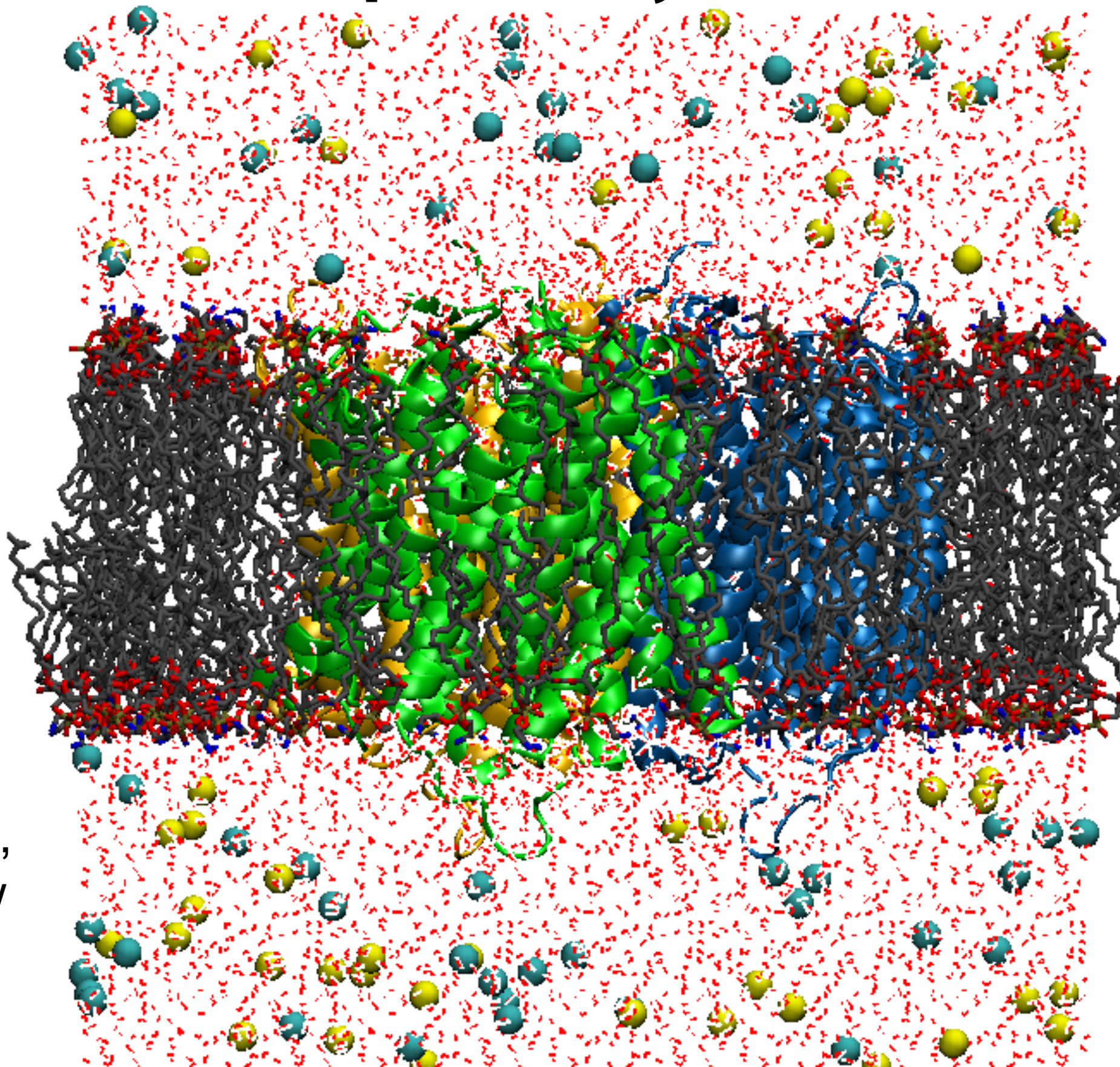
Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPE	170	176

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	4443.29007	4055.37973
Lipid Area	9996	10348.8
# of Lipids	170	176
Total Area	14439.29007	14404.17973
Protein X Extent	40.60	
Protein Y Extent	38.29	

Building a membrane-protein system (easier)

After a few more choices, a complete system is output (**step 5**)



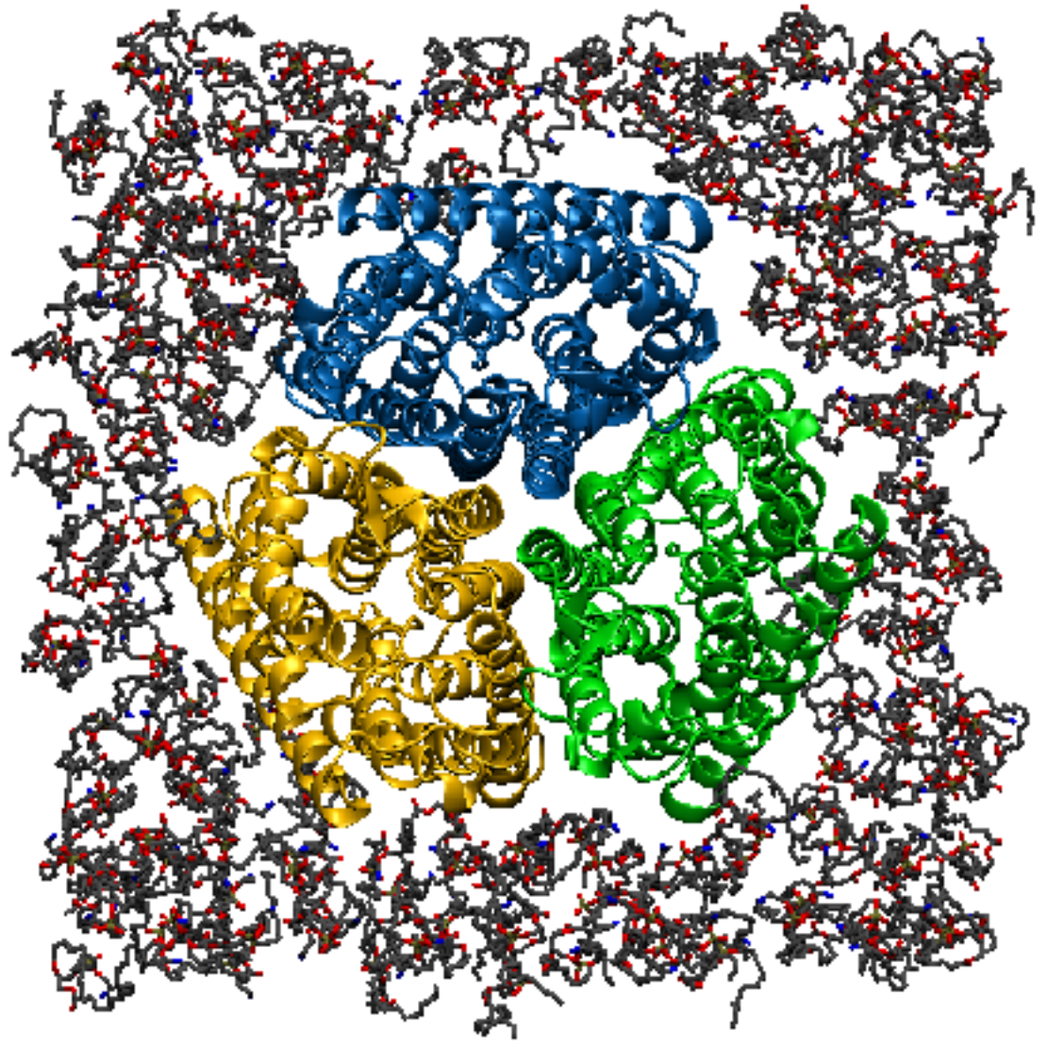
The system looks reasonable overall, but there are a few potential problems

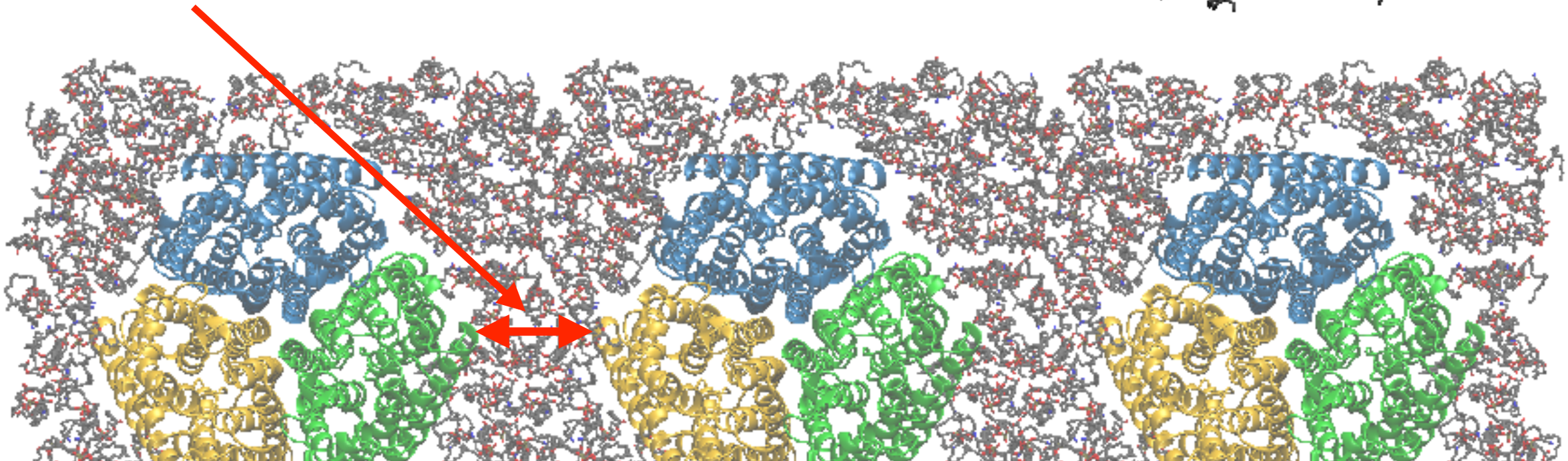
Size of the membrane?

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

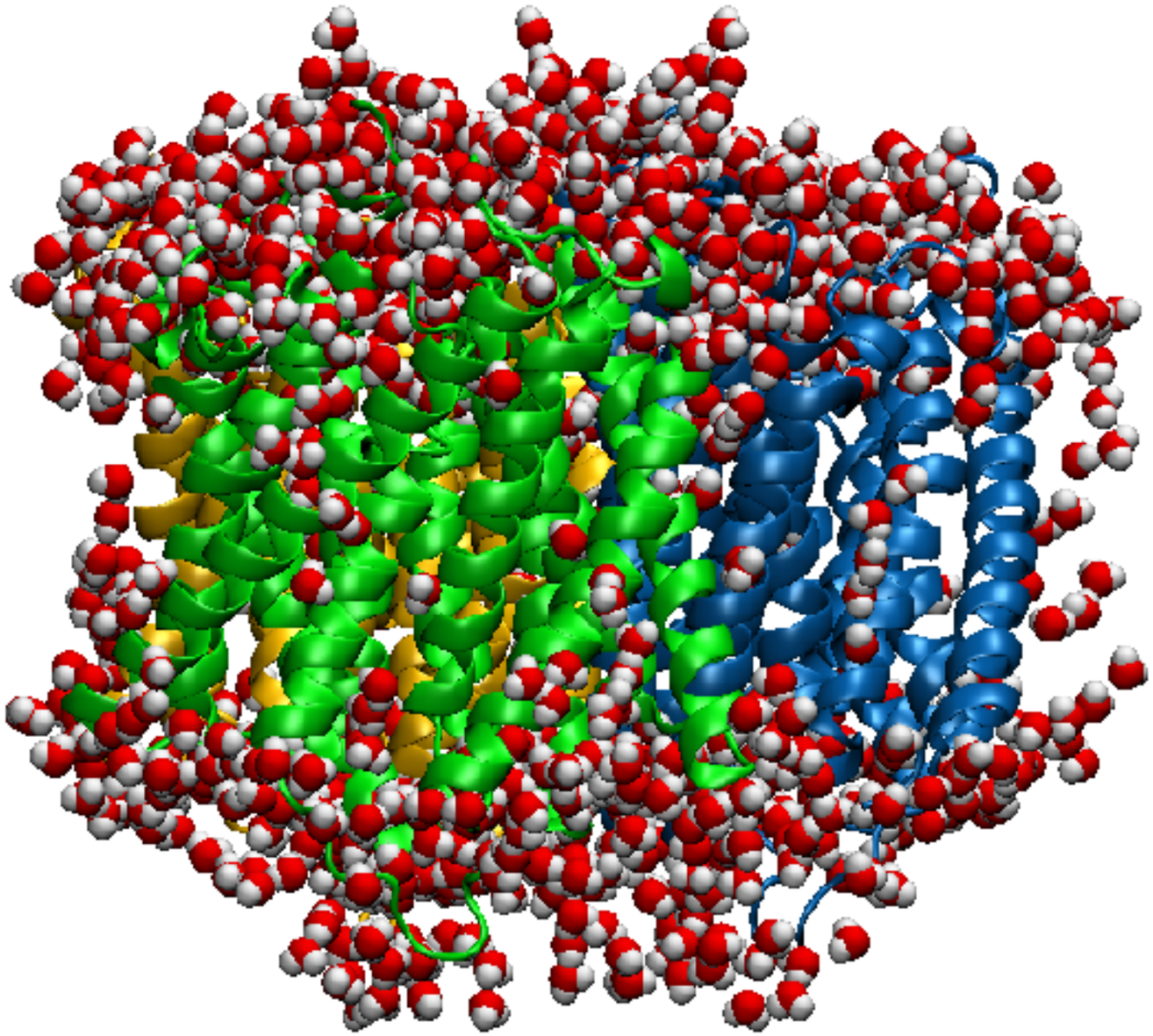
only 20 Å between images - will shrink after equilibration! (need 30 Å at least *after* eq.)





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 ☐ Read all models?

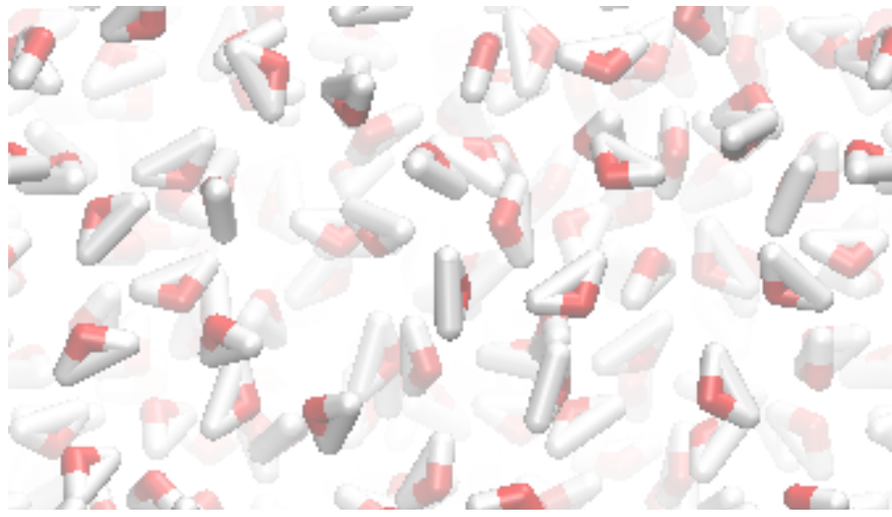
Type	SEGID	PDB ID	Residue ID First	Residue ID Last	Engineered Residues
☑ Protein	PROA	A	3	385	None
☑ Protein	PROB	B	3	385	None
☑ Protein	PROC	C	3	385	None
☐ Hetero	HETA	D			BOG
☑ Hetero	HETB	D			NH4
☑ Hetero	HETC	D			NH3
☐ Hetero	HETD	E			BOG
☑ Hetero	HETE	E			NH4
☑ Hetero	HETF	E			NH3
☐ Hetero	HETG	F			BOG
☑ Hetero	HETH	F			NH4
☑ Hetero	HETI	F			NH3
☐ Hetero	HETJ				DUM
☑ Water	WATA	D			
☑ Water	WATB	E			
☑ Water	WATC	F			

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???



```
RESI TIP3         0.000 ! tip3p water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT     -0.834
ATOM H1   HT      0.417
ATOM H2   HT      0.417
BOND OH2 H1 OH2 H2 H1 H2    ! the last bond is needed for shake
ANGLE H1 OH2 H2             ! required
DONOR H1 OH2
DONOR H2 OH2
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE
```

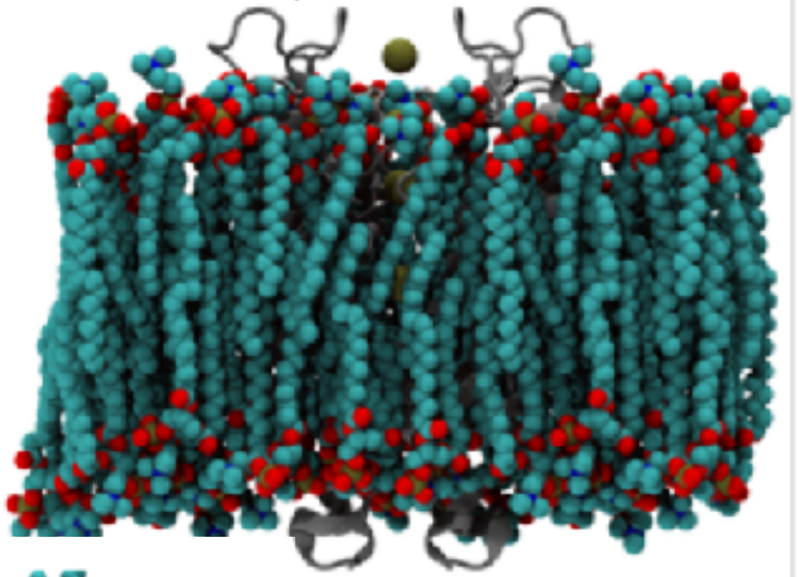
When adding water with VMD Solvate, that extra bond is commented out in the topology file
NAMD doesn’t care about it (it is harmless, just ugly!)

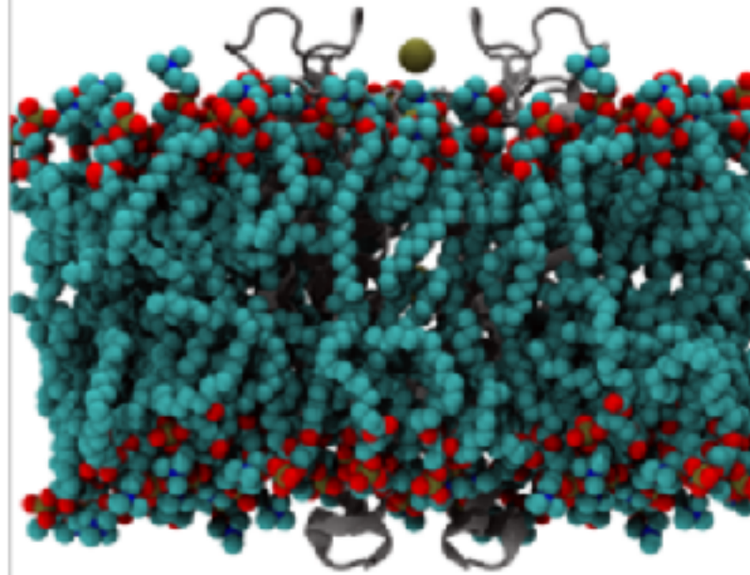
```
Warning: Ignored 19521 bonds with zero force constants.
Warning: Will get H-H distance in rigid H2O from H-O-H angle.
```

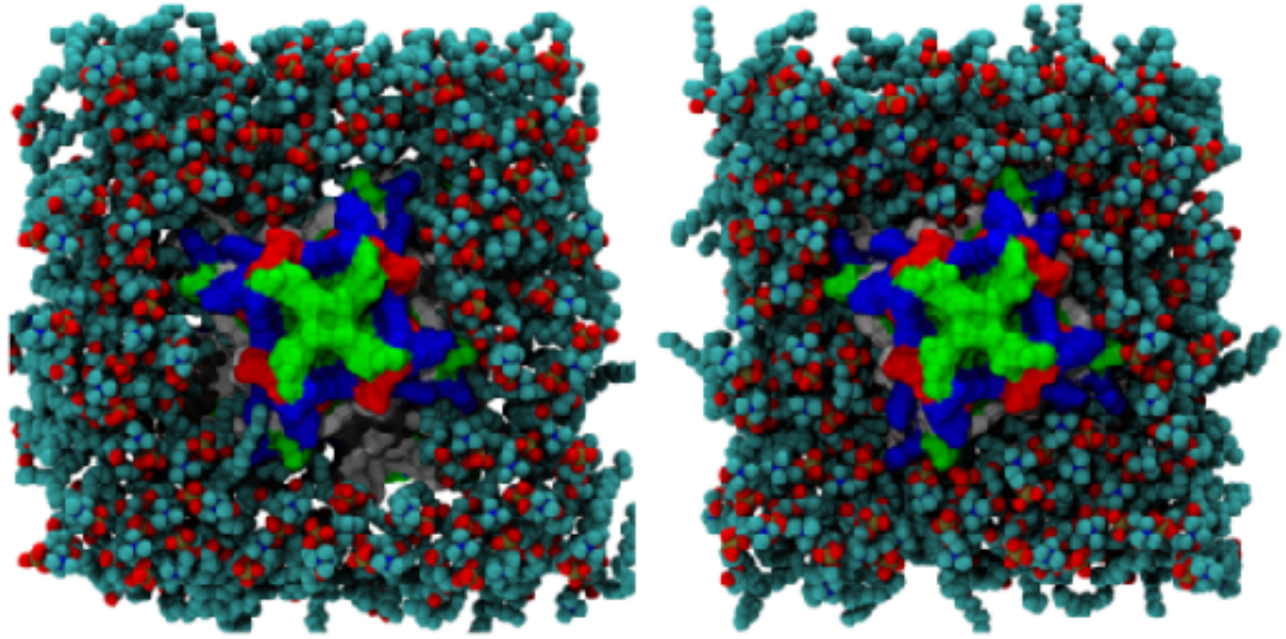
Equilibration stages

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!***