BIT2018 Virtual Workshop:

Using the CHARMM-GUI webserver for membrane protein molecular modeling

Krzysztof Kuczera, Department of Chemistry and Department of Molecular Biosciences, University of Kansas, kkuczera@ku.edu

June 30, 2018

1. Introduction

This exercise provides an introduction to a very powerful tool of molecular modeling, the CHARMM-GUI web server. CHARMM-GUI has been developed in the group of Wonpil Im at the University of Kansas and Lehigh University. The site allows deployment of the CHARMM modeling program for users at any level of experience through use of a graphical interface. The site has a wide array of functions, from processing Protein Data Bank structures to build missing residues or post-translational modifications, to continuum electrostatics and solvation, data analysis and inclusion of experimental restraints. It allows for construction of mixed solvent, ion, protein, lipid, glycolipid and polysaccharide systems. The end products are model 3D structures and sets of input files and scripts needed to run more advanced computer simulations with a range of programs like CHARMM, NAMD, GROMACS and others.

Our exercise involves constructing an atomistic model of a GPCR protein embedded in a lipid bilayer, suitable for a molecular dynamics simulation. This will provide an introduction to the many powerful features of CHARMM-GUI. We will start with downloading the 3D structure of a receptor, surround it with a lipid bilayer, water and physiological ion atmosphere and generate inputs for an MD simulation. After this, the generated files may be used on a high-performance computing system to generate an MD trajectory.

2. The system

G-protein coupled receptors (GPCRs) are large family of membrane-embedded eukaryotic proteins which bind ligands outside the cell and in response activate signal transduction pathways inside the cell. GPCR structures are very characteristic, involving seven trans-membrane helices, connected by loops of varying size. The ligands vary from small molecules to peptides and proteins. The neuropeptide Y receptor family (Y1-Y5) interacts with the peptide hormones neuropeptide Y (NPY), peptide YY (PYY) and pancreatic polypeptide (PP), and is involved in control of appetite, anxiety and circadian rhythms. Our focus will be the first member of this family, the Y1 receptor.

3. A map of this project

Here are the steps needed to complete this project:

1. Prepare initial protein structure
2. Insert Y1 receptor protein into a lipid bilayer and solvate with water and ions
3. Generate inputs for further MD simulation

4. Atomistic model set-up for GPCR receptor

4.1 Initial protein structure - PDB or sequence and modeling

In May 2018, there were two Y1 structures in the Protein Data Bank PDB: files 5ZBH and 5ZBQ. Choose one of them as a starting point – you can find out about the structures at 'www.rcsb.org.' – I worked with 5ZBQ.

Note: if you are interested in a receptor that is not in the PDB, you can build a model.

- First, download a sequence. E.g. on the NCBI web server 'https://www.ncbi.nlm.nih.gov' go to the protein section and locate the sequence for the receptor of interest. Download and save in FASTA format.
- Next, make a structural model. One powerful tool here is I-TASSER. Go the I-TASSER server 'https://zhanglab.ccmb.med.umich.edu/I-TASSER/' and submit your sequence for modeling. Once process is complete, extract the structural models – they should be called model1.pdb, … etc.
- View your structure with your favorite molecular graphics tool – e.g. rasmol, pymol, vmd or chimera. You should see a familiar seven-helix bundle. If you started with 5zbh.pdb or 5zbq.pdb – you will also have an inhibitor bound.

Some useful molecular graphics tools: VMD, PyMol, RasMol, Chimera, SPDBView.

4.2 Create membrane, solvent and ions – CHARMM-GUI

Go to www.charm-gui.org.

0. Initial step: method and starting structure selection.
   Bring your mouse over the 'Input generator' box and 'Membrane builder' menu item. In the membrane builder sub-menu choose 'Bilayer builder' and in that select

'Protein/Membrane System'. Type in your selected PDB id, e.g. '5zbq' and use 'OPM' as the file source to download pre-processed PDB file from the OPM server. Click on 'Next Step' in bottom right corner.

Note: If you are interested in OPM details, visit the Orientation of Proteins in Membranes server 'http://opm.phar.umich.edu/' website at the University of Michigan. Briefly, the server determines the protein transmembrane region, insertion depth, tilt and orientation (i.e. which way the N and C terminals should point). This orientation is not important our model membrane.

The program now embarks on 5 stages of preparation

1. Read PDB. Use defaults here, we do not need to do any manipulations. This will select the protein part for modeling. Our system also includes a ligand (named 9AO in 5zbq). This is treated as an 'engineered residue' by CHARMM-GUI because of the structure of this concrete PDB file. In other cases ligands will come up as heteroatom sections.

   Note: If you would like to model the ligand, you will need CHARMM topology and parameter files. These may be generated e.g. by CGENFF sever.

2. Orient. We have the structure correctly placed by OPM, so use PDB orientation and no translation. As you can see, several things are done by default
   - Terminal patching (default is NH_3+ for N-term and COO- for C-term)
   - Building in of missing residues
   - Addition of S-S bonds to topology

   Click 'Next Step'.

3. Determine system size. Chose 'homogenous lipid', default number of lipid layers surrounding protein and 'DPPC' as lipid type. This part determines how many lipid molecules will be used and the size of the box (with periodic boundary conditions) to contain the whole system.

   Click 'Next Step'.

4. Build components. Our system has 4 components: protein, lipid bilayer, solvent water and ions to neutralize the system and mimic a physiological salt concentration. The protein is already there. Now we actually add the lipids, add water molecules to fill up our box and replace some waters with ions.

   Select replacement method for lipid building and 0.15 M NaCl for ion atmosphere. At the end you will have atomic coordinates for all components. This step can take 10 min or more.
5. System assembly. The pieces are put together. You have the option to generate inputs for simulation with various programs – e.g CHARMM, NAMD, GROMACS, LAMMPS. Click 'Next Step' to complete the step.

   After every step is completed, a button in top right corner appears labeled 'download

.tgz'. With this tool you can download all files generated at every stage. Click on this button now to finish CHARMM-GUI work. Again, be patient and wait a few minutes for process to complete. A compressed tar archive named 'charm-gui.tgz' will be downloaded to your computer, containing files labeled 'step1…' to 'step6…'. To unpack it, go to your work area and type 'tar –xvzf ~/Downloads/charm-gui.tgz'. When you unpack the archive, here are some files you will find:

The files with 'step5_assembly' in their name have all the information about our system - .e.g.

step5_assembly.psf = CHARMM protein structure file of whole system – protein, lipids, water and ions

step5_assembly.pdb = starting structure

step5_assembly.str = box sizes, numbers of residues

folders named 'gromacs', 'namd' etc. with the required inputs.

Step 6. Equilibration. This step is not performed on the server. Instead, a table appears on the screen describing the detailed settings of the six stages of equilibration, and input scripts are generated for each stage – step6.1_equilibration.inp, …, step6.6_equilibration.inp, and for the production run without any restraints – step7_production.inp.

Conclusion

Our final product is an atomistic model and input scripts for the MD engines we selected. The rest of the work needs to be done on a computer system that has these programs installed – e.g. a large workstation, cluster, GPU or other high-performace computing system. That is it for the exercise. Happy modeling.

References

CHARMM-GUI:

S. Jo, T. Kim, V.G. Iyer, and W. Im (2008)
CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem. 29:1859-1865*

E.L. Wu, X. Cheng, S. Jo, H. Rui, K.C. Song, E.M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R.M. Venable, J.B. Klauda, and W. Im (2014)
CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem. 35:1997-2004*

S. Jo, 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*

S. Jo, T. Kim, and W. Im (2007)
Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE 2(9):e880*

PDB:

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242.

Peter W. Rose, Andreas Prlić, Ali Altunkaya, Chunxiao Bi, Anthony R. Bradley, Cole H. Christie, Luigi Di Costanzo, Jose M. Duarte, Shuchismita Dutta, Zukang Feng, Rachel Kramer Green, David S. Goodsell, Brian Hudson, Tara Kalro, Robert Lowe, Ezra Peisach, Christopher Randle, Alexander S. Rose, Chenghua Shao, Yi-Ping Tao, Yana Valasatava, Maria Voigt, John D. Westbrook, Jesse Woo, Huangwang Yang, Jasmine Y. Young, Christine Zardecki, Helen M. Berman, Stephen K. Burley. (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information *Nucleic Acids Research*, 45: D271-D281.

Y1 receptor structure with ligand:

1. Yang, Z., Han, S., Keller, M., Kaiser, A., Bender, B.J., Bosse, M., Burkert, K., Kogler, L.M., Wifling, D., Bernhardt, G., Plank, N., Littmann, T., Schmidt, P., Yi, C., Li, B., Ye, S., Zhang, R., Xu, B., Larhammar, D., Stevens, R.C., Huster, D., Meiler, J., Zhao, Q., Beck-Sickinger, A.G., Buschauer, A., Wu, B. Structural basis of ligand binding modes at the neuropeptide Y Y_1 receptor. (2018) Nature **556** 520-524.

PPM

Lomize M.A., Pogozheva I,D, Joo H., Mosberg H.I., Lomize A.L. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res., 2012, 40(Database issue):D370-6

I-TASSER
Y Zhang. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9: 40 (2008).

A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5: 725-738 (2010).

J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8 (2015).