Tighter Binding Spike Protein

## Learning Objectives: To visualize and explore the structures of the SARS-CoV-2 Spike protein, ACE2 receptor, and their binding using data from the Protein Data Bank and the RCSB Mol\* visualization tool

## Introduction

SARS-CoV-2 Spike protein makes a crown-like structure around the virus. These proteins attach to specific human cellular receptors (ACE2) to begin the infection. Exploring the structure of this protein can provide insights how the virus gains entry into host cells. By visualizing and understanding the nature of interactions between the spike and ACE2 proteins, options for preventing infection and treatment of COVID-19 can be designed. The Receptor Binding Domain is one of the main targets for the current vaccine development, antibodies from convalescent serum also bind to the Spike to prevent further reinfection of host cells.

## 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.

## Explorations

### Structure of SARS-CoV-2 Spike Protein

* Search for SARS-CoV-2 Spike protein structures in the PDB

|  |  |
| --- | --- |
|  | In the Top Search Box type “ Spike” and click on the magnifying glass or search icon |
|  | Once the results are returned - examine the source organisms listed in the left hand menu.  From this list select the name of SARS-CoV-2 - as shown.  Again click on the search icon to refine the search results to be only Spike proteins that belong to this organism. |

* Review the structures of SARS-CoV-2 Spike protein you have found. The protein adopts many conformations.

### A1. List the PDB IDs of one open and one closed state structure of the Spike protein? (Hint: look among structures of the Spike protein only)

* Go to RCSB PDB home page ([www.rcsb.org](http://www.rcsb.org)) >> type the PDB ID “6vxx” 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.

### A2. How many protein molecule(s) do you see?

### A3. What are the little blue boxes spread all over the structure?

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

|  |  |
| --- | --- |
|  | You can view and hide the polymer (protein chain), and carbohydrates 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). |

* Color the protein by secondary structure (i.e. all helices in one color and all sheets/strands in another color).

|  |  |
| --- | --- |
|  | 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 most of the beta sheets in the top part of the 3D-Canvas. |

### A4. Save an image, label the position of the membrane with a few lines representing the membrane width and include the figure below:

|  |
| --- |
|  |

### A5: Where are all the alpha helices located in the structure?

* In the top search box >> type the PDB ID “6vsb” in the top search box >> click enter. This should take you to the Structure Summary Page of this PDB structure.
* Review contents of the page and complete the following table about the entry.

|  |  |
| --- | --- |
| PDB ID |  |
| Author(s) of entry |  |
| Year when the structure was published/released |  |
| Structure determination method |  |
| Number of protein chains in the entry; protein name, Chain ID |  |
| Names and number of copies of ligands (Small Molecules) present in the structure |  |

* Click on the 3D view tab at the top of the structure summary page to view the structure interactively.
* 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.
* Orient the structure so that the C-termini (colored red) are positioned at the bottom and the N-termini (blue and green) are positioned towards the top.

### A6. Save an image, mark any part of the structure that looks different in this structure compared to that in PDB ID 6vxx. Include the figure below:

|  |
| --- |
|  |

* The UniProt entry for the Spike protein (<https://www.uniprot.org/uniprot/P0DTC2>) lists the Receptor Binding Domain (RBD) of the Spike protein (part of the protein that binds to the receptor protein ACE2) as the amino acids between 319 and 541. Locate this region of the protein in Chain ID A.

|  |  |
| --- | --- |
|  | 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 319 - 541 by clicking and dragging in the sequence panel. |
|  | Once these residues are selected they will be shown in bright green in the 3D Canvas.  Click on the paintbrush icon to activate the Theme options >> Select the color magenta >> Click on Apply theme. |

### A7: Save an image of the structure and label the Spike protein RBD?

|  |
| --- |
|  |

### A8. What is the difference in the RBD in the structures you have examined (PDB IDs 6vxx and 6vsb)?

### B. Structure of human ACE2 Receptor Protein

* In the top search box >> type the PDB ID “1r42” in the top search box >> click enter. This should take you to the Structure Summary Page of this PDB structure.
* Review contents of the page and complete the following table about the entry.

|  |  |
| --- | --- |
| PDB ID |  |
| Author(s) of entry |  |
| Year when the structure was published/released |  |
| Structure determination method |  |
| Number of protein chains in the entry; protein name, Chain ID |  |
| Names and number of copies of ligands (Small Molecules) present in the structure |  |

|  |
| --- |
| *Additional Notes about ACE2:*  ACE2 is a membrane bound carboxypeptidase of the renin-angiotensin hormone system. It is a critical regulator of cardiovascular homeostasis. It is expressed in a large number of cells including lung epithelial cells, enterocytes of the small intestine, as well as heart, kidney, testis, and gastrointestinal system. It is coexpressed with TMPRSS2 within lung alveolar type 2 cells. Note both these proteins play key roles in SARS-CoV-2 Spike protein binding and initiation of infection.  The extracellular region of the ACE2 enzyme has two domains - a zinc metallopeptidase domain (amino acid residues 19-611), and a C-terminus region (amino acid residues 612-740) that is similar to a protein called collectrin and is responsible for the trafficking of the neutral amino acid transporter.  The C-terminal domain was not fully ordered in the structure and only parts of it could be seen in the experimental data. Since the authors were unable to determine which amino acids were seen, they built in short peptide pieces composed of unknown amino acids (1 letter code X). For this discussion this region of the protein is not relevant so should be hidden or removed from the components. |

* Click on the 3D view tab at the top of the structure summary page to view the structure interactively. Hide the disordered Collectrin homology domain chains because they are not relevant to this structure exploration.

|  |  |
| --- | --- |
|  | Activate the selection mode by clicking on the arrow icon A picture containing table  Description automatically generatedin the Toggle panel >> change the selection level to Chain. |
|  | For each protein called “disordered segment of collectrin homology domain” click on the sequence (a number of Xs) to select it. The corresponding protein chain is highlighted in the 3D Canvas |
|  | Click on the minus icon to remove the chain from all components. This chain should now disappear from display. |

* The UniProt entry for the ACE2 enzyme (<https://www.uniprot.org/uniprot/Q9BYF1>) lists the catalytic residues as H374, H378 and E402.

|  |  |
| --- | --- |
|  | Activate the selection mode by clicking on the arrow icon A picture containing table  Description automatically generatedin the Toggle panel. Make sure that the selection level is set to Residue >> Select the specific residues from the sequence panel (H374, H378, and E402).  Create a component with this selection by clicking on the Cube icon A close up of a logo  Description automatically generated >> Select Representation Spacefill >> Label it “catalytic site” >> click on Create Component.  Click on the name “Catalytic site” in the component panel to focus on these residues. |

### B1. Orient the structure so that you are looking down at the 3 amino acid residues. Save an image, label the amino acids, and include it below. What is the purple sphere in the middle of these amino acids?

|  |
| --- |
|  |

* The Pathology and Biotech section of the Uniprot page lists 2 amino acids (K31 and K353) that if mutated can abolish the SARS Spike protein from binding. Using similar steps as above, visualize these amino acids, display them in spacefill representation.

### B2. Save an image of the ACE2 structure showing the relationship between the catalytic and SARS binding residues in ACE2.

|  |
| --- |
|  |

### C. Binding of SARS-CoV-2 Spike to human ACE2 receptor

* In the top search box >> type the PDB ID “6m0j” in the top search box >> click enter. This should take you to the Structure Summary Page of this PDB structure.
* Review contents of the page and complete the following table about the entry.

|  |  |
| --- | --- |
| PDB ID |  |
| Author(s) of entry |  |
| Year when the structure was published/released |  |
| Structure determination method |  |
| Number of protein chains in the entry; protein name, Chain ID |  |
| Names and number of copies of ligands (Small Molecules) present in the structure |  |

* Click on the 3D view tab at the top of the structure summary page to view the structure interactively.

### C1. Save an image of the structure you see and label the two proteins in the structure - Spike protein and ACE2. Include the labeled image below.

|  |
| --- |
|  |

### C2. Do the structures of the Spike protein and ACE2 receptor protein look like what you saw in the sections A. and B. of this exercise? Explain your answer.

* Examine the 2 amino acids critical for binding the SARS virus (the SARS Spike protein) in this structure.

|  |  |
| --- | --- |
|  | Make sure that the selection mode is not activated.  In the sequence panel identify the amino acid residue Lysine or Lys or K 31 in ACE2 and click on it.  In the 3D canvas this amino acid should be highlighted and all interactions in its neighborhood should be displayed. |
|  | To make sure that you are seeing all the non-covalent interactions in this region >> 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). |

### C3. Save an image of the non-covalent interactions around ACE2’s K31. Label any 2 different amino acids that forms a bond with this residue. Include the labeled image in the box below.

|  |
| --- |
|  |

* Examine the interactions of residue Lysine or Lys or K 353. Repeat the same steps as done for K31

### C4. Save an image of the non-covalent interactions around ACE2’s K353. Label any 2 different amino acids that forms a bond with this residue. Include the labeled image in the box below.

|  |
| --- |
|  |

### C5. Based on the above structure explorations do you think that SARS and SARS-CoV-2 bind to the same or different sites on ACE2? Explain your answer.

### D. Comparing binding of SARS-CoV and SARS-CoV-2 to the ACE2 Receptor

Both SARS (from the 2003 epidemic) and SARS-CoV-2 (from the 2020 pandemic) enter host cells by binding to the ACE2 receptor protein. However, the SARS-CoV-2 became a pandemic. Here we will compare the sequences, structures, and interactions of Spike RBDs from both viruses, when bound to ACE2 to see if we can gain some insights. We will use information from the PDB IDs 6m0j and 2ajf.

|  |
| --- |
| *Additional Notes about BLASTp*  BLAST or Basic Local Alignment Search Tool can be used to compare the sequences of nucleic acids (DNA, RNA) and proteins.  The BLASTp program takes a sequence of amino acids and compares this sequence to the existing database of millions of sequences to find a match.  In simple terms, the BLAST program uses an algorithm that searches ‘words’ of short amino acid sequences against the database. Matches are scored based on how similar the physicochemical characteristics of the corresponding amino acids are between the searched “word” and the prospective “match” word and then the search is repeated with another ‘word’.  In addition to finding sequences with similarity, the BLAST program will provide the alignment between two or more given sequences. The first sequence is referred to as the query and the sequence matched to it is called the subject. |

|  |  |
| --- | --- |
|  | Go to the NCBI BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and click the Protein Blast box. |
|  | In the new page that opens you can paste your query sequence. |
|  | Write 6m0j\_E in the top box. If a second box is not open, check on the align 2 sequences option and type in 2ajf\_E in the second box. Run the search by clicking on the BLAST button at the bottom of the page. |

* Examine the results page and click on the Alignment tab.

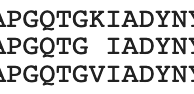
### D1. Copy the sequence alignment and paste it below. Make sure that you paste it using Courier or Courier New font, size 9.

|  |
| --- |
|  |

### D2. What is the level of sequence identity between these two proteins?

D3. Examine the regions where the sequences do not match and highlight amino acid residue pairs where a hydrophobic amino acid in one protein is aligned with a charged amino acid in the other protein. Mark at least two such pairs in the sequence alignment saved in Ans D1.

* One such pair that is different between the SARS and SARS-CoV-2 Spike RBDs is
  + K417 in SARS-CoV-2 (PDB ID 6m0j) and V404 in SARS (PDB ID 2ajf).
  + The neighboring residues are identical in both chains



* In the sequence panel for each of these structures’ Mol\* displays locate the residue and click on it to view its interactions. (Make sure that the selection mode is not activated)

D4. Save images of the interactions of K417 in PDB ID 6m0j and V404 in PDB ID 2ajf. Label the residues of interest and the ones it interacts with (as shown by the dashed lines).

|  |  |
| --- | --- |
| PDB ID 6m0j | PDB ID 2ajf |
|  |  |

### D5. In both the structures the viral Spike RBD is bound to ACE2. In their respective structures where are the residues of interest (i.e., K417 in PDB ID 6m0j and V404 in PDB ID 2ajf) located with respect to the ACE2 proteins.

### D6. If all other interactions were identical based on this analysis which of the proteins would bind ACE2 more tightly - SARS RBD or SARS-CoV-2 RBD?

**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)