**COVID-19: Molecular Basis Infection Revisited One Year into the Pandemic**

*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, as early as February 2020, governments of various countries around the globe started promoting social distancing, issuing stay home orders, and ordering lockdowns. By the end of March 2020, 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. Despite the great scientific advancements in our understanding of this virus and ongoing development of vaccines and therapies, Covid-19 still continues as a real global public health threat with a current (4/24/2021) total of 146 million people infected, 83.8 million recovered and 3.09 million dead.

***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. 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. Visit David Goodsell’s painting of the anatomy of the Coronavirus(<https://pdb101.rcsb.org/sci-art/goodsell-gallery/coronavirus>) and identify the above mentioned elements on the painting.

**TOPHAT QUESTION 1**

Although we hear constantly about the importance of wearing masks to locally trap water droplets containing viruses and curb the spread of virus, one very simple approach is often overlooked: Washing hands with soap and water. Watch the following video to remind yourselves on the biochemical basis of the effectiveness of this simple treatment method. <https://pdb101.rcsb.org/learn/videos/fighting-coronavirus-with-soap>

**TOPHAT QUESTION 2**

***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. Watch the first 2:43 min of the video “[What happens if you get Coronavirus?](https://www.youtube.com/watch?v=5DGwOJXSxqg)” (https://www.youtube.com/watch?v=5DGwOJXSxqg)

Review your understanding of what happens if you get coronavirus by studying the following figure from <https://www.nature.com/articles/nrmicro775.pdf> that summarizes the life cycle of a virus from infection (entry into host cells) to release of new viral particles.



***Molecular Basis of Covid-19 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). When the spike protein binds ACE-2 it gains entry into your cells. Once the virus enters your cell and deposits its genome into the cell, you are “INFECTED” (regardless of symptoms).

Since it is this binding step that is critical for the infectivity of this virus, most of the effort has gone into understanding the molecular details of this interaction with the hopes of abolishing it.

**STRUCTURE OF THE SPIKE PROTEIN:**

The SARS CoV-2 Spike glycoprotein is over 1200 amino acids long. Its structure was solved on 02/26/2020 using electron microscopy. Go to [www.rcsb.org](http://www.rcsb.org) and enter PDB ID: 6vsb in the top search box. Explore and investigate the molecular details of this protein to familiarize yourself with its 3D structure by clicking the 3D View tab to activate Mol\* visualization tool. Use the following questions to guide your investigations.

How many protein chains do you see? Based on your answer how can you characterize this protein (i.e. what is its quaternary structure?) Hint: Each chain is depicted with a different color.

What kinds of secondary structures do you see?

What do the blue cubes represent in the structure? Remember when you hover your mouse over it you can read off the element from the lower right hand of the window.

Where are the N- and C-termini of the chains? Orient the structure so that the C-termini of the protein chains are at the bottom of the page.

Let’s now focus on ONLY the receptor binding domain (RBD) of the spike protein (Residues R319 and F541). Using the displayed sequence on the upper part of your molecular canvas slect the region from R319-F541 in the A chain, i.e. the first set of amino acids with these numberings. Remember to be able to select the amino acids you first need to toggle onn the “selection” mode by clicking the arrow.





After selecting this region turn it into a component by clicking the cube icon, then selecting “cartoon” as the representation and adding the label “RBD” under the options. Then click create.

Finally, go to your components window on the right and “hide” all components except for the RBD you just created.



You should now be seeing only the RBD of the spike protein on your screen.

**STRUCTURE OF THE ACE2 RECEPTOR:**

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.

In a new browser tab go to [www.rcsb.org](http://www.rcsb.org) and enter PDB ID: 1r42 in the top search box which is the structure of human ACE2 determined on 10/07/2003 via X-ray crystallography

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.

Open the structure in the 3D Viewer and identify these residues on the ACE2 structure by displaying them in their ball and stick representations on the ACE2 structure. Also note the N- and C- termini.

Using the residues for spike protein binding, identify the face of the ACE2 protein that is expected to be in contact with the Spike protein’s receptor binding domain.

**STRUCTURE OF SARS CoV-2: ACE2 complex**

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.

In a new browser tab, go to [www.rcsb.org](http://www.rcsb.org) and enter PDB ID: 6m0j in the top search box which is the co-crystal structure of the ACE2 and SARS CoV-2 complex.

Examine this structure and note down the non-covalent interactions the residues you identified on the ACE2 protein make with residues within a 5A radius. Pay special attention to any interaction at the interface between SARS CoV-2 and ACE2 and any salt bridges that might be functionally important. Note that the two proteins are shown in different colors in the default view.

**HOW NOVEL IS SARS COV-2?**

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.

Let’s 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.

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 on to a word document. Make sure that you paste it using Courier font, size 10.

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;

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 |

Go back to the SARS CoV-2:ACE2 complex structure and locate these residues and the interactions they participate in. Locate corresponding residues and interactions in the SARS CoV:ACE2 complex structure with a PDB ID: 2ajf

Below are two close up figures from these interfaces:



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

**TOPHAT QUESTION 6**

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. Now you know what the reason for this difference is at the molecular level. Using the same molecular understanding, can you think of specific molecules scientist could develop to prevent infection by this virus.

**FIGHTING OFF COVID-19: Blocking virus: host interactions**

To understand the available and developing medical prevention and treatment options for Covid-19 let's review some immunology from a biochemists’ perspective and get familiar with what antibodies are:

**Video 1:** The Immune system (www.youtube.com/watch?v=zQGOcOUBi6s&t=294s)

**Video 2:**[Why are antibodies interesting to learn about?](https://drive.google.com/file/d/1dKzitd0Olw9ESRCRmpE4AHPpEC-thRoc/view?usp=sharing) (https://youtu.be/Q3eW7ROshjg)

**Video 3:**[The Antibody Structure](https://drive.google.com/file/d/166I_4jr1-1Ez5_VJ_1Fz_qhs3u-1BAg_/view?usp=sharing) (https://youtu.be/--7uqq9sMgc)

**Video 4:**[The Antigen-Antibody interaction and what it means for vaccine development and antibody treatment](https://drive.google.com/file/d/1IlNbrydCkCYxsnG0SenlTgODRoiFxdIT/view?usp=sharing) (https://youtu.be/Lh7jvfjJskk)

**Video 5:**[Physiological responses to Covid-19 Infection](https://drive.google.com/file/d/1tXVf0hsHbpfsUkRbCjf1GaaC_CfS9Rh0/view?usp=sharing) (https://www.youtube.com/watch?v=5DGwOJXSxqg) after the first 2:43 min

**Video 6:** [Beyond traditional Antibodies](https://www.youtube.com/watch?v=WjhbexLtYts&t=13s) (https://www.youtube.com/watch?v=WjhbexLtYts&t=13s)

**CORONAVIRUS VARIANTS – why should we care**

Viral genome – size, genes, central dogma.

Here is a resource that compiles all the viral genes encoded by the SARS-COV2 virus and the proteins they code for.

<https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html>

Read through this information and note down the answers to the following questions:

**TOPHAT QUESTION 7**

**TOPHAT QUESTION 8**

## CONNECTING GENETIC VARIATION AND PROTEIN STRUCTURE

Visit <https://www.nytimes.com/interactive/2021/health/coronavirus-variant-tracker.html> . Read through the compiled information on the current coronavirus variants to familiarize yourself with the nomenclature used to describe genetic variants of the coronavirus relevant to the current Covid-19 pandemic.

**TOPHAT QUESTION 9**

Make a table that lists the key amino acid mutations for each of these four variants of concern and the reason for concern. Then locate these residues on the SARS-Cov2 Spike: ACE2 complex you explored above (PDB ID 6M0J). You can also look for these residues using PDB ID 6M17 or on binding of the variant spike protein to one specific neutralizing antibody shown in the PDB ID 7K9Z. Based on the location and characteristics of these mutations make a prediction about their possible functional effect and compare your predictions with the concerns you have listed in your table.