# SUPPLEMENTARY DATA

## 1. Step-by-step to build a solution system of the HIV envelop protein

1. Go to *Solution Builder*.
2. Type 5FYL and click Next Step.
3. The antibody chain identifiers are B, C, E, and F, so unselect them and also CARN that is attached to chain E. Then, click Next Step. CHARMM-GUI automatically detect missing residues and disulfide bonds in the structure. Let's not touch them in this example. Scroll down to Glycosylation / Glycan Ligand(s). Users will learn how to edit them in the next example (2. Modeling N-glycans on a Fc heterodimer of immunoglobulin G), so let's not touch them here. Since we want to generate the biological trimeric envelop protein, we need to click Generation of Biological Functional Unit. This option will utilize the matrix information in the PDB to generate the remaining two copies of the envelop protein. Now, let's Next Step.
4. One can click view structure to check if the trimer with all N-glycans is correctly built. Now we can set up a few options for our solution system. Since we have flexible glycans, it would be better to increase the edge distance (i.e., the distance from the protein edge to the box) to 15 Å. Since we change the default system size, we need to click Calculate number of ions to update the numbers. Since we do not need specific DNA/RNA-ion interactions, we can change Ion Placing Method to Distance as the Monte Carbo method takes much longer time. Now, let's go to Next Step to build water box, generate ions, and then assemble all components. Due to a big system size (159 Å x 159 Å x 159 Å), it would take about 15 minutes for this step.
5. In this step, we do not need to change anything. So, just click Next Step to setup the periodic boundary conditions. This step includes a small number of minimization steps, so it would take about 6-7 minutes.
6. If users are interested in running their simulations using NAMD or OpenMM, for example, users can select either one. We do not need to change other options, so click Next Step to finalize the system building and input generation. Since *Solution Builder* generate all necessary restraint files for non-CHARMM programs, it would take about 3 minutes for this step.
7. Now, one can download "download.tgz" and visualize "step3_pbcsetup.pdb". One should read README in each directory of selected program name(s) to run equilibration and production.

**2. Step-by-step to model N-glycans on a Fc of immunoglobulin G1**

1. Download 5tps_mod.cif from http://charmm-gui.org/download/5tps_mod.cif.
2. Go to *Glycan Reader & Modeler.*
3. Upload 5tps_mod.cif, and click PDBx/mmCIF and then Next Step.
4. One can see that this modified PDB file contains only the first N-acetylglucosamine (GlcNAc) on chain B. Let's go to Next Step.
5. In Glycosylation / Glycan Ligand(s), let's first click edit to build the missing glycans on chain B in a new popup. One can load a glycan sequence in GRS format, which will be used in Step 6. For now, let's build the missing glycans by adding them in Glycan Sequence. First, click + and change the linkage and sugar type to $\beta$-14 and N-acetylglucosamine (GlcNAc), respectively. one can see that Sequence Graph is automatically updated to graphically guide the selected glycan sequence. Now, let's add mannose (Man) to GlcNAc with $\beta$-14 linkage. And, add Man to Man with $\alpha$-16 linkage. And, add GlcNAc to Man with $\beta$-12 linkage. And, add galactose (Gal) to GlcNAc with $\beta$-14 linkage. Now, we need to add a branch from the third Man, so we click + on the third Man and add Man with $\alpha$-13 linkage. This is the N-glycan on chain B of the original PDB. So, click Next Step to update the sequence.
6. Now, we are back to the PDB manipulation page. Let's add the N-glycan to Asn 297 on chain A by clicking Add Glycosylation and then edit on CARB in a new popup. First, we need to change Asn residue number to 297. This time, let's assume that we know the GRS format of the N-glycan on chain A:

   ```
   1 BGLCNA
   2 - 14B: BGLCNA
   3 - - 14B: BMAN
   4 - - - 13A: AMAN
   5 - - - - 12B: BGLCNA
   6 - - - 16A: AMAN
   7 - - - - 12B: BGLCNA
   8 - - - - - 14B: BGAL
   9 - 16A: AFUC
   ```

   Once users copy and paste it, click Apply. One can see that both Glycan Sequence and Sequence Graph are automatically updated. Once users confirm, click Next Step to update the sequence.
7. Since there is no further modification, we can click Next Step to build the N-glycan models.
8. Click view structure to check if both N-glycans are built properly. Note that one can repeat the same process to build a solution system of this glycoprotein.

**3. Step-by-step to build a membrane system of a cholera toxin B and ganglioside GM1 complex**

1. Go to *Membrane Builder* to start the building process.
2. Type 3CHB. We need to change Download source to RCSB because the OMP database for pre-oriented membrane proteins along the membrane normal centered at Z = 0 does not contain this PDB. Then, let's click Next Step.
3. The PDB file shows five CTB subunits and associated GM1 carbohydrates. Let's go to Next Step to build a complete CTB-GM1 complex.
4. In the Glycosylation / Glycan Ligand options, one can see that CARA and CARC need β-glucose (Glc) at the reducing end that should be linked to a ceramide. So, let's first click edit for CARA. In a new popup, we need to first add βGlc at the reducing end to galactose (Gal) with β-14 linkage. Then, we can add a ceramide tail by clicking None and CER180 to build a complete ganglioside GM1. Now, we can click Next Step to update the sequence. One can see that CARA now becomes a GM1 glycolipid. We can do the same for CARC. For CARB, CARD, and CARE, we can add CER180 to build a complete ganglioside GM1. Once everything is done, we can click Next Step to orient our model in a bilayer.
5. Let's first confirm if the CTB:GM1 complex model is built correctly by clicking view structure. A PDB file from RCSB is not generally oriented with respect to a membrane bilayer (like our model), so that we need to first orient it. CHARMM-GUI provides a few options and we will use Align the First Principal Axis Along Z, because we have a pentameric CTB whose principal axis can be easily defined and good to be used for this alignment. Since we have no idea where this aligned structure will be located with respect to a bilayer centered at Z = 0, we need to first click Next Step.
6. We can click view structure of "step2_orient.pdb" to see how our model is oriented in a bilayer. Note that the two planes represent an approximate hydrophobic region of a bilayer from -12 Å to 12 Å along the Z axis. Now, we can see that our model is oriented in a strange way. But, this can be understood as the long axis should not be the five-fold axis of the CTB structure due to its shape. Then, we can go back to the previous step and rotate it along the Y axis by -90 degree as the principal axis (i.e., the five-fold axis) was aligned along the X axis. Now, the five-fold axis is aligned along the Z axis (which is the membrane normal), but the protein is inside the membrane, although it should be above the membrane. So, we can go back to the previous page again and add an option to translate the molecule along Z by 40 Angstrom. This is an approximate guess obtained by trials and errors. Assuming that this orientation is acceptable, we can now start to build a bilayer part.
7. For #1 and #2 options, we can just take the default values. #3 is the most important option to determine the membrane dimension on the XY plane. The basic concept is to use the lipid area of each lipid to determine the XY dimension. So, we need to know an approximate XY length to use a lipid component ratio. Let's try this option in this example. We will simply use DMPC as a sole lipid type in this example. Type 100 in Length of X and Y and type 1 and 1 for upper and lower leaflet ratios in DMPC under PC lipids. Then, click Show the system info. *Membrane Builder* notices that the system size we chose is not big enough for our complex model. Then, we can adjust XY length to 150 and click Show the system info again. Now, we can go to Next Step to see if our XY

membrane dimension is reasonable. *Membrane Builder* is packing pseudo lipid spheres in the lower and upper leaflets. It would take 2-3 minutes.

8. In this step, it is very, very important to click view structure of step3_packing.pdb. It is highly recommended to have at least 4 spheres (i.e., lipid molecules) between the protein in the primary cell and those in the neighboring images. In our case, we seem to have too many lipids. Note that it should be up to users to decide which system size would be good (after having at least 4 spheres between images) depending on their research topic and resources. If necessary, one can go back to the previous step to adjust the XY dimension. In this example, we go back and adjust it to 120. Again, we check step3_packing.pdb and assume that this dimension is approximate for this example. Now, assuming that the default options for ions are reasonable, we can click Next Step to build a DMPC bilayer. The pseudo spheres are replaced by DMPC molecules in this step to avoid any bad contacts in the system. This is the most time-consuming step in *Membrane Builder*. In our example, it would take 35-40 minutes. When users have bigger systems, they should expect more time to finish this step. Therefore, one can use a bookmark link to check if their job is done later.

9. Now, let's click Next Step to build water box and ions. It would take 2-3 minutes in this example.

10. Now, we have all the components such as our CTB:GM1 complex model, a DMPC bilayer, water box, and ions. So, let's click Next Step to assemble them. It would take 3-4 minutes in this example.

11. If users are interested in running their simulations using GROMACS, they can select it. We do not need to change other options, so we can click Next Step to finalize the system building and input generation. Since *Membrane Builder* generate all necessary restraint files and performs some minimization for non-CHARMM programs, it would take about 5-8 minutes in this example.

12. Now, one can download "download.tgz". You should check "step5_assembly.pdb". One should read README in the gromacs directory to run equilibration and production.

## 4. Step-by-step to build a solution system of a heparin sulfate molecule

1. Go to *Glycan Reader & Modeler*.
2. Click Glycan Only System and then Next Step.
3. Let's build the glycan sequence of heparin sulfate with alternating α-glucose (Glc) and α-iduronic acid (IdoA) with 14 linkage:

   αDGlc (1→4) αLIdoA (1→4) αDGlc (1→4) αLIdoA (1→4) αDGlc

   Then, add chemical modifications of "sulfoamino" to the second position of all Glc residues and "sulfate" to the sixth position of all Glc residues and the second position of all IdoA residues. Then, click Next Step to build a heparin sulfate model.
4. Click view structure to check if the heparin sulfate molecule is built properly. From here, users just need to follow the typical *Solution Builder* steps (as we already did in the HIV envelop protein case).

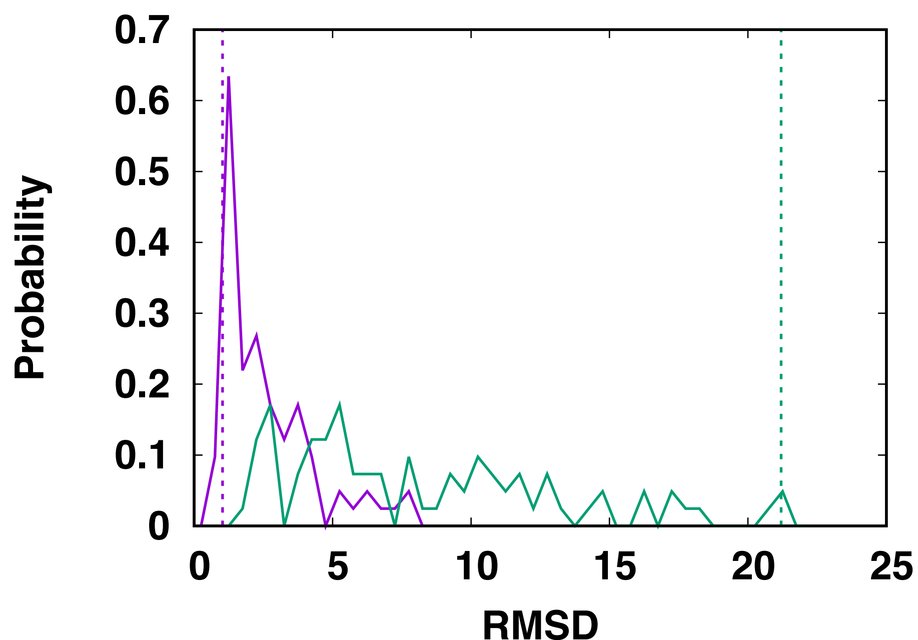


**Figure S1.** Benchmark result of *Glycan Modeler*. Purple represents the RMSD distribution with superposition between crystal and modeled glycan structures (i.e., RMSD of glycan conformations), whereas green represents the one without superposition (i.e., RMSD including glycan conformations and orientation on the protein surface). Dashed lines show the case with the highest gap between these two RMSD types (PDB: 4IIB).

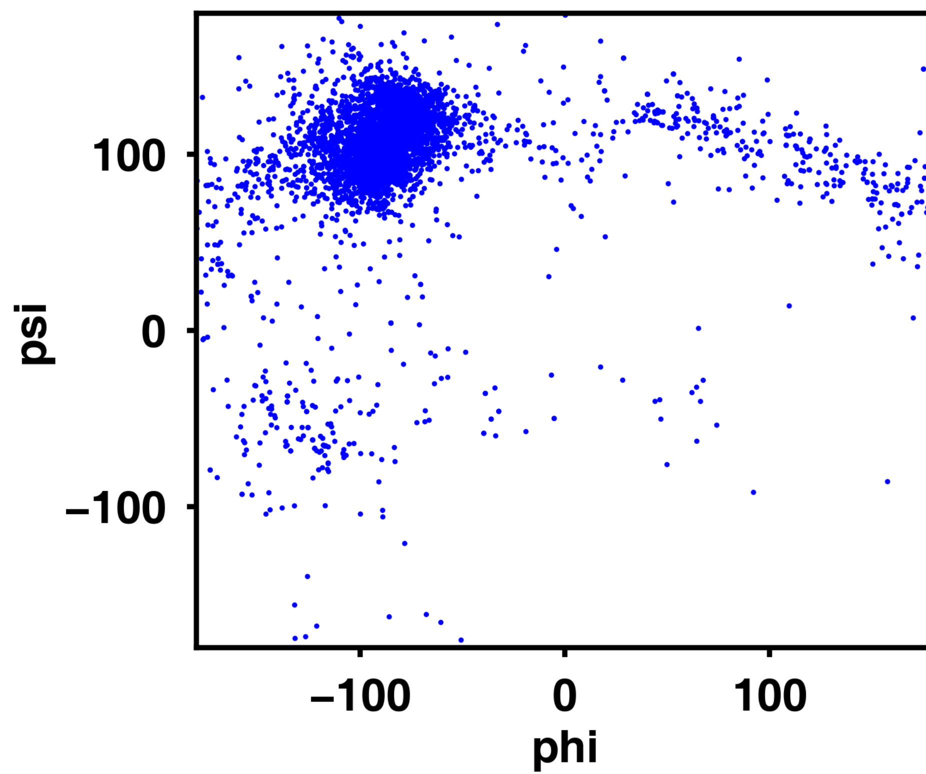


**Figure S2.** Distribution of φ and ψ dihedrals of Man(β1-4)GlcNAc in PDB glycan structures from GFDB search.

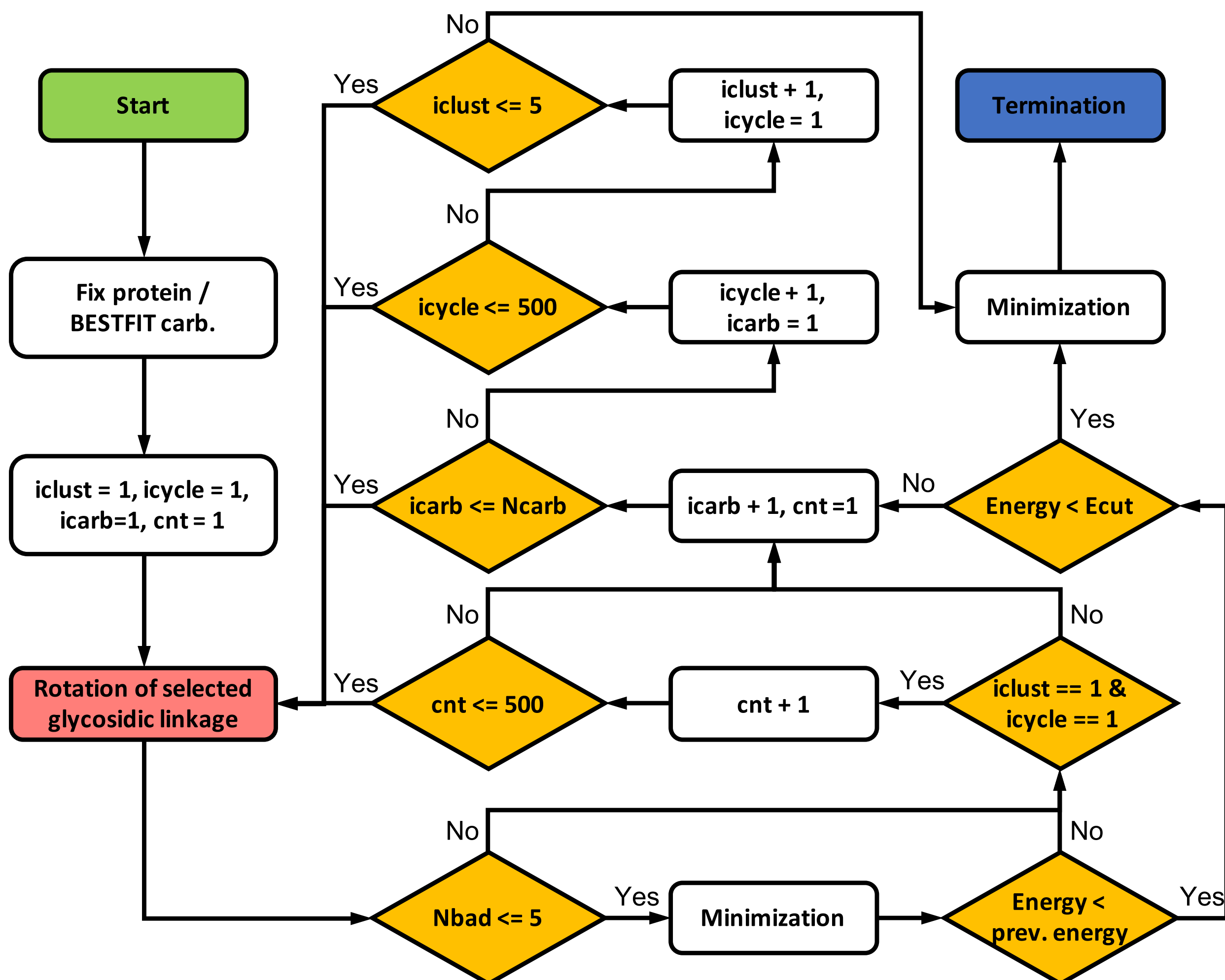


**Figure S3.** Workflow of *Glycan Modeler* for modeling N-/O-glycans. The variables are iclust (the cluster type from GFBD search), icycle (cycle step number), icarb (glycosylation site to be rotated), Ncarb (the total number of glycosylation sites), and cnt (step number for the first iteration to find an initial orientation). Protein coordinates are fixed and BESTFIT restraints applied to carbohydrates. 500 iterations of rotational search are performed to find an initial orientation of glycan structure for each glycosylation site, only when both icycle and iclust are 1. In the next step, another 500 iterations of rotation search are performed to find minimum energy orientation. Note that the whole process is terminated when the calculated interaction energy is lower than the cutoff energy at any point.

## Movie Captions

HIVenv_cryst.mpg: A movie of 2-μs simulation of HIV envelop protein starting from the glycan structures in the PDB:5FYL.

HIVenv_glyMod.mpg: A movie of 2-μs simulation of HIV envelop protein starting from the glycan structures modeled by *Glycan Modeler*.