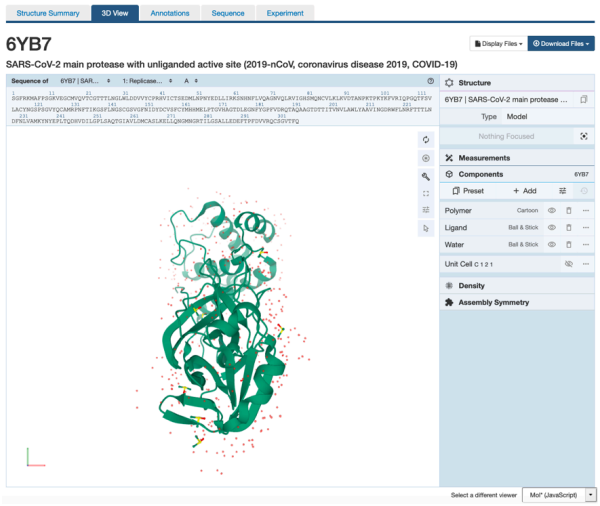
# Using RCSB PDB Mol*

*Download this document to use as a worksheet*

To save images, click on the camera (iris) icon  , Download and save a *.jpg file. Import the image in any image manipulation software of your choice (e.g., PowerPoint/ Photoshop) to add labels and additional text describing the images.
Some key commands and functions of Mol* are included in the Appendix at the end of this document.

## A) Explore the SARS-CoV-2 Main Protease Model

- Go to RCSB PDB home page (www.rcsb.org) >> type the PDB ID “6yb7” in the top search box >> click enter. This should take you to the Structure Summary Page. Click on the 3D View tab to view the structure in Mol*




There are 3 main areas on this screen:

1. Sequence panel (top left)
2. 3D-canvas (white space where the 3D structure is shown). Besides displaying the interactive 3D models, this space also offers
    a. Toggle panel (a series of buttons on the right) to enable various functions
    b. Log panel (at the bottom of the canvas) records actions taken
3. Control panel (blue column on right) with menus for Structure, Measurements, Components, etc.

- Use various mouse controls to rotate and translate the molecule you are viewing and answer the following questions. Note: Hovering the mouse over any object in the 3D canvas will display information about that item in the bottom right corner of the 3D canvas.

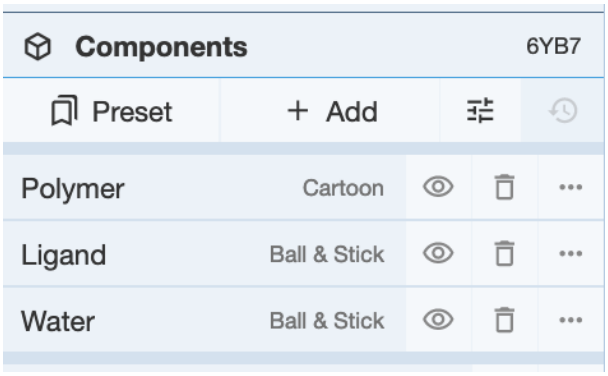
### A1. How many protein molecule(s) do you see?

---

### A2. What are the little red dots around the protein?

---

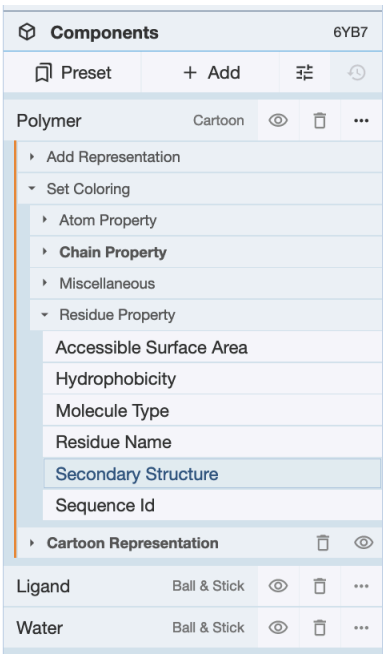
- Explore the structure you are visualizing by showing and hiding its various components*

| | |
|---|---|
|  | You can view and hide the polymer (protein chain), ligands, and waters by clicking on the eye icons in the component panel.<br><br>*Note a component is any grouping of atoms/ residues/ ligands/ chains etc. in the structure. In the default settings, all polymers (protein and nucleic acid chains) are grouped under the Polymer component; all ligands are included in the Ligands component and all water molecules are in the Water component. Additional components (groupings) can be made and added to this list (described later). |
|  | Hide the polymer and water components to reveal the ligands present in the structure. |

## A3. Which ligand(s) is/are present in the Ligand component in this structure? Why do you think it was included in the experiment?

---

- Color the protein by secondary structure (i.e. all helices in one color and all sheets/strands in another color).
- Display the Polymer component and hide the ligands and water components.




Click on the 3-dots in the far right for the polymer to open visualization options >> click on Set Coloring>> Residue properties >> Secondary Structure option in the menu

Orient the structure so that you can see 3 domains:

- Domain I composed of an alpha-beta structure on the top
- Domain II composed of a single beta sheet
- Domain III composed of only alpha helices in the bottom

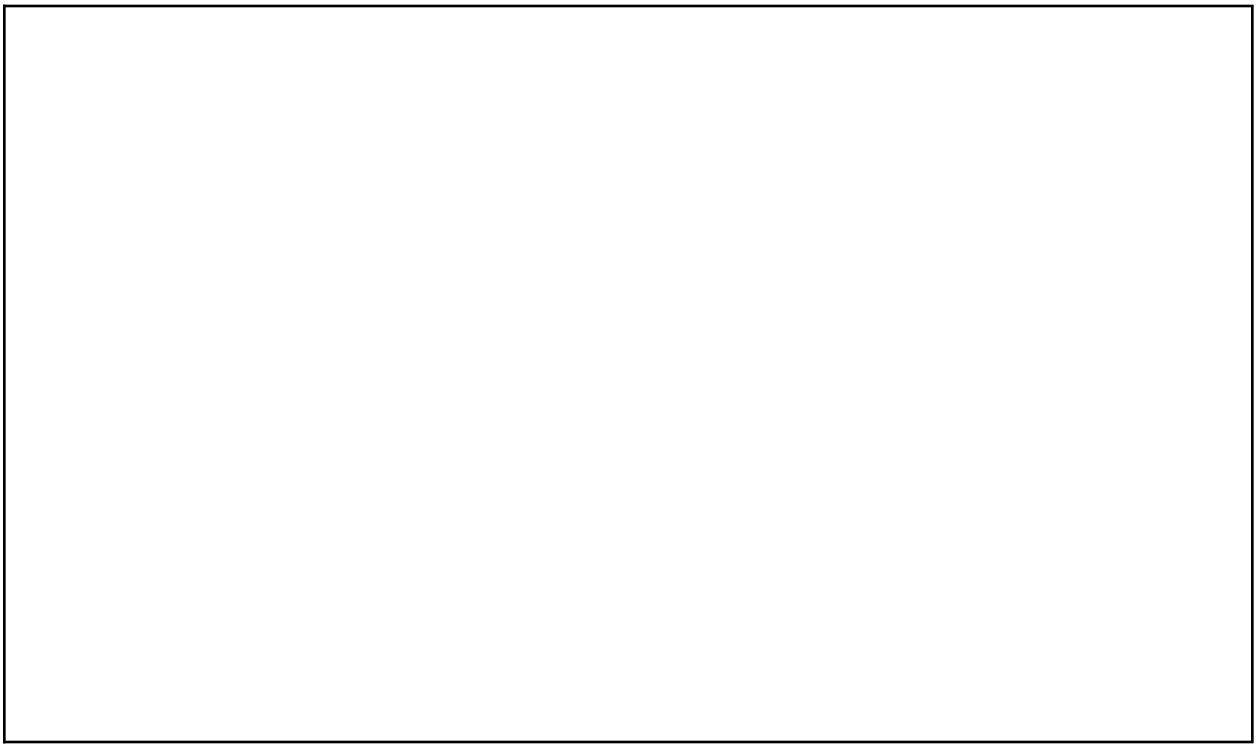
Save and image and include it below:

## A4: How many alpha helices are present in Domain III?

---

- Color the protein by sequence ID (i.e. the N- and C-terminal amino acids are colored blue and red respectively. By convention all amino acids in between the two termini are colored according to the rainbow colors. This coloring can be done using the following steps:
- Click on the 3-dots in the far right for the polymer Polymer … >> Set Coloring >> Residue property >> Sequence ID.

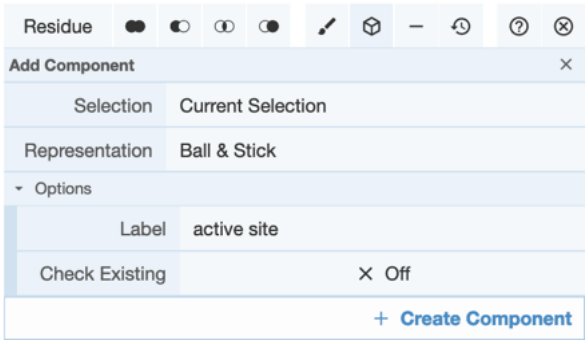
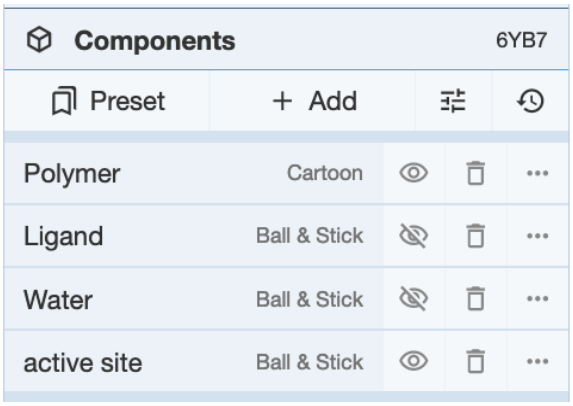
## Save an image of the N- to C- terminal coloring in the same orientation as above (i.e., showing 3 domains I - III). Label the N- and C-termini.



## A5: Which domain has most of the N-terminal residues and which has the C-terminal residues?

---

- Locate the enzyme's active site - i.e., catalytic residues – Cysteine C145 and Histidine H41 as well as a residue that defines the substrate binding pocket (Glutamate E166)

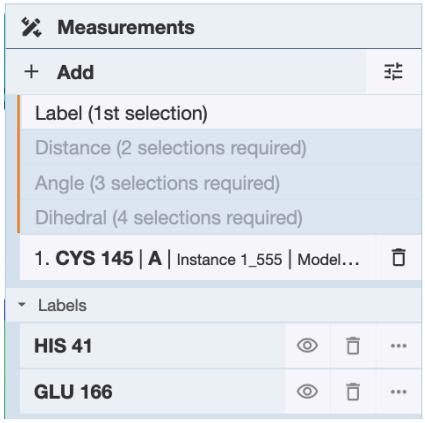
| | |
|---|---|
|  | Activate the selection mode by clicking on the arrow icon in the Toggle panel >> Select the specific residues from the sequence panel >> |
|  | Create a component with this selection by clicking on the Cube icon >> Select Representation Ball and Stick >> Label it “active site” >> click on Create Component. |
|  | This should create a new component “active site” in the Components panel.<br><br>Click on the name of the new component (active site) to focus on it. Click anywhere to deselect.<br><br>The 3 residues forming this component (Cys145, His41, and Glu166) should be displayed in ball and stick representation in the 3D Canvas. |

.

## A6. In which domain is the enzyme active site located?

---

- Label the 3 amino acids selected to create the component called “active site”

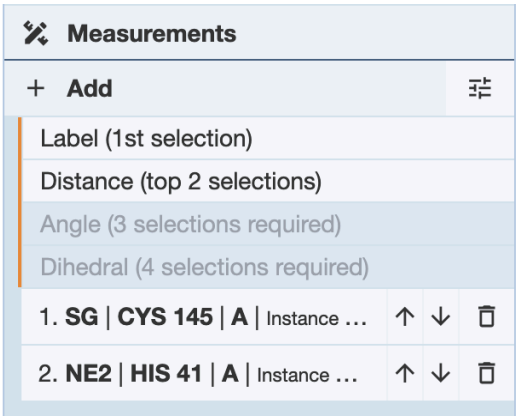
| | |
|---|---|
|  | in the 3D canvas, click on each residue (in the “active site” component) one at a time to label.<br><br>When one residue is selected (highlighted with a green halo) >> click on Measurements >> Add >> Label (1st selection in the selected list) to add the label for that residue<br><br>Repeat these steps for the other two residues too |

## Save the image with all 3 residues labeled and insert it here:

| |
|---|
| |

- Measure distance between residues forming the catalytic dyad - i.e., measure the distance between the S atom of Cys145 and NE2 of His41.

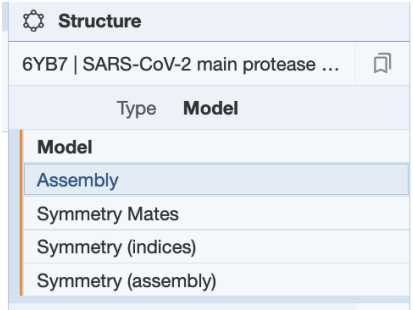
| | |
|---|---|
|  | Change the picking level of selection to Atom by clicking on the box that says "residue" and changing it to "Atom/coarse element" |

| | |
|---|---|
|  | Click on and select the specific atoms in Cys145 and His41 to measure the distance >><br>Click on Measurement >> Add >> Distance (Top 2 selections). |

## A7. What is the distance between the S atom of Cys145 and NE2 of His41?

---

## B) Explore the SARS-CoV-2 Main Protease Assembly

| | |
|---|---|
|  | Go to the Structure section of the Controls Panel >> click on the pull down menu next to the word Type >> select Assembly from the list |

| | |
|---|---|
|  | You should now see dimer loaded to the 3D canvas. |

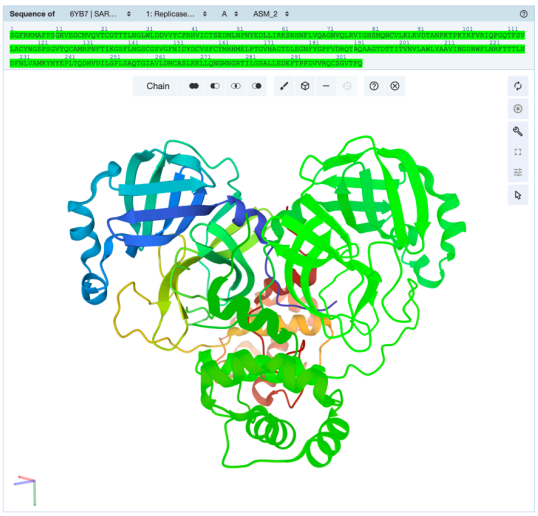
- Using the same steps as before hide the ligands and water components, color the polymer chains using the rainbow color scheme Polymer >> Set coloring >> Residue Property >> Sequence Id. Orient the molecule so that the Domain I is on the top, and Domain III is on the bottom.

## Save an image and paste it below. Label the domains I-III.

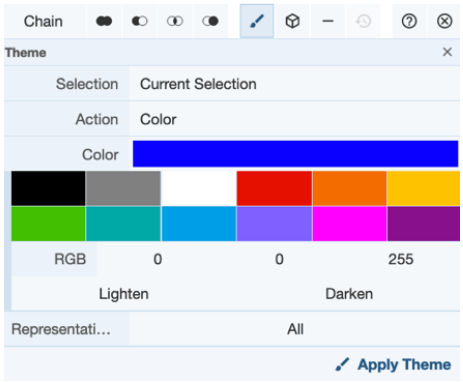
## B1. Which of the 3 domains participate in dimerization? Support your answer with a figure.

---

- In order to see the relationship between the two monomers in this dimer, color the symmetry-related polymer chain gray.




Click on selection button  and select the picking level to be chain >> in the sequence panel's 4th pull down menu select (ASM_2) >> Click anywhere in the sequence panel to select the entire symmetry related chain

| | |
|---|---|
|  | Click on the paintbrush icon to activate the Theme options >> Select the color gray >> Click on Apply theme. |

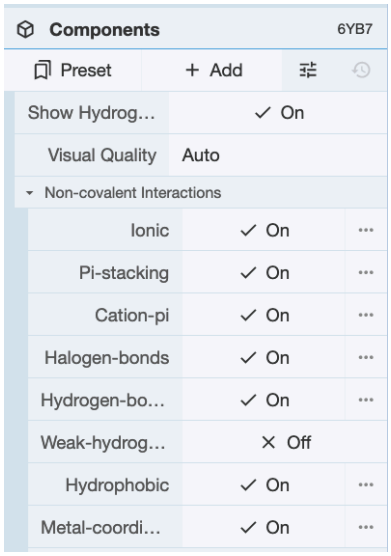
## Save an image here:

- Locate the enzyme active sites in the colored polymer chain (ASM_1) and display the residues Cys145, His41, and Glu166 (see selection and labeling steps described earlier).
  Note: you should change Picking Level back to “Residue”.

## B2. Where is the active site located in relation to the dimerization interface? Support your answer with an image.

- Examine the local interactions of the residue E166.

Return to the default mode by clicking on the arrow icon to end the selection mode >> Click on the residue Glu166 in the 3D canvas or in the sequence panel to show the neighboring residues and interactions between them. Glu166 will have a halo around it to indicate it is the residue of focus.




Click on the options icon in the Components Panel to select the types of non-covalent interactions to view:
Make sure that the types of non-covalent interactions shown include the figure (left).

## Save an image of the interactions:

B3. Identify residues in the vicinity of E166, whose side chains participate in the following types of interactions:

| Type of non-covalent interaction | Participating residues |
|---|---|
| Hydrogen bond | |
| Salt bridge | |
| Pi-stacking | |
| Hydrophobic interactions | |

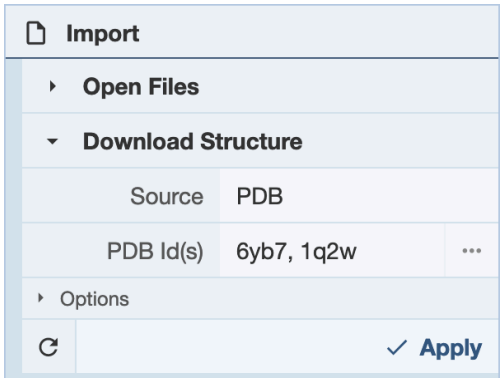
B4. The scientists who studied this structure (Zhang et al., 2020 DOI: 10.1126/science.abb3405) state that “Dimerization of the enzyme is required for catalytic activity.” Based on your analysis of the interactions, can you explain why?

B5. Following an infection, how do you think the first Nsp5 dimer is formed? Hint: See Chen et al., 2010 (doi: 10.1007/s13238-010-0011-4).

---

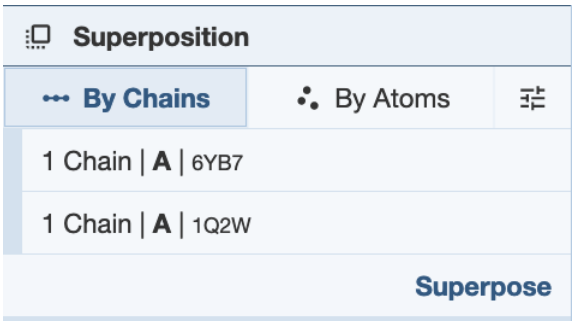
## C) Compare the structures of SARS-CoV (PDB ID 1q2w) and SARS-CoV-2 main protease structures (PDB ID 6yb7)

- Go to the https://www.rcsb.org/3d-view/

| | |
|---|---|
|  | In the Download Structure options >> Type in the PDB IDs in the box next to PDB Id(s) >><br>"6yb7, 1q2w"<br><br>Click on Apply |
|  | Both structures will show up in the 3D-Canvas<br><br>The 2 symmetry related chains of the SARS-CoV-2 protease (PDB Id 6yb7) are colored green<br>The 2 chains of SARS CoV (PDB Id 1q2w) are colored orange and purple |

- Select the chains from each of the structures for superposition as follows:
    - Activate the selection mode by clicking on the arrow icon 

- Change the picking level to “Chain” by clicking on the word “Residue” to see options in the pulldown menu
- Select Chain A in the PDB entry 6yb7 by clicking in the sequence panel
- Change the PDB entry to 1q2w in the menus at the top of the sequence panel
- Select chain A in this structure too
- In the 3D canvas, Chain A of both the structures should be highlighted with a green halo

| | |
|---|---|
|  | In the Superposition section of the control panel click on “By Chains” >> you should see the chains that you selected - chain ID A in both structures listed here<br>Click on Superpose<br>Once the superposition is done click anywhere in the white space in the 3D Canvas to clear selection. |

- Note the rmsd (root mean square deviation) between atoms in these chains listed in the log panel (bottom of the 3D Canvas).

## C1. Save an image of the superposed structures and record the rmsd value below. Do you think that the two structures are similar or different?