Teaching Enzymology with the Protein Data Bank: From Pandemic to Paxlovid

Exploring the SARS-CoV-2 Main Protease

Presenter: Dr. Paul Craig, RIT, NY.

Overall Learning Objective: To introduce or reinforce the use of a number of RCSB PDB resources for teaching protein structure and function. This handout has all of the links that you will need for the exploration that will be presented.

Enzymology Learning Objectives (exploring the structure-function relationship of enzymes).

- Visualize the enzyme's overall structure and active site
- Visualize the interactions between enzymes, substrates, and products.
- Explain how enzymes catalyze chemical reactions
- Evaluate the effects of different types of enzyme inhibitors and regulators.

Open the Structure Summary Page (SSP) for COVID-19 Main Protease PDB ID 6LU7

1. Review the elements of the COVID-19 Main Protease structure (PDB ID 6LU7) SSP
   a. Structure Summary
   b. Structure (visualize in Mol*)
   c. Annotations
   d. Experiment
   e. Sequence (including mapping of sequence-based annotations on the structure)
   f. Genome
   g. Versions

Poll question 1: Which of these tabs have you explored before?

- Structure Summary
- Structure
- Annotations
- Experiment
- Sequence
- Genome
- Versions

In the following sections we will explore the **Structure Summary**, **Structure** (or Mol*), and

Sequence (including annotations mapped on the 3D structure) tabs

2. Review the PDB ID 6LU7 Structure Summary tab
   a. Assemblies and options for viewing the 3D Structure, including Sequence Annotations (more on this later)
   b. wwPDB validation for structure quality
   c. Literature/PubMed
   d. Macromolecules, including EC Classifications
      - Mutations
      - Connection to UniProt
      - Entity Groups
      - Sequence Annotations
   e. Click on Biologically Interesting molecules to show interactions

3. Visualize the structure using Mol* in the PDB ID 6LU7 Structure tab
   a. Structure takes us to Mol*.
      - Click on any residue to show interactions in its neighborhood.
      - Click on the ligand to show interactions in its neighborhood
        - Hide the protein chains
        - Hide the waters
        - Right-click and drag to zoom in on one ligand and its surrounding residues
        - Click on the … next to [Focus] Surroundings (5…)
        - Hide and show Ball & Stick Representation
        - Hide and show Non-covalent Interactions Representation
        - Zoom in and hover over a non-covalent interaction to learn more details
      - Select the bound inhibitor (in selection mode) to change its representation
      - Electron density - click on a residue to see
      - Validation report provides information on clashes in Mol*
      - Global Symmetry 3D-View brings up the symmetry browser with symmetry axes
      - Go to Documentation > Exploring a 3D Structure > 3D Viewers > Mol* > Getting Started
   b. Find similar assemblies. Review refinements
   c. Finish with Documentation for Mol*

Poll question 2: Which features of Mol* have you explored before?
- Hide/Show Polymer/Ligand/Water
- Add Representation
- Set Coloring
- Measurements
- Structure Motif Search
- Select Atom/Residue/Chain

- Assembly Symmetry
- Screen Capture

4. In the Sequence Annotations view for PDB ID 6LU7 try the following to learn about the enzyme's structure and function
    a. Identify the active site residues: His-41 and Cys-145
    b. Explore the interaction between the two active site residues
    c. Find the covalent bond between the substrate analog and the catalytic residue
    d. Visualize the mechanism of catalysis in this enzyme

Poll question 3: What other features of the Sequence Annotation Viewer do you think you would use to examine/teach about this enzyme's catalysis?

- Secondary Structure
- Rotamer Outlier
- Uniprot Accession
- Mutated Residues
- CATH
- SCOP
- Evolutionary Classification of Protein Domains (ECOD)

5. On the structure summary page for PDB ID 6XHM find and visualize the inhibitor bound to the active site.
    a. Scroll down the structure summary page to Small Molecules
    b. On the right hand side, click on the Interactions drop down menu and select Focus chain C [auth A]. This will take you to a Mol* page that focuses on the ligand binding site.
    c. Find the 3-character ligand ID and enter it into the chat
    d. Visualize the ligand's interaction with the enzyme. Can you identify the catalytic residues? List them.
    e. Click the three dots on the right side of [Focus] Surroundings (5 A). Use the eye icon to turn the ball-and-stick side chains and the non-covalent interactions on and off.
    f. Save an image by clicking on the shutter icon on the Toggle Menu
    g. Select the Toggle menu again.
    h. Choose residue.
    i. Select the ligand, V2M.
    j. Paint it magenta
    k. Measure the distance for the hydrogen bond between His-41 and Cys-145.
        - Select Toggle
        - Choose Atom/Coarse Element Instance
        - Click on Measurements > Add
        - Select the yellow sulfur for Cys-145
        - Select the blue nitrogen for His-41
        - Click on Distance to see the distance displayed

Chat Q: What is the 3-letter code for the ligand

Poll question 4: Which of these tasks are you most likely to use in teaching enzymology?

- Protein visualization with Mol*
- Sequence and Structure relationships through Sequence Annotation Viewer
- Accessing literature sources through the RCSB Protein Data Bank website
- Exploration of the symmetry of protein structures
- Comparing enzyme mechanisms of closely related protein structures (e.g., mammalian serine hydrolases)
- Comparing enzyme mechanisms of different families of hydrolases (e.g., serine hydrolases, cysteine hydrolases, metalloproteases)

From Pandemic to Paxlovid, and Beyond …

Presenter: Dr. Shuchismita Dutta, Rutgers University, NJ.

Using prior knowledge to develop a drug for SARS-CoV-2

Open the Structure Summary Page (SSP) for COVID-19 Main Protease PDB ID 6LU7

6. Query for all other structures in the PDB with the E.C. # specified in the Macromolecules section of the SSP.

Chat Q: What is the E.C. number?

   a. How many structures did you find?
   b. What organisms are these enzymes derived from?

7. Browse the archive to find structures with the E.C. number identified above.
   a. How many structures did you find?
   b. What organisms are these enzymes derived from?
8. The compound PF-00835231 was designed as an IV inhibitor to treat SARS-CoV infection
   a. Type the drug name PF-00835231 in the top search box and run the search. Review the results returned to identify structures that include this molecule.
   b. Open the SSP of a structure with this ligand bound to figure out which of the small molecule identifiers corresponds to the inhibitor compound.

Chat Q: What is the ligand ID of the inhibitor?

Reflection: Are there other ways of identifying the Ligand ID of the drug molecule?

   c. How many PDB structures include this compound? List the PDB IDs and structure titles here.
   d. Examine the “Scientific Name of Source Organism” of the entries in the search results. List any two of them.

9. Begin by querying for the compound by typing the name PF-07321332 (or Paxlovid’s key ingredient - Nirmatrelvir) in the top search box (on www.rcsb.org).
   a. How many structures are bound to the compound PF-07321332?
   b. What is the ligand identifier for the compound PF-07321332? (Hint: look through the list of small molecules in each of the entries in the results list and find the common ligand).

Chat Q: What is the ligand ID of the compound PF-07321332?

   c. For the inhibitor bound structures, where were the proteases derived from?

Reflection: Are there patterns and trends in the binding of this drug molecule? What can we learn from that about this inhibitor/drug?

10. Compare using the tool https://www.rcsb.org/alignment:
    a. overall structures of the proteases from SARS-CoV and SARS-CoV-2 bound to PF-00835231 (Hint: use the PDB IDs 6xhl and 6xhm)
    b. interactions of SARS-CoV2 Main Protease with PF-00835231 and the new molecule PF-07321332 (Nirmatrelvir, active ingredient of Paxlovid). (Hint use PDB IDs 6xhm and 7rfw to examine and compare inhibitor binding)

Reflection: What did you learn about the structures of the target enzymes and the two inhibitors from these comparisons?

Preparing to deal with future SARS-CoV like pandemics

11. Search for all structures in the PDB that have the bound form of Nirmatrelvir in the structure. [Hint: use the ligand identifier for the Nirmatrelvir (bound form) that you identified earlier]
    a. Group the results of the search as follows:
        - Return options from “Structure” to “Polymer entities”,
        - “group by option” to “UniProt Accession” and
        - “display as” option to “Groups” before re-running the search

Reflection: Are these the only SARS-CoV-2 Main Protease structures in the PDB

    b. Explore the contents of one of the grouped search results - Select the group that corresponds to SARS-CoV-2 and has several structures (e.g., >30) See an example here.

Polling Q 5: Why do you think that the polymer sequences shown line up with only a small part of the complete UniProt sequence? Select the best answer.

- This is the only part of the structure that was visible in the experimental results (the rest of the structure was disordered)
- This is the active site of the enzyme
- This region of the sequence corresponds to MPro in the polyprotein
- The rest of the polymer sequence represents multiple repeats of the M-Pro protein.

    c. Identify any PDB structures with mutations (especially near the enzyme active site residues). List them.
    d. Explore the impact of this mutation on the structure and function of the enzyme(s). What does this tell you about possible resistance to Nirmatrelvir?