SARS-CoV-2 Life Cycle: A Webquest

## Learning Objective: To explore public biological data resources to learn about a topic of interest - Life Cycle of SARS-CoV-2

## Introduction

Many viral and host cell (human) proteins are involved in the SARS-CoV-2 life cycle (Figure 1).

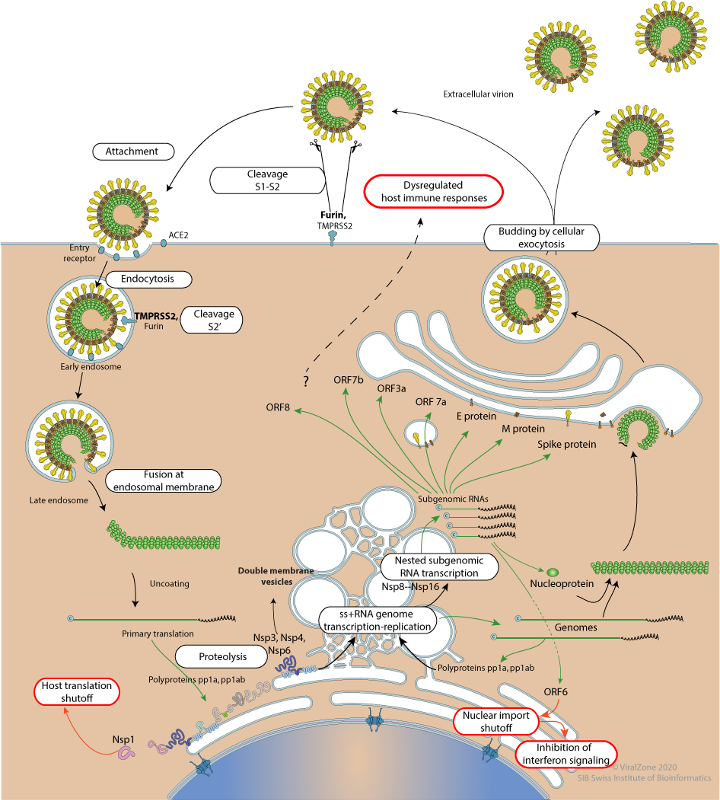


Figure 1: SARS-CoV-2 Life Cycle (from <https://viralzone.expasy.org/resources/Coronavirus_cycle.png>)

### Q1. In the above figure showing the SARS-CoV-2 life cycle mark the following proteins and reflect on their roles in the life cycle.

1. ACE2
2. Spike protein
3. TMPRSS2
4. Furin
5. Main Protease
6. PL-Pro
7. RNA-directed RNA polymerase
8. Nucleoprotein

Ans.

|  |
| --- |
|  |

Information about these proteins - their sequence, structure, and functions are available from various public data resources. Scientists use these resources to ask questions, design experiments, interpret results, and learn about these proteins and their interactions. In this webquest students will explore some of these public bioinformatics data resources to learn about names, shapes, and functions of proteins that play key roles in the SARS-CoV-2 life cycle.

## Data Resources Available for Exploration:

### RCSB PDB:

* This resource provides access to 3D structures of biological macromolecules and their various complexes.
* Go to <https://www.rcsb.org/> type the PDB identifier provided in the top search box to access information about that structure
* The Structure Summary Page that opens for the specific PDB ID provides a variety of information about the structure and a quick glimpse of the structure.

For a list of different kinds of information available from this page see the box below.

|  |
| --- |
| *Box 1: Navigating the Structure Summary Page*  1. **PDB ID and Title** - the PDB ID is a unique identifier for the structure. The title tells you what the structure is about  2. **Snapshot** - of what the structure of the molecule/complex looks like.  3. **Authors** – who solved the structure  4. **Publication** - when was this structure deposited and published  5. **Literature** –access the article that describes the structure. This section also includes links to PubMed page and the abstract of the article describing this structure, when available.  6. **Macromolecules** – All proteins and nucleic acids present in the structure are listed here. Each unique type of macromolecule or molecular chain is listed as a separate entity. There may be multiple copies of each molecule in the structure. This section also links to other resources that provide information about the sequence and functions of the protein.  7. **Oligosaccharides** - listing of all carbohydrate chains (if present in the structure) with two or more sugars linked together that are either floating in the solvent or linked to the protein or other molecule(s)  8. **Small molecules** – All ligands, ions, cofactors, inhibitors that are present in the structure are listed here. You can find links here to explore the interaction of this ligand with the target protein.  9. **Experimental details** – describe details about how the structure was determined  10. **Structure quality** – shows a slider that provides insights about the quality of the structure and its agreement with the experimental data and geometric standards.  See<http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction> for details |

* Click on the 3D View tab of the structure summary page to interactively explore the structure. There are 3 main sections on this screen
  + Sequence panel (top left)
  + 3D-canvas (white space where the 3D structure is shown). All protein chains are represented as different colored ribbons, nucleic acids as a spiral ladder, and saccharides as solid 3D shapes (spheres, cubes etc.).
  + Control panel (blue column on right) with menus for Structure, Measurements, Components, etc.
  + Use various mouse controls to rotate and translate the molecule you are viewing.
  + Hovering the mouse over any object in the 3D canvas will display information about that item in the bottom right corner of the 3D canvas.

### UniProt:

* This resource provides access to the protein sequence as well as information about the function, expression, interactions etc. of the protein
* Go to <https://www.uniprot.org/> and type the UniProt Identifier in the top search box.

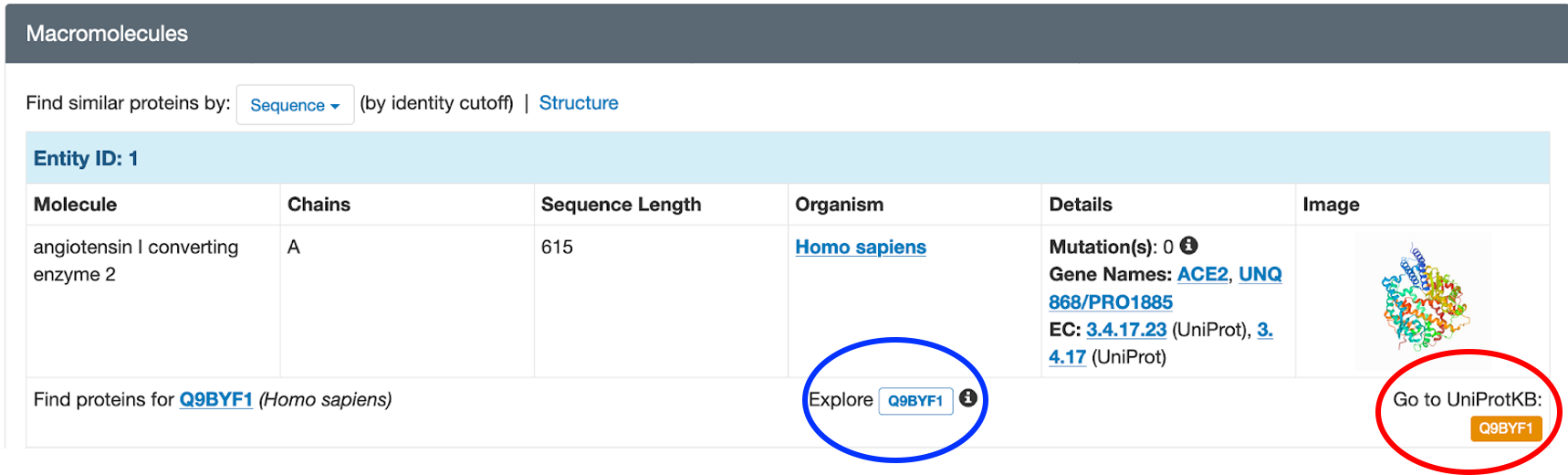
For a list of different kinds of information available from the UniProt page see the box below.

|  |  |  |
| --- | --- | --- |
| 1. Entry | 5. Subcellular location | 9. Interaction |
| 1. Function | 6. Pathology & Biotech | 10. Structure |
| 1. Name | 7. PTM\* / Processing | 11. Family & Domains |
| 4. Taxonomy | 8. Expression | 12. Sequence |

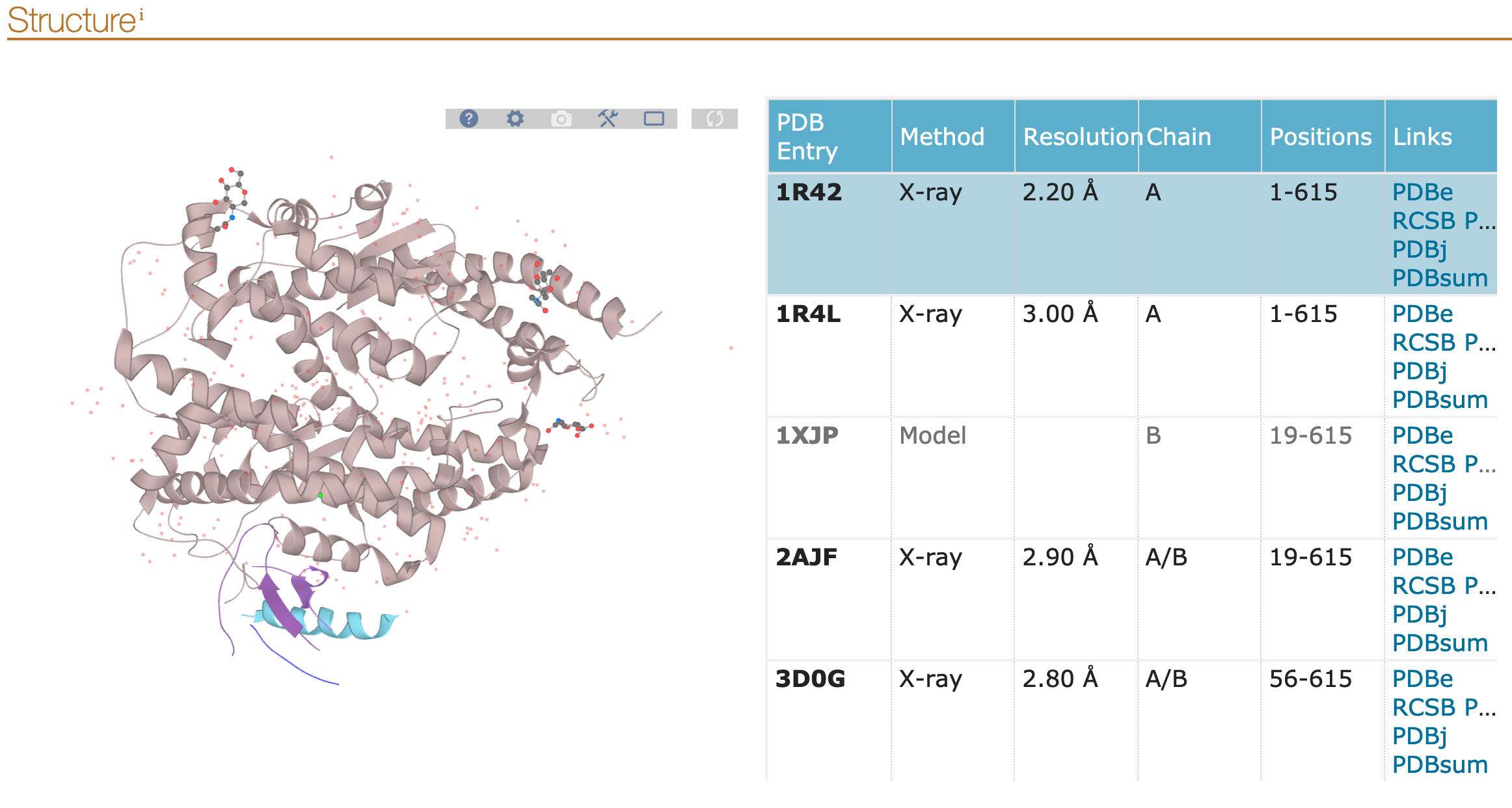
\* PTM: Post-translational Modification

### C. Cross-talk between data resources

* The PDB (archive for protein structures) and UniProt (archive for protein sequences) are connected to each other via hyperlinked identifiers.
* On PDB’s structure summary pages UniProt IDs are listed in the Macromolecules section. They provide a link to the UniProt entry page (orange box in red circle) or can be used to search the PDB archive for other structures with the same sequence (white box with Explore written next to it, blue circle).



* On the UniProt page, PDB structures are listed in the Structures section of the page



## Exploration:

A few proteins that play key roles in the SARS-CoV-2 life cycle (that you identified in Figure 1) are listed in the table below.

### Q2. Visit the PDB and UniProt data resources to gather information about these proteins and fill in the table.

The first row is completed for you as an example.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PDB ID | Snapshot | Name of ... | UniProt ID | Information from UniProt |
| 1r42 |  | The green colored protein is: angiotensin I converting enzyme 2  Gene name is: ACE2 | Q9BYF1 | In which organs/tissues is this protein expressed? (Hint: see Expression section). lung, small intestine, heart, kidney testis |
| 6vxx |  | The orange colored protein is: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Gene name is:  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_ | Which human enzymes cut this protein to form S1 and S2? (Hint see first paragraph in the function). \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 6lu7 |  | The purple colored protein is: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_ | What is the name of the polyprotein?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  List the names of another protease in this polyprotein?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 7jrn |  | The rainbow colored protein is: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_ | What is another name for this protein?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 7bzf |  | The green protein is: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  What is the main function of this protein?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_ | What is the name of the polyprotein?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  List the names of any other proteins that helps it with its function?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 6vyo |  | The magenta protein is: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  What is the Gene name? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_ | What is the function of this protein?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  In which host cellular compartment is it made?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Feel free to refer to the learning materials that describe the steps of the SARS-CoV-2 life cycle in detail.

### Q3. Using the knowledge that you have gathered above complete the following sentences describing the SARS-CoV-2 lifecycle.

1. Viral attachment and infection
   1. The \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ binds to human \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ to attach and infect \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ cells.
   2. Following this binding the virus is endocytosed and cleaved by the host \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ to activate fusion of the viral and endosomal/host cell membranes - leading to release of the viral genome into the host cell
2. Replication of viral genome
   1. The +ive strand genomic RNA is translated to produce \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ , which are cleaved by viral proteases such as the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. In double membrane vesicles the enzyme \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ copies the +ive strand genome to make -ive and +ive strands of the complete genome and sub-genomic RNA.
3. Viral assembly and release
   1. The sub-genomic RNA are translated to make Spike, Matrix, Envelope, and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ , proteins
   2. These proteins are processed in the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ compartment and help assemble new virus particles that are exocytosed.