# COVID-19: Molecular Basis of Infection Student Worksheet

Didem Vardar-Ulu

Chemistry Department, Boston University, Boston MA 02215
Shuchismita Dutta (contact: sdutta@rcsb.rutgers.edu)
Institute of Quantitative Biomedicine, Rutgers University, Piscataway NJ 08854

# Part 1: Quarantine, Social Distancing, and Stay Home Orders ...

## The COVID-19 Pandemic

On December 31, 2019, China reported a cluster of pneumonia cases in Wuhan, Hubei Province, caused by a novel coronavirus, later named SARS CoV-2, (World Health Organization, WHO). Within two weeks, reports of infection and resulting mortalities began coming in from Thailand, US, Japan, South Korea, Iran, and Italy. Concerned by the alarming levels of spread and severity of this infection, WHO declared this outbreak as the COVID-19 pandemic on March 11, 2020. In the first three months after COVID-19 emerged, nearly 1 million people were infected and 50,000 died.

Data from China, where the epidemic began, showed that quarantine, social distancing, and isolation of infected individuals can help contain the spread. So, governments of various countries around the globe started promoting social distancing, issuing stay home orders, and ordering lockdowns. By the end of March, most countries in the world had implemented travel bans and its citizens were in some form of lockdown. The goal of these community based measures was to mitigate the epidemic by "flattening the curve", i.e., delay the epidemic peak, reduce the number of infected individuals, and allow time for treatments and prevention strategies to be developed.

Optional: For a more detailed introduction read "Features, Evaluation and Treatment Coronavirus (COVID-19)" (https://www.ncbi.nlm.nih.gov/books/NBK554776/#article-52171.s15)

#### Anatomy of SARS CoV-2

Coronavirus is so named because it has an outer corona or crown formed by the Spike protein. The SARS CoV-2, was named after a similar virus that caused the Severe Acute Respiratory Syndrome (SARS) in 2002. The SARS CoV-2 is an enveloped virus, and its genetic material is a single positive-stranded RNA (Figure 1). The viral genome codes for (a) structural proteins such as the spike, matrix, envelope, and nucleocapsid proteins; (b) enzymes such as proteases, and RNA-dependent-RNA polymerase; and (c) 16 non-structural proteins that play different roles in infection, and evasion of host immune surveillance.

Q1. Visit David Goodsell's painting of the anatomy of the Coronavirus (<a href="https://pdb101.rcsb.org/sci-art/goodsell-gallery/coronavirus">https://pdb101.rcsb.org/sci-art/goodsell-gallery/coronavirus</a>) and use the provided information to annotate Figure 1. Your annotation should include

- labels for the Spike protein (S) proteins, Viral envelope (E), and any two other viral proteins
- a figure legend that describes the main functions of the elements you labeled within the virus.



Figure 1: Painting of SARS CoV-2 by David Goodsell, 2020 (https://pdb101.rcsb.org/sci-art/goodsell-gallery/coronavirus)

Q2. Watch the video <a href="https://pdb101.rcsb.org/learn/videos/fighting-coronavirus-with-soap">https://pdb101.rcsb.org/learn/videos/fighting-coronavirus-with-soap</a> and provide a 5-6 sentence summary about how soap treatment impacts virus structure and provides an effective prevention against coronavirus infection.

Make sure that your summary highlights the structure function relationship between the chemical properties and structure of the molecules involved and their function within the biological system.

#### Life Cycle of SARS CoV-2

Like any other virus the SARS CoV-2 virus does not have its own machinery to produce biological macromolecules (e.g., nucleic acids and proteins). It must infect a host cell and hijack its cellular machinery for replication.

Learn about the viral life cycle as follows:

- Watch the video "How does a virus replicate in a cell"
- Examine Figure 2 illustrating key steps in the coronavirus life cycle

Also review the life cycle at <a href="http://pdb101.rcsb.org/sci-art/goodsell-gallery/coronavirus-life">http://pdb101.rcsb.org/sci-art/goodsell-gallery/coronavirus-life</a>

<u>life-cycle</u>

Figure 2. Life cycle of a virus from infection (entry into host cells) to release of new viral particles. (See

https://www.nature.com/articles/nrmicro775.pdf for additional details).

Below is a table summarizing the information included in the video and the figure you studied in a way that highlights the roles of the virus and viral proteins and the roles of the host and its proteins in the COVID-19 viral infection.

Life Cycle Step	Viral Protein or role of the virus in the process	Host protein or role of the host in the process	Description of the Process
Viral attachment and infection	Spike Protein	ACE2 protein	The two proteins interact on the lung cells. Viral and host cell membranes fuse to release genetic material in cell.
Replication of viral genome	RNA-dep-RNA polymerase and other viral enzymes such as helicase, RNA primase, exo- and endo- ribonuclease	Translation of viral proteins	Positive strand genomic RNA serves as template for negative strand, which in turn is the template for the positive strand of mRNA within the host cell.
Viral assembly and release	Nucleocapsid Assembly	Host translation of viral proteins	The Spike, Matrix and other viral membrane proteins are synthesized in the ER; various viral enzymes are translated in the cytoplasm

The Central Dogma in biology describes the flow of genetic information from DNA to RNA (via transcription) to protein (via translation), while the DNA itself is maintained through generation via replication.

Q3. The genetic material of Coronavirus is a positive single stranded RNA. Since there is no DNA stage in the life cycle of this virus, what special abilities does this virus have to replicate its genetic material? Hint: Refer to the table above to review what components are necessary for each step of the viral life cycle and how they are provided to the system.

#### Part 2: Beginning of the Infection

The first step in the viral life cycle, infection, begins with the SARS CoV-2 Spike protein binding a host receptor protein (Angiotensin Converting Enzyme 2 or ACE2 protein on lung cells). In this section we will explore these proteins separately and then look at a structure of them interacting with each other to understand the molecular basis of this infection.

Watch the ACS Reactions video <a href="https://www.youtube.com/watch?v=gDY8pH6OWBc">https://www.youtube.com/watch?v=gDY8pH6OWBc</a> for an introduction to the two proteins that play a key role in the SARS CoV-2 infection – the viral Spike protein and the human ACE2 (receptor) protein.

## The Viral Spike Protein

The SARS CoV-2 Spike glycoprotein is over 1200 amino acids long. Explore the structure of the protein as determined by electron microscopy and discussed in the video you watched (PDB ID 6vsb). The focus in this exploration will be to:

- 1. Learn about the overall assembly and domain organization of the spike protein (domains).
- 2. Identify the domain within the spike protein that binds to the ACE2 protein.

Go to www.rcsb.org and enter PDB ID: 6vsb in the top search box.

Q1. Use the information presented on the Structure summary page that comes up to complete the following table:

PDB ID	6vsb
Author(s) of entry	
Year when the structure was	
published/released	
Structure determination method	
Number of protein chains in the	
entry (Note: This information is	
included under the	
"Macromolecules" Section)	
Names and number of copies of	
ligands (Small Molecules) present	
in the structure (Note: This	
information is included under the	
"Small Molecules" Section)	

- Visualize the structure of the Spike protein as follows:
  - o Go to the iCn3D website at https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html
  - Click on the button called File >> Retrieve by ID >> PDB ID so that a new window opens. Input the PDB ID of the structure you wish to visualize and click on Load.
  - o The structure opens in a new tab
  - Click on the tab "Windows" (next to help) and click "View sequences and Annotations". This will bring up an interface where you can access specific information about the components of the structure you are viewing on the left.
  - Use the displayed information and your own conclusions from rotating and examining the molecule to answer the following questions

Q2. How many protein chains do you see? How are they identified? (i.e. what is the label/name given to them, what color and depiction style- space filling/ stick/ ribbon is used for visualization?)

- Q3. Orient the structure so that the C-termini of the protein chains are at the bottom of the page. Take a screenshot of the structure and paste it below. Annotate the figure by labeling the termini and the different chains you see.
- To visualize only a single chain (A chain) of the Spike protein and explore it, you need to "hide" the parts of the molecule you do not want to visualize using the options under the different tabs at the top.
  - Click on the Select tab (next to "File" on the top) >> Defined sets
  - In the new window that opens Select the chains B and C and chemicals (i.e with the shift button pressed select from 6vsb\_B to 6vsb\_C and chemicals)
  - Hide the selected chains by clicking on Style>>Proteins >> Hide;
  - Hide the selected chemicals by clicking on Style>>Chemicals >> Hide;
  - Hide all side chains by clicking Style >> Sidechain >> Hide;
  - o Hide all disulfide bonds by clicking View >> Disulfide bonds >> Hide.
  - Clear selections by clicking on the button Select >> Clear Selection.
     Now you should only be viewing the A chain of the spike protein.
  - Select chain 6vsb\_A from the "Select Sets" window and color the chain using the Color
     Spectrum option. Note that the spectrum coloring option colors the molecule from violet (N-terminus) to Red (C-terminus)
  - Click on the "Details" tab of the Sequences & Annotations window displayed on the right of the "graphics" window that hosts the molecular view. By clicking and dragging on specific amino acids in the sequence window you can select and view specific amino acids in the graphics window.

UniProt of the Spike protein (<a href="https://www.uniprot.org/uniprot/P0DTC2">https://www.uniprot.org/uniprot/P0DTC2</a>) 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 R319 and F541.

- Q4. Make an annotated figure for the receptor binding domain (RBD) of the CoV-2 Spike Protein using the following instructions.
  - Select the receptor binding domain in the spike protein seen in the chain A (PDB ID 6vsb) by clicking on residue R319 on the displayed sequence in the "Sequences and Annotations" window and dragging your cursor over the entire sequence until you reach F541. Selected residues will be highlighted in yellow and will appear yellow on the graphics screen. Please note the following:
    - Clicking a second time on any residue will "unselect/ unhighlight" it.
    - You can always highlight fragments or single residues if you are having trouble selecting the entire range in one move.
    - You may realize that some of the residues are shown in lower case letters as opposed to uppercase and are not highlighted when you click on them. These are regions in the sequence for which accurate atomic coordinates are not available experimentally and hence cannot be visualized on the graphics window.
  - Color the selection magenta (click on Color >> Unicolor >> Magenta).
  - Orient this domain to be positioned at the top of the figure, save the image and paste it below. Label the following in the image
    - N- and C-termini
    - Receptor binding domain (RBD)
    - where you think the viral envelop (or membrane) is located in this image.

The Host Receptor: ACE2

The ACE2 protein is a membrane bound Carboxypeptidase, a protease that cleaves amino acids from the C-terminus of proteins, in the presence of a zinc ion. Explore the structure of the catalytic domain of this protein as determined by X-ray crystallography (PDB ID 1r42). The focus in this exploration will be to:

- 1. Learn about the overall structure of the ACE2 protease domain.
- 2. Identify the domain that binds to the SARS CoV-2 (and SARS) Spike proteins.

Go to www.rcsb.org and enter PDB ID: 1r42 in the top search box.

Q5. Use the information presented on the Structure summary page for PDB ID: 1r42 to complete the following table:

PDB ID	1r42
Author(s) of entry	
Year when the structure was	
published/released	
Structure determination method	
Number of copies, Name of	
protein chains, Chain IDs in the	
entry	
Names and number of copies of	
ligands (Small Molecules)	
present in the structure	

- Go to the iCn3D website at <a href="https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html">https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html</a> and retrieve ACE2 protein structure (1r42)
- Using the same steps as you did for the visualizing the spike protein, visualize the structure of the ACE2 protein structure in PDB ID 1r42 using iCn3D.
- To focus on only on the ACE2 chain, hide chains B-E which represent the disordered segment of collectrin homology domain.
  - Click on the Select tab >> Defined sets
  - o In the new "Select Sets" window, select the chains B through C by clicking on them while holding the shift button.
  - Go to Style>>Proteins>>Hide
  - Go to "Select Sets" window, select chemicals
  - Go to Style>>Chemicals>>Hide
  - Go to View>>Disulfide bonds>>Hide
  - Go to Select >> Clear Selection to clear all your selections.
  - Go to "Select Sets" window, select 1R42 A
  - Color the chain using the Color >> Spectrum option.
  - o Go to Select >> Clear Selection to clear all your selections.
  - Locate the two termini of the displayed chain and rotate the molecule so that the N-terminus is at the upper left corner and C-terminus is at the bottom of your graphics window. Note that the spectrum coloring option colors the molecule from violet (N-terminus) to Red (C-terminus)

UniProt lists the active site residues for the ACE2 enzyme as E375 and H505 (<a href="https://www.uniprot.org/uniprot/Q9BYF1">https://www.uniprot.org/uniprot/Q9BYF1</a>). It also lists 2 amino acids that if mutated can abolish the SARS Spike protein from binding (K31 and K353) in the Pathology and Biotech section.

- Click on the "Details" tab of the Sequences & Annotations window
- Locate and select the active site (E375-H505) and SARS spike protein binding residues (K31 and K353) by clicking and dragging on these residues in the sequence window.
- Display the side chains of the selected residues from the "Style" tab.
   Click Style >> Side chain >> Ball and Stick.
   Go to Select >> Clear Selection to clear all your selections.

Q6. Make an annotated figure showing the active site and theCoV-2 Spike Protein binding site of ACE2 by saving the image you created above and pasting a copy below.

- a. Make sure that your saved image clearly displays the side chains in the active site and the binding residues in ball and stick representation.
- b. Label the enzyme's active site by drawing a circle around the region occupied by the sidechains you turned on.
- c. Assuming that the SARS CoV-2 Spike protein binds in the same location as the SARS CoV Spike protein (a region including the two critical lysine residues you highlighted) indicate a likely binding interface for SARS CoV-2 Spike protein on ACE2 by draw an ellipse around that region and label the two known binding residues.

#### Viral Attachment and Entry

The first step in viral infection is attachment to the host cell receptor protein. In the case of SARS CoV-2, the viral Spike protein binds the ACE2 extracellular domain. Examine the structure of the SARS CoV-2: ACE2 complex (PDB ID 6m0j).

Go to www.rcsb.org and enter PDB ID: 6m0j in the top search box.

Q7. Use the information presented on the Structure summary page for PDB ID: 6m0j to complete the following table:

PDB ID	6m0j
Author(s) of entry	
Year when the structure was	
published/released	
Structure determination method	
Number of protein chains in the	
entry	
Names and number of copies of	
ligands (Small Molecules) present	
in the structure	

- Visualize the structure of the SARS CoV-2 Spike:ACE2 complex in PDB ID 6m0j using iCn3D.
  - Go to Select >> Defined Sets
  - Select in the "Select Sets" window the three blue 6MOJ-E entries.
  - Click Color >>Unicolor>>Yellow
  - Go to Select >> Clear Selection
  - Select in the "Select Sets" window "chemicals".

Go to Style>>Chemicals>>Hide

Now you will examine the charge based interactions between the Spike and ACE2 protein as follows:

- Click on Select >> Advanced
- In the new box that opens type exactly as written ":K,R" in the Select box and "KR" in the Name box to select all K &R residues in the structure
- Click on "Save Selections to Defined Sets" button.
- o Go to Style >> Side chain >> Stick to display the side chains.
- o Go to Color >> Unicolor >> Blue to show all K and R residues in blue
- Go to Select >> Clear Selection
   Repeat the process to select all the acidic side chains in the structure
- Click on Select >> Advanced
- In the new box that opens type exactly as written ":D, E" in the Select box and "DE" in the Name box
- Click on "Save Selections to Defined Sets" button.
- Go to Style >> Side chain >> Stick to display the side chains.
- Go to Color >> Unicolor >> Red to show all D and E residues in red
- Go to Select >> Clear Selection
- Go to View >> H Bonds & Interactions
- Unclick H-bonds and Click Salt Bridges in the first part of the window.
- o In the second part for "Select first set" click all the pink (6MOJ\_A related) entries and in the third part for "Select second set" click all yellow (6MOJ\_E related) entries.
- o In the fourth part, Click on Display.
- Close the window (using the x on the upper left)
- Go to Select >> Clear Selection
- Look for any Salt Bridges formed between the Spike and ACE2 proteins. Hint: Focus
  on the interface between the two protein chains. The interaction(s) is (are) shown as
  green dashed lines.
- Q8. Make an annotated figure using a saved image from your iCn3D visualization of the SARS CoV-2:ACE2 co-structure. How many salt bridges do you see between the SARS CoV-2 and ACE2 proteins? Circle the salt bridge(s) and the residues forming the salt bridge(s).
- Q9. What are some other types of non-covalent interactions that you may see between the SARS CoV-2 Spike protein and ACE2? (Hint: Think back at the hemoglobin case study. What were some other interactions you visualized?). List at least two of these interactions and briefly discuss which non-covalent interaction (including the salt bridge you considered for Q8) you expect to be the strongest at the virus-receptor interface and why.

#### Part 3: How novel is SARS CoV-2? Molecular Basis of the COVID-19 Pandemic

Both SARS coronavirus (SARS-CoV) and SARS CoV-2 begin the viral infection by binding to the same host receptor protein ACE2. The SARS-CoV caused a severe viral respiratory illness and led to an epidemic in 2002-2003. The SARS CoV-2 led to the COVID-19 pandemic. In the Science article published on March 13, 2020 by Wrapp et al. it was reported that SARS CoV-2 bound ACE2 10 fold tighter than SARS-CoV.

Here you will compare the amino acid sequences of the Receptor Binding Domains (RBD) of both viral Spike proteins to see if there are any significant differences between the two proteins that can account for the observed differences in binding affinities.

## Box 1: What is BLASTp?

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 (<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>) and click the Protein Blast box. In the new page that opens you can paste your query sequence. If the PDB entry ID and Chain ID is provided, NCBI BLASTp can fetch sequences from the PDB. Here we will compare the sequences of the SARS CoV-2 Spike RBD (PDB ID 6m0j, chain E) with SARS CoV Spike RBD (PDB ID 2ajf, chain E)
- 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.
- Q1. Examine the results page and click on the alignment tab.
  - a. Copy the sequence alignment and paste it below. Make sure that you paste it using Courier font, size 10.
  - b. Highlight in yellow any instances where a charged amino acid (aa) in the CoV-2 Spike protein aligns with a hydrophobic aa in CoV spike protein;
  - c. Highlight in blue any instances where 3 or more consecutive aas in the CoV-2 Spike protein does not align with the sequence in the CoV Spike protein.

Below is a table that shows a summary of a subset of these differences that correspond to regions of the protein at the RBD:ACE2 interaction interface.

Residue mismatch in ACE2 binding site	SARS CoV-2 amino acid number	SARS CoV amino acid number
K/V	K417	V404
KVG/TST	K444, V445, G446	T431, S432, T433

Using a similar workflow as you did for Part II of this worksheet you can visualize these amino acids at the binding interface using iCn3D. In figures 3 and 4, you will find representative screenshots of similar views taken from two such visualization sessions conducted using PDB IDs 6m0j (for structure of SARS CoV-2:ACE2 interface) and PDB ID 2ajf (for structure of SARS CoV:ACE2 interface). In both the figures the ACE2 protein is shown in blue, while the Spike protein is shown in magenta. Study these images and answer the following question:

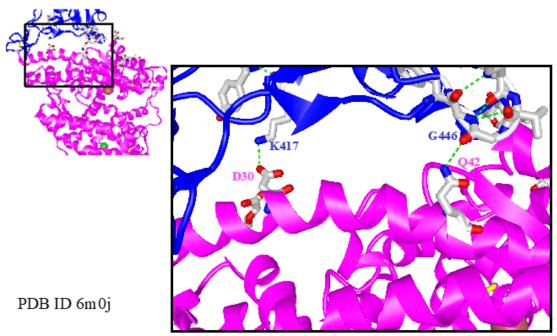


Figure 3: Close up of SARS CoV-2: ACE2 interface (based on coordinates in PDB ID 6m0j) highlighting the regions around K417 and K444-G446 where a sequence mismatch is identified between SARS-CoV-2 and SARS- CoV

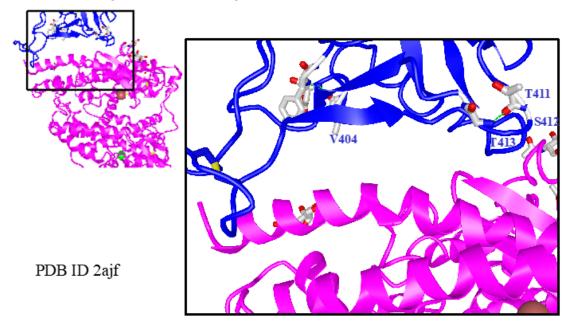


Figure 4: Close up of SARS CoV: ACE2 interface (based on coordinates in PDB ID 2ajf) highlighting the regions around V404 and T431-T433 where a sequence mismatch is identified between SARS-CoV-2 and SARS- CoV

Q2. Based on your study of the comparative molecular views of the interaction interface between RBD of the spike protein and ACE2 included above, which of the two Spike proteins (SARS CoV-2 or SARS CoV) is likely to bind ACE2 more strongly? Briefly justify your answer by including specific molecular details referencing the figures.

# Part 4: Fighting off COVID-19: Blocking virus: host interactions and designing a vaccine

So far the most effective measure to limit infection for COVID-19 has been social distancing. However, in order for the world to resume routine life, we need other alternatives.

Q1. Based on your molecular exploration in Part 3, list two ways you could prevent infection by blocking the SARS CoV-2 from binding ACE2. Please be specific by including details at the molecular level.

One proven approach to immunize individual's own body against viral infections is vaccination. Therefore, once a safe and effective vaccine is available for SARS-CoV-2, we should be able to fight off any possible infection of COVID-19. Armed with the knowledge of the SARS CoV-2 Spike protein structure and some prior foundational work on vaccine design many candidate vaccines began testing within a little over a month of the first report to WHO.

Listen to a <u>brief podcast</u> of the vaccination trial that started on March 16, 2020 and watch the <u>video</u> that describes the strategy that was used for making this vaccine.

Q2. Based on the information you gathered from the podcast and the video summarize how this vaccine now on trial is different from other traditional vaccines.

**Concluding remarks:** As you know from your course work, mRNA is an unstable molecule and can easily be destroyed in. To avoid this intrinsic degradation, the Spike protein mRNA is packaged in a lipid nanoparticle for this vaccine. It is this nanoparticle that is injected into the muscle to start the vaccination process.