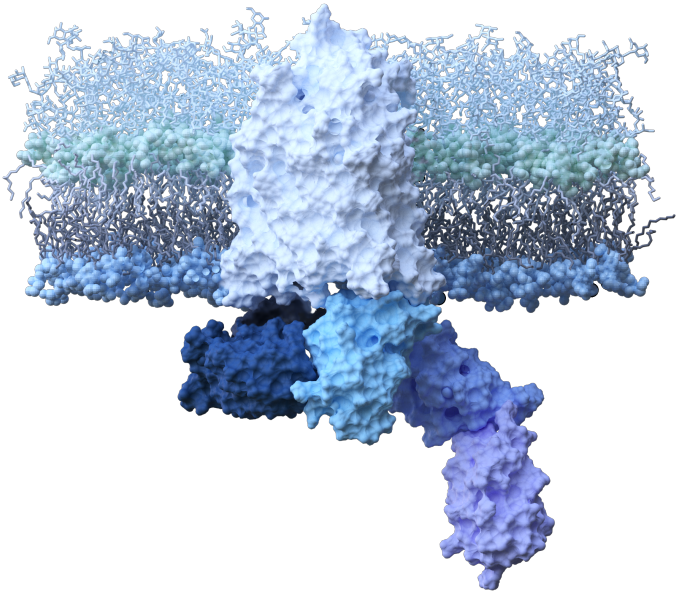
School of Physics
Georgia Institute of Technology

University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group

Membrane Protein Tutorial



Yupeng Li
Jinchan Liu
James C. Gumbart

Updated: Mar. 2023

A current version of this tutorial is available at
http://www.ks.uiuc.edu/Training/Tutorials/
Join the `tutorial-l@ks.uiuc.edu` mailing list for additional help.

Contents

1 Introduction **3**

2 Membrane-protein System Building in CHARMM-GUI **7**
2.1 Read protein coordinates and manipulate structure 7
2.1.1 Load PDB file 7
2.1.2 Manipulate PDB file 8
2.2 Orient the protein 8
2.2.1 Orient and position protein 9
2.2.2 Generate pore water 10
2.3 Determine the system size 10
2.4 Build the components 14
2.5 Assemble the components 16

3 System Equilibration in NAMD **17**
3.1 Melting of lipid tails 17
3.2 Equilibration with protein constrained 21
3.3 Equilibration with backbone constrained 23
3.4 Equilibration of the whole system 25

1 Introduction

Membrane proteins perform multiple functions and are vital to the survival of all organisms (1). It is estimated that the genes coding for membrane proteins make up 20-30% of the genomes of organisms (2). They serve as channels (3, 4), transporters (5–7), receptors (8), enzymes (9) and function in cell signaling (10, 11), translocation of substrates (12–14), energy transduction (15, 16) and cell-cell recognition (17–19). Due to their vital significance, advanced technological methods such as NMR (20), cryo-electron microscopy (cryo-EM) (21) and X-ray crystallography (22) have been developed, in part, to determine structures of membrane proteins. However, these experimental methods only provide a static state of proteins while molecular dynamics (MD) simulations are eligible to probe the dynamic behaviors of them with the high-resolution structure solved (23). Therefore, preparing a membrane-protein system at atomic resolution became a major concern of simulators.

Previous studies have emphasized the importance of building a native membrane: not only are the lipid-protein interactions responsible for regulating or stabilizing the conformation of membrane proteins (24–27), but also the composition of the membrane will influence their structure and function (28, 29). Consequently, one should be especially careful to select the appropriate membrane for a given membrane protein.

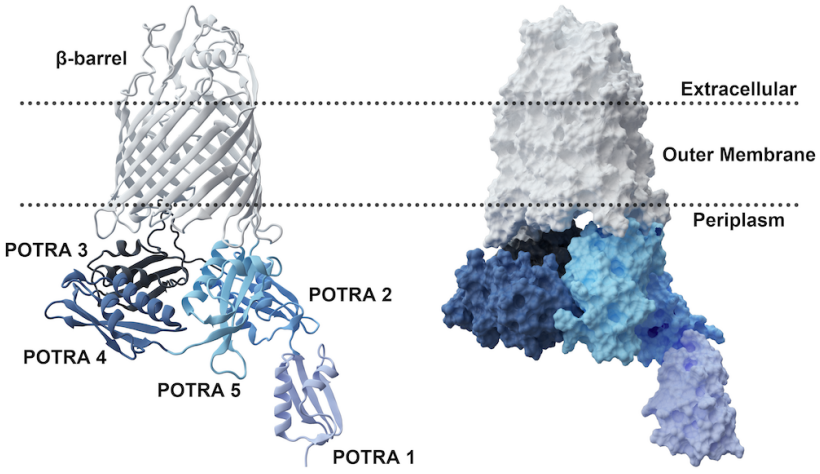


Figure 1: Structure of *E. coli* BamA.

To simplify and automate the building process of a native membrane-protein system for MD simulations, CHARMM-GUI (`http://www.charmm-gui.org`) (19, 30) provides a graphical user interface (GUI) of multiple modules for the biomolecular simulation program CHARMM (31). And Membrane Builder (19) is one of the modules in CHARMM-GUI, which offers users a relatively easy way to build complicated membranes with all types of lipids through a user-specified and automated process, including PDB loading, protein orientation, system size determination, generation for lipids, pore water, bulk water as well as ions, and components assembly (19).

In this tutorial, we will go through the process of preparing a membrane-protein system step-by-step using BamA as an example (Figure 1). BamA is the central component of BAM complex (32–35). It is an outer membrane protein (OMP) of Gram-negative bacteria, which is responsible for the folding and insertion of other OMPs (36). It contains a transmembrane β-barrel of 16 strands along with five periplasmic polypeptide-transport-associated (POTRA) domains (Figure 1). We are going to use the structure from *E. coli* (PDB ID: 5AYW (33)), which includes up to 5 POTRA domains.

The tutorial is divided into two units: we will first build the system in CHARMM-GUI (Figure 2a) and then equilibrate it in NAMD (Figure 2b) (37).

The tutorial assumes some basic knowledge of VMD and NAMD. For the accompanying VMD and NAMD tutorials, please see `http://www.ks.uiuc.edu/Training/Tutorials/`.

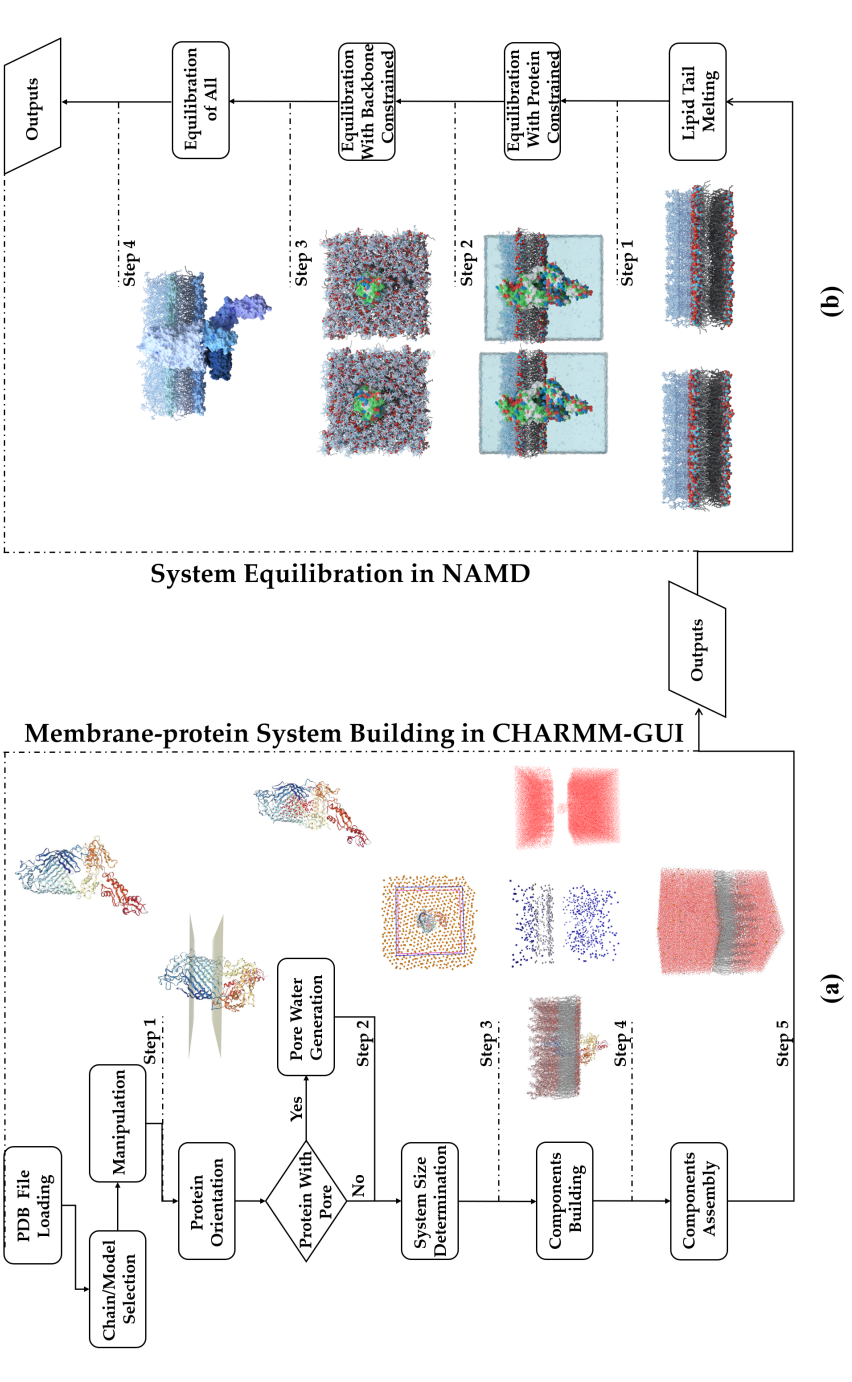


Figure 2: The preparation of a fully built and equilibrated membrane-protein system. (a) The building process of membrane-protein system in CHARMM-GUI. (b) System equilibration in NAMD.

Required programs

The following programs are required for this tutorial:

- In order to access CHARMM-GUI, a web browser such as Chrome, Firefox, etc. is required.
- **VMD:** The latest version of Visual Molecular Dynamics (VMD) is available at `http://www.ks.uiuc.edu/Research/vmd/`.
- **NAMD:** The latest version of NAnoscale Molecular Dynamics (NAMD) to run simulations is available at `http://www.ks.uiuc.edu/Research/namd/`.

Getting started

Files for this tutorial are provided along with example outputs for each step.

2 Membrane-protein System Building in CHARMM-GUI

2.1 Read protein coordinates and manipulate structure

The process of building a membrane-protein system via CHARMM-GUI starts with the loading of protein coordinates, followed by several alternative manipulation options, and finally generating a Protein Structure File (PSF). Users can upload a pre-oriented protein structure or specify a Protein Data Bank (PDB) ID to download PDB files directly from either the Research Collaboratory for Structural Bioinformatics (RCSB) database (38, 39) or the Orientations of Proteins in Membranes (OPM) database (40). Here, we will use BamA as an example.

2.1.1 Load PDB file

1 Open CHARMM-GUI (`http://www.charmm-gui.org/`) in a web browser. Select the menu item `Input Generator` → `Membrane Builder` on the leftmost part of the website.

2 Drag the scroll bar to the middle. Two options will appear on the screen: `Protein/Membrane System` and `Membrane Only System`. Choose the former one.

3 Enter `5ayw` (PDB ID of one conformation of BamABCDE Complex (33)) into the `Download PDB File` blank, meanwhile, selecting `OPM` as the `Download Source`. Then, click on the `Next Step: Select/Model Chain` button in the lower right corner.

Loading the PDB file. Users can either use PDB files from database by selecting `RCSB` or `OPM`, or upload their own pre-oriented PDB file. Options for `PDB Format` need to be chosen when using your own PDB file. Note that PDB files obtained from the OPM database have already been pre-oriented with respect to the membrane normal (Z axis by definition) while those from RCSB database need to be oriented manually by users themselves using VMD or in the subsequent step through CHARMM-GUI.

4 View `Model/Chain Selection Option`. This PDB file contains five proteins. Information such as type, segID, PDB ID, first and last residue ID of chains and engineered residues are listed here as well.

Users are able to view the constitutive segments already present in the PDB file, which mainly include protein chains, substrates, crystallographic water molecules, ions and crystallization detergents. They can also select whatever segments they want to use as well.

5 Here, we will focus on BamA alone. Check on the box of `PROA` only. Then, click on the `Next Step: Manipulate PDB` button in the lower right corner.

Beyond deciding which segments to include in their system, users also have the ability to select a subset of residues of a chain, rename every segment and remove engineered residues. These operations will not be used in this chapter.

2.1.2 Manipulate PDB file

In order to generate a PSF properly, extra manipulation options are required. CHARMM-GUI provides users with diverse options for manipulation to meet multiple demands, including terminal group patching, modeling missing residues, mutation, protonation, disulfide bonds, add lipidation, etc. Here, we will focus on terminal group patching and disulfide bonds manipulation options only. Readers can explore other options on their own.

6 Check the box labeled `Terminal group patching`. Select `NTER` for `First` and `CTER` for `Last`.

7 Check the box labeled `Disulfide bonds`. Set Pair 1 Residue ID to `690` while set Pair 2 Residue ID to `700`. This is an important disulfide bond in *E. coli* BamA.

Reading PDB structural information. Generally, CHARMM-GUI detects structural information automatically, such as missing residues, disulfide bonds and others if indicated by remarks in PDB files (Figure 3). However, depending on the source of the PDB, these remarks may have been written inadequately or even lost altogether. If that occurs, CHARMM-GUI cannot load those kinds of structure information, requiring users to add them manually in this step.

8 Click on the `Next Step: Generate PDB and Orient Molecule` button in the lower right corner. Users can view the loading structure in the next step by clicking on the `view structure` button on the top of the website.

You may notice another option called `Symmetry Operation Options` when you scroll down to the end of the page. This option is only supported when the PDB file contains the information about oligomerization, in which the protein oligomer is composed of two or more associating monomers with different or identical structures (41).

2.2 Orient the protein

After PDB loading and manipulation, the protein needs to be oriented and positioned properly relative to the membrane bilayer. This step consists of two subsections, i.e., orient and position protein and generate pore water.

```
SHEET   1 AA1 3 ASP A  28  GLU A  32  0
SHEET   2 AA1 3 THR A  83  GLU A  90  1  O  VAL A  88   N  GLU A  32
SHEET   3 AA1 3 PHE A  72  ARG A  79 -1  N  ARG A  76   O  GLN A  87
SHEET   1 AA2 3 THR A  93  SER A 100  0
SHEET   2 AA2 3 ARG A 162  VAL A 168  1  O  VAL A 163   N  ALA A  95
SHEET   3 AA2 3 LYS A 152  PRO A 157 -1  N  LYS A 152   O  VAL A 168
SHEET   1 AA3 4 GLU A 176  VAL A 183  0
SHEET   2 AA3 4 GLY A 253  THR A 261  1  O  ILE A 254   N  GLN A 178
SHEET   3 AA3 4 ASN A 239  LEU A 247 -1  N  GLN A 244   O  THR A 257
SHEET   4 AA3 4 SER B 193  LEU B 194  1  O  SER B 193   N  VAL A 245
SHEET   1 AA4 4 GLU A 294  LEU A 295  0
SHEET   2 AA4 4 LYS A 267  ASN A 276 -1  N  LEU A 268   O  GLU A 294
SHEET   3 AA4 4 THR A 334  ASP A 342  1  O  VAL A 335   N  LYS A 267
SHEET   4 AA4 4 ARG A 321  ASN A 329 -1  N  MET A 325   O  ARG A 338
SHEET   1 AA5 3 TYR A 348  GLU A 355  0
SHEET   2 AA5 3 GLN A 411  GLU A 420  1  O  VAL A 412   N  TYR A 348
SHEET   3 AA5 3 PHE A 395  ARG A 404 -1  N  GLU A 396   O  LYS A 419
```

Figure 3: Part of the structural information in the original PDB file.

2.2.1 Orient and position protein

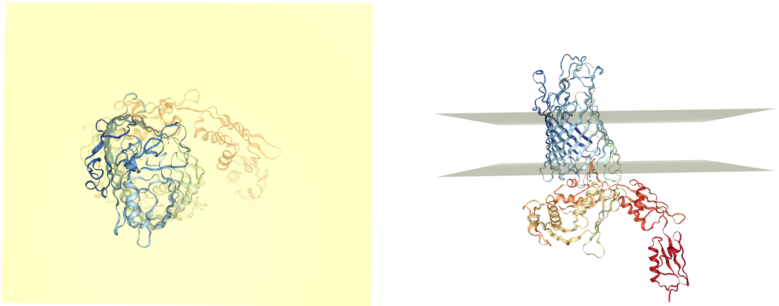
CHARMM-GUI's Membrane Builder defines the Z axis as the membrane normal and $Z = 0$ Å as the center of the membrane bilayer (19, 42). Therefore, to build a system with the proper protein orientation and position, it must be aligned with the Z axis and its hydrophobic region centered on $Z = 0$ Å. Since we use a pre-oriented protein from OPM, orientation and positioning are not necessary here.

1 Locate `Orientation Options`. Four options are provided here. Each option is labeled with the situation it is intended for.

Protein orientation. In CHARMM-GUI, the protein can be placed appropriately in the membrane by reorienting it via the alignment of its principal axis or a vector between two residues with the Z axis in `Orientation Options`, and repositioning it by means of the rotation with respect to the X or Y axis, or translation along the Z axis in `Positioning Options`. Users can also just utilize the original orientation and position information contained in the PDB file.

2 Subsequently, select `Use PDB Orientation`. Users can see the orientation file (Figure 4) in the next step by clicking on the `view structure` button on the top of the website.

Usually, proper orientation information is not available for PDB files from the RCSB database, such that proteins most likely need to be reoriented and repositioned in this step. Users can select `Use PDB Orientation` if they use PDB files from OPM database. Users can also use move and rotate commands in VMD to write a pre-oriented PDB file and then upload it to CHARMM-GUI.



(a) (b)

Figure 4: Protein orientation. The yellow sheets are the XY-planes of membrane. (a) Top view. (b) Side view.

2.2.2 Generate pore water

In general, proteins with pores, such as ion channels, transporters and porins, are able to accommodate water molecules inside their internal cavity. CHARMM-GUI provides a general approach for pore water generation.

3 Locate the `Area Calculation Options`.

4 Click on the box of `Generate Pore Water and Measure Pore Size`.

5 Select `Using protein geometry`.

Pore water generation. During the pore water generation process, CHARMM-GUI solvates the transmembrane region of protein with a water box and runs high temperature dynamics with the protein fixed and water restrained in the transmembrane region. Water molecules inside the pore will remain while water molecules outside the pore will evaporate (19) (Figure 5). Water staying close to the protein exterior due to strong interactions, can be removed by a refinement step in Section `1.4`.

6 Click on the `Next Step: Calculate Cross-Sectional Area` button in the lower right corner.

Note that the cross-sectional area of the protein will be calculated in this subsection to help determine the system size in the next step.

2.3 Determine the system size

According to the cross-sectional area of the protein calculated in the previous step and lipid surface areas from experiments, the system size in the XY-plane

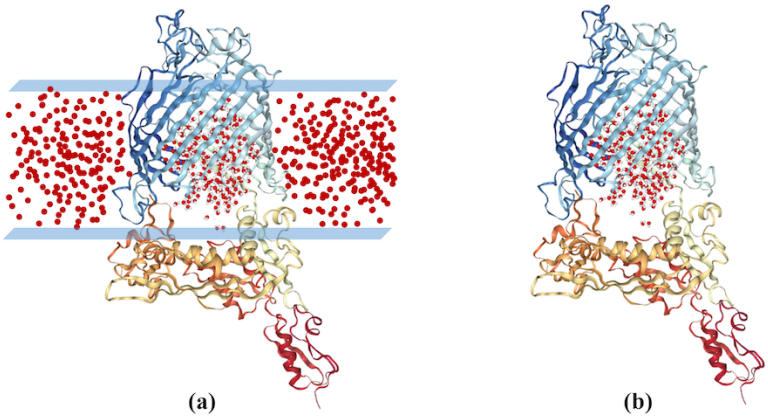


Figure 5: Pore water generation. (a) Solvating the transmembrane region with water. (b) Pore water remains after high temperature dynamics.

and along the Z axis can be determined by multiple user-specified parameters in System Size Determination Options, including lipid types, system shape, water thickness along the Z axis on the top and bottom of the membrane, and numbers or ratios of lipid components. Since we are building the membrane for BamA in *E. coli*, we will use an *E. coli* membrane. *E. coli* is a Gram-negative bacteria enveloped by two membranes, an inner membrane (IM) and outer membrane (OM). BamA resides in the OM. In Gram-negative bacteria, there is a special outer membrane component, lipopolysaccharide (LPS), which consists of lipid A and a polysaccharide, residing exclusively in the upper leaflet. The lower leaflet of the OM is a mixture of phospholipids. Here, we will use LPS for the upper leaflet while using PVCL2, PMPE, PMPG, PVPE and PVPG for the lower leaflet, with a ratio of 2 : 8 : 1 : 8 : 2 (43, 44).

1 Locate `System Size Determination Options`.

2 Select the `Heterogeneous Lipid option`.

Presently, the `Homogeneous Lipid` option is not supported, but users can select one type of lipid when using the `Heterogeneous Lipid` option to generate a homogeneous lipid bilayer.

3 Select `Rectangular` as the `Box Type`.

4 In the `Length of Z based on` option, select `Water thickness`. Change its initial parameter from 22.5 to 30 Å.

The scale of the entire system along the Z axis is determined by the height of the protein in Z and the thickness of the added water slabs (Figure 6). In

general, the default water thickness of 22.5 Å is sufficient. For a membrane-only system, users can select the `Hydration number` (number of water molecules per one lipid molecule) option to define the total number of water molecules (19).

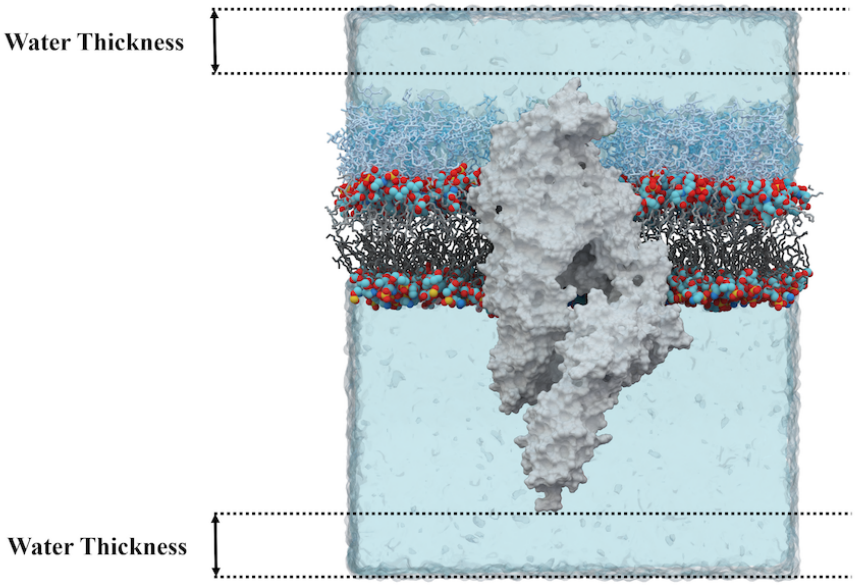


Figure 6: Water thickness of a membrane-protein system.

5 In the `Length of XY based on` option, select `Ratios of lipid components`.

Membrane Builder gives users two options to determine the system size in the XY-plane: `Ratio of lipid components`, which corresponds to the `Length of X and Y`, and `Numbers of lipid components`, which corresponds to the `XY dimension ratio`.

6 Go to `Lipid Type` column. In `CL (cardiolipin) Lipids`, set `PVCL2`'s `Lowerleaflet Ratio` as `2` and `Upperleaflet Ratio` as `0`. In `Bacterial Lipids`, set the `Lowerleaflet Ratio` of `PMPE`, `PMPG`, `PVPE` and `PVPG` as `8`, `1`, `8` and `2`, respectively, while keeping the `Upperleaflet Ratio` of all of them as `0`.

7 Locate `LPS (lipopolysaccharides)`. Set the `Upperleaflet Ratio` as `1` and the `Lowerleaflet Ratio` as `0`.

8 Click on `LPSA` button. In the pop-up, set all the parameters to match those shown in Figure 7. Then click on the `Next Step: Update LPS` button in the lower right corner.

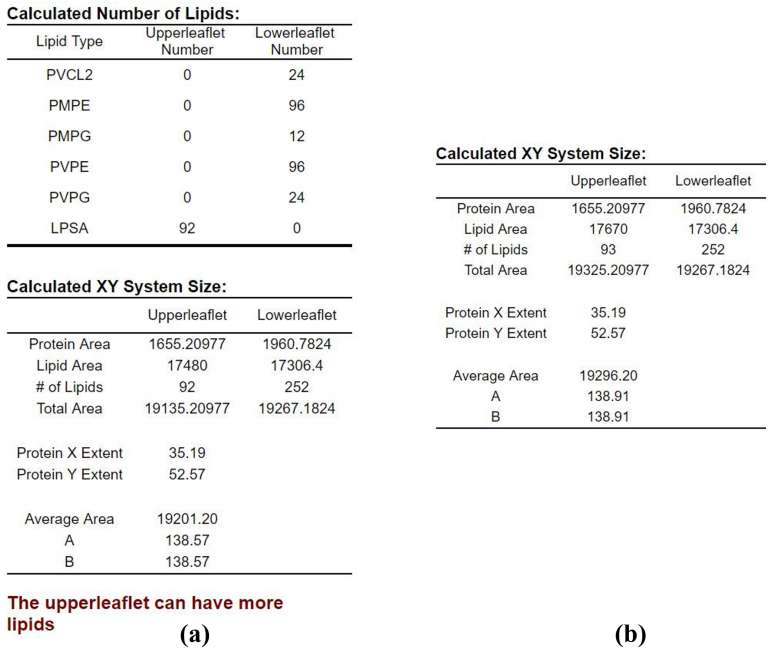
Figure 7: LPS type and core sequence.

Ideally, the types and numbers of lipids are chosen to match the native membrane. Users should search the literature to determine which species the protein is from as well as the composition of its membrane in advance.

9 Returning to the `Length of XY based on` option, enter 135 in the `Length of X and Y` blank as an initial guess. Then click on the `Show the system info` button and you should see the information shown in Figure 8a.

This situation is caused by the difference in areas between the upper leaflet and the lower leaflet of membrane. Generally, in order to solve it, we will use the `Ratio of lipid components` option first to determine the numbers of every membrane component under a certain initial guess. Then, use the `Numbers of`

`lipid components` option to fine tune the number of lipids according to the feedback.



Calculated Number of Lipids:

Lipid Type	Upperleaflet Number	Lowerleaflet Number
PVCL2	0	24
PMPE	0	96
PMPG	0	12
PVPE	0	96
PVPG	0	24
LPSA	92	0

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	1655.20977	1960.7824
Lipid Area	17480	17306.4
# of Lipids	92	252
Total Area	19135.20977	19267.1824
Protein X Extent	35.19	
Protein Y Extent	52.57	
Average Area	19201.20	
A	138.57	
B	138.57	

The upperleaflet can have more lipids

(a)

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	1655.20977	1960.7824
Lipid Area	17670	17306.4
# of Lipids	93	252
Total Area	19325.20977	19267.1824
Protein X Extent	35.19	
Protein Y Extent	52.57	
Average Area	19296.20	
A	138.91	
B	138.91	

(b)

Figure 8: Feedback information for determining the membrane size. (a) Only using ratio to determine the membrane size may lead to one leaflet having too few lipids. (b) Adjusting the lipid numbers slightly will eliminate this problem.

10 Select the `Numbers of lipid components` option. Change the upperleaflet lipid number of LPS from `92` to `93`. Click on `Show the system info` button and you will see the information in Figure 8b.

11 Click on the `Next Step: Determine the System Size` button.

2.4 Build the components

On the basis of the system size, the generation of individual components for the system, including the membrane, bulk water, and counter ions will be completed in this step.

1 Locate the `System Building Options`. Then select `Replacement method`.

Replacement method. Replacement method (Figure 9) distributes lipid-like pseudo atoms around the protein first, and then replaces them with lipid molecules selected randomly from a lipid molecule library, which contains 2,000 different conformations of lipids from MD simulations of pure bilayers (19). Note that `Insertion method` is no longer supported in CHARMM-GUI.

2 Move on to `Component Building Options`. Check the `Include Ions` box.

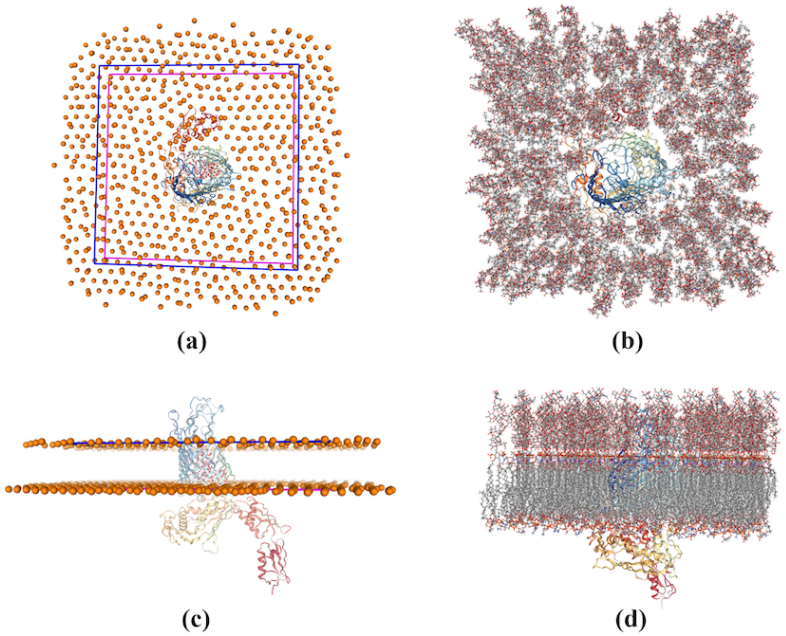


Figure 9: The replacement method uses lipid-like pseudo atoms to build lipids around the protein. (a,c) Lipid-like pseudo atoms around the protein. (b,d) Lipids around the protein. (a,b) Top view. (c,d) Side view.

3 Use `KCl` as neutralizing species with a concentration of `0.15 M`. Choose $\mathbf{Mg^{2+}}$ as the counter ions for both lipid A and core. Keep the Ion Placing Method as Distance.

Neutralization. In order to neutralize the system, Membrane Builder creates an appropriate number of ions based on the user-specified ion concentration and type. The initial configuration of ions is then determined through Monte Carlo simulations using a simplified model, i.e., van der Waals and scaled Coulombic interactions (19).

4 Go to `Pore Water Options`. Inappropriately placed water molecules can be removed here. Usually, there are no extra water molecules that need to be removed and this step can be skipped.

Refining pore water. water generated in `1.2.2` can be refined in this step, to ensure that no water molecules are left outside of the protein in the membrane hydrophobic core region. Users can download the structure file to verify whether those water molecules are removed and select the residue numbers of water molecules needing to be removed on the website.

5 Click on the `Next Step: Build Components` button in the lower right corner. The lipid bilayer will be generated first in this step.

6 To generate water molecules and ions, click on the `Next Step: Assemble Components` button in the lower right corner.

2.5 Assemble the components

Components generated in the previous steps will be assembled in this step. Users should check the system carefully and verify whether the system is built as intended. If not, go back to previous steps and re-generate the whole system.

1 Check carefully to ensure the system is built as intended. If no problem exists, then click on the `Next Step: Assemble Components` button in the lower right corner to complete the assembly. Otherwise, go back to rebuild the system.

2 Download all the output files by click on `download.tgz`.

So for now, the entire system containing protein, lipid bilayer, bulk water, and ions is generated completely through user-specified parameters and options in CHARMM-GUI. Users can load the system into VMD to see it in detail and begin the equilibration process with NAMD next.

3 System Equilibration in NAMD

Now, we have finished building the system, including protein (BamA in this case), membrane with LPS in the upper leaflet and phospholipids in the lower leaflet, water molecules, and ions in CHARMM-GUI. In this step, we are going to equilibrate the system using NAMD.

In general, we equilibrate this multiphase system step by step to speed up the equilibration process. The entire equilibration involves several minimization-equilibration cycles, fixing parts of molecules and relaxing the remaining components gradually. Releasing the whole system at once results in a rapid change of the system size as well as unfavorable conformations, typically causing the simulation to fail. You may see, for example, the following error in the log file:

```
FATAL ERROR: Periodic cell has become too small for original
patch grid!
```

Though readers can solve this problem by restarting the simulation, it will take much more time to fully equilibrate the system compared to doing it in multiple steps.

In this section, we equilibrate the system in four steps: (1) melting lipid tails, (2) relaxing the membrane and water with the protein constrained, (3) relaxing side chains with protein backbone constrained, and (4) relaxing the whole system.

3.1 Melting of lipid tails

In this step, the complete membrane-protein system excluding lipid tails will be fixed for the first simulation. Because the membrane is built in a nearly crystalline state, the aliphatic tails must be "melted" to achieve a more fluid-like state.

1 Change your current directory to `Chapter/Equilibration/Step1`.

2 Open the script file `getcnst_S1.tcl` in a text editor. In this file, we are going to set the beta value of lipid tails to 0, while setting the beta value of all others to 1, thus telling NAMD which atoms to restrain.

NAMD constraints. In the configuration file, a series of parameters related to constraints are given. In particular, the `conskcol` tells NAMD which column to use from the `conskfile` for the force constants. Atoms with a non-zero value in this column, which is beta in the example here, will be constrained during the simulations according to the potential $U(\vec{x}) = k\,|\vec{x} - \vec{x_0}|^{\mathrm{consexp}}$ where `consexp` defaults to 2; other atoms with 0 in this column are not constrained and, thus, can equilibrate.

Here, the atom selection relax1 stands for the head groups of lipid A while relax2 represents the head groups of phospholipid, respectively. Readers should specify the selections for head groups of lipid A and phospholipid on their own.

3 Close the text editor. Run the script `getcnst_S1.tcl` and produce a log file by typing the following commands in the terminal:

```
vmd -dispdev text -e getcnst_S1.tcl > getcnst_S1.log
```

`EcBamA_S1.cnst`,which is formatted as a PDB file, is generated. You can load `EcBamA.psf` file and add the cnst file you generated just now in VMD. Use the default line representation and color it by beta to confirm that you have set the beta values correctly.

4 Open the configuration file `EcBamA_S1.conf` in a text editor and go to the `Force Field Parameter File` section. Multiple parameter files are listed here. These files will be invoked as force field parameters when simulations are run.

```
paraTypeCharmm on
parameters ../../ParamFiles/par_all36m_prot.prm
parameters ../../ParamFiles/par_all36_na.prm
parameters ../../ParamFiles/par_all36_carb.prm
parameters ../../ParamFiles/par_all36_lipid.prm
parameters ../../ParamFiles/par_all36_cgenff.prm
parameters ../../ParamFiles/toppar_all36_lipid_bacterial.str
parameters ../../ParamFiles/toppar_water_ions_namd.str
parameters ../../ParamFiles/toppar_all36_lipid_lps.str
parameters ../../ParamFiles/toppar_all36_carb_imlab.str
```

Parameters files. Parameter files contain all of the numerical constants correlated with the determination of forces and energies. The `toppar_water_ions_namd.str` file includes optimized corrections for the interactions between ions and carbonyl oxygen atoms (45).

5 Go to the `Periodic Boundary Conditions` section. The size and center of the system need to be input here.

```
cellBasisVector1 145.0 0.0 0.0
cellBasisVector2 0.0 145.3 0.0
cellBasisVector3 0.0 0.0 193.0
cellOrigin 0.0 0.1 -17.4
wrapAll on
wrapNearest on
```

Note that `cellBasisVector` stands for the system size vectors along the X, Y and Z directions while `cellOrigin` represents the center of the system. Information about the system size and center can be obtained from VMD.

6 Open VMD and type the following commands in the `TK Console` to load the structure:

```
mol new ../../MembBuilding/EcBamA_memb.psf
mol addfile ../../MembBuilding/EcBamA_memb.pdb
```

7 Then type the commands below to get the center of the entire system:

```
set all [atomselect top all]
measure center $all
```

You can see the result in `TK Console` window:

```
-0.044672466814517975 0.10125688463449478 -17.42184066772461
```

The `cellOrigin` parameter specified in configuration file should be set to these values. Rounding to the tenth of an Å is sufficient.

8 In order to obtain the size of the system, type:

```
set wat [atomselect top water]
set min [lindex [measure minmax $wat] 0]
set max [lindex [measure minmax $wat] 1]
set length [vecsub $max $min]
```

Now, a list should appear on the screen:

```
144.0189971923828 144.28700256347656 192.01699829101563
```

Though VMD returns very precise values, readers can round them to a tenth of an Å.

Getting system size. Notice that the selection we have used here is `water` instead of `all`. This is because lipid tails can hang over the boundary, but only if the center of mass of a lipid crosses the periodic boundary will it be wrapped. Readers can avoid having an artificially large unit cell by choosing a selection without lipids, such as `water`

.

9 Exit VMD and scroll down to the `Constant Pressure Control` section. Notice the following line:

```
langevinPiston off
```

In the lipid tails melting step, this option is turned off on account of the fact that most of the system is fixed. If turned on at this step, the simulation may fail with the following error in the log file:

```
ERROR: Constraint failure in RATTLE algorithm for atom ID!
ERROR: Constraint failure; simulation has become unstable.
```

10 Now move to the `Constraints` section. It should read as follows:

```
constraints on
consref ../../MembBuilding/EcBamA.pdb
conskfile EcBamA_S1.cnst
conskcol B
margin 3
```

When running a simulation, constraints work in terms of the `conskcol B` for beta-coupling here. It could also be `X`, `Y`, `Z` or `O` (occupancy). Values contained in `conskfile` determine which atoms should be constrained. A detailed description on the role beta-coupling plays was introduced previously.

11 Go to the last section, i.e., `EXECUTION SCRIPT`. It reads:

```
minimize 2000
reinitvels 310
run 500000
```

This means NAMD will run 2000 steps of minimization first and then reset the velocities according to the chosen system temperature of 310 K, followed by a 500000-step equilibration. Each step takes 2 fs, making the total 1 ns for equilibration.

Minimization. Because of the possibility of tensile or compressive deformations in bonds and angles or interactions between components possessing high energies, each step starts with minimization, which involves searching the energy landscape of the atomic positions using the MD force field to achieve a local minimum. Equilibrating directly without minimization may lead to the unnecessary release of extremely high energies stored in improper structures, probably causing drastic motion, which further gives rise to the simulation behavior that is incongruous with the real behavior of the system in solution.

12 Close the text editor and run your simulation on a supercomputer if possible. Or type the following command in the terminal to run the simulation on your own computer or laptop:

```
namd2 EcBamA_S1.conf > EcBamA_S1.log &
```

We don't recommend you run this on your own machine unless you have a fast GPU and are using the GPU-accelerated version of NAMD. If you do not have computing resources available, example output is given.

13 Once the simulation is done, open VMD and load the trajectory file `EcBamA_S1.dcd` on top of the psf file `EcBamA.psf`.

14 Play the trajectory to see the changes of the system during the first step of equilibration. Lipid tails have become more disordered as desired (Figure 10).

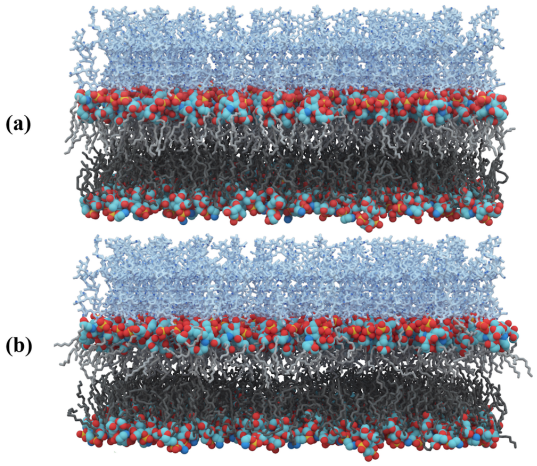


Figure 10: Lipid tails melting process (protein not shown). (a) The initial system. (b) System with tails melted.

3.2 Equilibration with protein constrained

Starting from the result of the last step, the whole system will be further equilibrated with only the protein constrained.

1 Change your directory to `Chapter/Equilibration/Step2`.

2 Open the script file `getcnst_S2.tcl` in a text editor. Commands shown below set the beta value of the protein to 1 while setting those of all others to 0.

```
set all [atomselect top all]
set protein [atomselect top protein]
$all set beta 0
$protein set beta 1
```

3 Close the text editor. Type the following command in the terminal window to run the script:

```
vmd -dispdev text -e getcnst_S2.tcl > getcnst_S2.log
```

`EcBamA_S2.cnst`,which is formatted as a PDB file, is generated now. You can load `EcBamA.psf` file and add the cnst file you generated just now in VMD. Use the default line representation and color it by beta to confirm that you've set the beta values correctly.

4 Open the configuration file `EcBamA_S2.conf`. You may notice some parameters are different from the configuration file in the previous step. Go to the `Input` section:

```
binCoordinates ../Step1/$name_S1.restart.coor
binVelocities ../Step1/$name_S1.restart.vel
extendedSystem ../Step1/$name_S1.restart.xsc
```

Since lipid tails were melted in the previous step, we want to continue the next simulation on the basis of that result. The commands listed above are used to restart the simulation from where it ended in the last step. More specifically, `binCoordinates`, `binVelocities` and `extendedSystem` invoke files containing position data, velocity data, and the periodic cell, respectively.

```
firsttimestep [get_first_ts ../Step1/$name_S1.restart.xsc]
```

`Firsttimestep` is the number of the first step when a simulation is running, normally used when the simulation is a continuation of another one. For most purposes, it doesn't affect the dynamics. Exceptions include steered MD, among other time-dependent forces. Here, we will use a customized function to get it from the xsc file of the last step. Go to the `Get Firsttimestep` section. The function get_first_ts is defined by the following commands:

```
proc get_first_ts {xscfile} {
  set fd [open $xscfile r]
  gets $fd; gets $fd
  gets $fd line
  set ts [lindex $line 0]
  close $fd
  return $ts
}
```

You can also open the xsc file in a text editor to see how these commands work.

5 Go to `Periodic Boundary Conditions` section:

```
wrapAll on
wrapNearest on
```

Because the xsc file contains the periodic cell parameters, there is no need to reset `cellBasisVector` and `cellOrigin`.

6 Scroll down to the `Constant Pressure Control` section. Notice the following commands:

```
useFlexibleCell yes
langevinPiston on
```

Since most components of the whole system are able to move in this step, `useFlexibleCell` should be used here. `LangevinPiston` is activated as well in this step to control the pressure of the whole system.

useFlexibleCell. The three orthogonal dimensions of the system are allowed to fluctuate independently when `useFlexibleCell` is enabled. Generally, this option is used for systems with a membrane, while it is not typically suitable for a protein solvated in a water box.

7 Go to the last section, i.e., `EXECUTION SCRIPT`. The number of steps is set to 5000000 (10 ns) rather than the original value 500000 (1 ns) to give the membrane more time to relax.

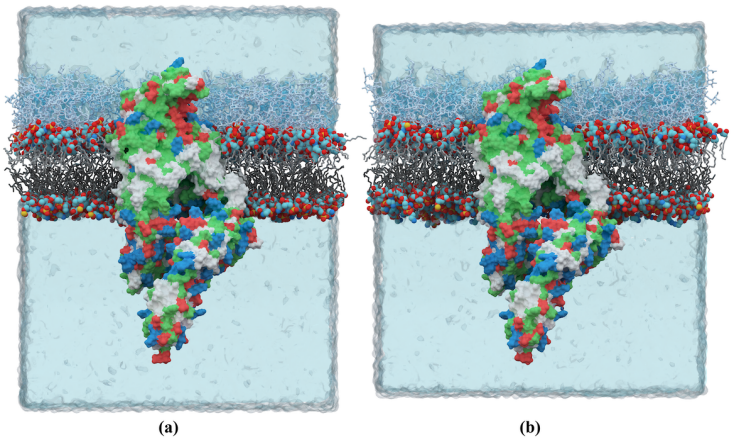


Figure 11: Equilibration with protein constrained. (a) (b) Side view of the system before (a) and after (b). The system is compressed along the Z axis as water packs around the membrane and protein.

8 Close the text editor. Run your simulation on a supercomputer if possible (highly recommended). Or type the following command in the terminal window:

```
namd2 EcBamA_S2.conf > EcBamA_S2.log &
```

9 Once the simulation is finished, load the trajectory file `EcBamA_S2.dcd` on top of the psf file in VMD. You will find the entire system seems to experience a compression along the Z axis (Figure 11), owing to the packing of water around the protein and membrane under constant pressure.

3.3 Equilibration with backbone constrained

After system relaxation with the protein constrained, we have obtained a membrane-protein system in which lipids are well packed around the protein, while water molecules have not entered into hydrophobic regions. We will further release the side chains of the protein in this step.

1 Change your directory to `Chapter/Equilibration/Step3`.

2 Open the script `getcnst_S3.tcl` in a text editor. Commands shown below set the beta values of the protein backbone to 1 while setting those of all others to 0.

```
set all [atomselect top all]
set backbone [atomselect top "protein and backbone"]
$all set beta 0
$backbone set beta 1
```

3 Close the text editor. Type the following commands in a terminal window to run the script:

```
vmd -dispdev text -e getcnst_S3.tcl > getcnst_S3.log
```

The `conskfile EcBamA_S3.cnst` is generated. You can load `EcBamA.psf` file and add the cnst file you generated just now in VMD. Use the default line representation and color it by beta to confirm that you've set the beta values correctly.

4 Run your simulation on a supercomputer if possible (highly recommended). Or type the following command in the terminal window:

```
namd2 EcBamA_S3.conf > EcBamA_S3.log &
```

5 After the simulation is done, load the trajectory file `EcBamA_S3.dcd` on top of the psf file in VMD. Lipids are well packed around the protein now (see Figure 12) because of the equilibration of the interactions between lipids and side chains of protein in this step.

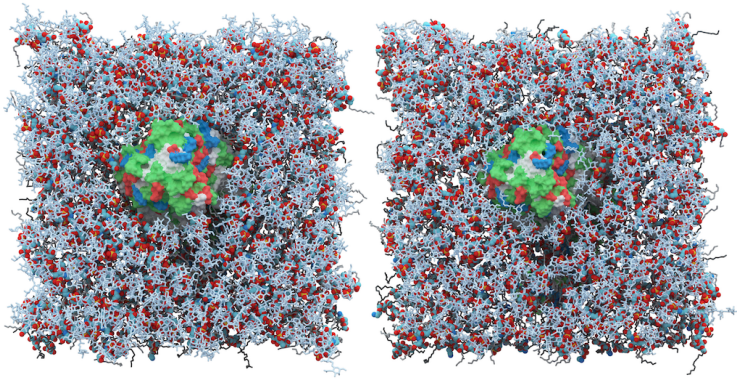


Figure 12: Equilibration with backbone constrained. (a, b) Top view of the system before (a) and after (b). Lipids are well packed around the protein after this step.

3.4 Equilibration of the whole system

In the previous step, the side chains of the protein were released. We will proceed to equilibrate the system without any constraints now.

1 Change your directory to `Chapter/Equilibration/Step4`.

2 Since no constraints are needed in this step, the script to set beta values is unnecessary.

3 Open the configuration file. Go to the last section. Notice that the minimization step is eliminated.

4 Close the text editor and run your simulation on a supercomputer if possible (highly recommended). Or type the following command in the terminal window:

```
namd2 EcBamA_S4.conf > EcBamA_S4.log &
```

After the simulation is finished, the membrane-protein system should be fairly well equilibrated. However, sometimes a larger system might need more time to equilibrate. Readers may wish to utilize hydrogen-mass repartitioning (HMR) (46, 47) to accelerate the simulation. You can view the changes of the system during the simulations through trajectory files.

Acknowledgements

This work was supported by the National Institutes of Health (R01-GM123169). Computational resources were provided through the Extreme Science and Engineering Discovery Environment (XSEDE; TG-MCB130173), which is supported by NSF Grant ACI-1548562.

References

[1] Almén, M. S., K. J. Nordström, R. Fredriksson, and H. B. Schiöth. 2009. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* 7:1–14.

[2] Krogh, A., B. Larsson, G. Von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567–580.

[3] Hidalgo, P. and R. MacKinnon. 1995. Revealing the architecture of a K+ channel pore through mutant cycles with a peptide inhibitor. *Science.* 268:307–310.

[4] Agre, P., L. S. King, M. Yasui, W. B. Guggino, O. P. Ottersen, Y. Fujiyoshi, A. Engel, and S. Nielsen. 2002. Aquaporin water channels–from atomic structure to clinical medicine. *J. Physiol.* 542:3–16.

[5] Rahman, K. S., G. Cui, S. C. Harvey, and N. A. McCarty. 2013. Modeling the conformational changes underlying channel opening in CFTR. *PLoS One.* 8:e74574.

[6] Fairman, J. W., N. Noinaj, and S. K. Buchanan. 2011. The structural biology of β-barrel membrane proteins: a summary of recent reports. *Curr. Opin. Struct. Biol.* 21:523–531.

[7] Gadsby, D. C. 2007. Ion pumps made crystal clear. *Nature.* 450:957–959.

[8] Shukla, A. K., A. Manglik, A. C. Kruse, K. Xiao, R. I. Reis, W.-C. Tseng, D. P. Staus, D. Hilger, S. Uysal, L.-Y. Huang, et al. 2013. Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature.* 497:137–141.

[9] Cojocaru, V., K. Balali-Mood, M. S. Sansom, and R. C. Wade. 2011. Structure and dynamics of the membrane-bound cytochrome P450 2C9. *PLoS Comput. Biol.* 7:e1002152.

[10] Jordan, J. D., E. M. Landau, and R. Iyengar. 2000. Signaling networks: the origins of cellular multitasking. *Cell.* 103:193–200.

[11] Hunter, T. 2000. Signaling–2000 and beyond. *Cell.* 100:113–127.

[12] Fu, D., A. Libson, L. J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science.* 290:481–486.

[13] Khademi, S., J. O'Connell, J. Remis, Y. Robles-Colmenares, L. J. Miercke, and R. M. Stroud. 2004. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science.* 305:1587–1594.

[14] Yellen, G. 2002. The voltage-gated potassium channels and their relatives. *Nature.* 419:35–42.

[15] Dong, J., G. Yang, and H. S. Mchaourab. 2005. Structural basis of energy transduction in the transport cycle of MsbA. *Science.* 308:1023–1028.

[16] Elston, T., H. Wang, and G. Oster. 1998. Energy transduction in ATP synthase. *Nature.* 391:510–513.

[17] Medley, Q. G., N. Kedersha, S. O'Brien, Q. Tian, S. F. Schlossman, M. Streuli, and P. Anderson. 1996. Characterization of GMP-17, a granule membrane protein that moves to the plasma membrane of natural killer cells following target cell recognition. *Proc. Natl. Acad. Sci. USA.* 93:685–689.

[18] Brooks, J. M., S. P. Lee, A. M. Leese, W. A. Thomas, M. Rowe, and A. B. Rickinson. 2009. Cyclical expression of EBV latent membrane protein 1 in EBV-transformed B cells underpins heterogeneity of epitope presentation and CD8+ T cell recognition. *J. Immunol.* 182:1919–1928.

[19] Jo, S., T. Kim, and W. Im. 2007. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One.* 2:e880.

[20] Hong, M., Y. Zhang, and F. Hu. 2012. Membrane protein structure and dynamics from NMR spectroscopy. *Annu. Rev. Phys. Chem.* 63:1–24.

[21] Earl, L. A., V. Falconieri, J. L. Milne, and S. Subramaniam. 2017. Cryo-EM: beyond the microscope. *Curr. Opin. Struct. Biol.* 46:71–78.

[22] Garavito, R. M., D. Picot, and P. J. Loll. 1996. Strategies for crystallizing membrane proteins. *J. Bioenerg. Biomembr.* 28:13–27.

[23] Gumbart, J., Y. Wang, A. Aksimentiev, E. Tajkhorshid, and K. Schulten. 2005. Molecular dynamics simulations of proteins in lipid bilayers. *Curr. Opin. Struct. Biol.* 15:423–431.

[24] Lee, A. G. 2011. Biological membranes: the importance of molecular detail. *Trends Biochem. Sci.* 36:493–500.

[25] Landreh, M., E. G. Marklund, P. Uzdavinys, M. T. Degiacomi, M. Coincon, J. Gault, K. Gupta, I. Liko, J. L. Benesch, D. Drew, et al. 2017. Integrating mass spectrometry with MD simulations reveals the role of lipids in Na+/H+ antiporters. *Nat. Commun.* 8:1–9.

[26] Dawaliby, R., C. Trubbia, C. Delporte, M. Masureel, P. Van Antwerpen, B. K. Kobilka, and C. Govaerts. 2016. Allosteric regulation of G protein–coupled receptor activity by phospholipids. *Nat. Chem. Biol.* 12:35–39.

[27] Dowhan, W. and M. Bogdanov. 2011. Lipid–protein interactions as determinants of membrane protein structure and function. *Biochem. Soc. Trans.* 39:767–774.

[28] Vigh, L., P. V. Escribá, A. Sonnleitner, M. Sonnleitner, S. Piotto, B. Maresca, I. Horváth, and J. L. Harwood. 2005. The significance of lipid composition for membrane activity: new concepts and ways of assessing function. *Prog. Lipid Res.* 44:303–344.

[29] Saeedimasine, M., A. Montanino, S. Kleiven, and A. Villa. 2019. Role of lipid composition on the structural and mechanical features of axonal membranes: a molecular simulation study. *Sci. Rep.* 9:1–12.

[30] Jo, S., T. Kim, V. G. Iyer, and W. Im. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comp. Chem.* 29:1859–1865.

[31] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. a. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

[32] Gu, Y., H. Li, H. Dong, Y. Zeng, Z. Zhang, N. G. Paterson, P. J. Stansfeld, Z. Wang, Y. Zhang, W. Wang, et al. 2016. Structural basis of outer membrane protein insertion by the BAM complex. *Nature.* 531:64–69.

[33] Han, L., J. Zheng, Y. Wang, X. Yang, Y. Liu, C. Sun, B. Cao, H. Zhou, D. Ni, J. Lou, et al. 2016. Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat. Struct. Mol. Biol.* 23:192.

[34] Iadanza, M. G., A. J. Higgins, B. Schiffrin, A. N. Calabrese, D. J. Brockwell, A. E. Ashcroft, S. E. Radford, and N. A. Ranson. 2016. Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. *Nat. Commun.* 7:1–12.

[35] Bakelar, J., S. K. Buchanan, and N. Noinaj. 2016. The structure of the β-barrel assembly machinery complex. *Science.* 351:180–186.

[36] Noinaj, N., A. J. Kuszak, J. C. Gumbart, P. Lukacik, H. Chang, N. C. Easley, T. Lithgow, and S. K. Buchanan. 2013. Structural insight into the biogenesis of β-barrel membrane proteins. *Nature.* 501:385–390.

[37] Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* 26:1781–1802.

[38] Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.

[39] Rose, P. W., A. Prlić, C. Bi, W. F. Bluhm, C. H. Christie, S. Dutta, R. K. Green, D. S. Goodsell, J. D. Westbrook, J. Woo, et al. 2015. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* 43:D345–D356.

[40] Lomize, M. A., A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg. 2006. OPM: orientations of proteins in membranes database. *Bioinformatics.* 22:623–625.

[41] Ali, M. H. and B. Imperiali. 2005. Protein oligomerization: how and why. *Bioorg. Med. Chem.* 13:5013–5020.

[42] Jo, S., J. B. Lim, J. B. Klauda, and W. Im. 2009. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* 97:50–58.

[43] Pandit, K. R. and J. B. Klauda. 2012. Membrane models of *E. coli* containing cyclic moieties in the aliphatic lipid chain. *Biochim. Biophys. Acta Biomembr.* 1818:1205–1210.

[44] Lind, T. K., H. Wacklin, J. Schiller, M. Moulin, M. Haertlein, T. G. Pomorski, and M. Cárdenas. 2015. Formation and characterization of supported lipid bilayers composed of hydrogenated and deuterated *Escherichia coli* lipids. *PLoS One.* 10:e0144671.

[45] Noskov, S. Y., S. Berneche, and B. Roux. 2004. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature.* 431:830–834.

[46] Hopkins, C. W., S. Le Grand, R. C. Walker, and A. E. Roitberg. 2015. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theor. Comp.* 11:1864–1874.

[47] Balusek, C., H. Hwang, C. H. Lau, K. Lundquist, A. Hazel, A. Pavlova, D. L. Lynch, P. H. Reggio, Y. Wang, and J. C. Gumbart. 2019. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *J. Chem. Theor. Comp.* 15:4673–4686.