**Drugging SARS-CoV-2**

By: Mallory Bliss\*, Alexzandria Khan\*, Nandni Verma\*, and Kasandra Riley

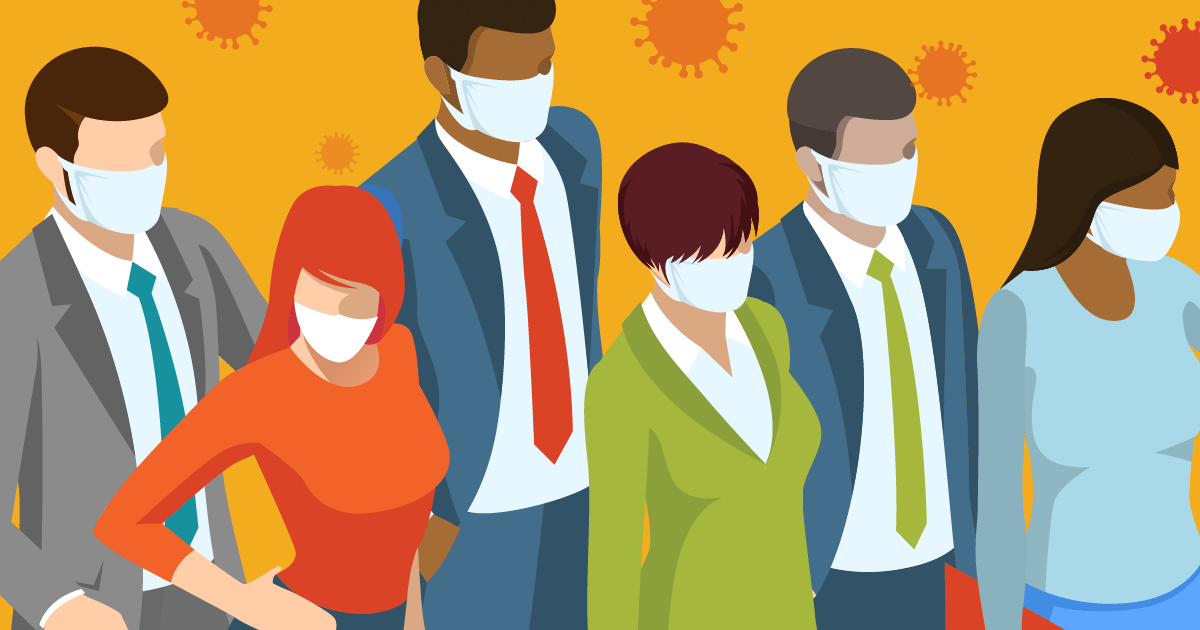
Department of Chemistry, Rollins College

\*These authors contributed equally to this manuscript

**Part 1: Introduction to COVID-19**

Learning Objective: This worksheet introduces students to the COVID-19 pandemic exploring its origin through phylogenetic analysis and its viral life cycle.

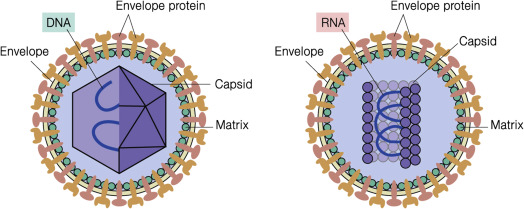
*The Pandemic*

****

https://www.insperity.com/blog/covid-19-pandemic/

Viruses, which are composed of either DNA or RNA, are infectious agents that hijack components of living cells in order to survive (Lodish 1970) (**Figure 1**). In December 2019, a new viral coronavirus pathogen, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), caused an outbreak of a respiratory illness (a diseased termed “COVID-19”) in the city of Wuhan, China. The most common symptoms of COVID-19 included fever, shortness of breath, and dry cough. In the span of three short months, the virus rapidly spread around the world through human-to-human transmission (Li et al. 2020). On March 11, the World Health Organization (WHO) declared the outbreak a pandemic (Zhang et al. 2020). Governments responded with extreme responses including social isolation, business and school closures, as well as international travel bans. Genetic comparisons demonstrated that SARS-CoV-2 is similar to viruses causing two previous coronavirus outbreaks: Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). As of March 2020, there were not any drugs, therapeutics, or effective treatment options for COVID-19 and the race to find a drug or vaccine was on (Li et al. 2020).

**Figure 1.** Schematic of a typical enveloped virus particle with a helical capsid. A part of the nucleocapsid (purple) is uncoated to show the viral genome inside (RNA). Three kinds of virus structural proteins are denoted: envelope glycoproteins (pink and orange), capsid protein (purple), and matrix protein (green) (adapted from Ryu 2016).

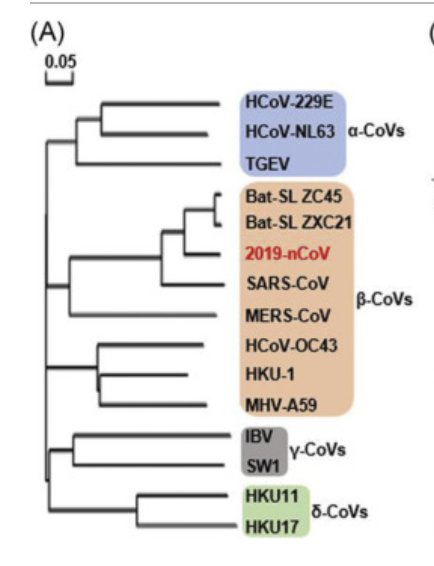


It is March 2020, and you are the head drug developer at the prestigious pharmaceutical company RollyColly. You have been directed by the United States Centers for Disease Control (CDC) to lead your team in discovering and developing a drug to bind to a biomolecule expressed by SARS-CoV-2. To begin the challenging task, you explore the fundamentals of the virus. Because it is early in the SARS-CoV-2 pandemic, you wonder if you can use information about MERS or SARS-CoV-1 to supplement your understanding. Which virus was the immediate ancestor of SARS-CoV-2? What are the relationships between SARS-CoV-2 and other previously described coronaviruses? Answering these two questions are vital for designing strategies to combat the virus and for understanding the life cycle of SARS-CoV-2. To save yourself some work, you dig through peer-reviewed literature to research the origin of SARS-CoV-2. The phylogenetic tree offers a unique perspective to examine a species over time. The closer the lines (clades) are on the tree the closer their relationship is.

Phylogenetic Trees

*Part A. Origin of SARS-CoV-2*

From your literature search, you find a published phylogenetic tree showing the divergence of coronaviruses over time (**Figure 2**). You learn about the four main sub-groupings of coronaviruses: 𝛂 (blue), 𝝱 (orange), 𝞬 (grey), and 𝝳 (green) (“Coronavirus” 2020). Alphacoronaviruses and beta-coronaviruses are believed to have originated from bats, while gamma-coronaviruses and delta-coronaviruses have an avian origin (Jaimes et al. 2020). Understanding the unique genetic and phenotypic structure of SARS-CoV-2 compared with the other coronaviruses is important for the production of efficient drugs.

****

**Figure 2.** Phylogenetic tree of coronaviruses. The four main sub-groupings of coronaviruses are highlighted: 𝛂 (blue), 𝝱 (orange), 𝞬 (grey), and 𝝳 (green). SARS-CoV-2 (named 2019-nCoV here) is shown in red (Mousavizadeh et al. 2020).

Based on **Figure 1**, what main sub-group of coronaviruses is SARS-CoV-2 in? With what coronavirus(es) does SARS-CoV-2 share the most common ancestor with?

* Complete answers will consist of:
* The main sub-group of coronaviruses that SARS-CoV-2 is in is 𝝱 -CoVs OR betacoronaviruses

AND

* SARS-CoV-2 shares the most common ancestor with the coronaviruses Bat-SL ZC45 and Bat-SL ZXC21

Which virus, SARS-CoV or MERS-CoV, is more closely related to SARS-CoV-2? What evidence led you to draw this conclusion?

* Complete answers will consist of:
* SARS-CoV is more closely related to the 2019nCoV.

AND

* As seen in the phylogenetic tree, SARS-CoV has a more recent common ancestor to 2019nCoV as compared to MERS-CoV

OR

* The SARS-CoV branching is more closer to SARS-CoV-2 compared to MERS-CoV

Based on **Figure 2**, do you think a drug designed for SARS-CoV could also effectively treat SARS-CoV-2. **Explain** your answer.

* Answers may vary, but the student should mention **Figure 2** to support their statements. For example:
* Yes, a drug designed for SARS-CoV is likely to be beneficial for SARS-CoV-2 since both coronaviruses are beta-coronaviruses and appear in **Figure 2** to have not diverged too long ago. It is probable that the two coronaviruses are still similar enough to respond to the same drug.

OR

* No, a drug designed for SARS-CoV-2 is unlikely to be beneficial for SARS-CoV since the two are not as closely related as Bat-SL ZC45 and Bat-SL ZXC21 as seen in **Figure 2.** It appears that the two species diverged sooner and are not similar enough to respond to the same drug.

OR

* There is not enough information provided in **Figure 2** such as the two viruses genomes to be able to deduce just how divergent they are to make such conclusions. There was no provided quantitative data to make conclusions.

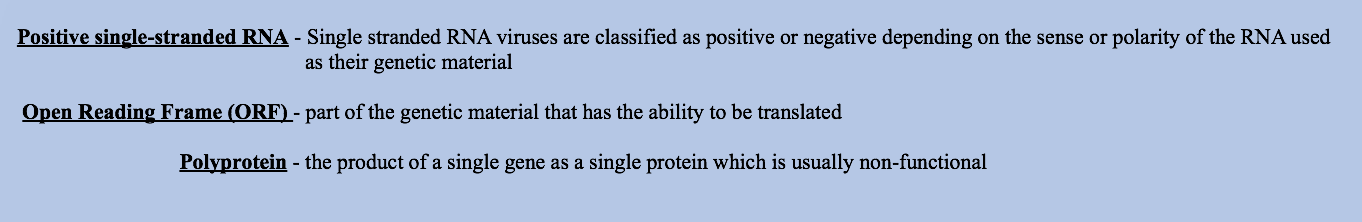
What is the driving force behind the evolution of viruses into new strains?

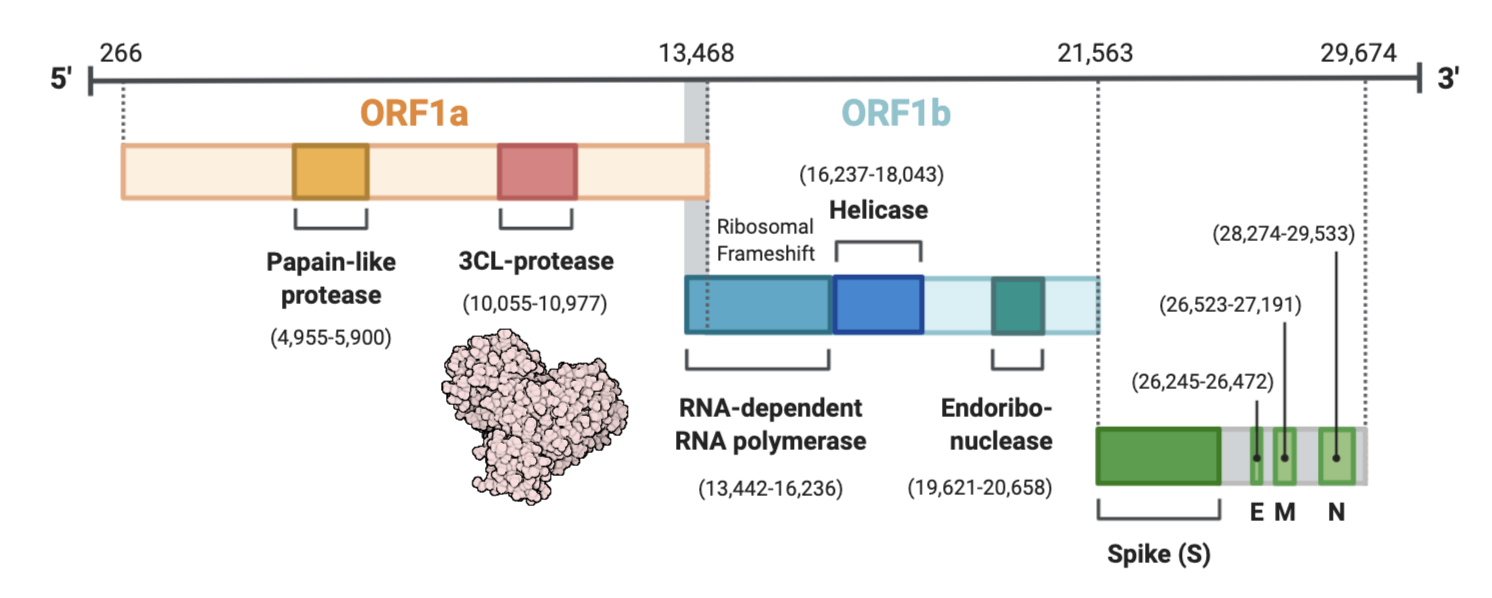
* Acceptable answers must mention the role of mutations in altering viruses.

For example:

* Mutations can occur during replication through an error prone polymerase, typically found in viruses. Mutations within the virus genome might lead to a virus changing its host species or ability to infect other species. In fact, viral RNA polymerases have a higher rate of variation to allow mutations to occur.

*Part B: Viral Life Cycle of SARS-CoV-2*

To develop therapies to impede SARS-CoV-2 infection, you research how the non-living virus hijacks the host during the course of infection and apply this knowledge towards designing both new drugs and repurposing existing ones.



Created with BioRender.com

**Figure 3.** Schematic of the positive single-stranded RNA viral genome of SARS-CoV-2 (adapted from Cascella et. al 2020). The genome is approximately 30 kb in length with a 5’ cap structure and a 3’ poly A tail. The open reading frame 1a (ORF1a) is highlighted in orange as well as the two encoded proteases: Papain-like protease (yellow) and 3CL-protease (red). The structure of the 3CL-protease is pictured below (pink). A frameshift between ORF1a and ORF1b is represented by a grey line. The ORF1b is highlighted in blue as well as the three encoded proteins: RNA-dependent RNA polymerase, Helicase, and Endoribonuclease. The four structural proteins (S, E, M, and N) are highlighted in green.

In your research you come across the genome of SARS-CoV-2. As seen in **Figure 3**, the 5’ open reading frame (ORF) 1a / ORF1b encodes for two polyproteins (pp1a and pp1ab). A frameshift between ORF1a and ORF1b directs the synthesis of the two polyproteins (Cascella et. al 2020). Both polyproteins encode a chain of 16 non-structural proteins (Corum, 2020). The chain of proteins are then cut by two viral proteins 3C-like protease (3CLpro), also called Main Protease (Mpro), and papain-like protease (PLpro) that cut the links between the different proteins (Zhang et al. 2020). Proteases catalyze proteolysis, which is the breakdown of proteins into smaller polypeptides or single amino acids. The cut proteins are then free to perform their functions. The polypeptides encode non-structural effector proteins that form the replicase / transcriptase complex (RTC) playing a vital role in viral replication (Cascella et. al 2020). The remainder of the viral genome is transcribed into a nested set of subgenomic mRNAs, which encodes four structural proteins (Corum, 2020). You also find an article that describes the different proteins encoded by SARS-CoV-2[.](https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html)

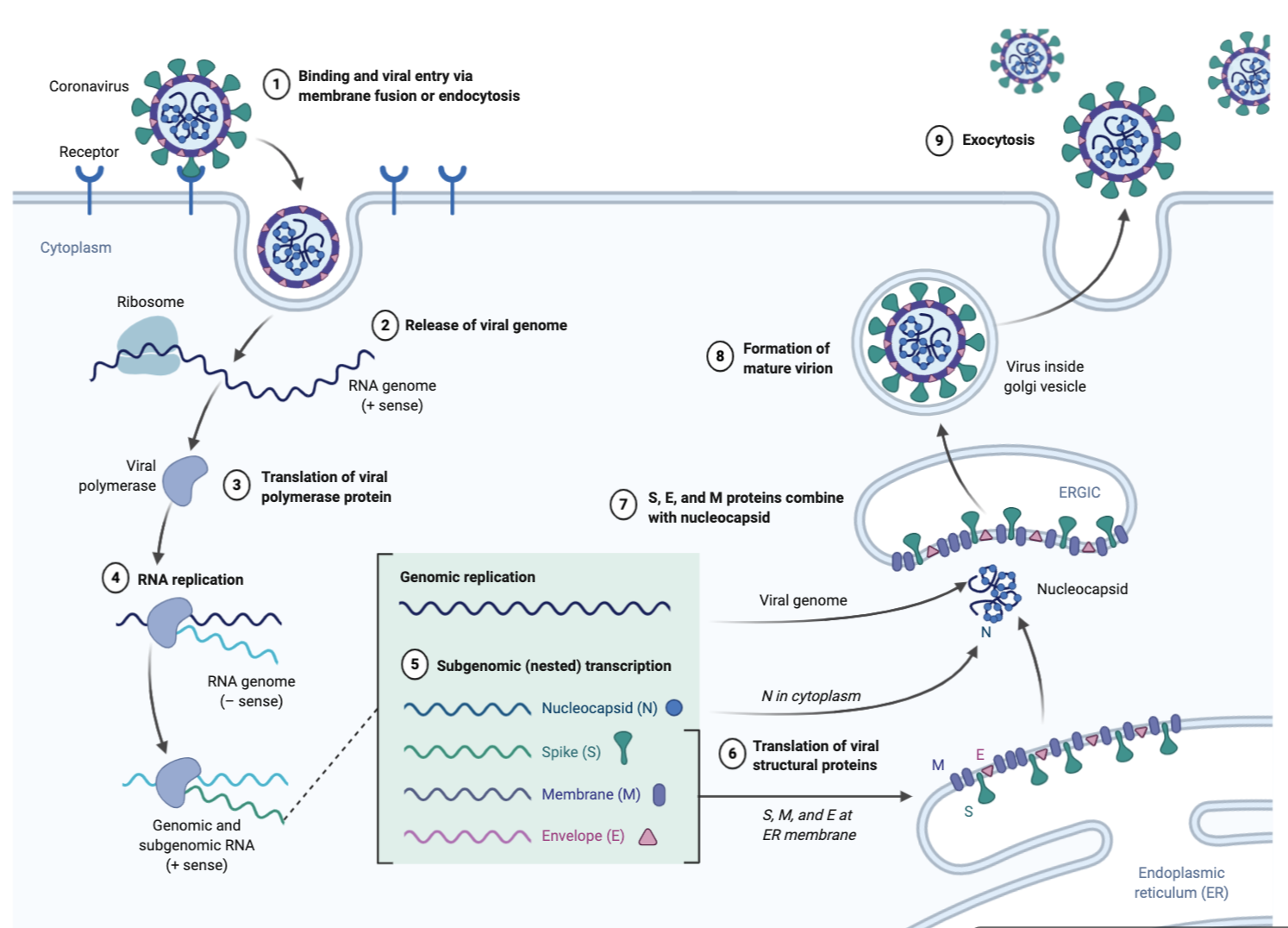
Based on **Figure 3** and the *Bad News Wrapped in Protein: Inside the Coronavirus Genome* [article](https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html) (right click to open link), briefly describe three potential proteins within the genome we could selectively bind with a drug?

Answers may vary as students can select any of the proteins that are discussed in both the figure and article as well as an explanation for how a drug will inhibit the protein.   
**Overall:** The student must show that inhibiting any of these 3 proteins will lead to instability of the virus to propagate within a host.

Answers may include:

* **Papain-like protease**- drug will prevent binding of ATP, a cofactor, block ligand binding site and inhibits the protease’s ability to splice up the polyproteins.
* **Spike Protein**- drug will prevent the virus from attaching to the host cell.
* **Nucleocapsid Protein-** drug binds to protein to alter the protein’s shape and ability to bind and protect the virus’ RNA
* **Envelope Protein**- drugs binds to protein which helps prevent the virus from turning host genes on/off as needed
* **Membrane Protein**- drug binds to outer portion of protein and blocks entry for transport
* **RNA-dependent RNA polymerase**- drug will prevent the virus from being able to replicate its DNA
* **Helicase**- drug will block the unwinding of DNA so that the DNA cannot be accessible for replication to occur
* **Endoribonuclease-** drug will inhibit cleavage of RNA, thus gene expression cannot be altered??

Ultimately, your team at RollyColly wants to identify a druggable protein that is disease-modifying. In the drug discovery field, a target protein is considered “druggable” if it is known to or is predicted to bind with high affinity to a drug. If changing the function of the target protein influences the disease phenotype, it is considered a “disease-modifying” protein target. Binding affinity is dictated by the combination of covalent and non-covalent interactions that occur between the protein and the drug. You have compiled a list of proteins encoded by SARS-CoV-2, however, drugs selectively bind one protein that is both druggable and disease-modifying. Disease modification is defined as a treatment that affects the underlying pathophysiology of the disease and results in a beneficial outcome. To aid in your decision of selecting one site of the virus to inhibit, you research the viral life cycle (**Figure 4**). In general, viruses are not equipped to survive on their own. The only way a virus can ensure its survival is through replicating in a host, which often results in mutations due to the error prone polymerase.



Created with BioRender.com

**Figure 4.** Schematic of the viral replication of SARS-CoV-2 in a host cell (adapted from Cascella et. al 2020). To enter the host, the virus contains an external protein that can be recognized by the host cell **(step 1)**. Once the protein binds the host cell, the virus undergoes endocytosis and is “absorbed” by the host cell. Then the RNA is released into the cytoplasm **(step 2)**. The viral genome is translated by hijacking the host ribosome. The two polyproteins and two proteases are translated. The proteases cleave the polyproteins to construct the polymerase **(step 3)**. Nextthe virus uses the host’s transcription and translation machinery **(steps 4-7)** to produce viral proteins to build more viruses to be secreted and spread to other host cells **(steps 8-9)**.

How will the biochemical properties of a drug targeting a protein differ extracellularly as shown in **step 1** versus intracellularly as shown in **step 4?**

* Complete answers will consist of stating the polarity of the drug
* **Intracellular:**
  + Drug must be small enough to fit through the plasma membrane or be specific for an entry channel

OR

* + Drug would be hydrophobic/nonpolar to pass through plasma membrane
  + could result in a signaling cascade with secondary messengers

AND

* **Extracellular:**
  + Drug may outcompete the virus to bind the host receptor first.
  + Drug would be hydrophilic/polar to bind to the surface of the host cell and not pass through the hydrophobic center of the plasma membrane

Based on **Figure 3** and **Figure 4**, how would the absence of the proteases (3CLpro and PLpro) affect the virus from undergoing its life cycle in its host?

* Complete answers will include how the polyproteins will not be cut AND the effect on the function of the life cycle
* For example: The absence of these proteases would affect the virus from undergoing its life cycle because it would not be able to cut the polyproteins into their individual forms. Therefore, the uncut proteins would not be able to complete their functions in the life cycle and form the RTC necessary for viral replication.

As potential inhibition sites of SARS-CoV-2, what more information would you want to learn about these proteins?

* Answers will vary but students should include at least one answer such as:
  + Primary sequence

OR

* + Secondary and tertiary structures

OR

* + Active site

OR

* + Specific cleavage site

**Part 2: Mpro****of SARS-CoV-2**

Learning Objective: This worksheet explores the function of SARS-CoV-2's Mpro.

To devise an antiviral drug against SARS-CoV-2, we must select a site of SARS-CoV-2 that employs chemical and biological disciplines. From a biological perspective, a good place to inhibit is a biological pathway that can be intercepted to confer a therapeutic outcome. From a chemical standpoint, a good inhibitor is one that can intercept the biological pathway with an active small organic molecule. The most common drug binding sites include enzymes, receptors, and ion channels (Bull 2015).

You were given a grant to examine the protease function of Mpro in SARS-CoV-2 as a potential area to be inhibited by a drug. For various diseases, proteases are promising areas to inhibit as approximately 5-10% of all drugs under development bind proteases (Bull 2015). For example, the Food and Drug Administration (FDA) approved drug, atazanavir (Reyataz), is a HIV protease inhibitor (Orrick and Steinhart 2004). The modes of inhibition often employed by drugs are reversible (competitive, uncompetitive, and mixed) or irreversible. Since most drugs are competitive inhibitors of specific enzymes, you decide to pursue a competitive inhibitor against Mpro (“Libretexts” 2019).

Describe how competitive inhibitors function. How do the competitive inhibitors specifically bind?

* Complete answers should include:
  + How competitive inhibitors resemble the substrate and bind to the active site of the enzyme therefore preventing the substrate from binding

AND

* + To bind the active site, the drug needs to be complementary in shape, size, and charge to the active site in order to bind and inhibit the site. Competitive inhibitors often have a higher affinity to bind to the active site to out-compete other substrates.

Since most competitive inhibitors normally bind the active site, you decide to research the active site of the four major classes of proteases.

List the active site residues of the four major classes of proteases.

|  |  |
| --- | --- |
| **Class of protease** | **Active site residues** |
| Cysteine protease | Cysteine and Histidine |
| Serine protease | Aspartate, Histidine, Serine |
| Aspartic protease | 2x Aspartate |
| Metalloprotease | Zinc, Glutamate |

**Table 1:** The active sites of the four major classes of proteases.

How would you determine what type of protease Mpro is? Explain.

* Correct answers can be:
  + Primary, secondary, and/or tertiary WITH explanations to support their answers such as:
  + You should look at the primary, secondary, and tertiary sequence of the protease to determine the type of protease. The primary and secondary sequences can help identify the active site through consensus. The tertiary structure can show the overall shape of the protein to identify where the active site is located.

**Part 3: Structural Analysis of Mpro****of SARS-CoV-2**

Learning Objective: This worksheet explores the primary, secondary, and tertiary structure of SAR-CoV-2’s Mpro. Specifically, it highlights the important amino acids within the active site and their binding interactions.

Part A. Primary Structure

Why is it important to look at the primary sequence of a protein along with the tertiary structure?

* Acceptable answers:
  + Structure-function relationship as the amino acids of the primary structure dictates the tertiary structure shape and function.

OR

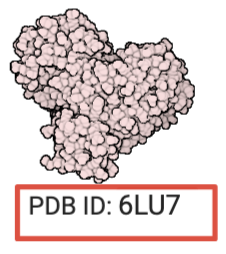
* + Primary structure also allows for direct comparison of differences in amino acids that may be advantageous for viral infectivity.

OR

* + The primary structure is important to the protein's unique three-dimensional structure, its mechanism of action, and its relationship to other proteins with similar physiological roles as it dictates the 3D structure of the protein.

BLASTp

Using Protein BLAST (BLASTp), you examine the primary, secondary, and tertiary structure of Mpro to try and determine what type of protease it is. Looking at the 3D structure offers a perspective of how the protein would function and interact based on its folding of tertiary and quaternary structures, which is critical for drug design. By viewing the primary structure you can identify areas of consensus within the Mpro active site. The greater the level of consensus within the active site will allow you to design an efficient drug that will work for people in different populations.

1. Go to the NCBI BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) You should see a page with the following search options (**Figure 5**).



**Figure 5.** Image of the NCBI BLAST website for obtaining the primary sequence of Mpro (PDB ID: 6LU7). <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

2. Click on Protein BLAST to do a protein-to-protein BLAST search.

3. Enter “6LU7” into the “Enter accession number(s), gi(s), or FASTA sequence(s)” box.

4. Scroll down the page and click on BLAST to begin the search.

5. Scroll down to find the top 100 sequences that were identified as a match to the query sequence by this search.

6. Note that many of these sequences are identified as “Severe Acute Respiratory Syndrome Coronavirus 2” (under Description). This is because there are many sequences from the same organism that have been submitted to the sequence database. To eliminate sequences that are identical (we are looking for similar sequences) you will filter the identical sequences from the results. Go to the top of the BLASTp results page and select “Edit Search”.

7. Your sequence should still be visible in the “Enter accession number(s), gi(s) or FASTA sequence(s)” box.

* 1. Below the sequence entry, there is a box labeled “Organism”. In this box, type the sequences you want to eliminate (“Severe acute respiratory syndrome coronavirus 2”).
  2. To the right of the box is a button for ‘exclude’. Click on that button and a checkmark should appear.
  3. Repeat the BLAST search.

Multiple sequence alignment

1. On the left hand side there is a checked “select all” button. Click that checkmark off.

2. Scroll down in the window and you will see the descriptions of the identified sequences.

* 1. Select five of these sequences. *One of the sequences you select must be SARS-CoV.* *You can select any sequences that you find interesting for the other four.*
  2. Next to the select all button (which should not be selected), you should now see “5 sequences selected”.
  3. Click Download and select “FASTA (Aligned sequences)”. This will create a .txt identified as “seqdump.txt”. *If you have done this a number of times, the file name may be “seqdump (#).txt”.*
  4. Download the 6LU7 sequence by returning to the top of the results page, find the Query ID and click on 6LU7. A page will open with the sequence of 6LU7. Click on FASTA, copy the FASTA formatted text and paste into your text document.

3. Create a .txt document that has the 6LU7 and the five sequences that were just downloaded.

* 1. Save this file which will be used for the multiple sequence alignment.

4. Open up CLUSTAL Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

5. Click on “Protein” underneath “Enter or paste a set of”

6. Copy and paste the FASTA sequences from your .txt file into the sequences window on Clustal Omega.

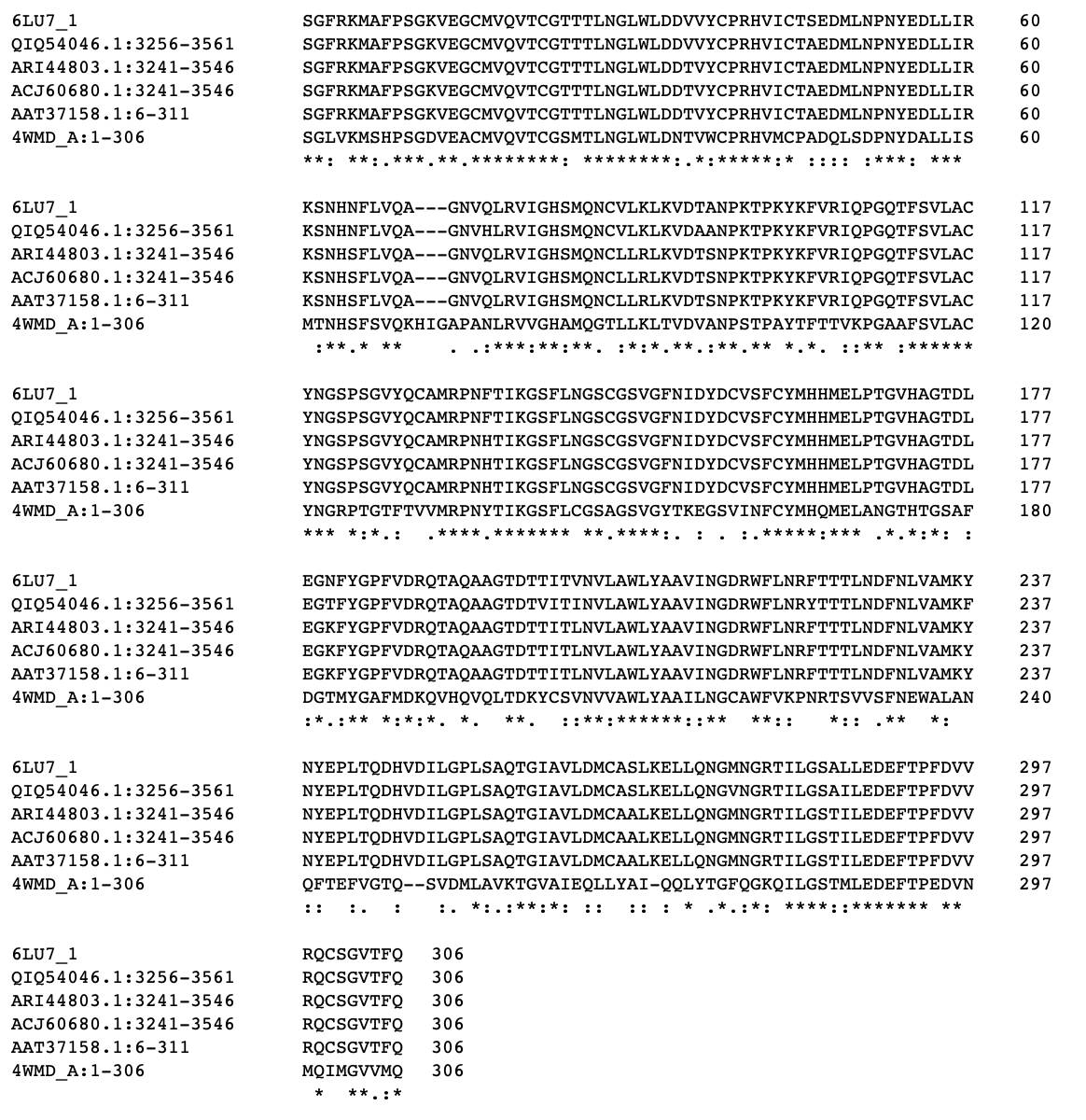
14.Scroll down to see “OUTPUT FORMAT” and click on “More options…”

15. In the expanded “OUTPUT FORMAT” window, select “ORDER” and pull down “input”.

16. Scroll further down and click “SUBMIT” to run the program.

17. Copy the sequences and paste below.

Example:



Given are the identities of the following symbols:

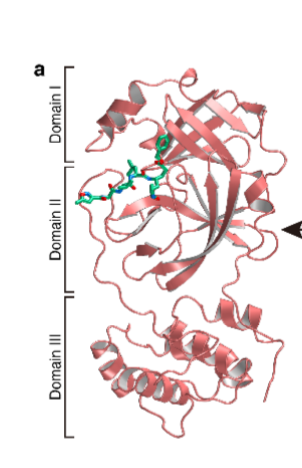
|  |  |
| --- | --- |
| **\*** | identical amino acid in all aligned sequences |
| **:** | amino acids with strongly similar biochemical properties |
| **.** | amino acids with weakly similar biochemical properties |

Are the six sequences identical, similar, or unrelated? Briefly justify your answer.

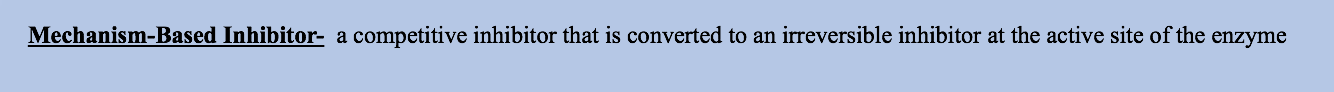
* Answers depend on the student’s primary sequence alignment and what strains they picked.
* Answers should include a deduction of the conservation of the amino acids with an explanation using the chart above for support.
  + For example: Comparing the six sequences, the primary structure of these proteins seem very conserved. There are more \* present in my sequence alignment representing that there were a lot of identical amino acids.

Part B: Mpro of SARS-CoV-2 Secondary Structure

After learning about the conserved primary sequence of Mpro, you decide to examine how it dictates the secondary and tertiary structure of the protein. The SARS-CoV-2 Mpro forms a homodimer. Each protomer is composed of three domains as shown in **Figure 6**. Domains I (residues 8–101) and II (residues 102–184) have an antiparallel β-barrel structure. Domain III (residues 201–303) contains five α-helices arranged into a largely antiparallel globular cluster, and is connected to domain II (residues 185–200) by means of a long loop region (Jin et al. 2020).



**Figure 6.** Domains of one protomer of the dimeric Mpro-inhibitor complex. (Jin et al. 2020)

Part C. Tertiary Structure of Mpro  


1. To visualize the tertiary structure, go to<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>. This website will allow you to visualize and highlight different structures within the protein structure file.
2. In the PDB ID input box (left of LOAD) in the iCn3D start-up window, replace the existing ID# with 6LU7. Click LOAD and a new window with the protein structure will open.

Take a screenshot of the structure and paste it below.

Example:



1. Set a timer for five minutes and explore the structure by rotating it freely to view the different perspectives of this protein complex, zoom in and out using your mouse or touchpad, color in the various domains, see if you can label some of the residues. Then select “View > Reset > All”.
2. After exploring the structure, go to the pulldown menu under “Windows > View Sequences and Annotations”. *This will open a window with information about the proteins and chemicals present in the structure.*
3. Scroll down in the “View Sequences and Annotations” window to see the different proteins that make up the ternary protein complex you are viewing.

Fill out **Table 2** with the identities of the proteins corresponding to each label.

|  |  |
| --- | --- |
| **Subunit** | **Protein** |
| 6LU7\_A | SARS-CoV-2 main protease |
| 6LU7\_C | N-[(5-METHYLISOXAZOL-3-YL)CARBONYL]ALANYL-L-VALYL-N~1~-((1R,2Z)-4-(BENZYLOXY)-4-OXO-1-{[(3R)-2-OXOPYRROLIDIN-3-YL]METHYL}BUT-2-ENYL)-L-LEUCINAMIDE |

**Table 2.** Protein components of the PDB file (6LU7) structure of the Main Protease (Mpro) of SARS-CoV-2.

What is 6LU7\_C? Why did the researchers include it? (hint: see paper in which the structure was published)

* 6LU7\_C is a small molecule that binds to the active center of the protease AND could be used as a blueprint for an inhibitor.

1. Click "Assembly" in the menu "View" to switch between asymmetric unit and biological assembly (**2** asymmetric unit).
2. Examine the neighborhood of the 6LU7\_C and 6lu7\_A to explore their interactions. Click on the Select button >> by Distance >> a new window opens up >> input distance 4 angstrom and select the chain ID >> click on Display. This should highlight the neighboring residues in yellow. Close the new window.
3. Show the side chains of these amino acid residues (click on Style button >> Side chains >> Ball and Stick).
4. Color the select amino acids and other ligands by clicking on the Color button >>Unicolor >>White.
5. Label the select amino acids by clicking on View>>Label>>per Residue
6. Focus in on the selected residues by clicking on View >> Zoom in Selection.
7. Take a screenshot of these residues and include the image below (with labels showing key amino acids and key interactions)

Image should depict highlighted key amino acids with labels as shown below:



1. The protein structure paper (Jin *et al*, 2020) explores the active site of SARS-CoV-2’s Mpro. The researchers determined the crystal structure of SARS-CoV-2 virus Mpro in complex with a mechanism-based inhibitor, N3 to identify new drug possibilities (“leads”). The goal was to use virtual drug screening and high-throughput screening of real molecules that might bind SARS-CoV-2’s Mpro. Briefly read the section “**The crystal structure of COVID-19 virus Mpro in complex with N3”** to uncover which two amino acids comprise the catalytic dyad of the protease.

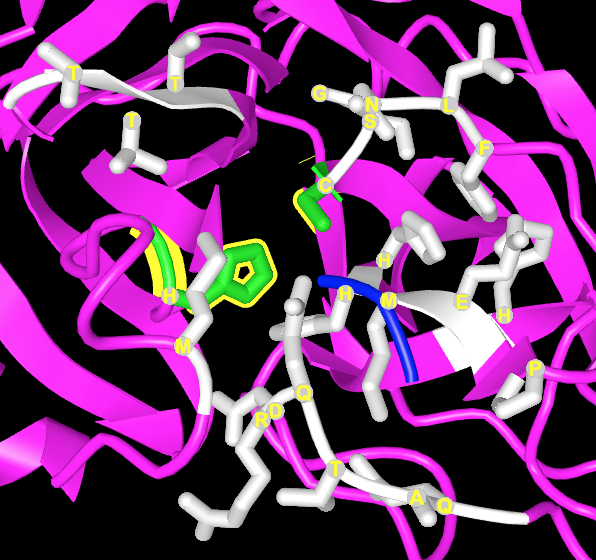
What are the two amino acids within the catalytic dyad? (name and number)? What class of protease is Mpro?

* + The two amino acids within the catalytic dyad are Histidine 41 and Cysteine 145.

AND

* + The protease is a Cysteine protease.

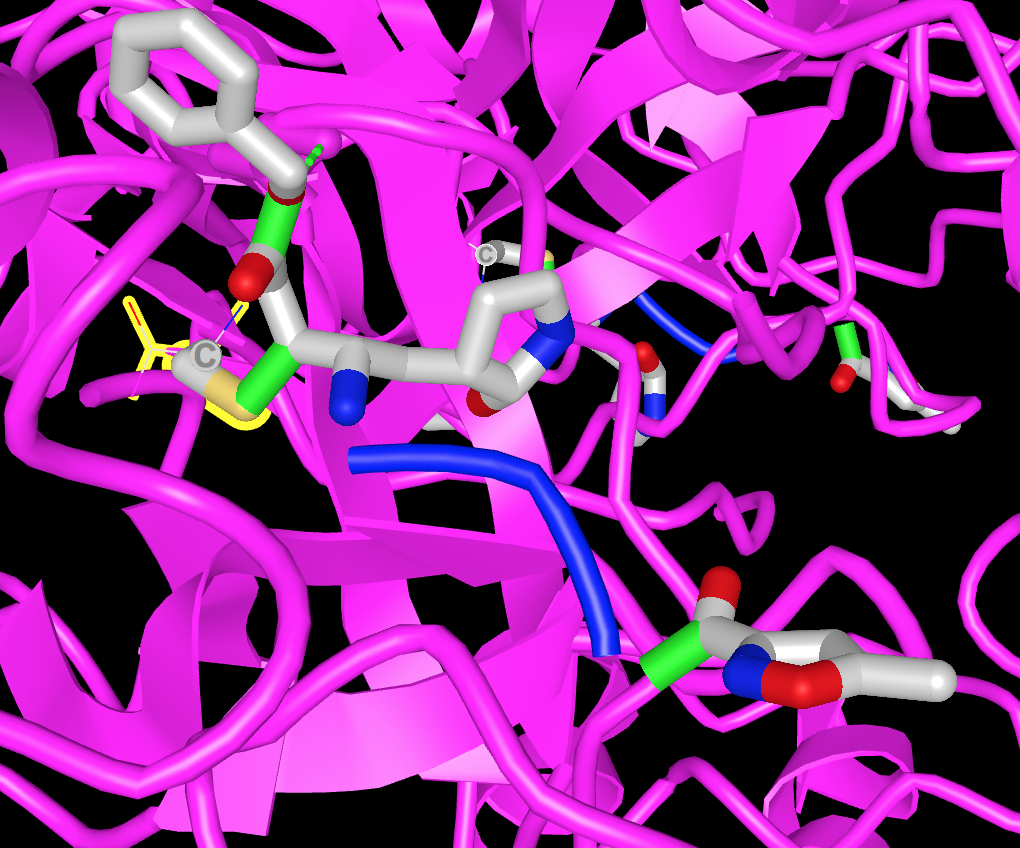
1. Click on the “Windows > View Sequences and Annotations” and select the first amino acid of the catalytic dyad within the 6LU7\_C sequence (*using a mouse, draw a box around the amino acid- it should turn yellow).*
2. Click on the Style button >> Side chains >> Stick. Now the side chain of the residue is visible. To make it more prominent, color it by clicking on the button called Color >> Unicolor >>Green.
3. Repeat for the second amino acid of the catalytic dyad.
4. Take a screenshot of these residues and include the image below (with labels showing key amino acids and key interactions)



**18.** Click on the “Windows > View Sequences and Annotations” and select cysteine 145 (*using a mouse, draw a box around the amino acid- it should turn yellow).*

19. Click on the View button >> Cross-Linkages >> Show.

20. Based on the chemical interactions, what type of inhibitor (reversible or irreversible) is N3?The image should show that N3 is an irreversible inhibitor since it forms a covalent bond with C145.

**

**Part 4: Compound Screening for Mpro of SARS-CoV-2**

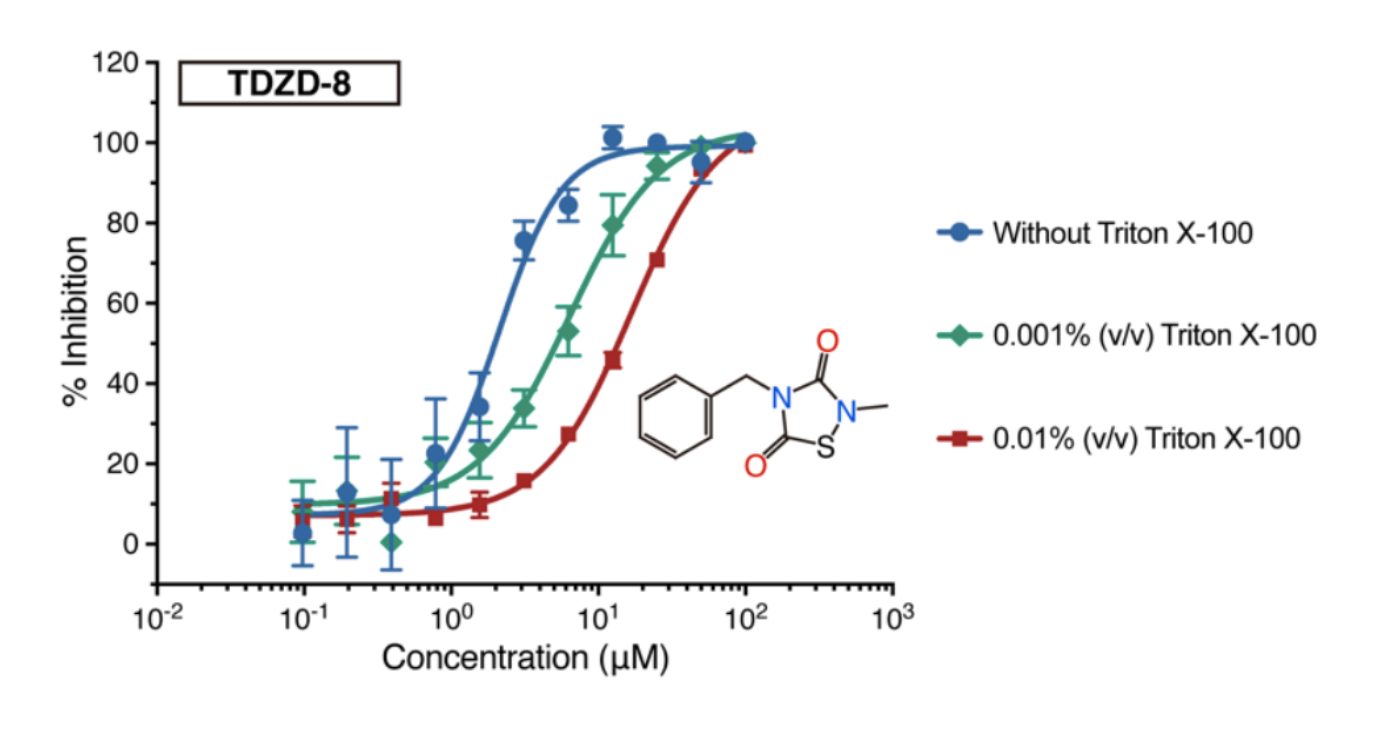
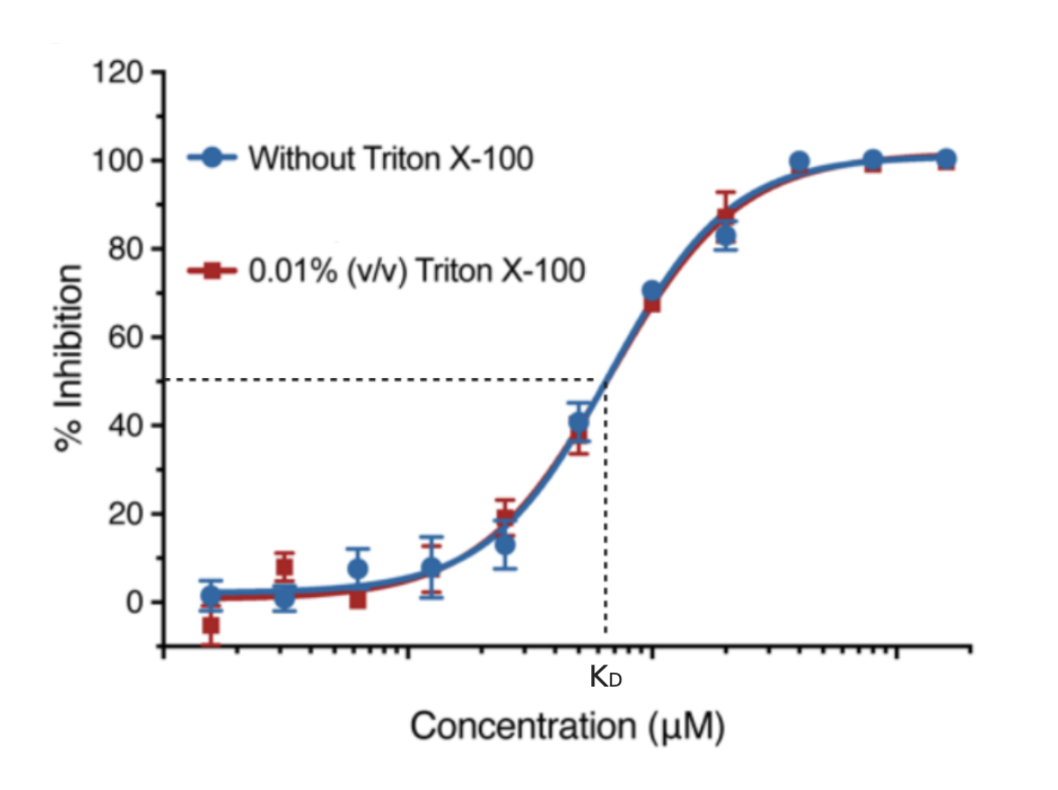
Learning Objective: This worksheet explores the primary hits, potential successful drugs, identified through *in silico* screening of Mpro.

*Congratulations!* After your head bosses at RollyColly evaluated your research on Mpro for potential drugs to bind it and inhibit SARS-CoV-2, they granted you approval to move forward with the next step in drug development. “Screening” is the initial assay used to identify chemical matter as complementary in size, shape, and charge to bind the active site. You measure the binding affinity of thousands of existing drugs *in silico* to identify which drugs to bind to and inhibit the active site of the Mpro.

|  |
| --- |
| ***In silico*** means screening compounds using computers and docking software. Molecules are docked using sophisticated modeling software, which utilize force fields and energy minimizations to simulate binding to protein. |

The experiment is performed in the presence and absence of 0.01% (v/v) Triton-X-100 and the KD is measured. Triton-X-100 is used as a detergent in antiviral drug experiments as a control.

The following inhibition trend was observed:



**Figure 7.**  Sigmoidal graphs indicating the half maximal inhibitory concentration (KD) needed to irreversibly bind in the active-site of Mpro (Jin et al., 2020).

The following five drugs have been FDA-approved or are being considered for clinical trial drug candidates for the Mpro active site. Pharmaceutical companies like RollyColly want to identify a drug that can bind the active site and produce the same binding affinity without Triton-X-100. The substance Triton-X-10 contains ecotoxic ingredients commonly found in antiviral drugs. The following KD values were measured:

|  |  |  |
| --- | --- | --- |
| **Drug** | KD with Triton-X-100 **(µM)** | KD without Triton-X-100 **(µM)** |
| Ebselen | 0.67 | 0.66 |
| Tideglusib | 1.67 | 1.70 |
| Carmofur | 1.35 | 1.82 |
| PX-12 | 21.81 | 21.85 |
| TDZD-8 | 16.92 | 2.15 |

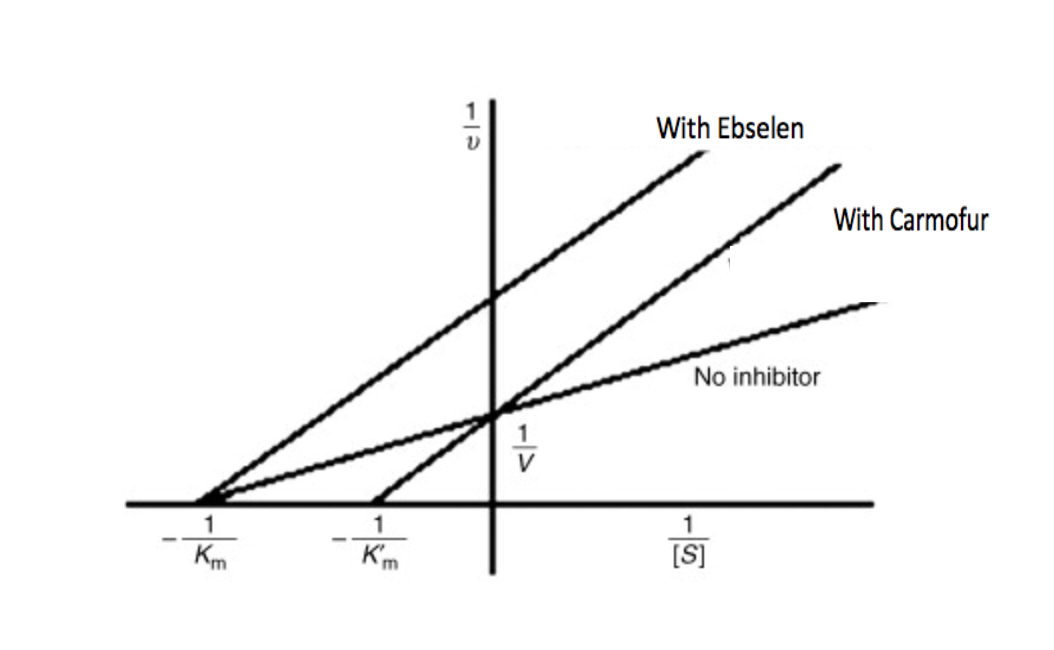
**Table 3.** KD values of the primary drug targets against the Main Protease (Mpro) of SARS-CoV-2 with and without the presence of Triton-X-100. (Jin et al., 2020)

Propose a plausible explanation for why there is a difference between the KD values (with vs without Triton-X) for TDZD-8?

* TDZD-8 has a stronger binding affinity to the Mpro active site, therefore it will function the best as an antiviral drug inhibitor. TDZD-8 works best without Triton-X-100 which is rare. This is highly advantageous, because Triton-X-100 may have harmful side effects to the person taking the drug. Usually we see drugs with higher KD values (lower affinity) without triton-X-100 as seen with Carmofur, Tideglusib, and PX-12.

Using the data in **Table 3**, what drug do you find of greatest and least interest to further explore as potential drug, and what was your reasoning behind the decisions you made?

|  |  |  |
| --- | --- | --- |
|  | Drug | Explanation |
| Greatest interest | Ebselen  (Tideglusib) | * Most effective 3CLPro inhibitor as it had the lowest KD value (0.67 µM) * Downfall that it might covalently bind too tightly * (only if they argue Ebselen might be irreversible due to is extremely low KD value) |
| Least interest | PX-12 | * least effective drug as it had the highest KD value (21 µM) |

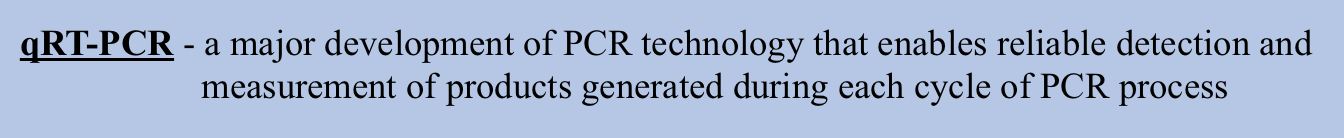


**Figure 8.** Lineweaver-Burk Plot for enzyme activity of Mpro with potential drugs, Ebselen and Carmofur (created using Biorender).

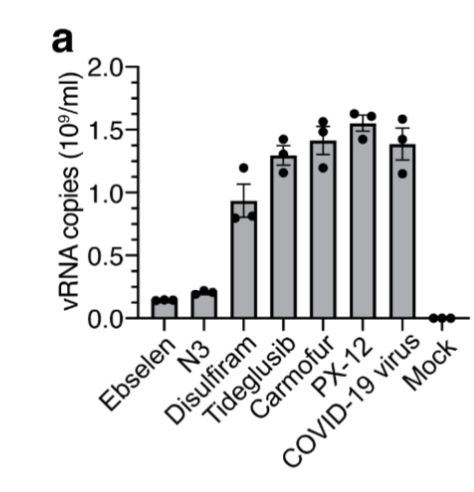
Based on **Figure 8**, what type of inhibitor is Ebselen? Make sure to highlight your answer.

A. Noncompetitve  
B. Competitve  
C. Mixed  
D. Uncompetitve

1. Ebselen is non-competitive.



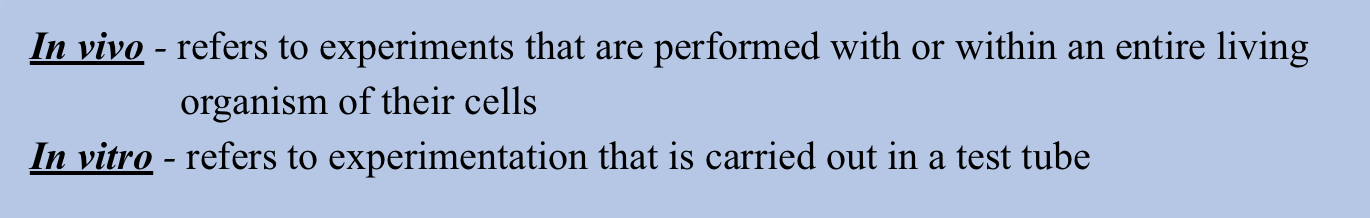
After comparing KD values of the primary drug targets, you want to further support the enzymatic inhibition results *in vitro*. Therefore, you performed a cell-based assay to assess if the compounds could prevent viral replication. The results are displayed in **Figure 9**.



**Figure 9.** Antiviral activities of the primary drug targets against the Main Protease (Mpro) of SARS-CoV-2 virus. The quantification of absolute viral RNA copies (per ml) in the supernatant at 72 h post infection (p.i.) determined by qRT-PCR analysis. Data are mean ± s.e.m., *n* = 3 biological replicates (Jin et al., 2020).

Based on all the data presented, which drug would you pursue to treat SARS-CoV-2 infected cells? Why?

Ebselen showed the strongest antiviral effects against SARS-CoV-2 infected cells. **Figure 10** showed the use of Ebselen yielded the least amount of viral replication.



What are the advantages and disadvantages of doing an *in vivo* vs *in vitro* experiment? Why must an *in vivo* experiment be tested next? What would be a good model organism? (Human cells, mice, etc.)

* Complete answers must consist of:
* Advantages and disadvantages for in vitro and in vivo (at least 2 bullet points for each)

For example:

* Advantages for *in vitro*:
  + More controlled setting of the experiment
  + Repeatable
  + Not harming animals/ ethics
  + Quality control
  + Fast
* Disadvantages of *in vitro*:
  + Cannot apply conclusions to a living organism
  + May not being a living subject tested, therefore results may be significantly different in an organism.
  + Might be other proteins/ interactions not accounted for that occur in an organism
* Advantages of *in vivo*:
  + Simulate real body conditions
    - Off drug effects?
    - How will the drug react in whole system vs isolated system
  + Study of complex interactions
  + Clinically relevant
* Disadvantages of *in vivo*:
  + Expensive
  + Time consuming
  + Ethical issues
  + Does not always correlate to humans

AND

*In vivo* explores the effects on a whole organism.

MDCK cells/ lung epithelial cells from humans- for better context

The results from applying Ebselen to *in vivo* experiments utilizing bat cells are favorable. The drug performed similarly to your *in vitro* results in terms of potency and efficacy. To further eradicate the disease, you perform *in vivo* experiments with bat cells to determine how to biochemically decrease the virulence of SARS-CoV-2. RollyColly went on to conduct clinical trials with Ebselen. By performing a BLASTp, areas of consensus were identified and verified the success of Ebselen for people from different populations suffering from SARS-CoV-2. The drug went on to be tested in the three phases of clinical trials. After several years of monitoring the drugs in human volunteers, the FDA approves Ebselen to be a feasible drug to inhibit SARS-CoV-2. The world is forever grateful for your team’s efforts and success in discovering this opportunistic drug!



https://www.freepik.com/free-vector/people-celebrating-together-illustration\_6525004.htm

**References:**

Bull, S. C. and Doig, A. J. (2015). Properties of Protein Drug Target Classes. Plos One *10*.

Cascella, M., Rajnik, M., Cuomo, A., et al. (2020). Features, Evaluation and Treatment Coronavirus (COVID-19). StatPearls [Internet].

Clark, Allan. (2016). Caspase Allostery and Conformational Selection. Chemical reviews. 116. 10.1021/acs.chemrev.5b00540.

Coronavirus (2020). Centers for Disease Control and Prevention.

Corum, J., Zimme, C. (2020). Bad News Wrapped in Protein: Inside the Coronavirus Genome. The New York Times.

COVID-19, MERS & SARS. National Institute of Allergy and Infectious Diseases. U.S. Department of Health and Human Services.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 1–10.

Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K. and Whittaker, G. R. (2020). Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. Journal of Molecular Biology.

Jin, Z., X. Du, Y. Xu, Y. Deng, M. Liu, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, D. Discovery, S. Advanced, C. Clinic, B. Mega-science, and M. Biosciences. 2020. Structure of M pro from COVID-19 virus and discovery of its inhibitors.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., et al. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine *382*, 1199–1207.

Lodish, H. (1970). Viruses: Structure, Function, and Uses. Molecular Cell Biology. 4th edition*.*

Mousavizadeh, L. and Ghasemi, S. (2020). Genotype and phenotype of COVID-19: Their roles in pathogenesis. Journal of Microbiology, Immunology and Infection.

Orrick, J. and Steinhart, C. (2004). Atazanavir. Annals of Pharmacotherapy *38*, 1664–1674.

Ryu, W.-S. (2016). Virus Structure. Molecular Virology of Human Pathogenic Viruses, 21-29.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K. and Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science.

**Professor Instructions**

* This case study is designed to be handed out **one** part at a time. Please verify student answers with the key before each student proceeds to the next step. We recommend this assignment be given as group work.
  + If the student does not answer correctly the first time, have the student relook at the question and try again before handing out the next part.
  + We estimate the case study to take 3-4 hours.
* The questions for the students to answer are highlighted in yellow while the answers are colored red.
* We recommend providing a brief background on viruses to help your students better understand the context of the case study.
* It is highly recommended that the case study is given after students have been taught about amino acids, *in vitro* and *in vivo* experiments, Kd, protein structural levels, and enzymes: inhibitors, proteases, active sites, Lineweaver-Burk Plot, and chymotrypsin.
* For **Part 6**, *TDZD-8* serves as a distractor but it tests the student’s understanding of Triton-X-100 as a detergent from lab

QUESTIONS:

1. Have you learned more about the final due date?
2. Should we start with Figure 1 for each section or keep the numbering consistent throughout the whole document?
3. For Part 1 A and B, we have a total of 8 questions. Should we trim down/ is the section too long?-should we tell the professors to distribute the sections separately or together?
4. Is there a standard way of providing definitions for students? Underlining? Adding an asterisk? Should the definitions be on the side, in a footer, or at the top of each section?
5. Part 1B question 1... is the role for the endoribonuclease correctly described?
6. Instructions for sequence alignment are replicated from the Happy Blue Baby case study. Is that ok? Or do we need to alter the instructions more?
7. Should we cut out Part 3B?
8. For the very last question, is it too broad? Is it a good question to end on?