

Term Project: Design an antidote for CoVid19.

SARS-CoV-2 is the virus responsible for the current worldwide CoVid-19 pandemic. It is highly contagious and kills somewhere between 1 and 21% of those it infects (exact value depends on the current degree of underreporting of infections, which is not known due to insufficient testing.) It has killed 114,248 so far (4/13/20) and models predict a final death toll between 200,000 and 1.5 million people over the coming months, globally. Every country in the world has at least one confirmed case, and the US currently has 560,433 cases, with 22,115 dead. The rapidity of its spread means that developing a vaccine (a 12 to 18 month process) will be too slow. We have to cure it.

Crystallographers and cryo-electron microscopists have already isolated and solved the structures of many proteins from the virus, including the spike glycoprotein, paving the way for a molecular modeling response to the pandemic. The spike is responsible for viral entry into the cell. It binds to the angiotensin converting enzyme 2 (ACE2) a membrane bound enzyme which controls vasodilation in the lungs, located on the extracellular surface of the alveolar epithelial cells type 2. Virtually the entire exposed surface of the spike glycoprotein is covered with and blocked by glycans, which are hydrophilic, mobile and highly variable carbohydrates that do not present a viable drug target. Thus the target region for an antiviral drug is the region of the spike that binds the ACE2 receptor. The receptor binding domain (RBD), residues T475-K670, binds ACE2 using loop V587-P641 and surrounding residues. This is the surface we will target for binding of a protein-based drug.

I hereby deputize you to help design an antiviral drug!

In short, you will explore candidate spike binder proteins, docking each to the spike protein in one of its several conformations. You will dock to the ACE2-binding surface, and then design the loops and sidechains of the ligand to fit the RBD surface shape and charge. We will be taking advantage of the 3-fold symmetry of spike to triple our affinity by using a trimeric template. We will start with only human proteins to guard against an immune response to the drug.

Step 1 : Find templates.

Search for a trimeric, human, small known protein structure using the following NCBI query (<https://www.ncbi.nlm.nih.gov/protein>)

Search => "homo sapiens"[organism] AND trimer

Set filter Source Database => PDB

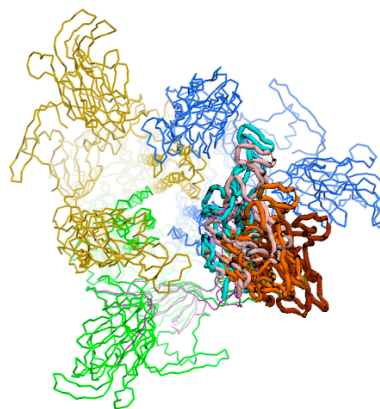
Set filter Molecular Weight => 10000, 20000 kilodaltons

The result will be 783 proteins of known structure that have the word "trimer" in them. Not all of them are really trimers. Read the PDB code and chain ID and open the structure in MOE. Look at the structure to see if it really is a compact, tightly-bound trimer, and whether it is the right size to fit in the space between the RBDs. As soon as you find a trimer of the right size, go on to step 2. Come back to this step only if your trimer fails in Step 2. **As soon as you have made a choice, please email us the PDB code.** I would like everybody to work on different trimers. Do this no later than Apr 21.

Step 2 : Dock trimeric protein ligands to the spike protein.

Download a set of "all-up" spike trimers from the course website. (TBA)

These structures have the RBD, only, in a variety of different conformations. For each ligand, find the spike trimer with the best size of the opening around the 3-fold axis. You must dock along this axis. The RBD models are compatible with 6vsb. If you want to make images with the full length spike protein included, open 6vsb and display atoms that are not the RBD. For that reason, don't move the RBD's.



Make sure you are docking from the outside of the spike protein not from the inside! If you are not sure, display the full spike protein. You want to dock around the loop V587-P641.

We want three-fold binding, so we must look for a pose that lines up the 3-fold symmetry axis of the ligand with the 3-fold symmetry axis of the spike protein. This eliminates 4 of the 6 degrees of freedom in docking. You can only adjust the rotation angle around the 3-fold and the height of the ligand relative to the spike protein. You can also flip the ligand and try the other side to see if that side works better. If you dock on the 3-fold symmetry axis your ligand will not encounter any of the glycans; one less thing to worry about.

Dock by selecting the ligand and moving it using alt-middlemouse (out of plane rotation), shift-alt-middlemouse (translation). Rotate in-plane by using alt-middlemouse just inside the edge of the window.

Dock it to create as much "touching" surface as possible, where "touching" means minimum backbone atoms distances are in the range 5-8Å but no closer. Use ribbon rendering for docking. To refine the orientation, display ligand atoms and near residues on spike. Make only minor adjustments.

Ignore any side chain collisions created at this point. We'll fix them by mutating.

If you have searched and cannot find a good fit for any of the RBD models, reject the template. Go on to the next ligand template from Step 1. Keep trying ligand templates until you find one that gives the most buried surface area.

Save images of your initial docking. Display a surface of the interface.

Step 3 : Loop adjustments if necessary.

Avoid having to do any loop search if possible by avoiding backbone atom collisions during Step 2. If you must do a loop search it should have no insertions or deletions, or at most one residue inserted or deleted. Using Loop Modeler or Homology Model.

Any changes you make in this step or Step 4 should be done on just one of the monomers. You can copy all of the changes to the other 2 monomers later using the copy/paste/superpose method (as in Homework 3 part 2).

Step 4 : Design your ligand.

Most of your time should be spent on this step. Use Protein Builder, gizmin.svl, EPUSIEPF, along with Geometry and Clash. Choose rotamers with a low E_{rot} whenever possible. Minimize each side chain that you mutate. Draw from your experience doing Homework 4. Pay attention to charged side chains especially. Make salt bridges. Don't accidentally mutate the spike protein! If you accidentally mutate spike, just mutate it back to wild type.

Step 5 : Humanize the sequence.

Since we have added mutations, the sequence is no longer immunologically self, maybe. To check, select each 5-residue piece that has at least one designed mutation in it. Run protein BLAST on that 5-residue sequence (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Set BLAST to search Organism *homo sapiens* only. If BLAST finds an identical 5-residue sequence in any human protein, then we are going to say that your mutations in that 5-mer are likely to be *immunologically silent*, since it is identical to a human sequence. If you find no hits with 100% identity, then consider mutating side chains so that it matches one of the near-identity BLAST hits.

Step 6 : Waters.

Add waters one by one using Builder during Step 4, or all at once using the freeze-dry method. Add only pocket waters and cavity waters. Try to make sure that all buried hydrogen bonding acceptors and donors are satisfied, if not by protein then by waters. Consolidate all of your waters into one chain in SEQ by Edit | Split/Join | Join.

Step 7 : Molecular dynamics

Before going to Step 7, make two copies of the monomer you designed and the RBD closest to it and superpose the copies on the other two monomers (see Homework 3 part 2.), then delete the originals. Use the ligand residues to superpose both designed ligand model and RBD it binds to. (Give the monomer and its RBD chain the same tag name to keep them together when superposing.) If you were perfect in aligning the 3-fold axis in Step 2 then the new copies of the RBD will sit nicely on top of the originals. But if not, don't worry. We are assuming that the RBD is *very* flexible. It is. Delete the originals. You should have 6 chains again (plus waters).

Now your ligand model is 3-fold symmetrical. Proceed.

Energy minimize the complete trimeric, 6-chain (plus waters) structure bit by bit using the EPUSIEPF method, gradually expanding the unfixed zone. Finally, energy minimize the whole ligand trimer, then the ligand plus near residues of the spike RBD.

Next, run a 100 ps MD simulation. Heat from 100 to 150K for 50 ps, then cool from 150 to 100K for 50 ps. Fix the spike protein backbone but leave side chains unfixed. Energy minimize the final structure. Check for anything falling apart. This usually means the atoms were too tightly packed or you had too many un-preferred rotamers, or bad phi/psi angles.

This MD simulation will take hours. When done, open the dbv file and Browse the data. Mark loops that move the most and mention them in your talk. Check to see if secondary structure hydrogen bonds are breaking or staying put. Look down the 3-fold axis. Does the whole ligand rotate at all? Rise or sink? Does it stay aligned with the 3-fold of the spike?

The final frame of MD, after minimizing, will be the structure you present. For context, add the non-RBD parts of the spike protein using your Homework 3 part 2 model, or use 6vsb.

Step 7 : Presentation

We will be presenting term projects Tuesday April 28 at class time. Please send us your slides with presenter notes included at that time. Thank you. The presenter notes content should be approximately what you say in your talk.

In your talk, walk us through the design process specific to your model. Point to the changes you made. Show the quality of the model. Show the interface electrostatic surface. Make sure your images are easy to see. You will have to hide most atoms.

Talk about the native function of your ligand template. Would its function interfere with its use as a therapeutic? Does it have any required ions or cofactors? Any enzyme activity?

Working in teams

We will split into teams of 2. Probably the best way to work on this project as a team is to "tag team" the design. That is, avoid working in parallel because you will have to combine your work. Don't step on each others toes. Pass MOE files back and forth by email. Make sure you know who has the active copy. Take turns. Number and initial your file names to keep organized. Set your schedule ahead of time. Don't let your partner do all the work. Be equitable and considerate. Thanks. Don't forget the literature work. That can be done in parallel.

Final thoughts

This is serious. We're not just doing a term project here. Any successful designs will be considered for testing in the lab, and maybe in clinical trials. Wouldn't it be nice if you were the one who made the molecule that saved millions of lives (no pressure!). I know this is a difficult term project, but in this time of crisis people who can do should do. I think you, and MOE, can do it.