



SARS-CoV-2 PROTEIN MODELING AND DOCKING

Atishya Mahesh Jain*¹, Shaivee Tamrakar², Narmadha Devi S³, C Dharshini⁴

Karunya Institute of Technology and Sciences, Coimbatore

ABSTRACT

COVID-19 is one of the fatal pandemic throughout the world. In this project we have constructed a 3D structure of a SARS-CoV-2 protein sequence using homology modeling and perform molecular docking of drugs against this protein molecule and infer protein-drug interaction. We have accomplished it in by completing each task in the project which includes - Model protein structures from sequence data - Process proteins and ligands for docking procedure - Molecular docking of drugs against protein molecules. We have also utilized various bio informatics and protein visualization tools such as Auto dock Vina, PyMOL, protein plus and access databases such as protein data bank, PDB and PubCHEM

INDEX TERMS— Covid, Docking, Pandemic , SARS

1. INTRODUCTION

COVID-19 is a deadly disease, where the infection is caused by SARS-CoV-2. The corona virus particles are spherical in shape having spike proteins around them. These proteins are responsible for virus replication in human host cells. Spike proteins after attaching with human cells, undergo structural changes, which results in a fusion of viral particle membrane with human host cell membrane. Thus, the viral RNA enters into the host cell and produces more viruses after copying its genome. SARS-CoV-2 spike proteins bind to the receptor proteins, on the host cell surface, known as angiotensin converting enzyme 2 (ACE2). The molecular level structure of SARS-CoV-2 spike protein has a Receptor Binding Domain (RBD) for binding to host human cells. Receptor Binding Domain (RBD) of spike glycoprotein interacts with ACE2 receptor in Protease Domain (PD) of the host human cell, causing viral infection.

Firstly, all protein molecules are chains of amino acids or amino acid monomers, which further coil, turn and twist to form a secondary structure. These second restructures, combined with other amino acid chains and form tertiary or even Quaternary protein molecules, now consider you have an amino acid sequence or a protein sequence, which might look like a long stretch of random alphabets where each alphabet corresponds to an amino acid. Now to study docking, you will need a 3D structure of this sequence. But now how can we convert this string of alphabet into a 3D representation? Yes, the answer is homology modeling.

Modeling is a procedure where we fit, or align an unknown amino acid sequence with a previously known sequence and try to predict its structure. Using the known sequence as a template, we will learn how to perform homology modeling in our next task. Next is molecular docking. Molecular docking- It's a type off bio informatics, modeling an essential tool in structural molecular

biology and in drug design. The purpose of using this technique is to predict the most likely binding site or binding scenarios between the drug and the protein molecule, virtually given their three dimensional structures. Now, you might ask, Why do we need to perform molecular docking? The answer is pretty straightforward. The stages that followed. The design or identification of a new drug are both costly and time consuming. The entire process of drug development can take up to 12 to 15 years and cost billions of dollars. But in-silico molecular docking studies have been seen to both speed up the discovery rate and reduce, although not eliminate, the need for expensive lab work. Hence, molecular docking studies are very crucial in speeding up drug discovery and screening processes in times such as the current Covid 19 pandemic.

2. PROCEDURE

2.1 Model protein structures from sequence data.

We will model an unknown protein from an amino acid sequence using an online tool Swiss model which is an fully automated protein structure - Homology Modeling server Firstly, click on start modeling button. You will see a modeling project interface where you can input your unknown sequence.

To do this, click on the green button which says upload Target sequence file and from the Pop-up Navigate to the Downloads folder and select the fasta File named R -CSB PDB - fasta and click on open. This should load the sequence, which we want to model next add a project title of your choice, or you can type in unknown protein modeling and leave the rest to their default state and click on build model Option SWISS model will construct a 3D model of our sequence by aligning and fitting are unknown sequence against an already known protein structure using



homology or similarity based prediction on the left section off.

The model result you'll be presented with the details and the parameters employed along with the quality of the protein. But what we should give importance to is the sequence identity, which should ideally be above 90% on the right. You'll see a predicted interactive 3D structure of the model of our protein. We will be using this protein for our docking studies. From the description, it's evident that our predicted model matches with Corona Viral spike glycoprotein 2 Ghv. We can download the structure by clicking on this model 01. Under this, you can click on PDB format. PDB, short for protein databank format, is a file format to store sequence and structural 3D coordinate details of our protein molecules in case if the structure of the protein molecule, which you want to dock, is already present. And available. You can download and access the PDB file from the RCSB PDB database. Now we have a protein sequence to fetch our drug molecule, which we will be using to dock. Visit PubChem database type in pubchem and select the first option. And in the search box type in chloroquine which'll be the drug molecule which we will be using and pick the first option. Chloroquine is one of the drugs employed in COVID 19 treatment.

2.2 Process proteins & Ligands for docking procedure.

In the previous task, we learned how to model an unknown protein from an amino acid sequence using an online tool Swiss model and downloaded the drug molecule, which we will be using for docking procedure. In this task, we will process these same files to remove water molecules, add hydrogen and charges for both the drug and protein molecules to begin with. Go to pymol and click on open. Now from here, select the files option from the menu bar and click on Open. Navigate to the Downloads folder and from here click on the file name, which says confirm 3D CID 2719 SDF, which is a ligand or drug molecule and click on open, this will open up a 3D structure of the drug to proceed further with Docking. We need the drug to be in the PDB format. We will use Pymol to convert the SDF format of the drug, which we loaded here into PDB. To do this, navigate to the files option from the menu and click on export molecule from here. Select all and click on Save. Rename this molecule as Ligand and set the file format to be PDB, which is the second option and click on Save. After this, navigate to the Autodoc tools. Click on the folder icon and navigate to the Downloads folder and select the Ligand dot PDB file, which we just created and click on Open. It should load the PDB format of the ligand or drug molecule. From the sub menu.

Click on the Ligand Option from the sub menu. From here, go to input and choose, select the ligand by clicking on it and click on Select Molecule for AutoDock four and click on OK to select the file or the current molecule, which is open next. Click on Ligand again this time go to output and click on Save as PDBQT here. Navigate to the Downloads folder, and erase everything before ligand dot pd BQT and click on Save. This will save the Ligand in the PDBQT format, which is necessary for docking. Now it's time to modify the protein molecule. Firstly, we will have to delete this Ligand or drug by right clicking on the ligand here and clicking on delete option. We will load the

protein molecule using the same procedure as we did with the Ligand or the drug that is selecting the folder icon. And from here, select the protein named 2GHV, which is the protein which we modeled and click on open. Next, we need to remove water molecules since they might interfere with the drug docking procedure. Next, add hydrogen to compensate for the lost water molecules, followed by addition of Kollman charges to balance the charges. These three steps are essential in performing an accurate protein ligand or protein drug docking. To do this, navigate to edit option and from here, select delete water. This should remove all the water molecules next, go to the edit option again. This time around, select hydrogen and click on add and click on OK.

This will add hydrogens to the protein to compensate for the lost water molecules. To add charges, go to edit again followed by going to charges and select add Kollman charges. And click on OK, this will add Kollman charges balancing the charge on our protein. Now to save this as a PDBQT file, as we did with Ligand, go to the sub menu and select GRID and from the drop down, select macro molecule and pick choose From here. Click on the protein name and click on Select Molecule to select the molecule, which we have over here. Now click on OK, a pop up with the location to save will be displayed again. Go to the downloads folder, clear everything before the protein name, which is 2GHV, everything before that and click on save.

2.3 Create grids & write configuration files for docking.

We will create various configuration files which will guide the docking process. Firstly, delete the protein molecule from the dock console or the AutoDock console by right clicking on the protein and selecting delete option, loading the protein file that is 2ghv with the dot PDBQT format we saved in the previous task. To do this, select the folder icon, and from the pop up select 2ghv, but with a pdbqt format and click on open. First, we will create a grid box. Navigate to the option called grid, and from here, select the option macro molecule and click on Choose and select the protein, followed by select molecule option and click on Yes and okay, now go to grid option again. From here, select grid box. A dialog box with different coordinates will pop up, and if you observe, along with our protein, there is a box. This box is known as a grid box. Grid box restricts the binding of the drug to a specific region that is, we can control as to where or which region of the protein we want to perform. Docking of the drug molecule usually identifies the region's of interest called pockets, which are pockets or cave like regions which protein molecules form and place the great box at these pocket regions from the pop up menu, set the X Y coordinates, which are these three dials to 60.

You can do this by right clicking on it and slowly dragging it to you right or left. Similarly, do it for the next two coordinates as well, by dragging the color dials to change the size of the grid box. And we can set the position of the great box using the X Y Z center position, but we will leave them at the default positions now. Go to file from the pop up menu and select Output Great Dimension File. Name this as CONFIG file and make sure you're in the Downloads folder and click on save. Now open this directory and go to the downloads folder from here,



right Click on the configuration and click on Open With and From Here Select Open with text editor. We're going to edit this file. Firstly We will specify the receptor Which is our 2ghv And along with this, we will specify our ligand, which is a drug molecule, and this will be ligand dot pdbqt file . Don't forget to add the extension for our protein as well, which was 2ghv dot pdb qt. here the line, which specifies the spacing value after this hit enter To add a new line here, we will specify our center values. That is the position center underscore X in the next line Center underscore Y and center underscore Z. These values or positional values correspond to the values off the center, which is in the last line, So our X value will be 5.779. Y will be minus 22.064 and Z will be 14. 495. hit enter to add new line. After this, we will have to specify the size. Size underscore X, Similarly size Y and size Z. The values for this will correspond to the values of the npts which is set to 60. So the first is the X value. Second is the Y, and third is Z, so we'll just add in 60 everywhere now delete the last two lines. After this, add another line and here type in exhaustiveness and set this to eight. Here, Receptor and Ligand are the molecules which we processed Size X Y Z are the grid box sizes and center X Y Z are the position off the grid box exhaustiveness mentions the iteration of times. The docking procedure will be repeated now to save this file, navigate to this settings icon and click on Save as Add Number two after config and let it as it is and click on Save.

2.4 Docking of drug against the protein molecule.

We will make use of this file to run the docking. Firstly close the Autodock program and go to the terminal ubuntu terminal and here type in CD and Downloads add in a forward slash and hit enter This takes us to the downloads folder Now we will type in Vina space - - config And here we will pass in the file that we created Config2 Dot TXT space - - log space log dot TXT space hyphen hyphen out again space output dot PDBQT here Vina is an AutoDock command or Autodock tool, which performs docking hyphen, hyphen, config parameter takes in the configuration file with the parameters which we created in task four hyphen hyphen log option writes the result to a file named lag dot txt, followed by hyphen hyphen out option which saves the docked positions of the drug to an output file named output dot PDBQT Hit enter to run this command A progress bar will be displayed along with an ongoing task. Now the docking is complete Once the docking is complete, you will be able to see two new files in the directory. Namely, Output dot pdbqt and log dot txt which constitutes the docking results right Click on the log dot TXT file and click on open with text Editor here we need to observe two things firstly affinity mentioned in kilocalories per mole and the RMSD score Short for root mean square deviation which should be the lowest amongst the other entries in the column. And if you see the First Post has an affinity of minus 5.5 and an RMSD score of 0.0 which signifies that this post is the near perfect fit. Compared to the rest.

Let's have a look at the output file to do this Go to pymol After this navigate to file and click on open and from the downloads folder, Select the protein named 2GHV dot PDB and click on Open after this navigate to file again and click on open. But this time, open the output dot PDBQT, which has our docked

result and click on Open Now. You should have your protein molecules along with the Ligand or drug on the pymol console. To visualize this better go to the editing panel on the left and right besides the protein name. Click on this s and as and select surface. After this, click on the letter C and go to green and select the first color. this will give our protein molecule or green color. To red goto the output option, which constitutes our ligand or drug and click on C and go to red and select red. Now both the drug and protein molecules can be easily distinguished by default. The pose of the drug is set to one, as indicated by the state on the bottom left panel, which shows one out of nine you can cycle between all the nine poses using the arrow keys below the state option by clicking on these arrow buttons.

If you see you can switch between the different poses Now Let's go back to the first position By clicking on this button. We'll have to save this protein molecule and the drug as one file by going to the files section and selecting export molecule. And from here, click on the drop down menu Select. All followed by Click on Save This will give a Popup Name this molecule as Doc underscore file and change the format to PDB, which is the second option. And again, make sure in the Downloads folder now click on save. Now we can close these programs to free up the memory. Goto the browser again. Open a new tab here type in Proteins dot Plus and hit enter and upload The file. The docked File, which we just created by clicking on the browse option and navigating to the Downloads folder and selecting the docked file dot PDB and click on Open Protein. Plus is an online bioinformatics tools to visualize interaction of molecules. now click on Go, as you can see from the result, our drug molecule is interacting with isoleucine to visualize this. Better scroll down and from the appearance, change the surface to on and the color option to yellow. Now, if you go back and zoom in, you should be Able to see a drug molecule is interacting with the Ligand Visualized the specific interaction. Go to your right hand side from here. Pick the option pose View. pose view creates a 2D or two dimensional diagrams of these interactions and click on the Pose view button and in this field, type in the name, which is displayed here and click on Calculate.

3 RESULTS

Our Drug molecule(Fig 3.1) is interacting with the two of the residues from the protein, which are isoleucine and tyrosine. These are the two amino acids in the pocket region of our protein.

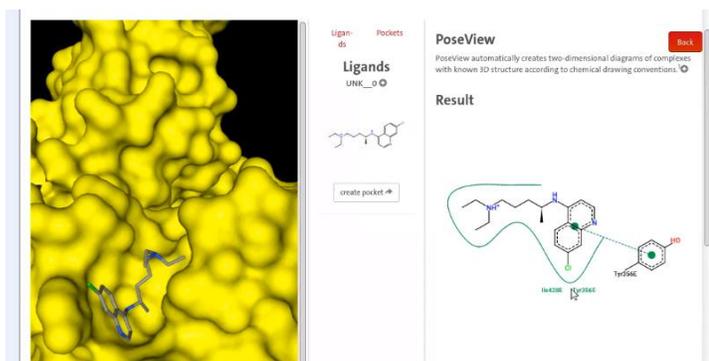


Fig 3.1 Final Result

CONCLUSION

Based on these interactions and docking scores, we can further filter and compare multiple drugs, to identify a handful of potential candidates to proceed with the clinical tests

ACKNOWLEDGMENT

We wish to thank our family , teachers and friends.