COVID-19 Virus
Drug Discovery Computational Research

-By SAAI SUDARSANAN D
SASTRA Deemed to be University, Thanjavur

INTRODUCTION:

SARS-CoV2-2 is the scientific name and its full form is **Severe Acute Respiratory Syndrome**.

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.

Most people infected with the **COVID-19** virus will experience **mild to moderate respiratory illness** and recover without requiring special treatment. **Older people, and those with underlying medical problems** like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to **develop serious illness**.

The COVID-19 virus spreads **primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes**.

At this time, there are **no specific vaccines or treatments for COVID-19**. However, there are many ongoing clinical trials evaluating potential treatments. WHO will continue to provide updated information as soon as clinical findings become available.

Genome Sequence:

- https://www.ncbi.nlm.nih.gov/nuccore/NC_045512

The crystal structure of COVID-19 main protease in complex with an inhibitor N3:

Protease:

- A protease (also called a peptidase or proteinase) is a Trypsin that catalyzes (increases the rate of) proteolysis, the breakdown of proteins into smaller polypeptides or single amino acids.
- Proteases are involved in many biological functions, including digestion of ingested proteins, protein catabolism (breakdown of old proteins) and cell signalling.
- Without additional helping mechanisms, proteolysis would be very slow, taking hundreds of years. Proteases can be found in all forms of life and viruses.

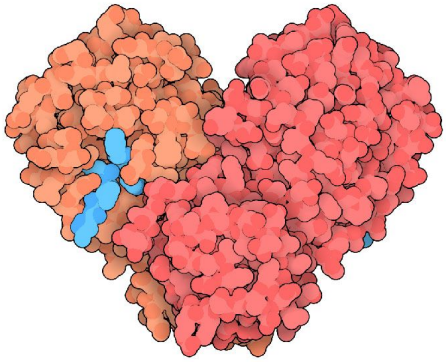
Protease Inhibition:

- Protease inhibitors (PIs) are a class of antiviral drugs that are widely used to treat HIV/AIDS and hepatitis C. Protease inhibitors prevent viral replication by selectively binding to viral proteases (e.g. HIV-1 protease) and blocking proteolytic cleavage of protein precursors that are necessary for the production of infectious viral particles.
- Also refer to https://www.rcsb.org/ for advanced structures of the virus.

3D View:

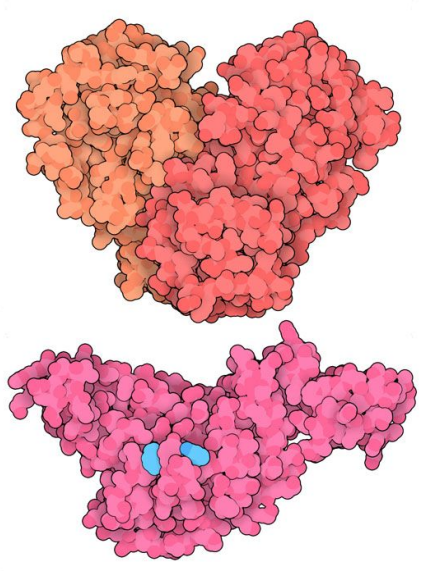
- https://www.rcsb.org/3d-view/6LU7?preset=electronDensityMaps
- https://www.rcsb.org/3d-view/6LU7/1

The **main protease** of coronavirus makes most of these cuts. The one shown here is from the **SARS-CoV-2 (2019-nCoV) coronavirus**. It is a dimer of two identical subunits that together form **two active sites**. The protein fold is similar to serine proteases like **trypsin**, but a **cysteine amino acid and a nearby histidine perform the protein-cutting reaction** and an **extra domain stabilizes the dimer.** This structure has a **peptide-like inhibitor** bound in the active site.



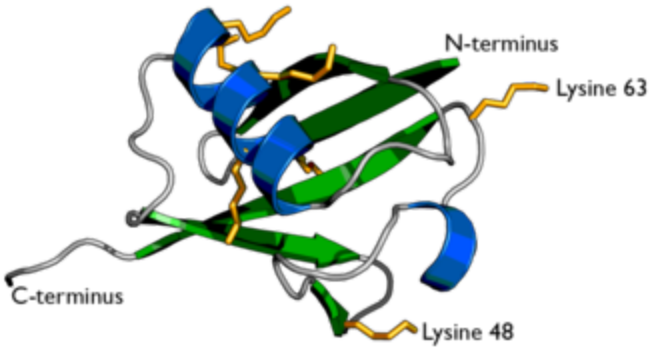
SARS and Papain-like Protease:

- Note that the SARS-CoV2 Protease is similar to a combination of these two.



It makes **three specific cuts** in the SARS polyproteins, and also **clips several proteins in the infected cell**, including **removing ubiquitin** from **ubiquitinated proteins**.

Ubiquitin is a small (8.6 kDa) **regulatory protein** found in most tissues of **eukaryotic** organisms, i.e., it is found *ubiquitously*.



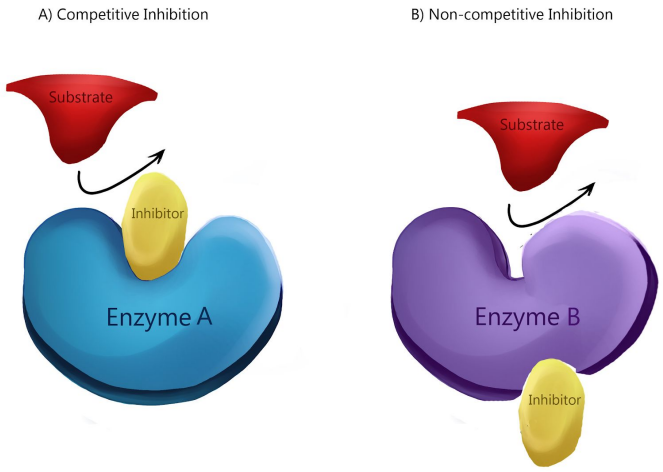
Deubiquitinating enzymes (DUBs) oppose the **role of ubiquitination** by **removing ubiquitin** from **substrate proteins**. They are **cysteine proteases** that cleave the amide bond between the two proteins.
SARS-CoV2 virus has cysteine amino acid unlike the original SARS CoV1 and I believe it is this property that makes it hard to find an inhibitor, since it deubiquitinases the substrate proteins and interferes in the production of interferons.

Interferons are a group of **signaling proteins** made and **released by host cells** in response to the presence of several viruses as a part of the host's innate immunity.
In a typical scenario, a virus-infected cell will release interferons causing nearby cells to heighten their antiviral defenses.

After a small study about how interferons are released, I have come to the conclusion that interferons undergo **ubiquitination-mediated stimulation**. Which makes ubiquitin a necessity.
Artificially injecting interferons might work(but, I am no doctor).

What should we do?

Bind - Find a Binding Site in the Virus's protease, which has already been found, but the N3 inhibitor is not effective.
Inhibit - Find a proper inhibitor molecule
Synthesise - Synthesise the Inhibitor molecule
Treat - Treat patients after various clinical trials.
We have to find the right ligand, so that the activation sites of the protease of the virus are blocked off before it attaches to the cell.



In our case: Competitive Inhibition
A Competitive Inhibitor of the right shape(so that it can bind to the main protease of the SARS-CoV2 Virus).
The best possible ligand(inhibitor) is found using a docking software that measures the connectivity of the ligand to a binding site.

Globally, as of **12:44pm CEST, 15 August 2020**, there have been **20,995,433 confirmed cases** of COVID-19, including **760,774 deaths**, reported to WHO.
More COVID19 Data : https://covid19.who.int/

Haste makes waste, but even that waste is useful in such a dire situation.

DATA:

- https://opendata.ncats.nih.gov/covid19/
- https://www.ncbi.nlm.nih.gov/home/develop/
- https://www.ncbi.nlm.nih.gov/home/download/
- https://en.wikipedia.org/wiki/Interferon
- https://pdb101.rcsb.org/motm/242

Methodology:

We use **Molecule Transformer-Drug Target Interaction (MT-DTI)**, developed and tested by bioRxiv.
The Model is **trained and tested on drug-target interactions of different compounds** and predicts the required drug for the CoronaVirus. The above model is a **transformer based model**.
We are also required to perform Molecular Docking tests and obtain data, in order to generate the dataset.

(or)

We use the LSTM_Chem model(https://github.com/topazape/LSTM_Chem) and train in the ChemBL(https://www.ebi.ac.uk/chembl/) drug database. We score their binding ability to the virus using a docking program called, PyRX(Computational Drug Design Screening Software: https://pyrx.sourceforge.io/) and then use it to find out which drug has the maximum binding ability to the virus's binding site.

SMILES (Simplified Molecular Input Line Entry System) is a chemical notation that **allows a user to represent a chemical structure** in a way that **can be used by the computer**. SMILES is an **easily learned and flexible notation**. The SMILES notation requires that you learn a handful of rules.

Rules:
https://archive.epa.gov/med/med_archive_03/web/html/smiles.html#:~:text=SMILES%20(Simplified%20Molecular%20Input%20Line,learn%20a%20handful%20of%20rules.

C*1*C*C*C*C*C1 is Benzene in SMILES format.

SMILE formats of 30 prospective candidates were taken and their structures were plotted using the RDkit python library.

References

1. Cui, J., Li, F., Shi, Z.L. (2019) Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17, 181-192.
2. 4yoi: St John, S.E., Tomar, S., Stauffer, S.R., Mesecar, A.D. (2015) Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4-The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS). Bioorg.Med.Chem. 23: 6036-6048
3. 4ow0: Baez-Santos, Y.M., Barraza, S.J., Wilson, M.W., Agius, M.P., Mielech, A.M., Davis, N.M., Baker, S.C., Larsen, S.D., Mesecar, A.D. (2014) X-ray Structural and Biological Evaluation of a Series of Potent and Highly Selective Inhibitors of Human Coronavirus Papain-like Proteases. J.Med.Chem. 57: 2393-2412
4. Hilgenfeld, R. (2014) From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. FEBS J. 281,4085-4096
5. 1q2w: Pollack, A. (2003) Company says it mapped part of SARS virus. New York Times, July 30, section C, page 2.
6. Mol* (D. Sehnal, A.S. Rose, J. Kovca, S.K. Burley, S. Velankar (2018) Mol*