SARS-CoV-2 Main Protease

## Learning Objectives: (a) To visualize and explore the structure of the SARS-CoV-2 Main Protease (or Nsp5) using data from the Protein Data Bank and the visualization tool RCSB Mol\*. (b) To compare the structures of the Main Proteases from SARS and SARS-CoV-2

## Introduction

SARS-CoV-2 is a + strand RNA virus. As soon as it enters the host cells it translates polyproteins - i.e., a string of proteins with various activities that are translated as long proteins. The shorter polyprotein includes two protease enzymes that cut themselves out of the translated product and then cut the polyprotein at a number of locations to liberate the other proteins. One of these enzymes is called nonstructural proteins 5 or Nsp5 which makes 11 cuts in the polyproteins. It is also called the Main Protease or the 3C-Like (3CL) protease, since it is like another viral protease (i.e., picornavirus 3C protein).

Exploring the structure of this protein can provide insights into the function of the molecule. If this protein is specifically and effectively inhibited the whole viral life cycle can be stopped.

## Tools

* The exercise will use data from the Protein Data Bank (PDB) and use the RCSB molecular visualization tool Mol\* for the visualization and analysis of SARS-CoV-2 Nsp5.
* To save images, click on the camera (iris) icon A screenshot of a social media post

  Description automatically generated, 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.

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

* Go to RCSB PDB home page ([www.rcsb.org](http://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\*

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | 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\*

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | You can view and hide the polymer (protein chain), ligands, and waters by clicking on the eye icons in the component panel.  \*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). |

|  |  |
| --- | --- |
| A screenshot of a social media post  Description automatically generated | 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.

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | 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 |

### A4. Save an image, label Domains I-III and include the figure below:

|  |
| --- |
|  |

### A5: How many alpha helices are present in Domain III?

### A6: How many strands are present in Domain II? Is the sheet in this domain parallel or antiparallel?

* Color the protein by sequence ID (i.e. by convention the N- and C-terminal amino acids are colored blue and red respectively. 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.

### A7. 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.

|  |
| --- |
|  |

### A8: 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).
* To select the active site residues and display them on the 3D Canvas

|  |  |
| --- | --- |
|  | Activate the selection mode by clicking on the arrow icon A picture containing table  Description automatically generatedin the Toggle panel >> Select the specific residues from the sequence panel (H41, C145, and E166). |
| A screenshot of a cell phone  Description automatically generated | Create a component with this selection by clicking on the Cube icon A close up of a logo  Description automatically generated >> Select Representation Ball and Stick >> Label it “active site” >> click on Create Component. |
| A screenshot of a cell phone  Description automatically generated | This should create a new component “active site” in the Components panel.  Click on the name of the new component (active site) to focus on it. Click anywhere to deselect.  The 3 residues forming this component (Cys145, His41, and Glu166) should be displayed in ball and stick representation in the 3D Canvas. |

### A9. In which domain is the enzyme active site located?

Between the Domains I and II

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

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | in the 3D canvas, click on each residue (in the “active site” component) one at a time to label.  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  Repeat these steps for the other two residues too |

### A 10. Save the image with all 3 residues labeled and insert it here:

|  |
| --- |
|  |

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

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | 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 |
| A screenshot of a social media post  Description automatically generated | 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.

### B1. Save an image, label the domains I-III, and paste it below.

|  |
| --- |
|  |

### B2. Which of the 3 domains participate in dimerization? Support your answer with a labeled figure.

|  |
| --- |
|  |

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

|  |  |
| --- | --- |
| A picture containing screenshot  Description automatically generated | Click on selection button A picture containing table  Description automatically generated 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 |
| A screenshot of a cell phone  Description automatically generated | Click on the paintbrush icon to activate the Theme options >> Select the color gray >> Click on Apply theme. |

### B3. Save an image and paste it below.

|  |
| --- |
|  |

* 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”.

### B4. Where is the active site located in relation to the dimerization interface? Support your answer with a suitably labeled image.

|  |
| --- |
|  |

* Examine the local interactions of the residue E166 to see which amino acid residues are located in its neighborhood and how they interact with each other.

Return to the default mode by clicking on the arrow icon to end the selection mode.

Now click on the residue Glu166 to zoom in and visualize the non-covalent interactions in the neighborhood. You can either click on the residue in the 3D canvas or in the sequence panel. Glu166 will have a halo around it to indicate it is the residue of focus and the non-covalent interactions in the neighborhood can be seen. All the non-covalent interactions will be indicated as different colored dashed lines.

|  |  |
| --- | --- |
|  | If you mouse over any of these lines, the type of interaction and residues involved in the interaction are listed at the bottom right corner of the 3D Canvas page. |

* Make sure you are visualizing all the relevant non-covalent interactions (see below)

|  |  |
| --- | --- |
| A screenshot of a cell phone  Description automatically generated | 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). |

### B5. Save an image of the interactions and paste it below.

|  |
| --- |
|  |

### 

### B6. 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 |  |

### B7. The scientists who studied this structure (Zhang et al., 2020 [DOI: 10.1126/science.abb3405](https://doi.org/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?

### C) Compare the structures of the Main Protease from SARS-CoV and SARS-CoV-2

* In this section we will superpose the structures of the main protease from SARS-CoV (PDB ID 1q2w) and SARS-CoV-2 (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) >>  “6yb7, 1q2w”  Click on Apply |
|  | Both structures will show up in the 3D-Canvas  The 2 symmetry related chains of the SARS-CoV-2 protease in PDB Id 6yb7 (Chain A ASM\_1 and Chain A ASM\_2) are colored green  The 2 chains of SARS CoV in PDB Id 1q2w (Chains A and B) - 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 A picture containing table

    Description automatically generated
  + Change the picking level to “Chain” by clicking on the word “Residue” to see options in the pulldown menu

The goal here is to align the chains labeled A from PDB ID 6yb7 and and PDB ID 1q2w.

* + 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  Click on Superpose  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? What does this mean in terms of evolutionary relationships between SARS-CoV and SARS-CoV-2?

RMSD 0.78 Å;

the structures are very similar suggesting that the protease is highly conserved in these two viruses

|  |
| --- |
|  |

**Appendix: Mol\* Quick Reference**

**1. Navigate the 3D Canvas:**

***a. Rotate***

● Press left mouse button and move OR use Shift + left mouse button and drag.

***b. Translate***

● Press right mouse button and move OR use Control + the left mouse button and move. On a touchscreen device, use a two-finger drag.

***c. Zoom***

● Use the mouse wheel. On a touchpad, use a two-finger drag. On a touchscreen device, pinch two fingers.

***d. Center and zoom***

● Use right mouse button to click on the part of the structure you wish to see.

***e. Change clipping planes***

● Use Shift button + the mouse wheel. On a touchpad, use the Shift button + a two-finger drag.

**2. Select:** first open Selection Mode and change the Picking Level (if needed)

***a. Select Picking Level***

● Click on objects in the 3D canvas – such as atoms, residues, chains, etc.

***b. Select object in 3D canvas***

● Click on residues, chains etc. in the 3D canvas based on picking level

***c. Select object from Sequence Panel***

● Click on residues, ligands, or entire chain in the Sequence Panel

***d. Custom Select combinations***

● Use the Set Operations Menu in the Selection Mode toolbar

**3. See or Hide:**

***a. To add representations***

* Create a component of the region you wish to see/hide → Go to the Components Panel and press the “eye” icon next to the component you create

***b. To hide/remove from view***

● Select region you wish to hide → Click on the subtract/hide icon in the Selections toolbar

**4. Color:**

***a. N-terminus to C-terminus (rainbow)***

● Components → Polymer → Set Coloring → Residue Property → Sequence Id

***b. Heteroatom***

● Components → Polymer → Set Coloring → Atom Property → Element Symbol

***c. Secondary structure***

● Components → Polymer → Set Coloring → Residue Property → Secondary Structure

***d. Hydrophobicity***

● Components → Polymer → Set Coloring → Residue Property → Hydrophobicity

***e. Domain***

● Select domain → Selections Menu → Apply Theme to Selection → Color → Apply Theme

**5. Compare Structures:** first upload two or more structures at *rcsb.org/3D-view*

***a. By chains***

● Select 2 or more polymer chains/residues → Superposition → By Chains → Superpose

***b. By atoms***

● Select 1 or more atoms → Superposition → By Atoms→ Superpose

**6. Make Measurements:**

***a. Distance***

● Make 2 or more selections → Measurements → Add → Distance (for first 2 selections)

***b. Angle***

● Make 3 or more selections → Measurements → Add → Angle (for first 3 selections)

***c. Dihedral***

● Make 4 or more selections → Measurements → Add → Dihedral (for first 4 selections)