**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)

**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.