**COVID-19: Molecular Basis of Infection**

Didem Vardar-Ulu1 and Shuchismita Dutta2\*

1Biological Chemistry, Boston University, Boston MA

2Institute of Quantitative Biomedicine, Rutgers University, Piscataway NJ 08854

\*contact: sdutta@rcsb.rutgers.edu)

**Preparation:**

As homework and prior to the case discussion in class, get acquainted with the case

* Read the following sections A-D and explore the articles/ links included
* Answer the questions 1-3 in preparation for the case discussion.

**Quarantine, Social Distancing, and Stay Home Orders …**

*A. 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>)

*B. Anatomy of SARS-CoV-2*

Examine David Goodsell’s painting of the anatomy of the Coronavirus below and read a description of the virus.



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

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. In the Figure 1 (above) label the following:

* Spike protein (S) proteins,
* Viral envelope (E), and
* any two other viral proteins

Describe the main functions of the proteins that you labeled within the virus.

*C. Washing Hands with Soap*

Watch the video <https://pdb101.rcsb.org/learn/videos/fighting-coronavirus-with-soap>.

Q2. How does soap impact virus structure to provide an effective prevention against coronavirus infection? Highlight the structure function relationship and chemical properties of the key molecules involved.

Ans. Hydrophobic interactions of the soap molecule’s lipid tails with the membrane phospholipids and viral spike protein disrupt the viral envelope and destroy the virus.

*D. Viral Replication*

Learn about the viral life cycle by watching the video “[How does a virus replicate in a cell](https://www.youtube.com/watch?v=QHHrph7zDLw)”

Q3. What is the role of infection in the viral life cycle? Why do you think viruses infect only specific cells in specific organisms?

**Part 1: 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. Examine Figure 2 illustrating key steps in the coronavirus life cycle.



Figure 2. Life cycle of a virus from infection (entry into host cells) to release of new viral particles. (See <https://viralzone.expasy.org/resources/Coronavirus_cycle.png>).

Q1. Based on what you learned from studying the figure above, complete the following table about key steps in the SARS-CoV-2 life cycle and identify the key players at each step.

|  |  |  |  |
| --- | --- | --- | --- |
| Life Cycle Step | Viral protein/ process | Host protein/ processes  | Comment/ Description |
| Viral attachment and infection |   |   |   |
| Replication of viral genome |   |   |   |
| Viral assembly and release |   |   |   |

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. The genetic material of Coronavirus is a +ive single stranded RNA.

Q2. 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?

**Part 2: Beginning the Infection**

The first step in the viral life cycle is infection and it begins with the SARS-CoV-2 Spike protein binding a host receptor protein (Angiotensin Converting Enzyme 2 or ACE2 protein on lung cells). In the following 3 sections of this part of the case 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 <https://www.youtube.com/watch?v=gDY8pH6OWBc> 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.

*A. 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.
* Open the structure explorer page for the entry by entering the PDB ID in the top search box on [www.rcsb.org](http://www.rcsb.org).

Q1. 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 |   |
| 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. Take a screenshot of the page and answer the following question:

Q2. How many protein chains do you see?

* To visualize and explore this structure of the SARS-CoV-2 Spike protein in detail:
	+ 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.
	+ The structure opens in a new tab – rotate the molecule and examine the overall structure.

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.

* Visualize a single chain of the Spike protein and explore it as follows:
	+ Click on the Select button >> Defined sets
	+ In the new window that opens Select the chains B and C with the shift button pressed (select from 6vsb\_B to 6vsb\_C and chemicals)
	+ Hide the selected chains/molecules by clicking on the button Style >> Proteins >> hide; Style >> Sidechain >> Hide; View >> Disulfide bonds >> Hide. Clear selections by clicking on the button Select >> Clear selection.
	+ Select chain 6vsb\_A and color the chain using the Color >> Spectrum option.
	+ You can view and select specific amino acid residues by clicking on the Windows >> Sequences & Annotations >> Details (in the new window that opens up). By clicking and dragging on specific amino acids in the sequence window you can select the amino acids in the graphics window.

UniProt of 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.

Q4. Select the receptor binding domain in the spike protein seen in the chain A for PDB ID 6vsb and color it magenta (click on Color >> Unicolor >> Magenta). Orient this domain to be positioned at the top, save an image and paste it below. Label the following in the image:

* + N- and C-termini
	+ RBD
	+ where you think the viral envelope (or membrane) is located in this image.

*B. 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.
* Open the structure explorer page for the entry by entering the PDB ID in the top search box on [www.rcsb.org](http://www.rcsb.org).

Q5. Review the contents of the page and complete the following table with information about the entry.

|  |  |
| --- | --- |
| PDB ID |  |
| 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 |  |

* Visualize the structure of the ACE2 protein structure in PDB ID 1r42 using iCn3D. Hide chains B-E which represent the disordered segment of collectrin homology domain. Select the ACE2 protein and color it, using the Spectrum option. Clear selections.

UniProt lists the active site residues for the ACE2 enzyme as E375 and H505 (<https://www.uniprot.org/uniprot/Q9BYF1>). 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. Visualize this information on the structure of the ACE2 enzyme to understand the location of the enzyme active site with respect to the Spike binding site.

* Open the sequences and annotation window (using the same steps as when you explored the Spike protein), locate and select the enzyme active site residues and the residues to which the SARS spike protein binds by clicking and dragging on these residues in the sequence and annotation window. Display the side chains of the selected residues by clicking on the button Style >> Side chain >> Ball and Stick.

Q6. Save an image of this structure and paste a copy below. Label the enzyme’s active site on the image. Assuming that the SARS-CoV-2 Spike protein binds in the same location as the SARS-CoV Spike protein, draw a circle around that region and label that location in your figure as SARS-CoV-2 binding site.

*C. Viral Attachment to Host*

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 a structure of the SARS-CoV-2: ACE2 complex (PDB ID 6m0j).

* Open the structure explorer page for the entry by entering the PDB ID in the top search box on [www.rcsb.org](http://www.rcsb.org).

Q7. Review the 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 |  |
| 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.

The video that you watched mentioned that the binding between SARS-CoV-2 Spike protein and ACE2 is stronger. Of all the non-covalent interactions that facilitate protein-protein interactions salt bridges are one of the strongest.

* Examine the structure to see if there are any charge based interactions (salt bridges) between the Spike and ACE2 protein using the following steps in iCn3D:
	+ Click on the Select button >> Defined sets
	+ In the new window that opens Select the chains E (6moj\_E)
	+ Examine any salt bridges between this selected chain and the other chain in the structure by clicking on View >> H Bonds & Interactions >> check to turn off all the box next to different types of interactions and select the box next to Salt Bridges >> click on Display.
	+ Look for any Salt Bridges formed between the Spike and ACE2 proteins.

Q8. How many salt bridges do you see between the SARS-CoV-2 and ACE2 proteins? Make a figure using an image from your iCn3D visualization of the SARS-CoV-2:ACE2 co-structure where you circle the salt bridge(s) and Save an image and label the residues forming the salt bridge(s).

(Note: Mouse over any residue in the graphics window to see the residue number. Convert that NCBI reference number to the PDB/UniProt number by reading off the corresponding number from the Sequences and Annotations window).

**Part 3: 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. However, the COVID-19 caused by SARS-CoV-2 led to a pandemic. Here we will compare the amino acid sequences of the Receptor Binding Domains (RBD) of both viral Spike proteins to see if they are any significant differences that can account for the 10-fold difference in binding affinity with ACE2, discussed in the Wrapp et al., 2020 paper.

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 (<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. 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.
* Examine the results page and click on the alignment tab. Copy the sequence alignment and paste it below. Make sure that you paste it using Courier font, size 10.

Review the alignment and highlight in yellow any instances where a charged amino acid (aa) in the CoV-2 Spike aligns with a hydrophobic aa in the CoV Spike protein.

Q1. In the above sequence alignment, list the SARS-CoV-2 charged amino acids that are non-polar in the SARS-CoV spike protein.

Locate the amino acids identified above in the structure 6moj as visualized in iCn3D. In the Sequence and Annotations window, click and drag on the specific amino acids (use the NCBI numbering to match the numbers from the sequence alignment).

* Locate and select residues K99, E153, and E166.
* Show side chains of the residues - Click on Style >> Side chain >> Stick
* Color them to stand out – Click on Color >> Unicolor >> Cyan
* Examine interactions of the selected residues – Click View >> H-Bonds & Interactions >> Click off the Contacts and interactions box >> 3D display interactions
* Click View >> Zoom in Selection to see a closeup of the interacting residues

(Note: Mouse over any residue in the graphics window to see the residue number. Convert that NCBI reference number to the PDB/UniProt number by reading off the corresponding number from the Sequences and Annotations window).

Q2. Are any of the amino acids that you have listed in Answer 1 located at the Spike: ACE2 interface? What do these residues interact with? Support your answer with a suitable image with the residues labeled.

Q3. Based only on your explorations of the interactions of the residues that are mismatched in SARS-Cov-2 and SARS-Cov binding to ACE2, which of the Spike proteins is likely to bind ACE2 strongly? Explain how the differences in these interactions may have a role in SARS-CoV-2 causing a pandemic.