RCSB PDB
PROTEIN DATA BANK

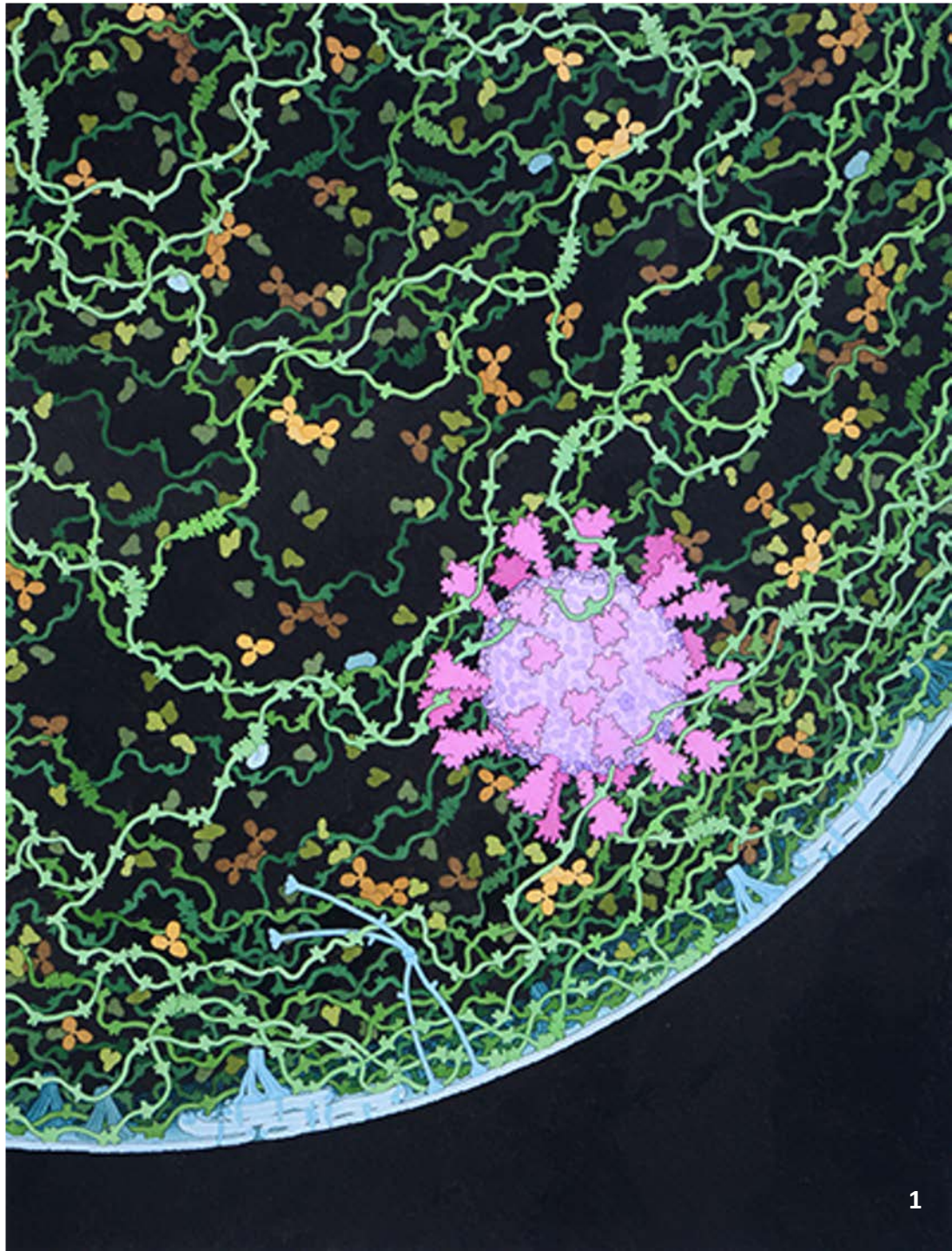
rcsb.org

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

Stephen K. Burley, M.D., D.Phil.
Director, RCSB Protein Data Bank, Rutgers University, NJ

Paul Craig, Ph.D.
Rochester Institute of Technology, NY

Shuchismita Dutta, Ph.D.
RCSB Protein Data Bank, Rutgers University, NJ

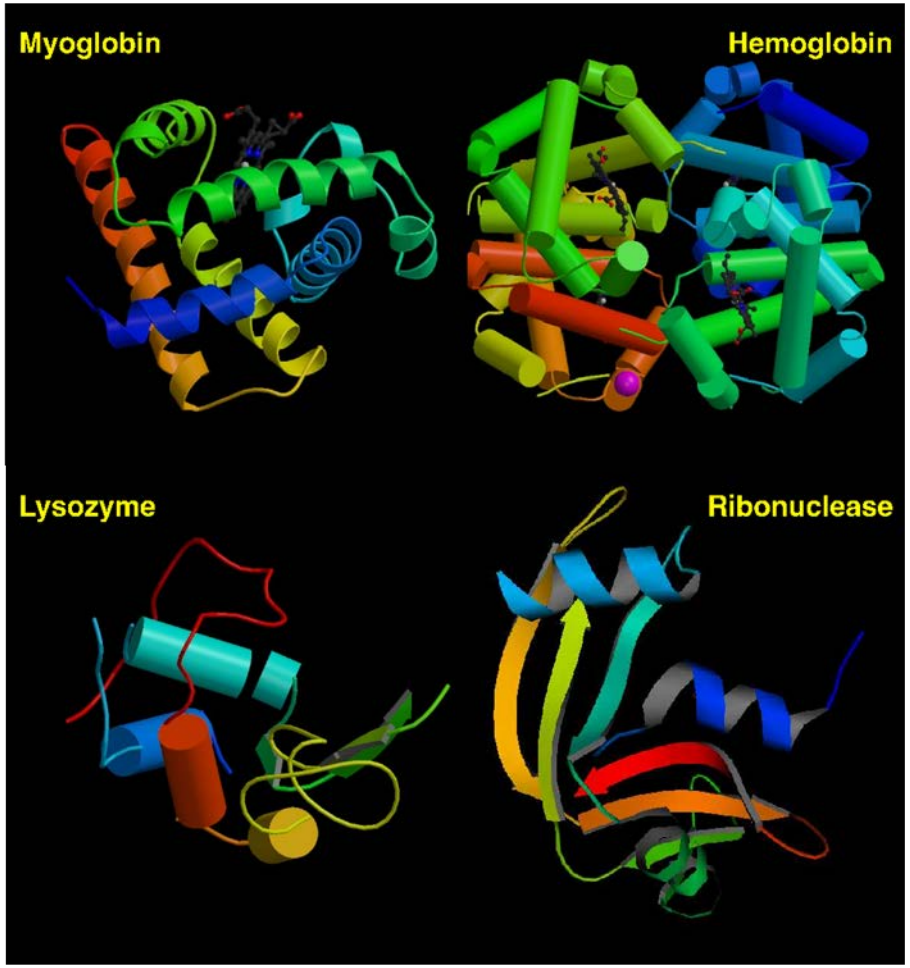


Today's Agenda

- Introduction to the Protein Data Bank and SARS-CoV-2
  - Stephen K. Burley, M.D., D.Phil. - Director, RCSB Protein Data Bank
- Exploring the SARS-CoV-2 Main Protease structure using RCSB.org
  - Paul Craig, Ph.D. - Rochester Institute of Technology
- Making connections using RCSB.org
  - Shuchismita Dutta, Ph.D. - RCSB Protein Data Bank
- Discussions

Protein Data Bank (Established 1971)

- PDB 1st online Open Access digital data resource in all of biology
- Founded 1971 with 7 protein structures
- Single global **archive** for protein and DNA/RNA experimental structures
- **Open Access to >222,000 structures!**
- wwPDB Partnership founded in 2003
- Members: RCSB PDB (US), PDBe (EMBL-EBI), PDBj (Japan), and PDBc (China); plus EMDB (3DEM) and BMRB (NMR)

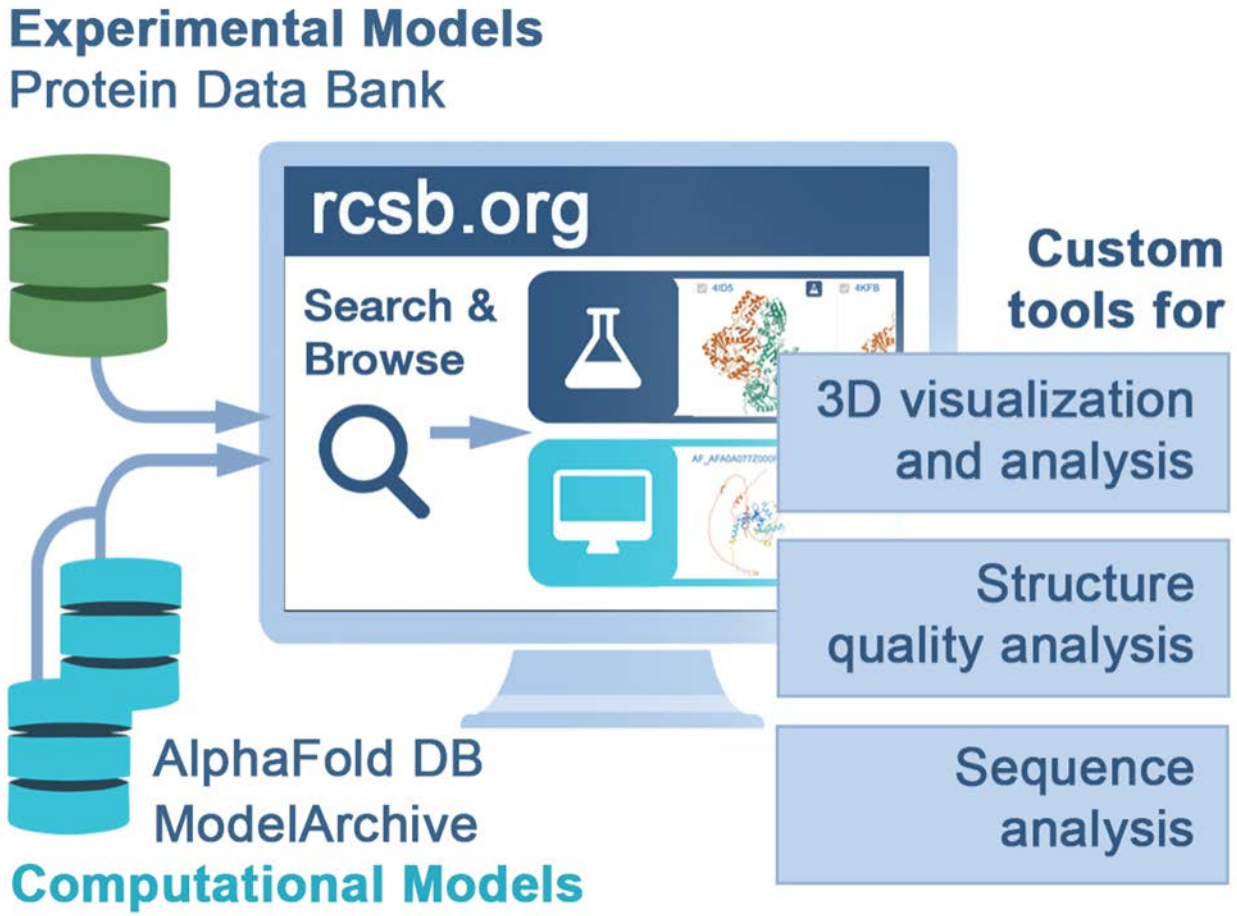


Structures that Inspired Launch of the PDB

Protein Data Bank (1971) *Nature New Biology 233,* 223.
Worldwide Protein Data Bank (2019) *Nucleic Acids Research 47*, D520–D528.

RCSB.org Research-focused Web Portal: One-Stop-Shop for Public 3D Biostructure Data

- RCSB.org delivers
  - >222,000 PDB structures
  - >1 million Computed Structure Models (CSMs) from AlphaFold DB and the ModelArchive
- RCSB.org data exploration and visualization tools used by many millions of researchers, educators, and students worldwide
- Provenance/reliability of both data types are clearly identified



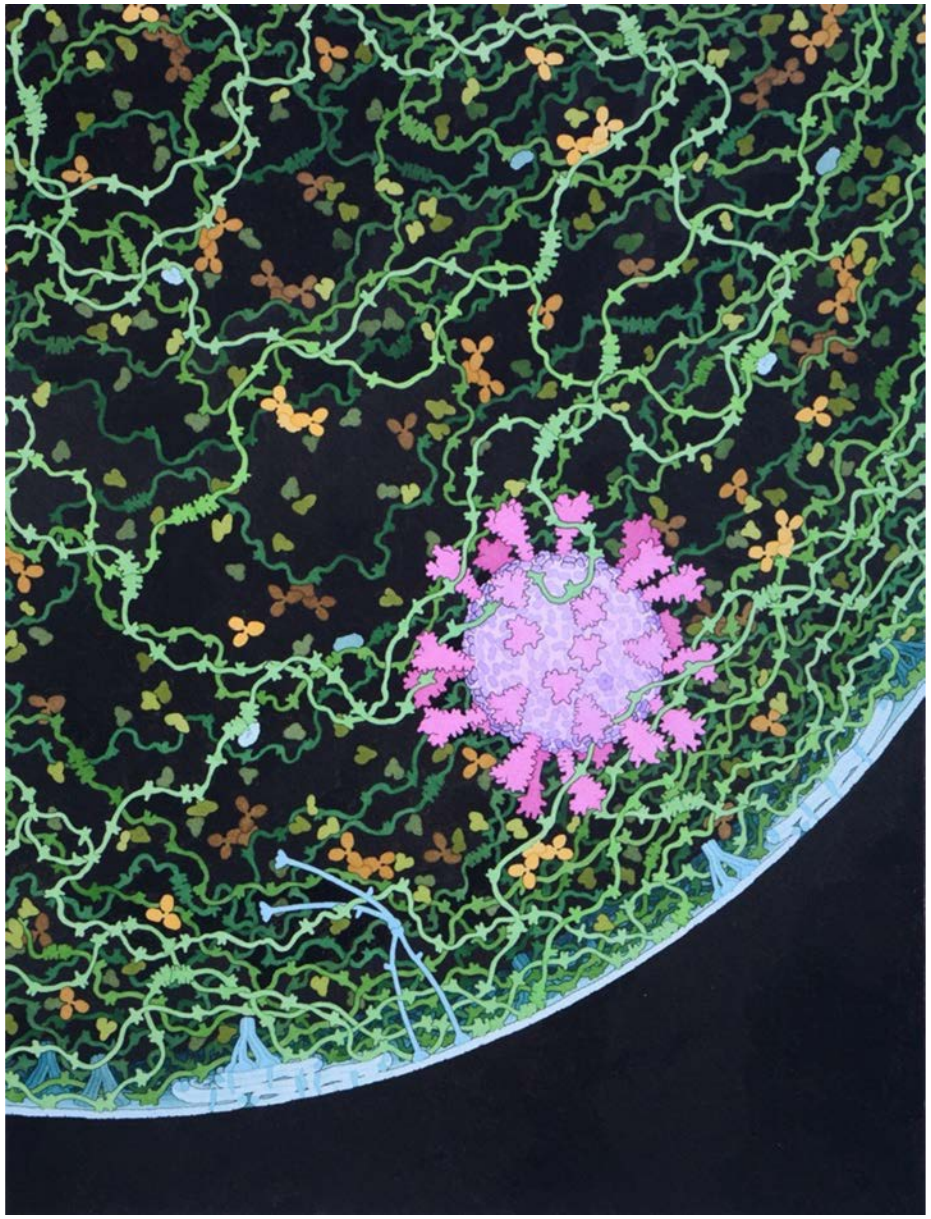
Burley *et al*. (2023) *Nucleic Acids Research 51*, D488-D508.

RCSB.org Opt In for Computed Structure Models

PDB Essential for Responding to Emerging Viruses

- SARS-CoV Epidemic 2002
  >240 SARS-CoV structures in the PDB
- MERS-CoV Epidemic 2012
  >170 MERS-CoV structures in the PDB
- COVID-19 Pandemic 2019
  >4,300 SARS-CoV-2 structures in the PDB

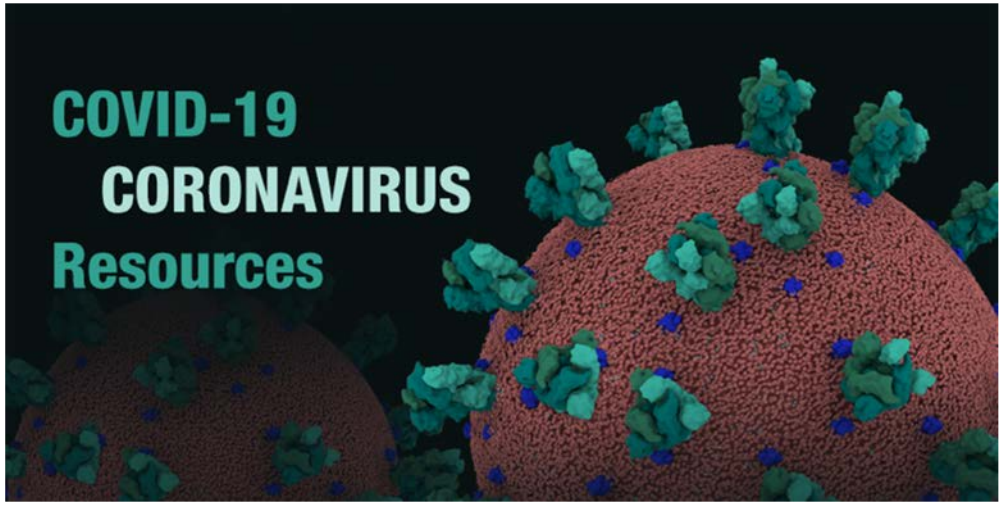
- Effective mRNA vaccines designed and antiviral agents discovered/develped using PDB structures of SARS-CoV, MERS-CoV, and SARS-CoV-2 proteins

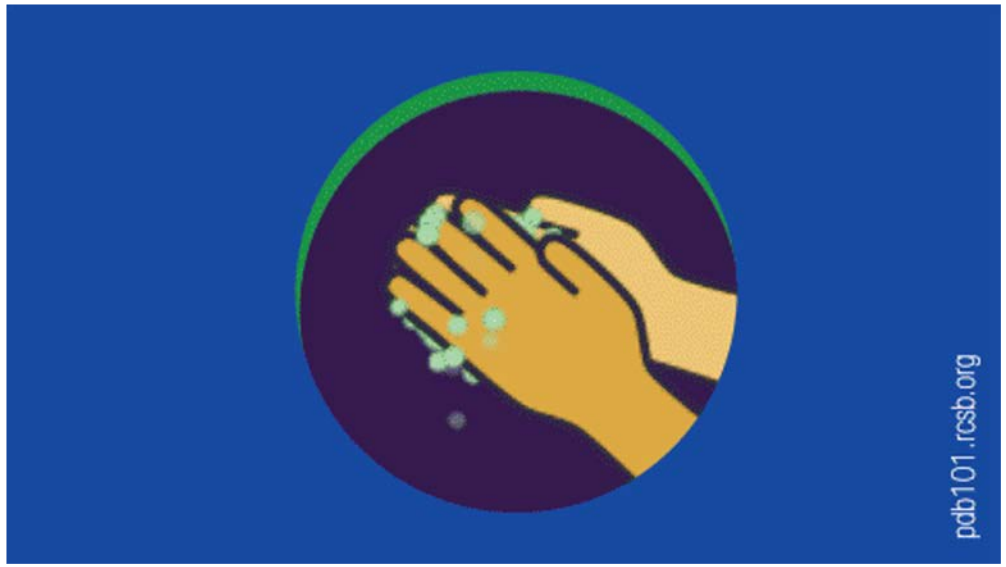


Respiratory Droplet, 2020; David S. Goodsell

RCSB PDB Response to COVID-19

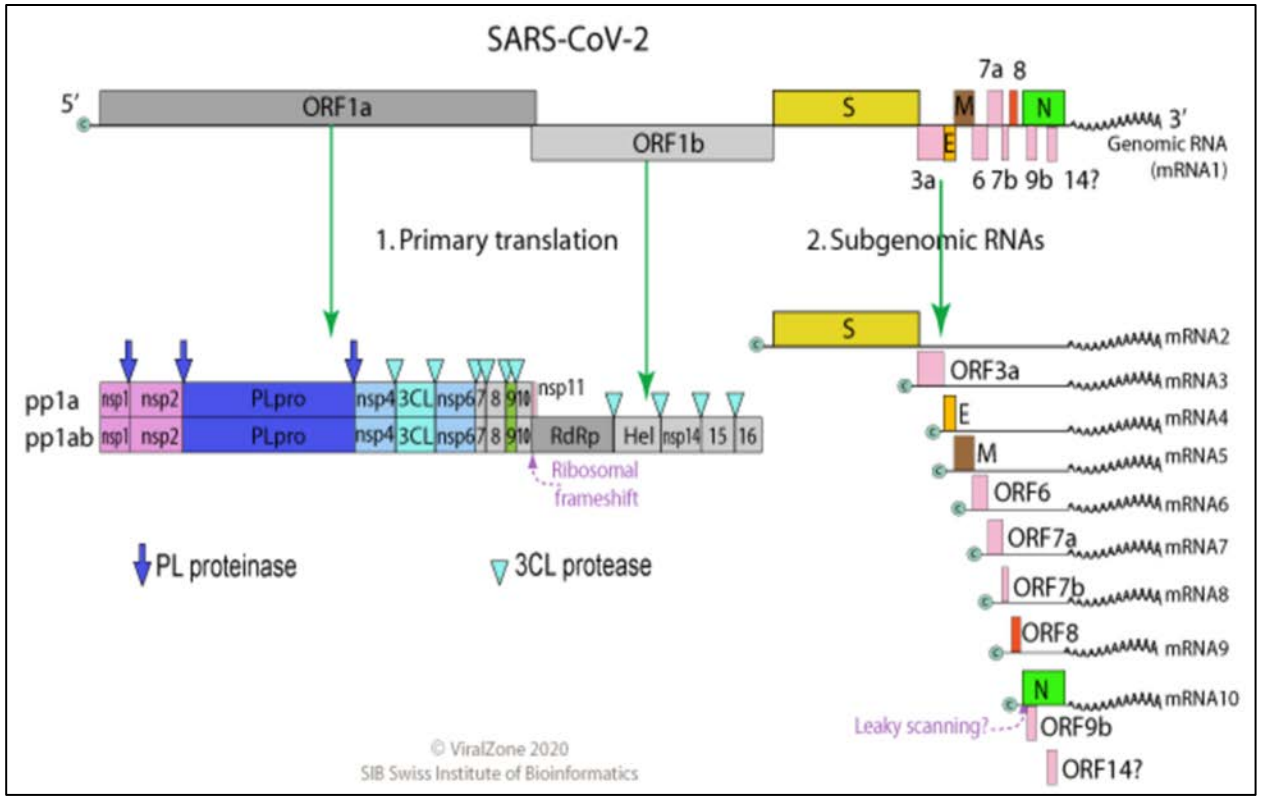
- Biocuration of COVID-19 structures prioritized, including post-release revisions (e.g., citation updates)
- PDB depositors strongly encouraged to release COVID-19 structures immediately
- Consistent taxonomy name/ID
  - Severe acute respiratory syndrome coronavirus 2; 2697049
- Consistent UniProt referencing
  - P0DTD1, P0DTC1, P0DTC2, P0DTC9
- Released structures and educational resources updated at https://RCSB.org/covid19





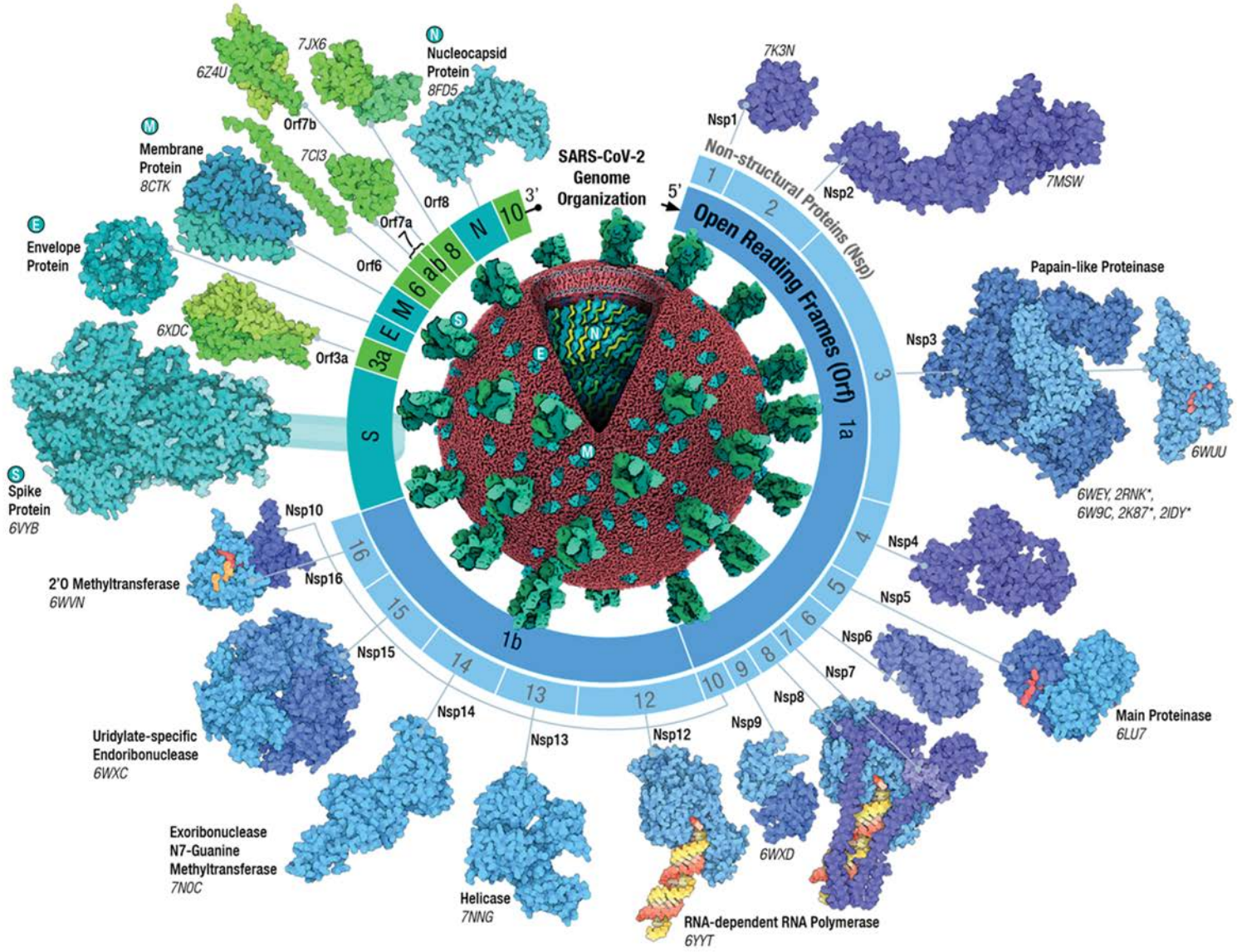
Coronavirus (SARS-CoV-2) Genome Organization

- Viral genome is a single-stranded, +ve-sense,
  5’-capped, 3’ polyadenylated messenger RNA
- Non-structural proteins expressed as polyproteins requiring enzymatic cleavage by
  1. Main Protease (Mpro) and
  2. Papain-Like Proteinase (PLpro)



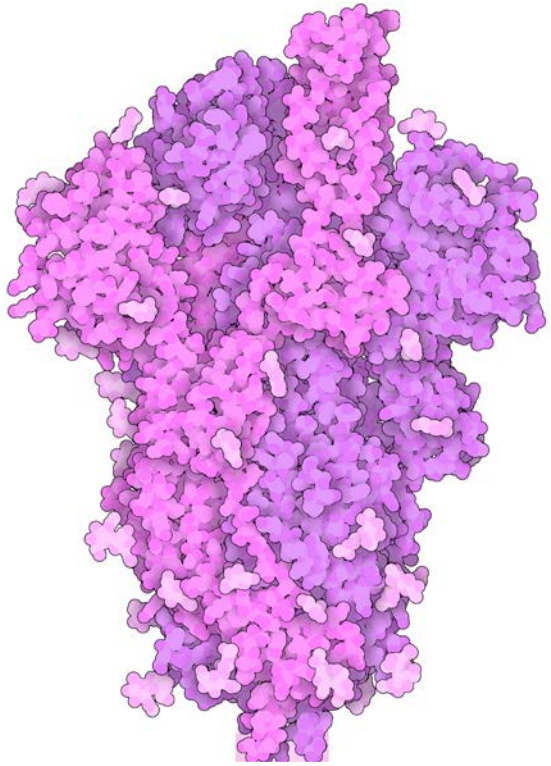
Near Complete SARS-CoV-2 Parts List in 3D

SARS-CoV-2 Fusion, 2020;
David S. Goodsell



Structure-Based Vaccine Design: Spike Protein

- Spike Protein
  - ~1,800 3DEM/Crystal structures
  - All Down and 1 Up/2 Down Trimers
  - Post-fusion Trimers
  - Complexes with ACE2, Fabs, etc.
- mRNA vaccine design relied on PDB structures of SARS-CoV and MERS-CoV spike proteins
- >5 billion vaccinated worldwide!
- Tens of millions of lives were saved!
- Hundreds of millions spared serious illness, hospitalization, etc!



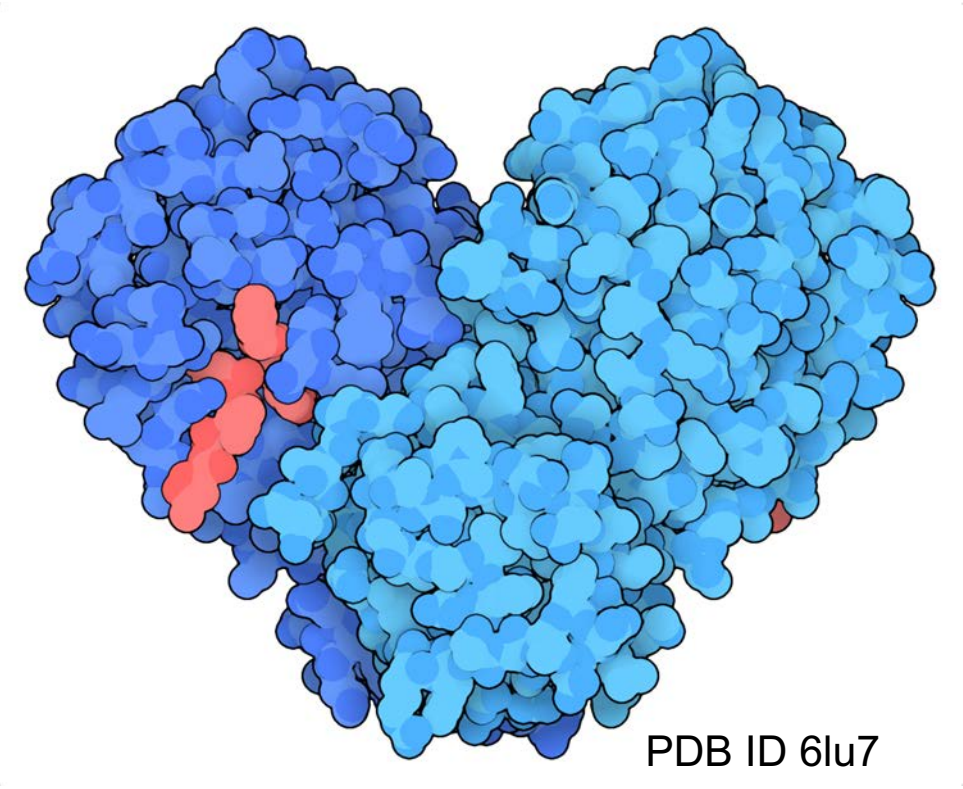
PDB ID 6vsb

Vaccine Discovery and Antibody Discovery Target

Wrapp *et al*. (2020) *Science 367*, 1260-1263.

Main Protease: Achilles Heel of SARS-CoV-2

- Nsp5/Main Protease (Mpro)
  - >1,450 Apo/Co-crystal structures
  - Target of Pfizer's nirmatrelvir (+ritonavir=Paxlovid)
- Paxlovid is approved for outpatient treatment of individuals infected with SARS-CoV-2



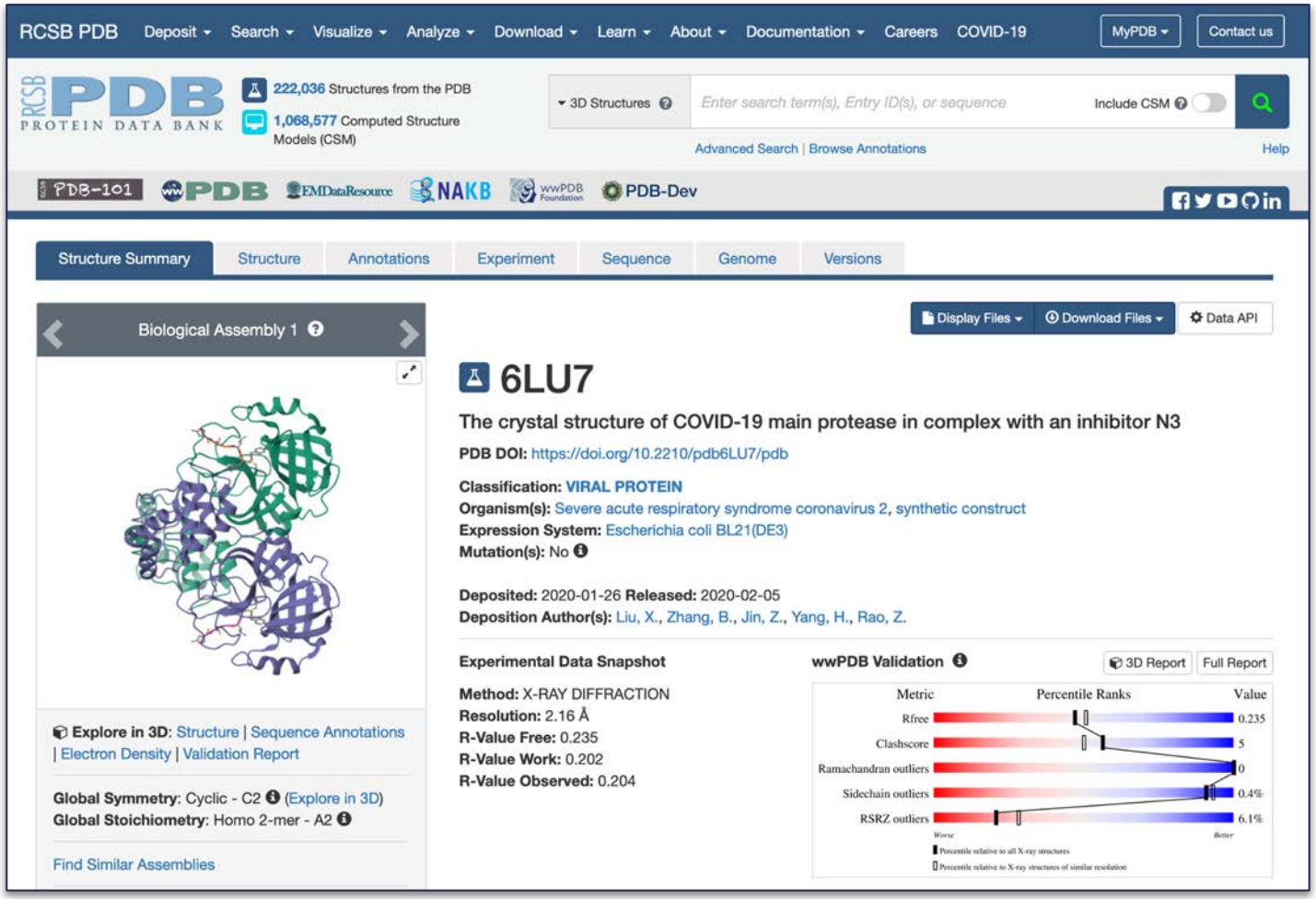
Drug Discovery Target
Symmetric Homodimer; Two Active Sites

Jin *et al*. (2020) *Nature 582*, 299-293.

Today's Agenda

- Introduction to the Protein Data Bank and SARS-CoV-2
  - Stephen K. Burley, M.D., D.Phil., Director, RCSB Protein Data Bank
- Exploring the SARS-CoV-2 Main Protease structure using RCSB.org
  - Paul Craig, Ph.D., Rochester Institute of Technology
- Making connections using RCSB.org
  - Shuchismita Dutta, Ph.D., RCSB Protein Data Bank
- Discussions

The Structure of COVID-19 Main Protease



https://www.rcsb.org/structure/6lu7

To follow along in this section, go to Exploring the SARS-CoV-2 Main Protease

SSP: Experiment, Validation, and Literature

Display Files | Download Files | Data API

6LU7

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

PDB DOI: https://doi.org/10.2210/pdb6LU7/pdb

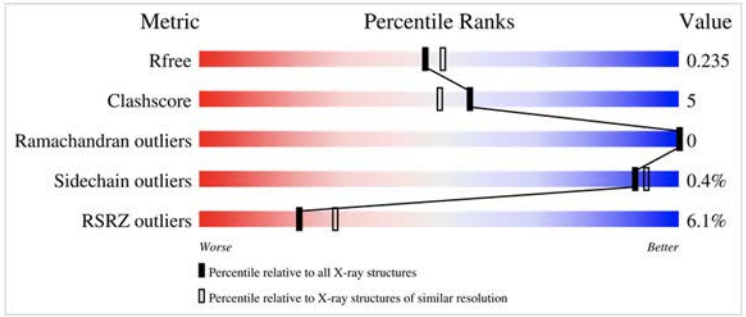
Classification: VIRAL PROTEIN
Organism(s): Severe acute respiratory syndrome coronavirus 2, synthetic construct
Expression System: Escherichia coli BL21(DE3)
Mutation(s): No

Deposited: 2020-01-26 **Released:** 2020-02-05
Deposition Author(s): Liu, X., Zhang, B., Jin, Z., Yang, H., Rao, Z.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.16 Å
R-Value Free: 0.235
R-Value Work: 0.202
R-Value Observed: 0.204

wwPDB Validation 3D Report | Full Report



Literature | Download Primary Citation

Structure of Mprofrom SARS-CoV-2 and discovery of its inhibitors.

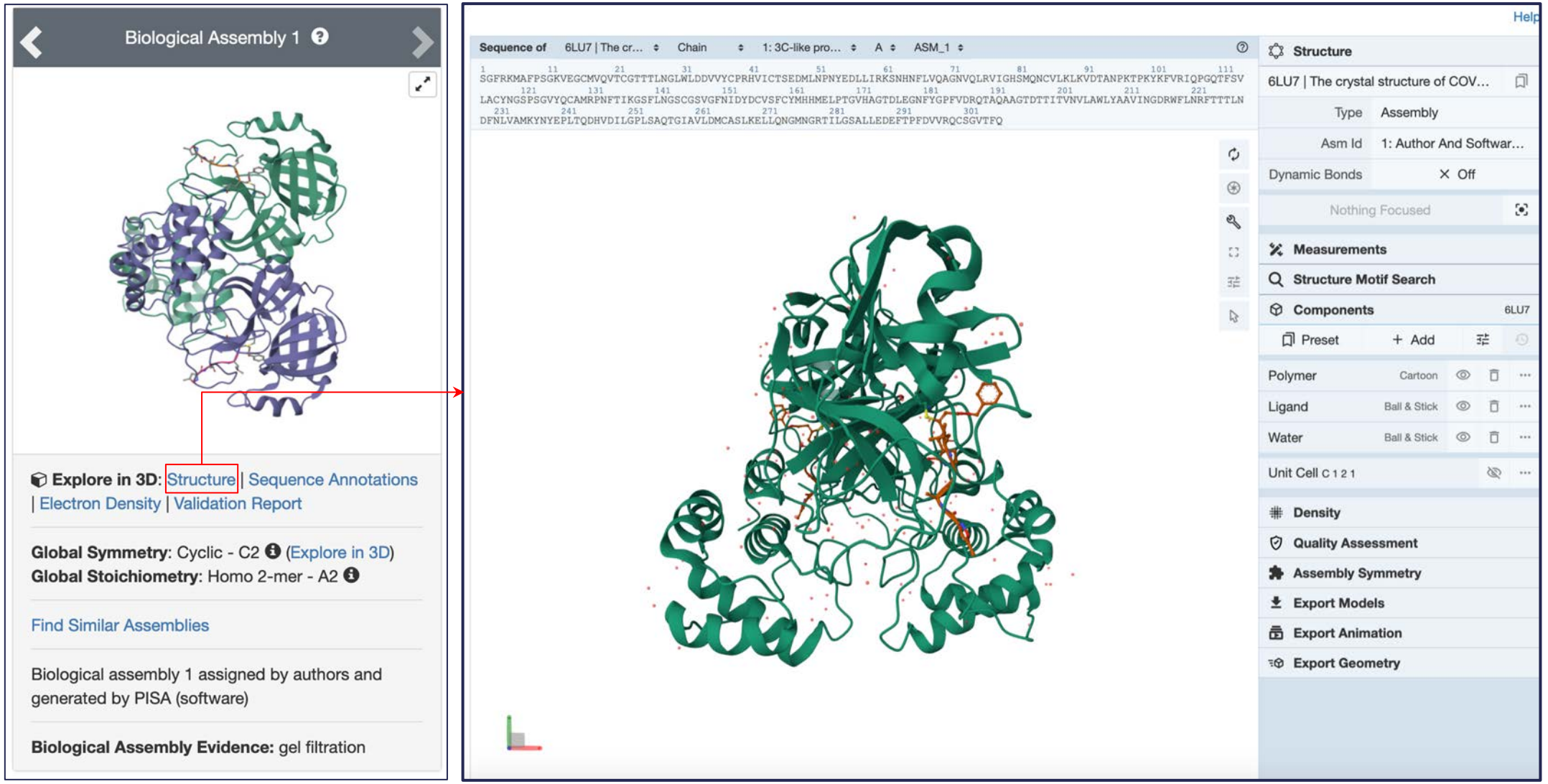
Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., Yang, H.
(2020) Nature **582**: 289-293

PubMed: 32272481 Search on PubMed | Search on PubMed Central
DOI: https://doi.org/10.1038/s41586-020-2223-y
Primary Citation of Related Structures:
6LU7, 7BQY

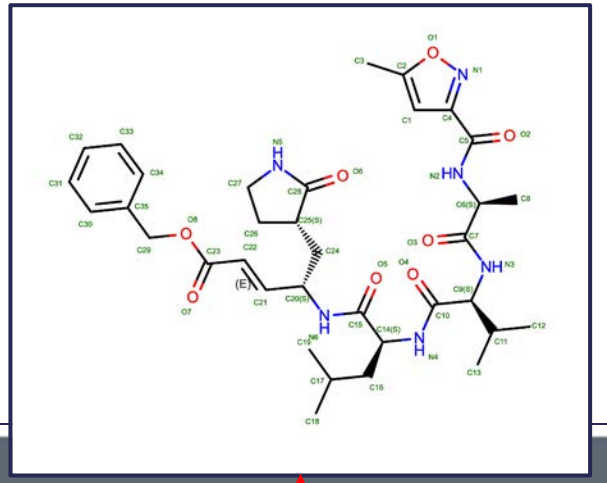
PubMed Abstract:
A new coronavirus, known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the aetiological agent responsible for the 2019-2020 viral pneumonia outbreak of coronavirus disease 2019 (COVID-19) [1-4]. Currently, there are no targeted therapeutic agents for the treatment of this disease, and effective treatment…
View More

Organizational Affiliation:
Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai, China.

SSP and Structure: Visualize and Explore



Structure: Interactions

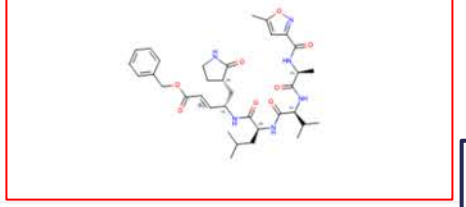


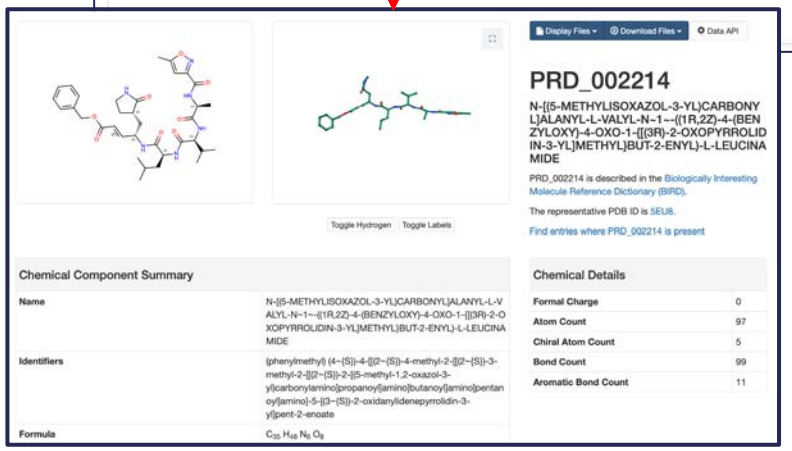
2D diagram

Biologically Interesting Molecules (External Reference) **1 Unique**

Entity ID: 2

ID	Chains	Name	Type/Class	2D Diagram	3D Interactions
PRD_002214 Query on PRD_002214	B [auth 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	Peptide-like / Inhibitor		Interactions Interactions & Density

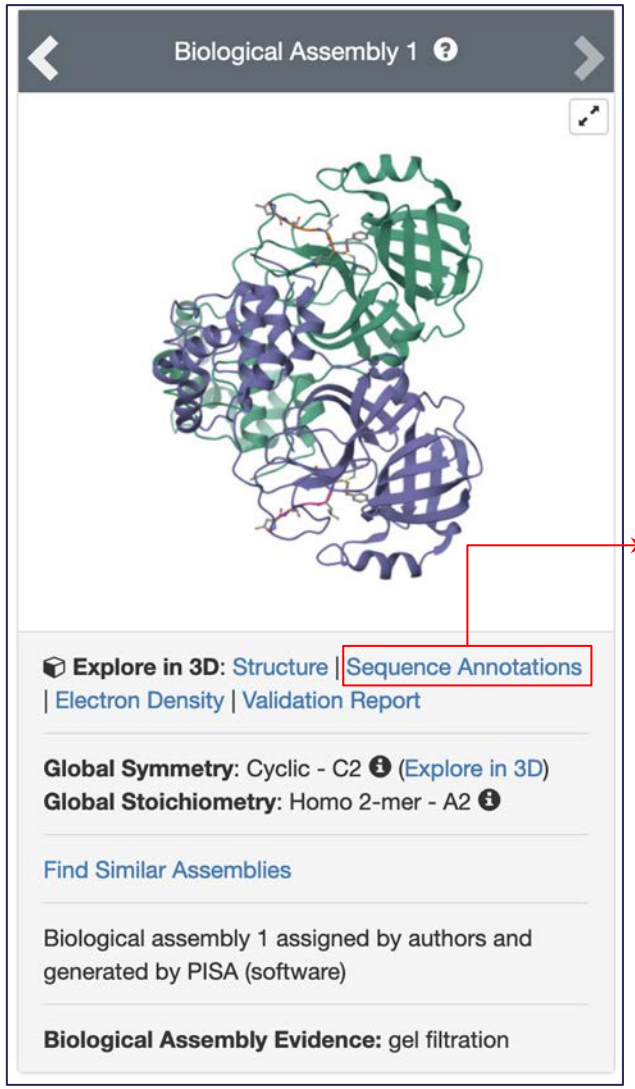


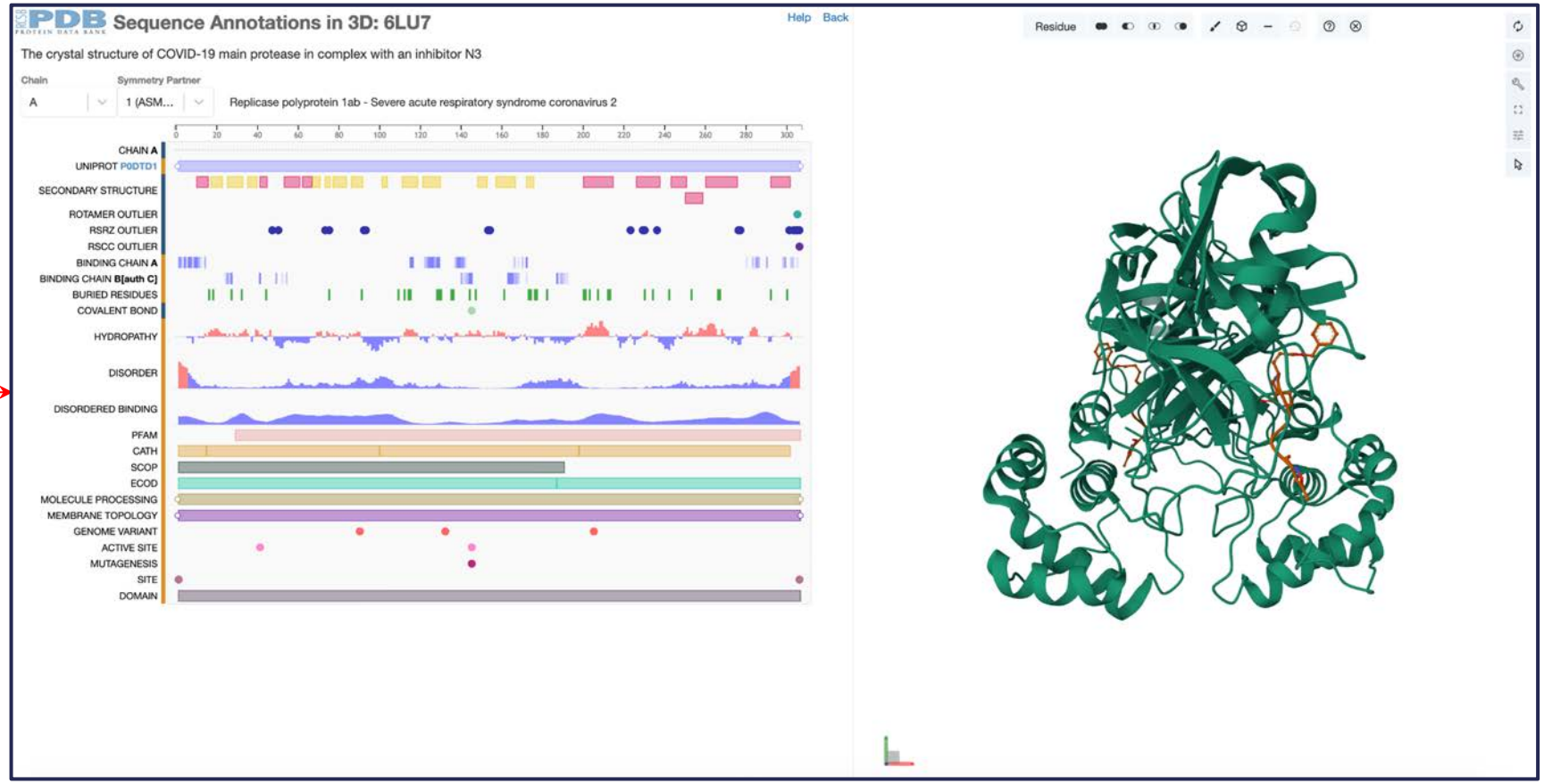


Biologically Interesting Reference Dictionary Summary Page

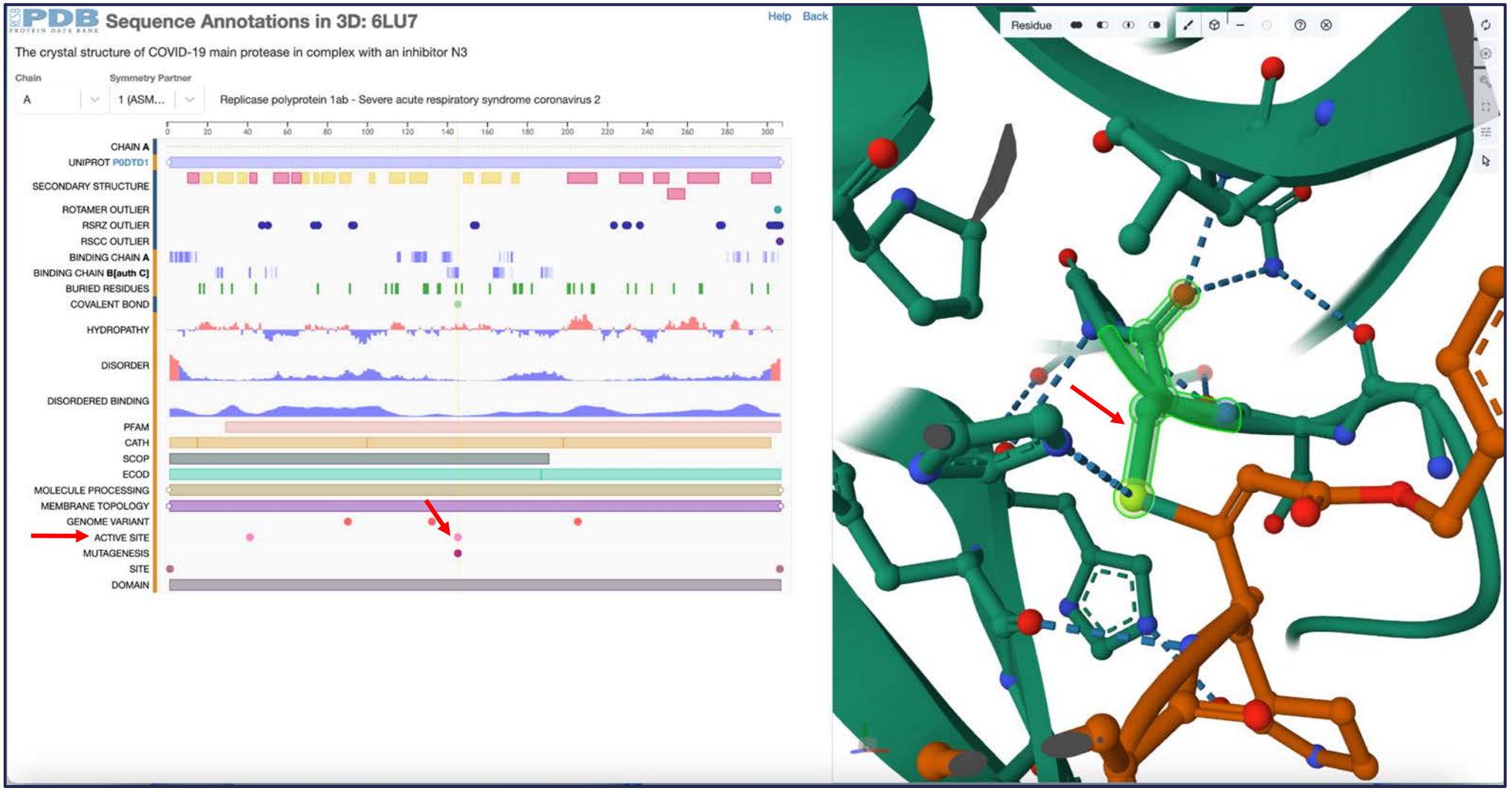
Ligand Interaction

SSP: Symmetry, Similar Assemblies and more



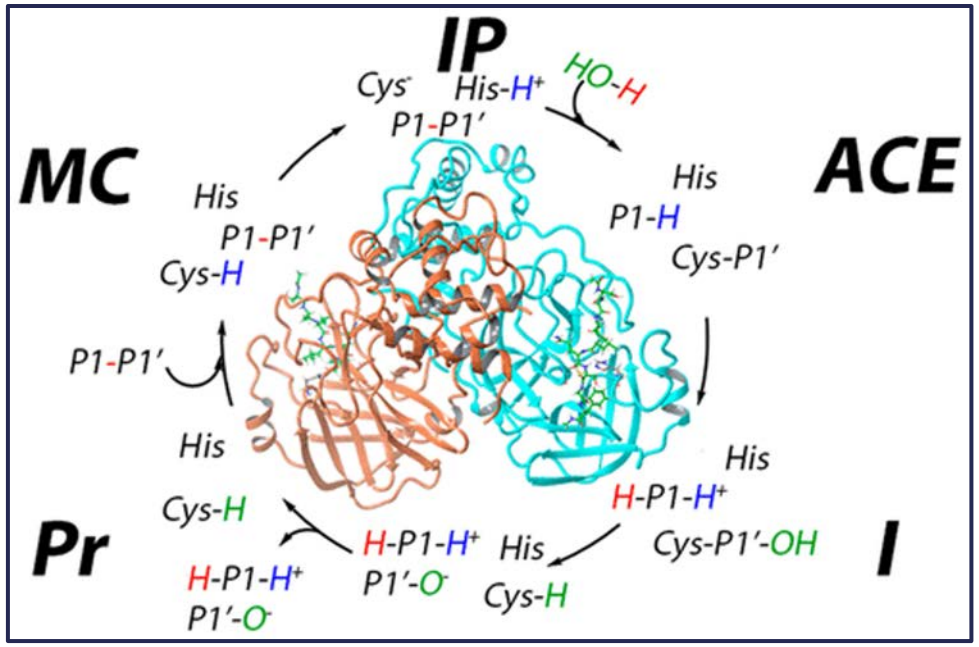


Sequence Annotations

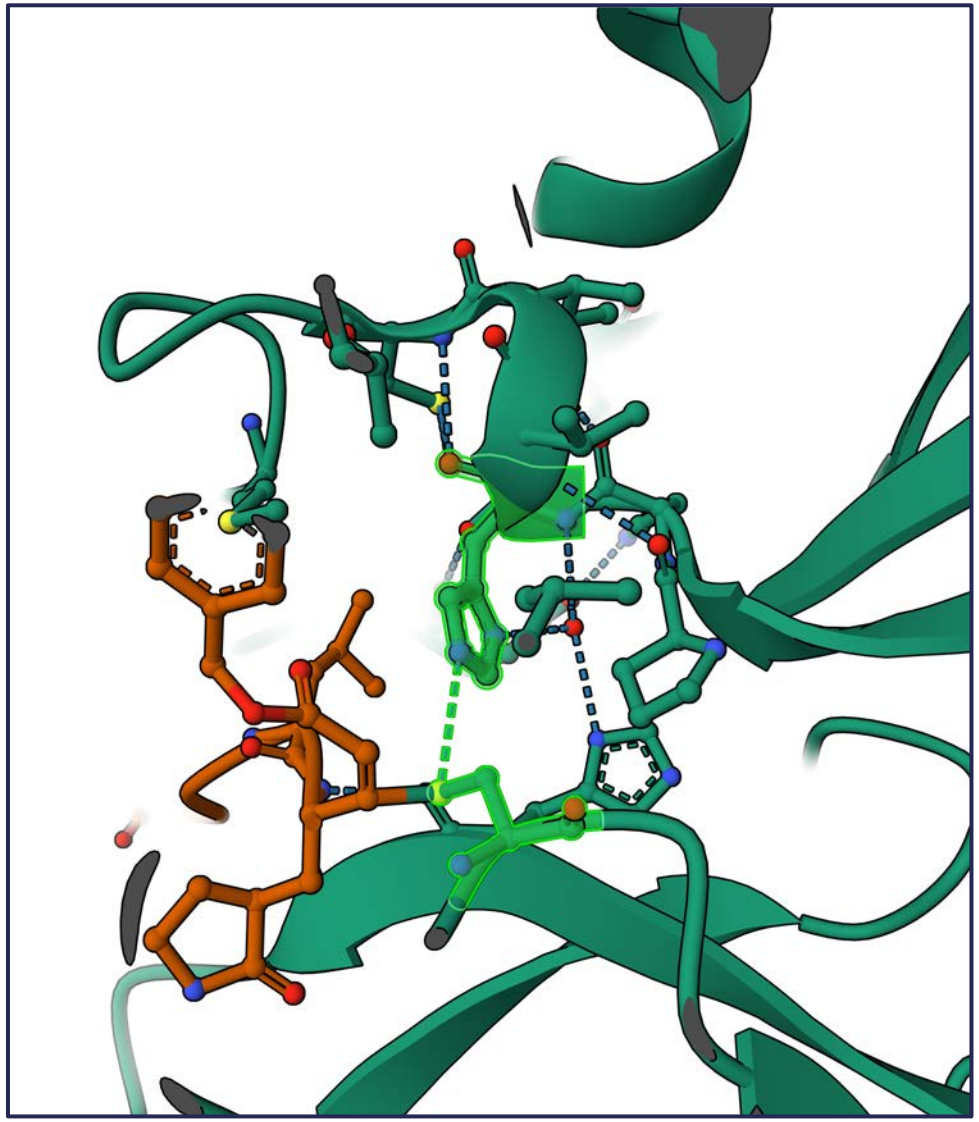


Teaching Enzymology: Basic Concepts

- Overall shape substrate binding
- Active site & catalytic residues
- Mechanism of enzyme catalysis



https://doi.org/10.1021/acscatal.0c03420



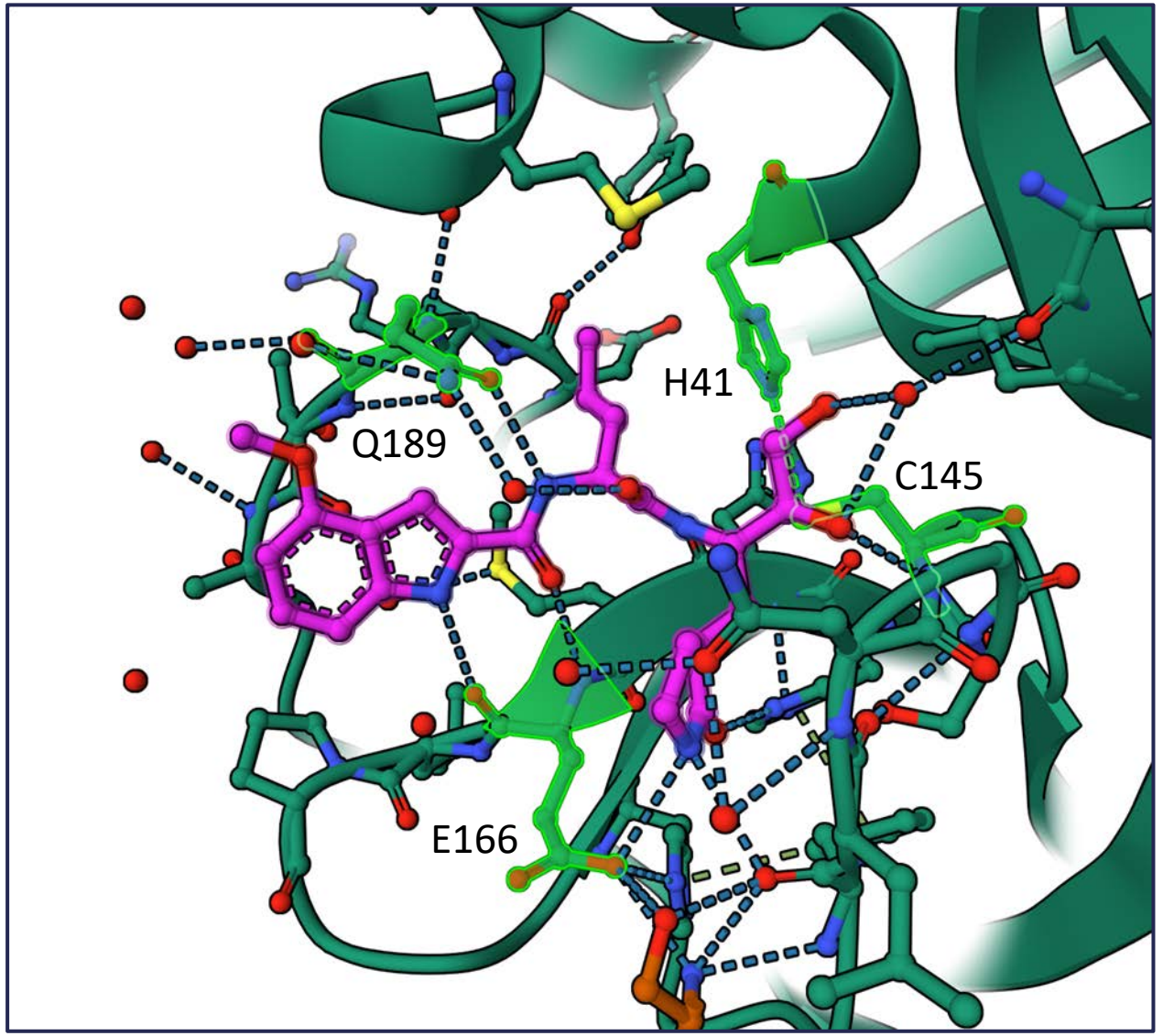
https://www.rcsb.org/3d-sequence/6LU7?assemblyId=1

Blocking Catalysis with a Covalent Inhibitor

- PDB ID 6xhm
  SARS-CoV-2 Mpro
  bound to PF-00835231

- Catalytic residues
  - His 41
  - Cys 145

- Other interactions
  - Glu 166
  - Gln 189



https://www.rcsb.org/3d-view/6XHM

Today's Agenda

- Introduction to the Protein Data Bank and SARS-CoV-2
  - Stephen K. Burley, M.D., D.Phil. - Director, RCSB Protein Data Bank
- Exploring the SARS-CoV-2 Main Protease using RCSB.org
  - Paul Craig, Ph.D. - Rochester Institute of Technology
- Making connections using RCSB.org
  - Shuchismita Dutta, Ph.D. - RCSB Protein Data Bank
    - Explore the structures of other enzymes like SARS-CoV-2 MPro
    - Use prior knowledge to develop a drug for SARS-CoV-2
    - Prepare to deal with future SARS-CoV like pandemics
- Discussions

Learning About the Enzyme (Target)

Macromolecules

Find similar proteins by: Sequence (by identity cutoff) | 3D Structure

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
3C-like proteinase	A	306	Severe acute respiratory syndrome coronavirus 2	Mutation(s): 0 Gene Names: rep, 1a-1b EC: 3.4.22.69	

UniProt

Find proteins for P0DTD1 *(Severe acute respiratory syndrome coronavirus 2)* Explore P0DTD1 Go to UniProtKB: P0DTD1

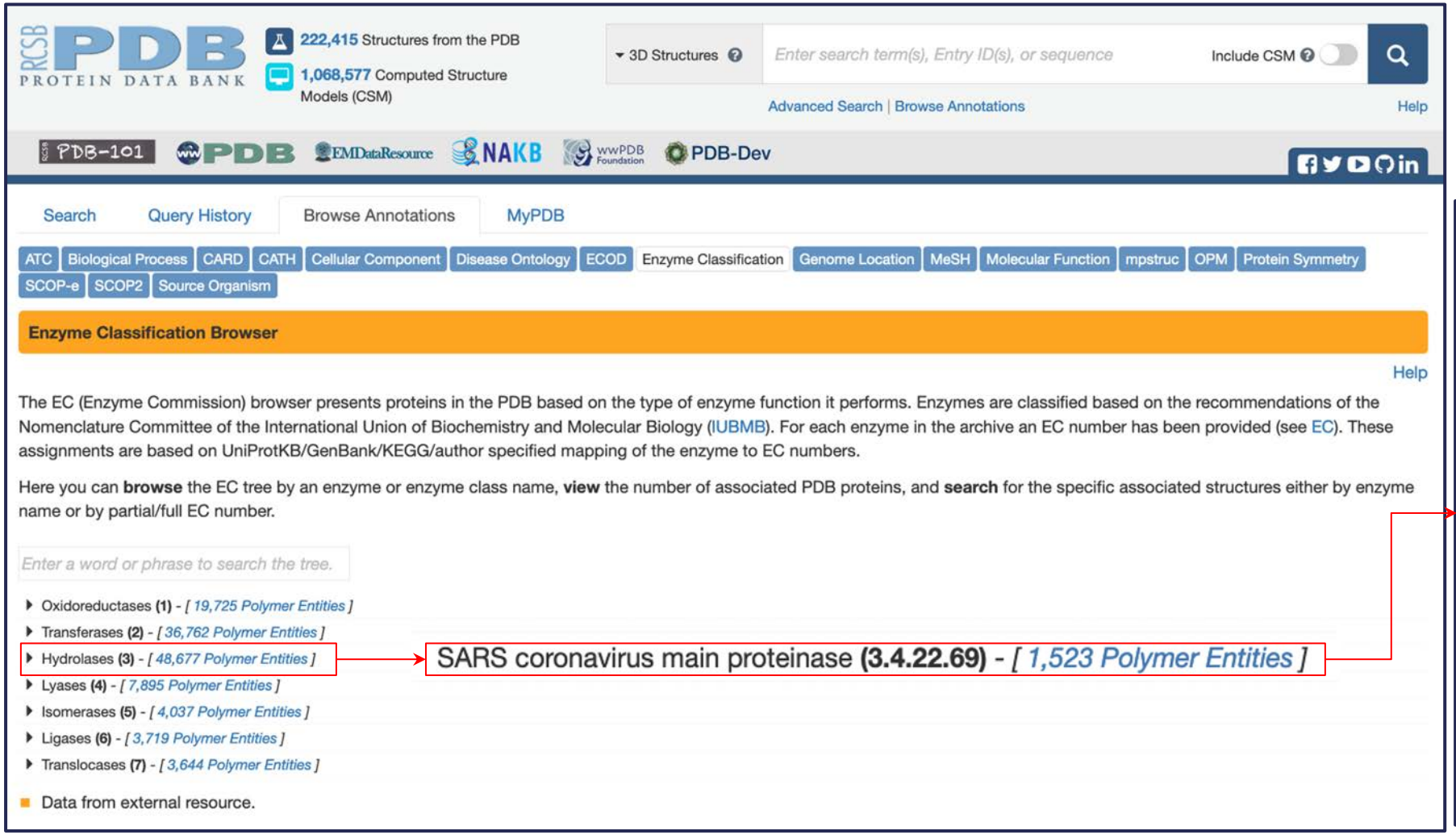
Entity Groups

Sequence Clusters	30% Identity, 50% Identity, 70% Identity, 90% Identity, 95% Identity, 100% Identity
UniProt Group	P0DTD1

Scientific Name of Source Organism

- Severe acute respiratory syndrome coronavirus 2 (1,421)
- Severe acute respiratory syndrome-related coronavirus (50)
- synthetic construct (35)
- Homo sapiens (16)
- Middle East respiratory syndrome-related coronavirus (16)
- Severe acute respiratory syndrome coronavirus (10)
- Bat SARS CoV Rf1/2004 (2)
- Betacoronavirus England 1 (2)
- Camelus bactrianus (2)
- Murine hepatitis virus strain A59 (2)
- Mus musculus (2)
- SARS coronavirus Sin2774 (2)
- SARS coronavirus Urbani (2)
- Tylonycteris bat coronavirus HKU4 (2)
- Bos taurus (1)
- Escherichia coli (1)
- Human coronavirus OC43 (1)
- Rousettus bat coronavirus HKU9 (1)
- Streptomyces exfoliatus (1)
- Streptomyces roseus (1)

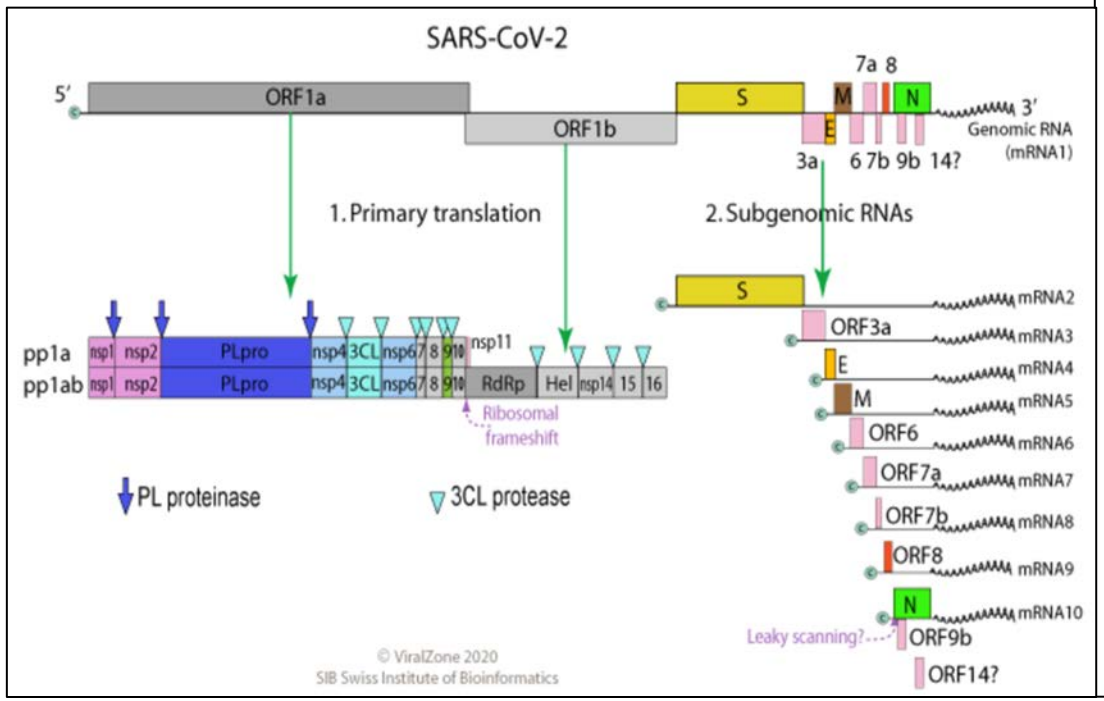
Learning About the Enzyme (Target)

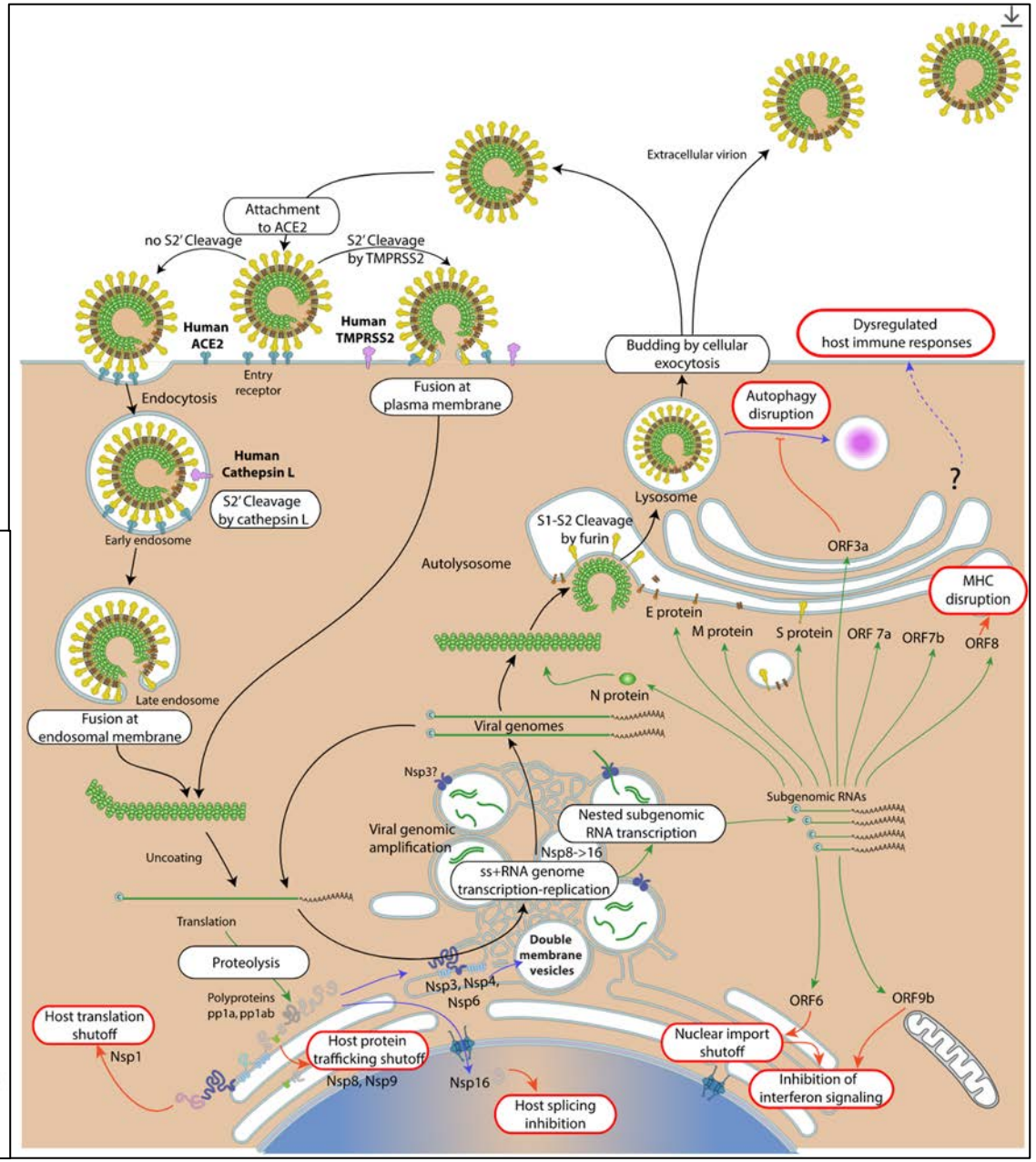


Scientific Name of Source Organism

- Severe acute respiratory syndrome coronavirus 2 (1,431)
- Severe acute respiratory syndrome-related coronavirus (50)
- Middle East respiratory syndrome-related coronavirus (16)
- Severe acute respiratory syndrome coronavirus (10)
- SARS coronavirus Urbani (3)
- Bat SARS CoV Rf1/2004 (2)
- Betacoronavirus England 1 (2)
- Human coronavirus OC43 (2)
- Murine hepatitis virus strain A59 (2)
- SARS coronavirus Sin2774 (2)
- Tylonycteris bat coronavirus HKU4 (2)
- Rousettus bat coronavirus HKU9 (1)

A Conserved Target





From Pandemic to Paxlovid (Nirmatrelvir)

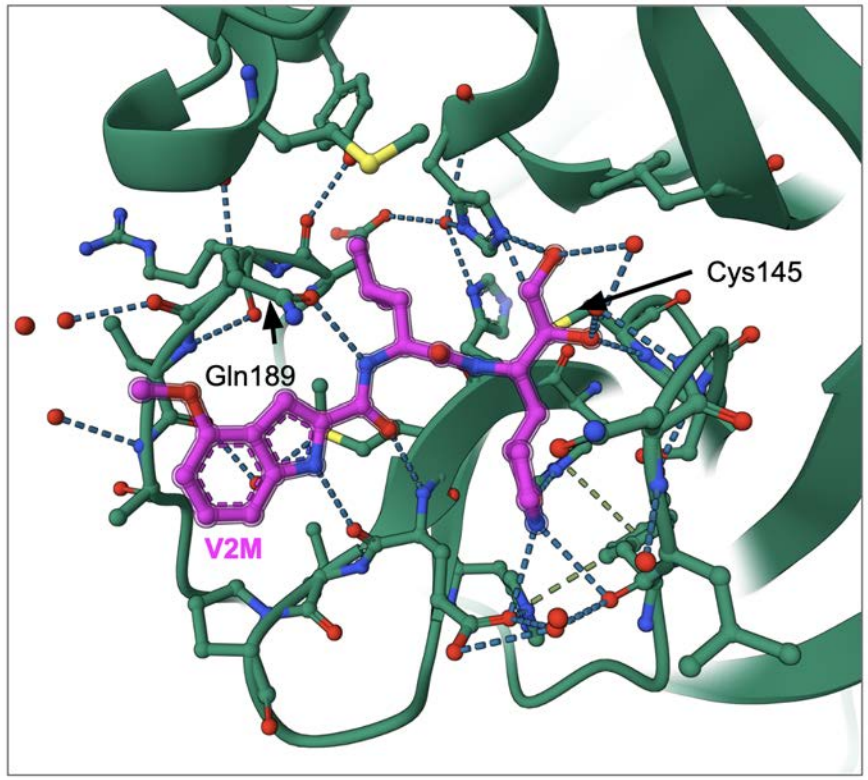
Q: *When did you first get the inspiration to look at leads from previous antiviral programs that you were a part of?*
Dr. Owen's reply: PF-835231 (PF-00835231) was the culmination of our SARS program from 2003/4. It was designed for IV dosing and yet thankfully the SARS outbreak had been effectively contained by the time we had the molecule ready for evaluation in the clinic in 2004. There were no subjects for a clinical trial, so we were not able to clinically evaluate the compound. Following the outbreak of Covid-19, the protein sequences from the SARS-CoV-2 viral genome were in the public domain by February 2020. Given Pfizer's experience in viral protease research, our leadership planned and proposed an oral protease inhibitor program. The critical SARS-CoV-2 main protease catalytic site, when compared to SARS main protease from 17 years earlier, was identical. We quickly showed that PF-835231 (PF-00835231) was therefore a potent in vitro inhibitor of the SARS-CoV-2 main protease and it became the starting point for designing an oral protease inhibitor, specifically for Covid-19.

Source: An Interview With The Team Leader For The Scientific Discovery Of Pfizer's Covid-19 Drug, Paxlovid, John LaMattina, Jan 6, 2022

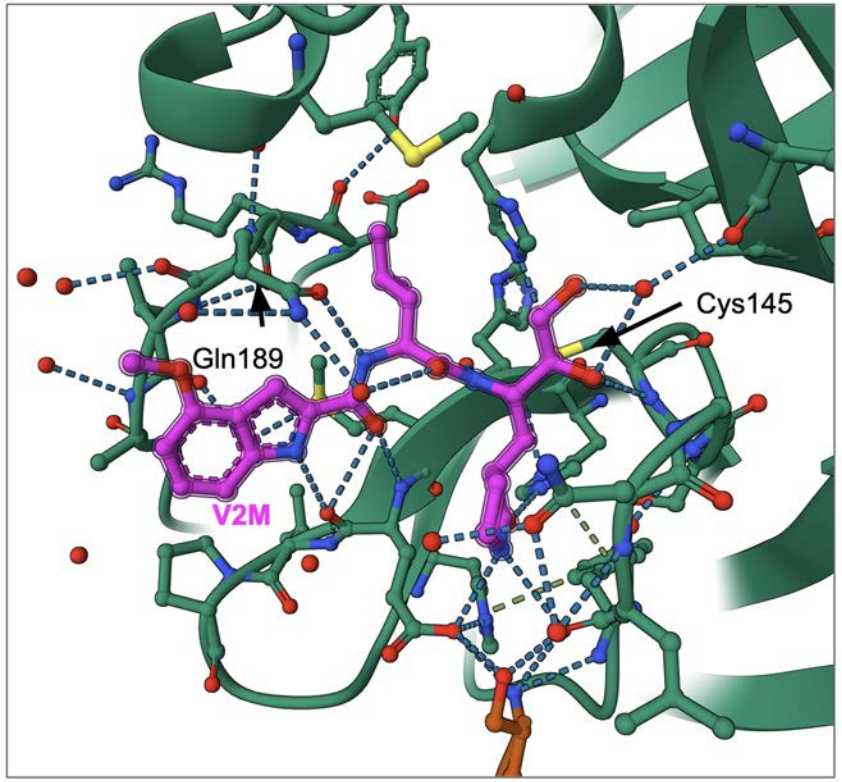
Team of Pfizer scientists who developed Paxlovid, led by Dr. Dafydd Owen.
Source: Pfizer - 2022 Heroes of Chemistry, https://youtu.be/e2rRGoSyC5U.

On Friday March 13th 2020, our CEO published a five-point plan for our response to Covid-19. One of those was 'marshalling our people'. That was the day I was asked to plan some specifics for the potential program. I was asked '**What would you need and how would you prosecute an oral protease inhibitor program**?'.

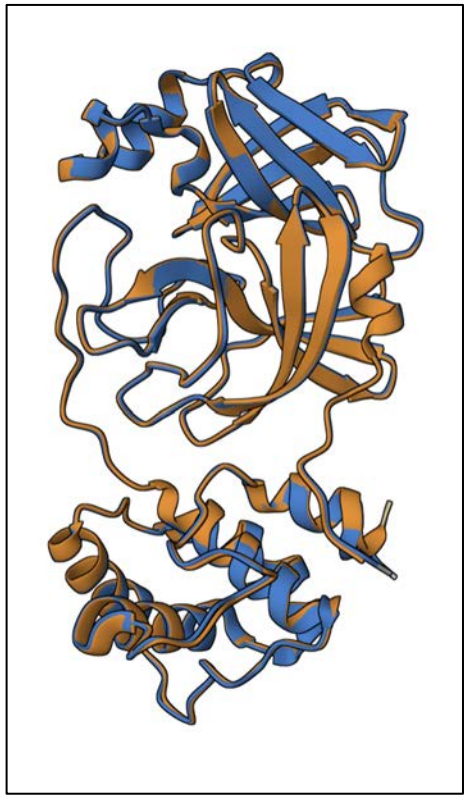
Inhibitor PF-00835231 Binding CoV MPros



PF-00835231 bound to SARS-CoV main protease (PDB ID **6xhl**)

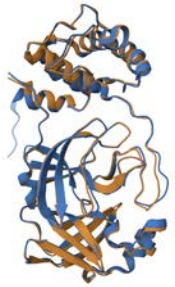


PF-00835231 bound to SARS-CoV-2 main protease (PDB ID **6xhm**)

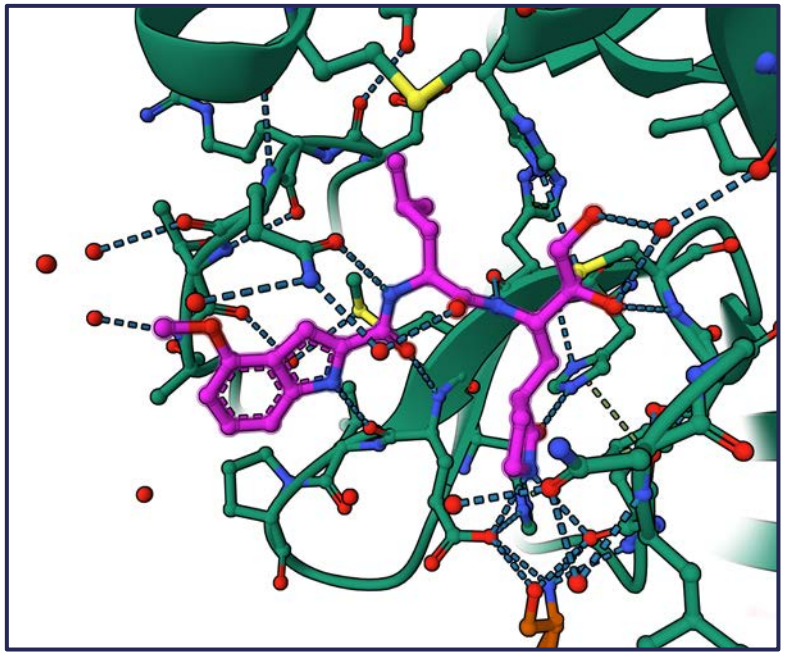


Compare main proteases (PDB IDs **6xhl** and **6xhm**)

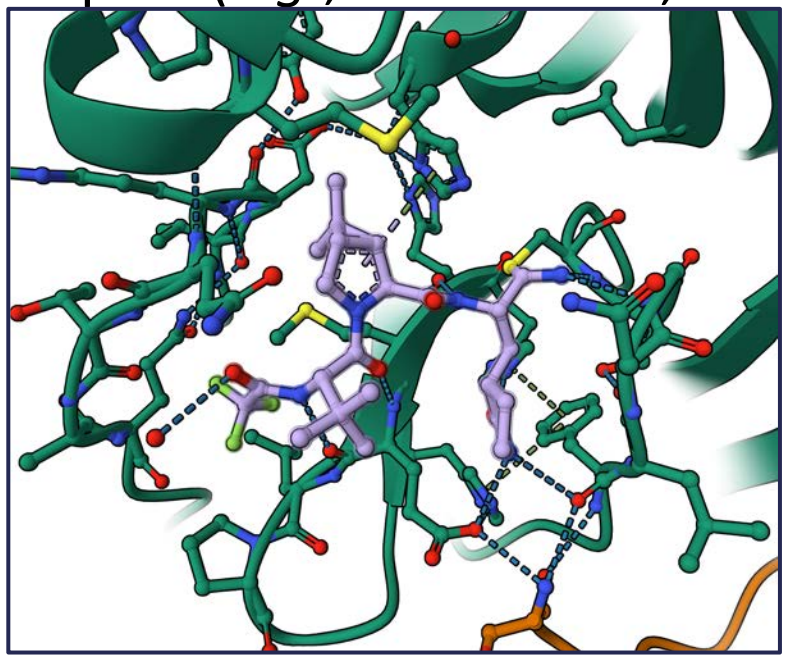
An oral SARS-CoV-2 Mpro inhibitor



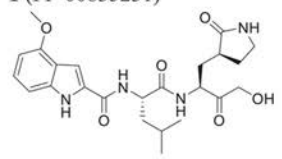
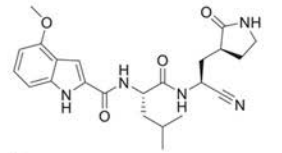
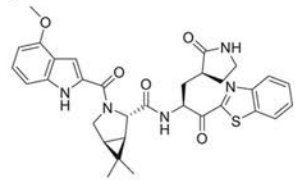
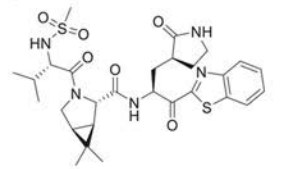
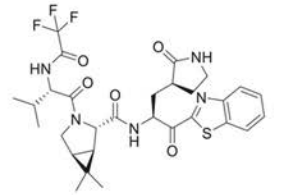
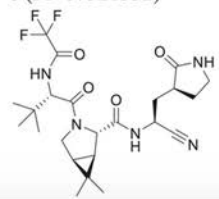
- PF-00835231 (PDB ID 6xhm, left)
  - Intravenous administration only
- PF-07321332 or Nirmatrelvir (PDB ID 7rfw, right)
  - Administered orally, good selectivity and safety profiles
  - Inhibits other coronavirus Mpros (*e.g.*, SARS-CoV-1, MERS)



https://www.rcsb.org/3d-view/6XHM?preset=ligandInteraction&label_asym_id=C



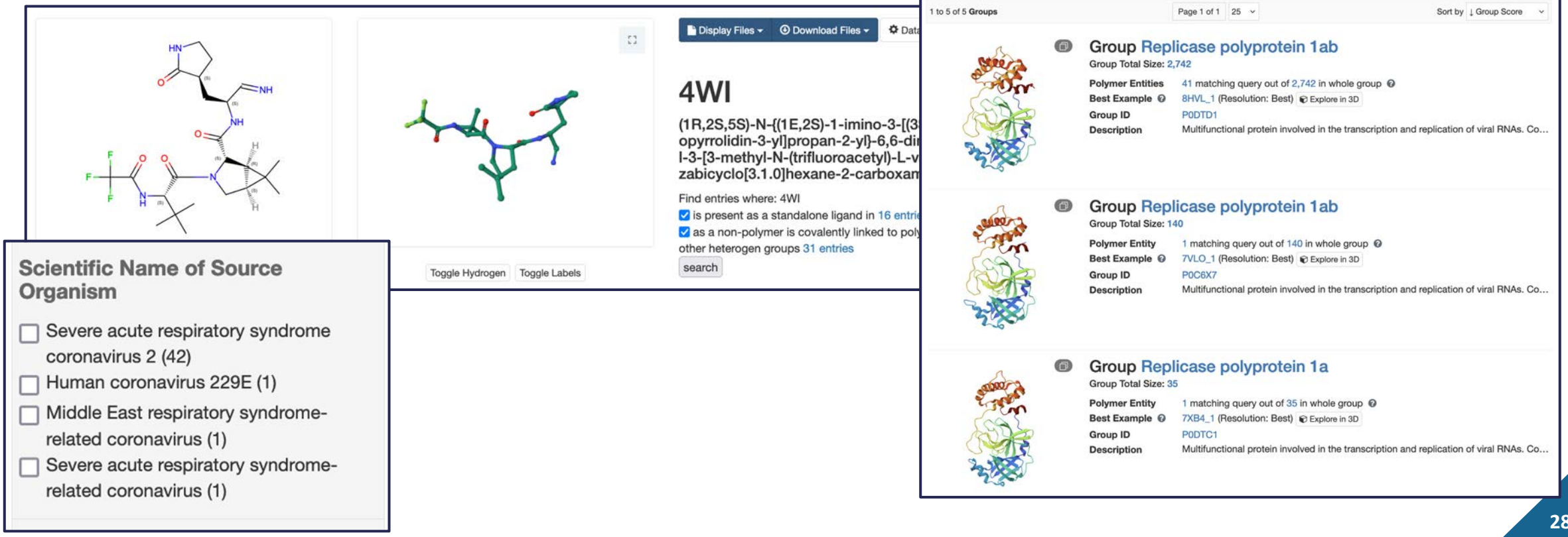
https://www.rcsb.org/3d-view/7RFW?preset=ligandInteraction&label_asym_id=B

Number/Structure	SARS-CoV2 M^{pro} K_i (nM)*
1 (PF-00835231) 	0.271 (0.155 – 0.471, n=6)
2 	27.7 (18.4 – 41.7, n=5)
3 	230 (181 – 292, n=4)
4 	7.93 (3.62 – 17.4, n=5)
5 	12.1 (8.05 – 18.1, n=7)
6 (PF-07321332) 	3.11 (1.47 – 6.59, n=6)

https://doi.org/10.1126/science.abl4784

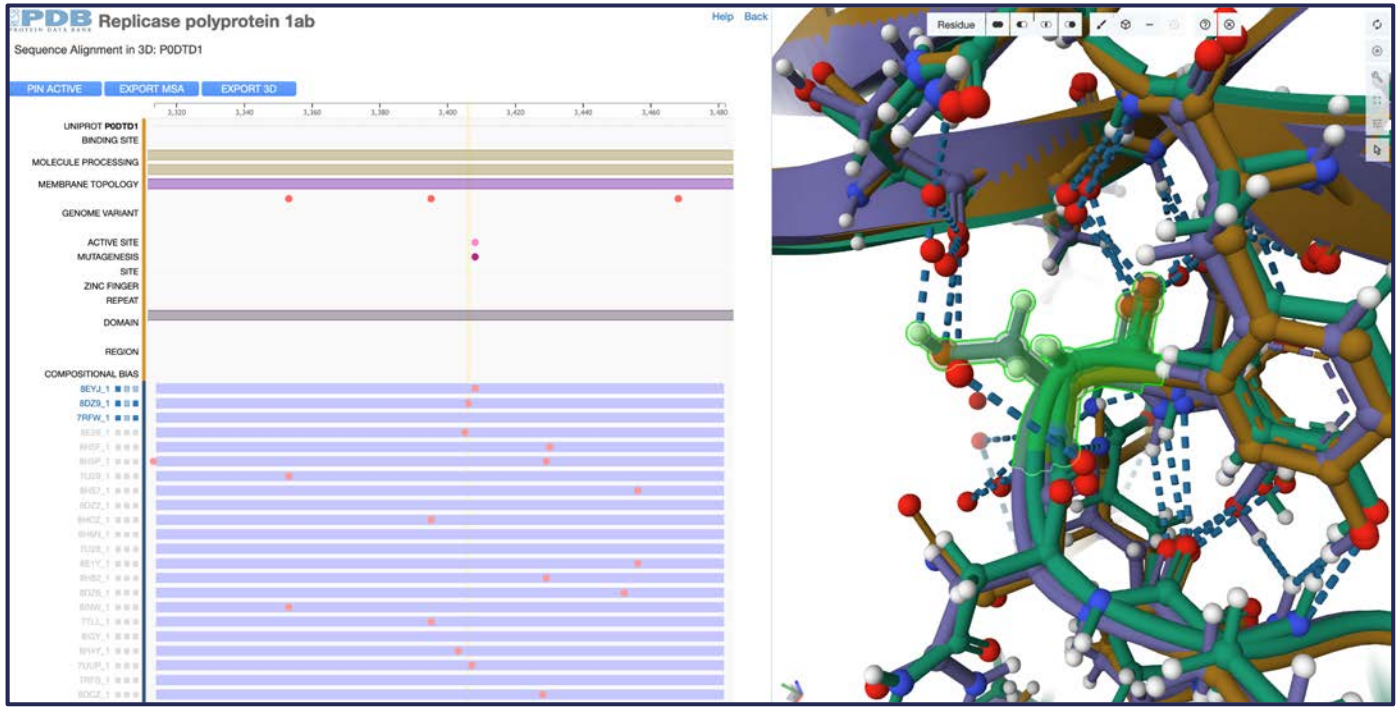
Preparing for a Future Pandemic

- Search for Nirmatrelvir (bound) containing structures
- Group polymers in result by UniProt



Teaching Enzymology: Advanced Concepts

- Exploring conformational changes during enzyme activity
- Understanding mechanisms of enzyme action
- Designing inhibitors and allosteric regulators
- Understanding the impact of mutations/variants on enzyme activity



How would you use rcsb.org to explore/learn/teach about developing resistance, designing new drugs, and more?

Summary

- Introduction to the Protein Data Bank and SARS-CoV-2
  - Stephen K. Burley, M.D., D.Phil. - Director, RCSB Protein Data Bank
- Exploring the SARS-CoV-2 Main Protease using RCSB.org
  - Paul Craig, Ph.D. - Rochester Institute of Technology
- Teaching enzymology using RCSB.org
  - Shuchismita Dutta, Ph.D. - RCSB Protein Data Bank
- Discussion

Exit Survey and Certificates of Completion

Please take the exit survey (https://go.rutgers.edu/eenu6dq8) to

- share what you liked about this webinar
- how we can improve this webinar
- tell us what other webinars you would be interested in

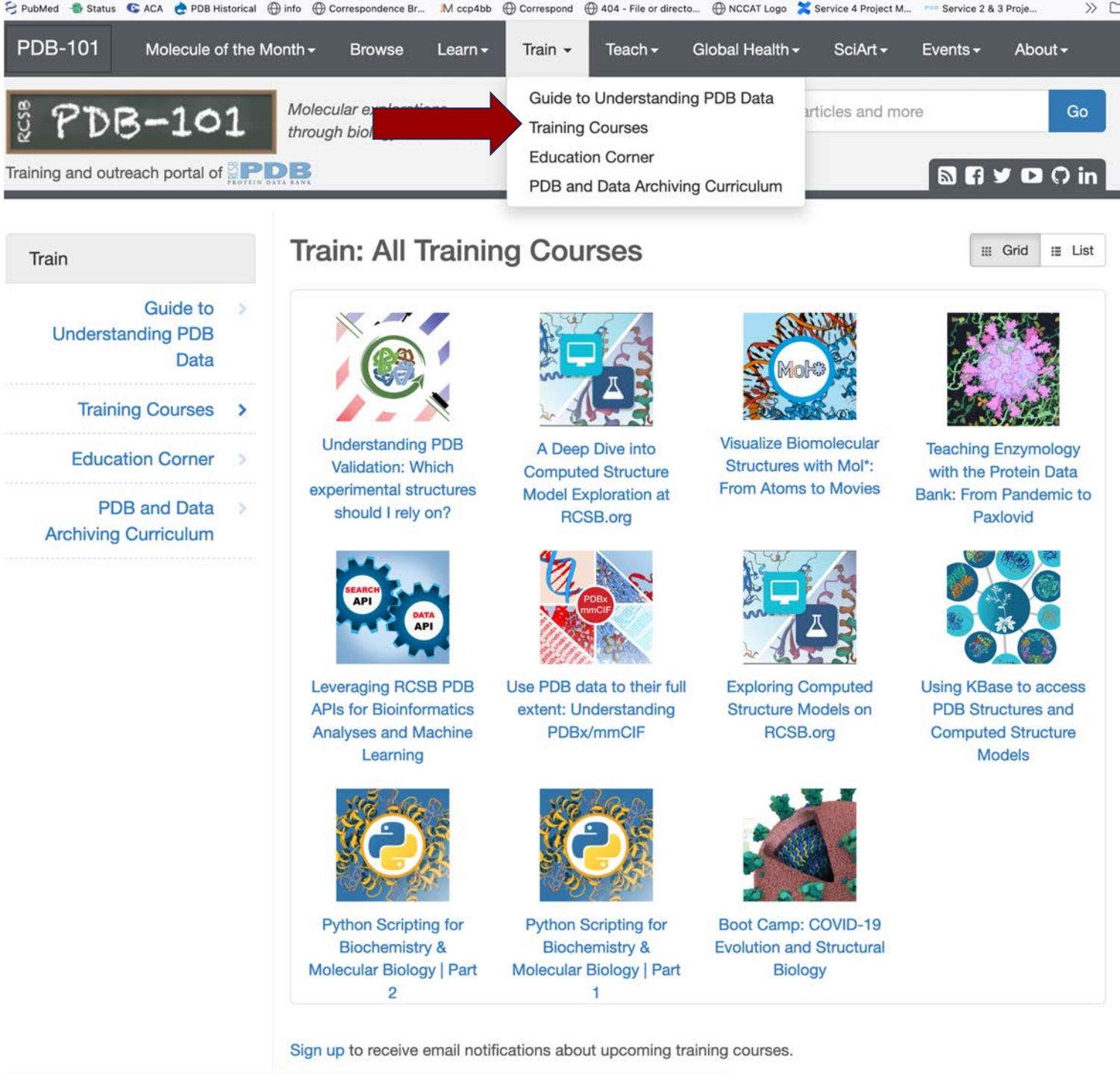
This survey will be closed on Tuesday, **August 6, 2024**.

Want to receive a certificate of completion?
You MUST complete the exit survey.
Please allow 4 weeks to receive the certificate.

Recordings will be added to PDB-101 in the fall (pdb101.rcsb.org)

Course Materials

Course recordings and presentations will be published at PDB-101.rcsb.org sometime in the Fall.
They will not be emailed separately.



RCSB PDB Team

RCSB.ORG
info@rcsb.org

Core Operations Funding

US National Science Foundation (DBI-2321666), National Institute of General Medical Sciences, National Institute of Allergy and Infectious Disease, and National Cancer Institute (NIH R01GM133198), and the US Department of Energy (DE-SC0019749)

Management

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

UC San Diego

SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF University of California San Francisco

Member of the Worldwide Protein Data Bank (wwPDB; **wwpdb.org**)

Follow us

Paul

Craig

Training Resources on PDB-101

pdb101.rcsb.org > *Train*

Materials to help effectively use **RCSB.org** tools for searching, visualizing, and analyzing 3D biostructure data

- Guide to Understanding PDB Data
- Training Courses
- Education Corner
- PDB & Data Archiving Curriculum

Sign up for Training Event Announcements