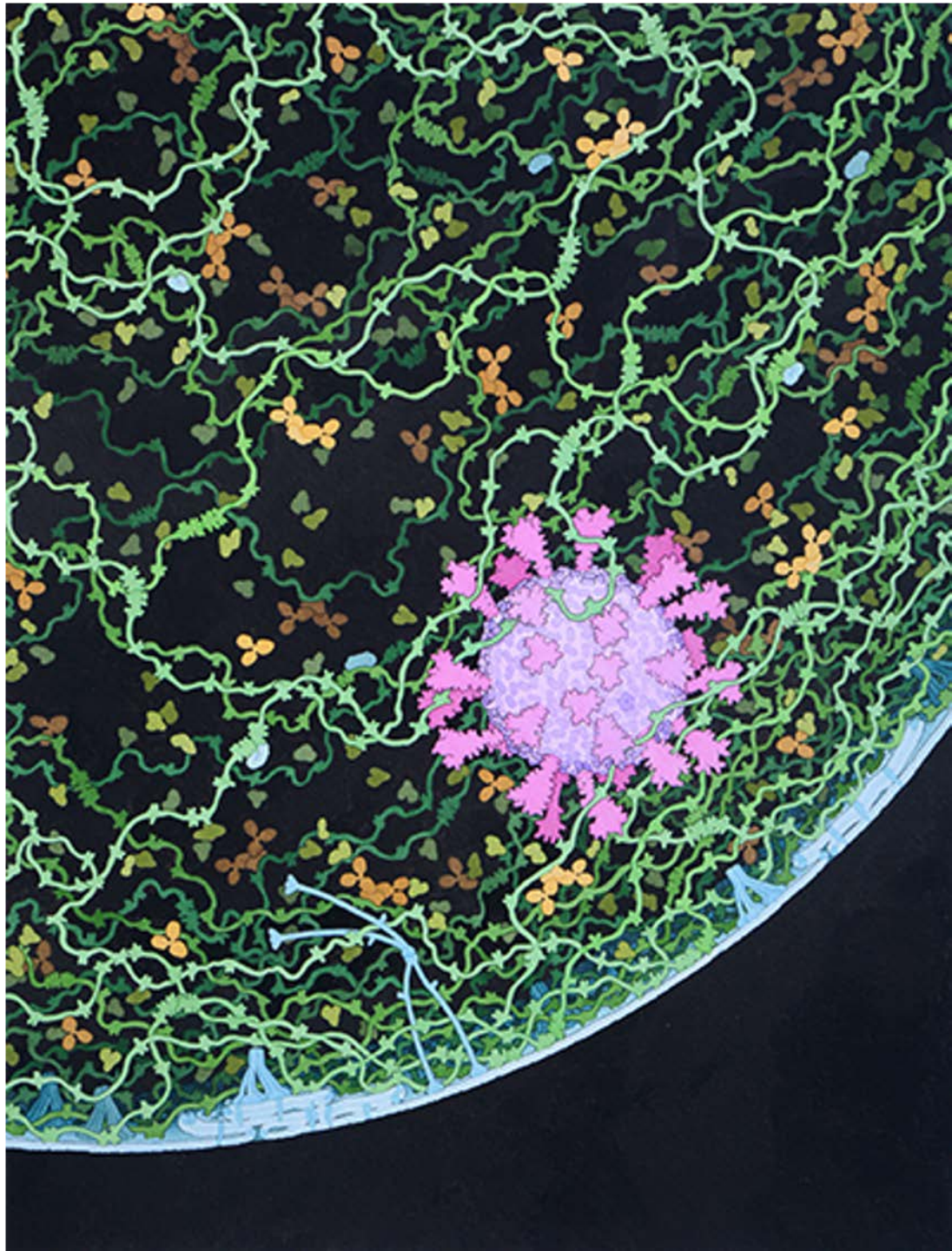
RCSB PDB PROTEIN DATA BANK

ASBMB American Society for Biochemistry and Molecular Biology

rcsb.org

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

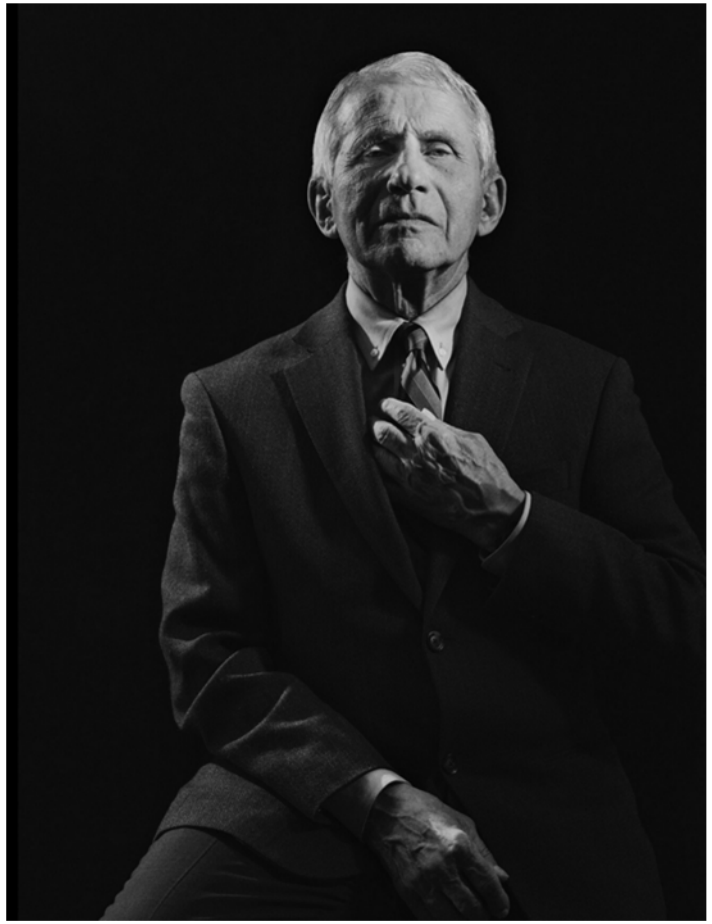
Stephen K. Burley, M.D., D.Phil.
Director, RCSB Protein Data Bank



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
- Discussion

Dr. Anthony Fauci Loves the PDB



Wallace-Wells: And what about pandemic preparedness more generally? Let's say we're working from scratch and designing the system at a white board. What reforms are needed?

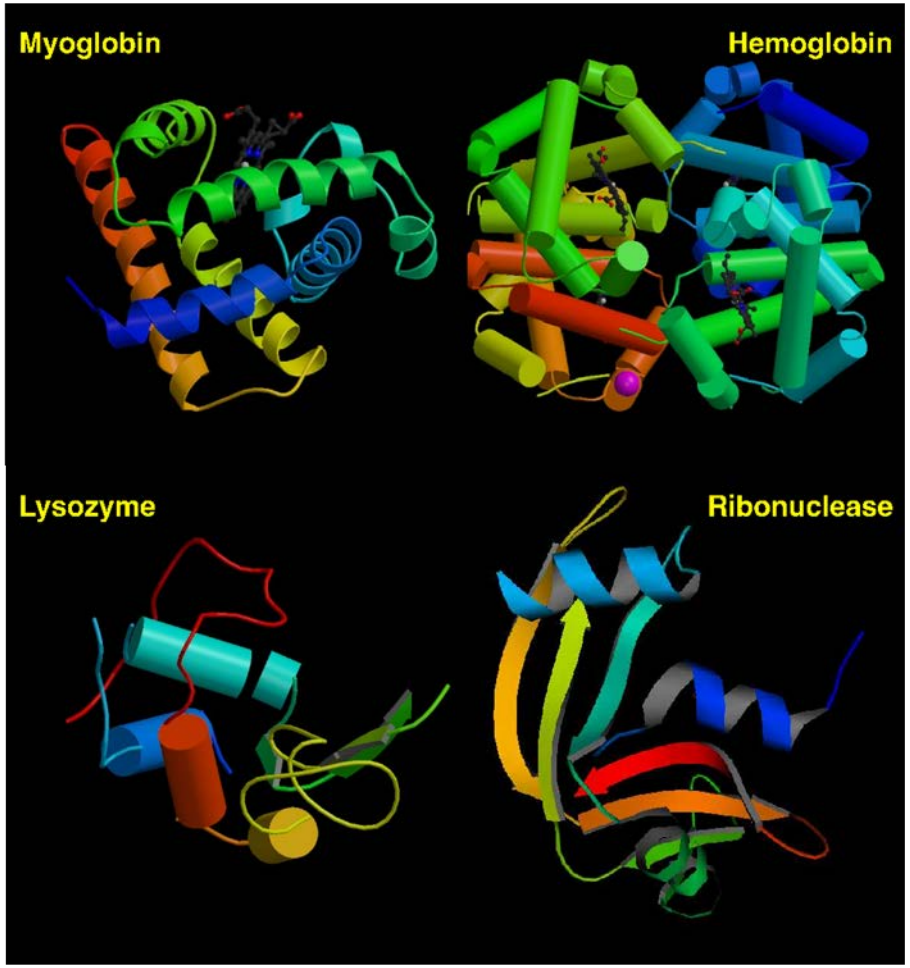
Fauci: Do you have two weeks to talk?

If you look at what worked for us, it was on the science side: the extraordinary investments that were made for decades before the emergence of SARS-CoV-2. First, the work in platform technology that led to essentially a revolution in how we make vaccines. No.2 is structure-based immunogen design. That helped with antiviral design, too — that has been the most underrated part of our response. I mean, show me a person who's vaccinated, got infected, took Paxlovid and died. I can't find anybody.

https://www.nytimes.com/interactive/2023/04/24/magazine/dr-fauci-pandemic.html

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 >210,000 structures!**
- wwPDB Partnership founded in 2003
- Members: RCSB PDB (US), PDBe (EMBL-EBI), PDBj (Japan), and PDBc (China); plus EMDB (3DEM) and BMRB (NMR)

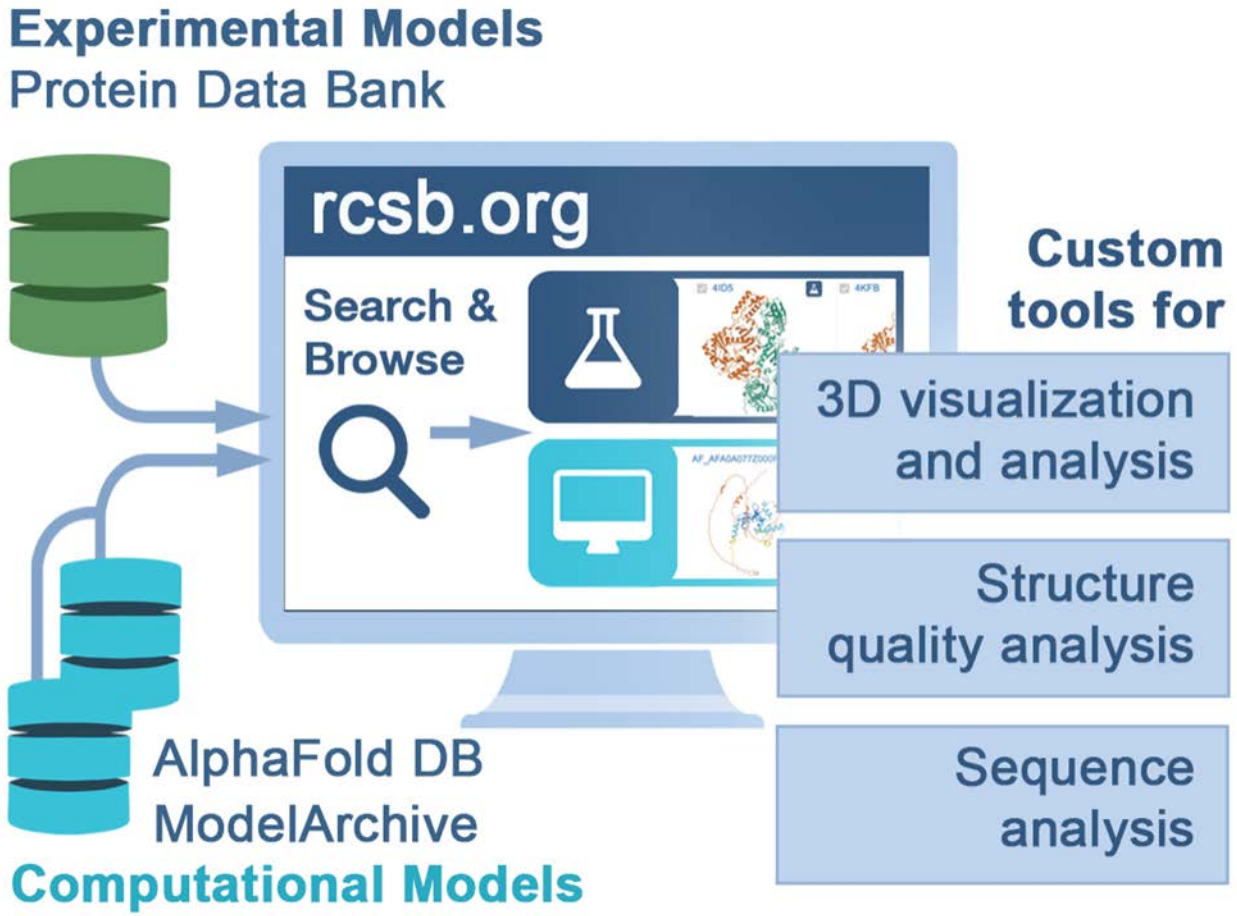


Structures that Inspired Launch of the PDB

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

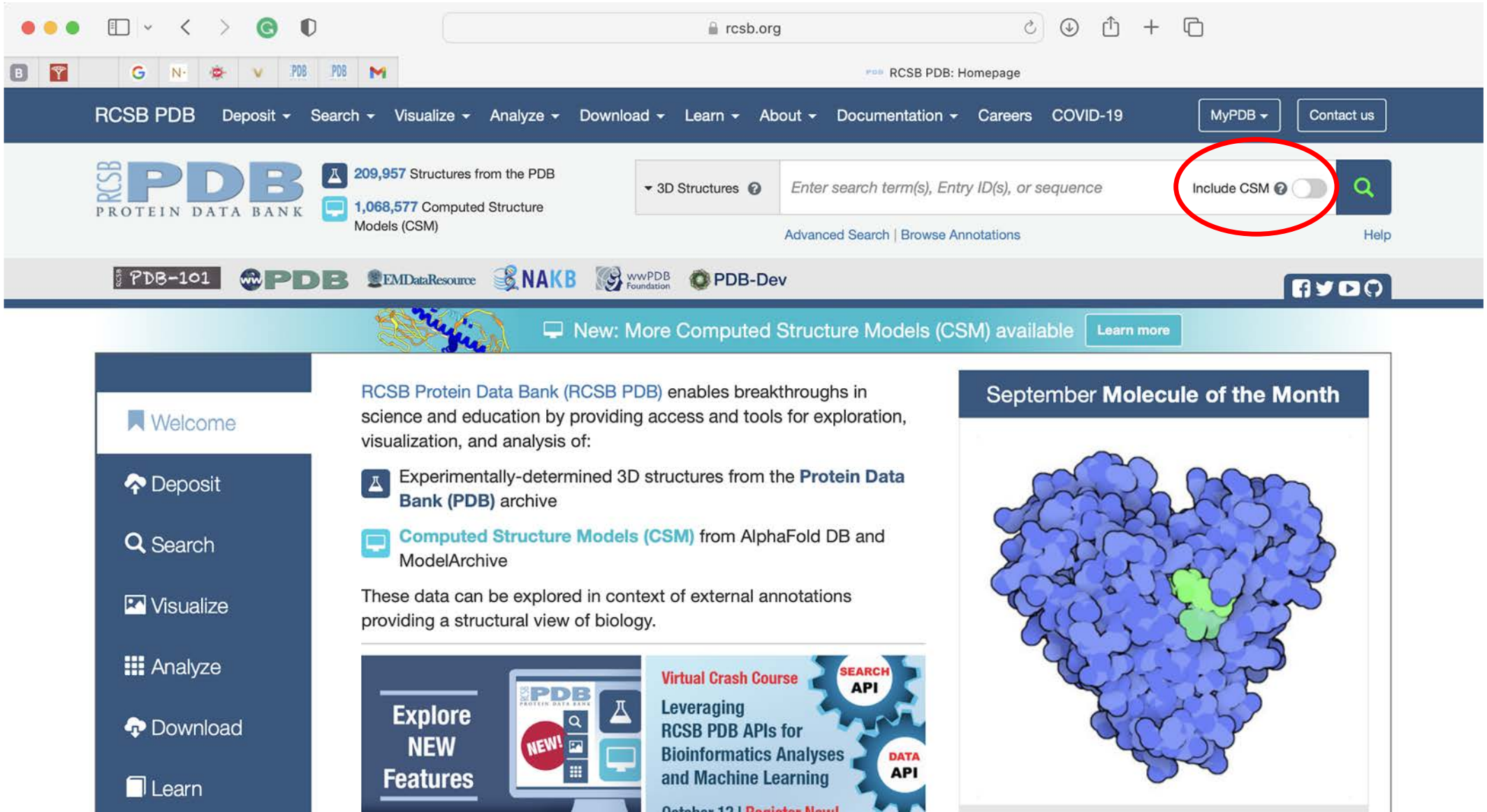
RCSB.org Research-focused Web Portal: One-Stop-Shop for Public 3D Biostructure Data

- RCSB.org delivers
  - >210,000 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



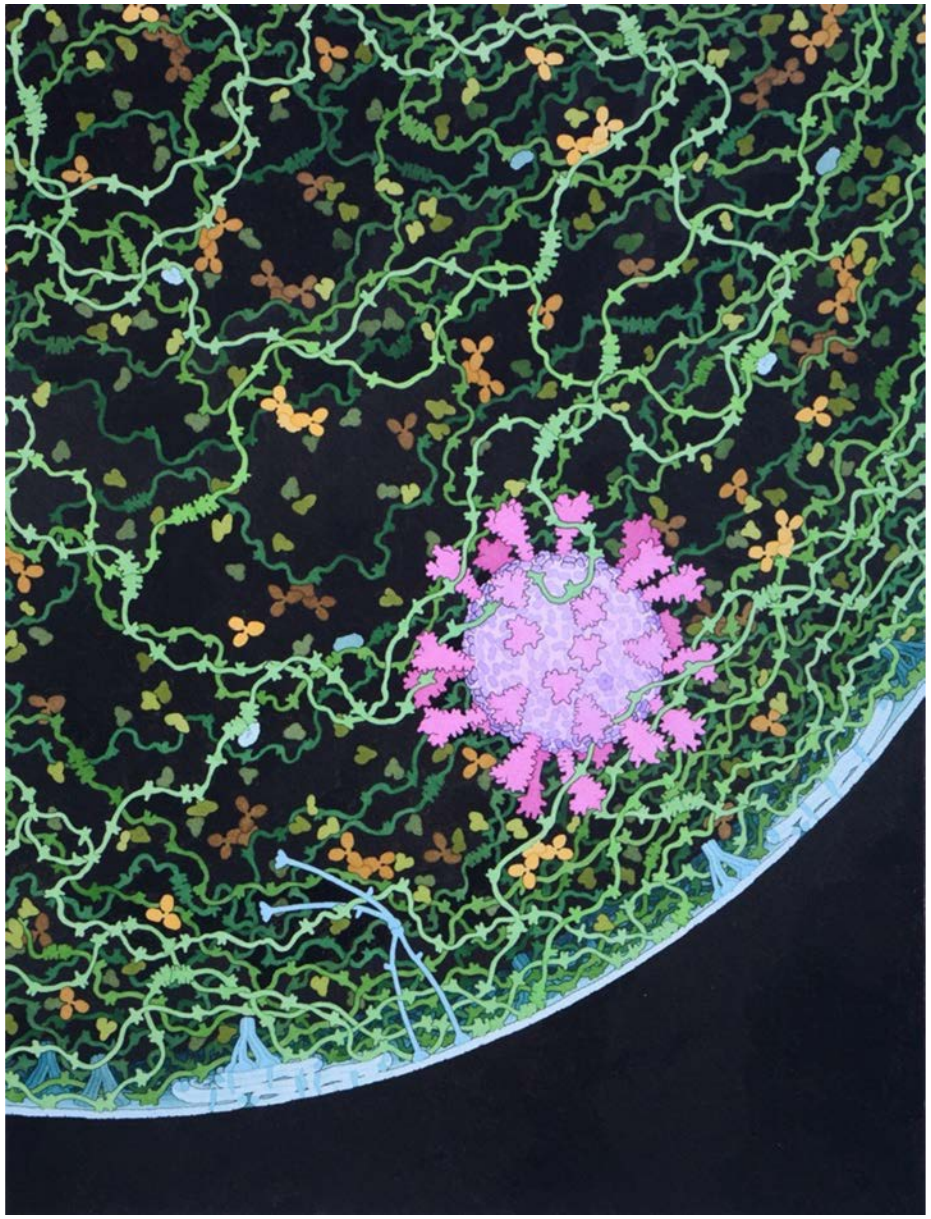
Burley *et al*. (2023) *Nucleic Acids Research 51*, D488-D508.

RCSB.org Opt In: Computed Structure Models



PDB Essential for Responding to Emerging Viruses

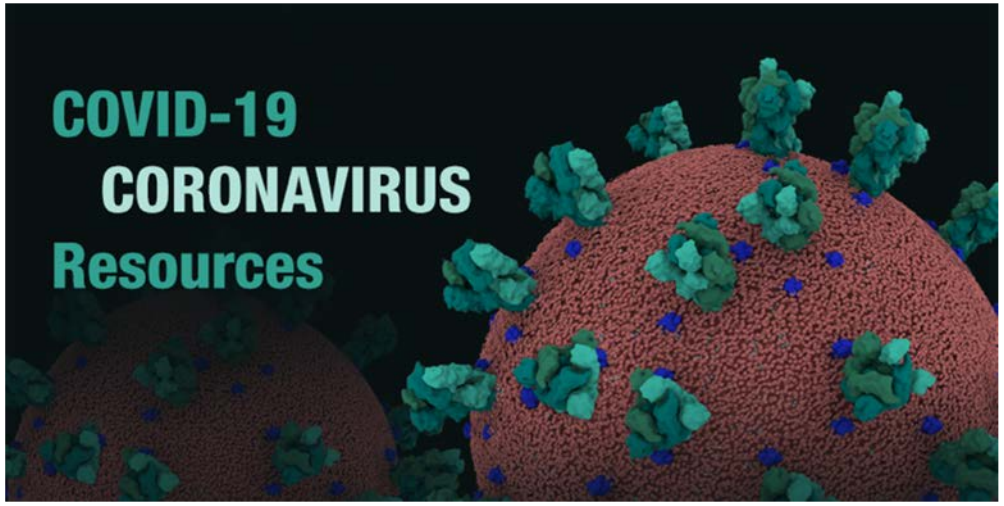
- SARS-CoV Epidemic 2002
  >170 SARS-CoV structures□PDB
- MERS-CoV Epidemic 2012
  >100 MERS-CoV structures□PDB
- COVID-19 Pandemic 2019
  >3,000 SARS-CoV-2 structures□PDB
- Effective mRNA vaccines designed and antiviral agents discovered using PDB structures of SARS-CoV, MERS-CoV, and SARS-CoV-2 proteins

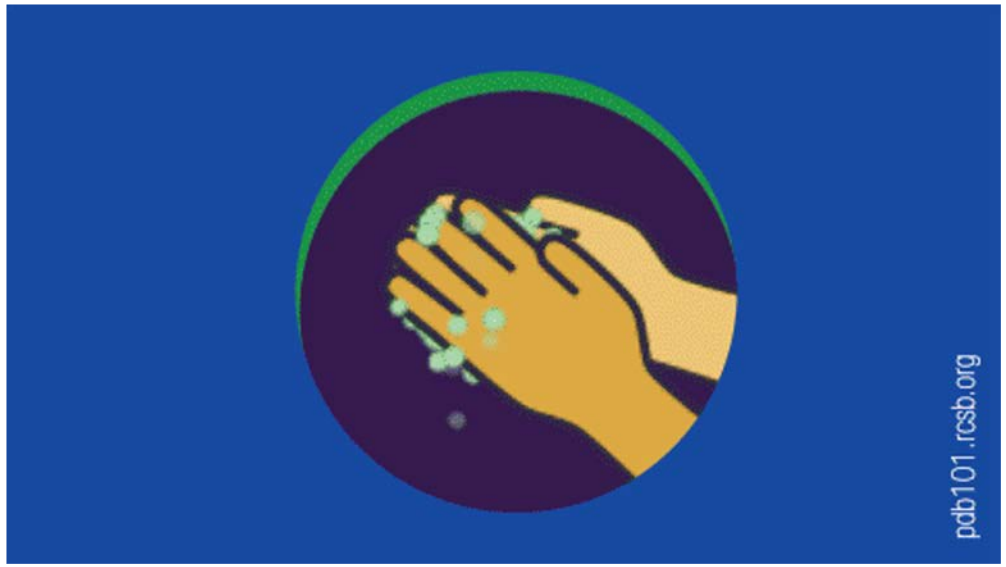


Respiratory Droplet, 2020; David S. Goodsell

RCSB PDB Response to COVID-19

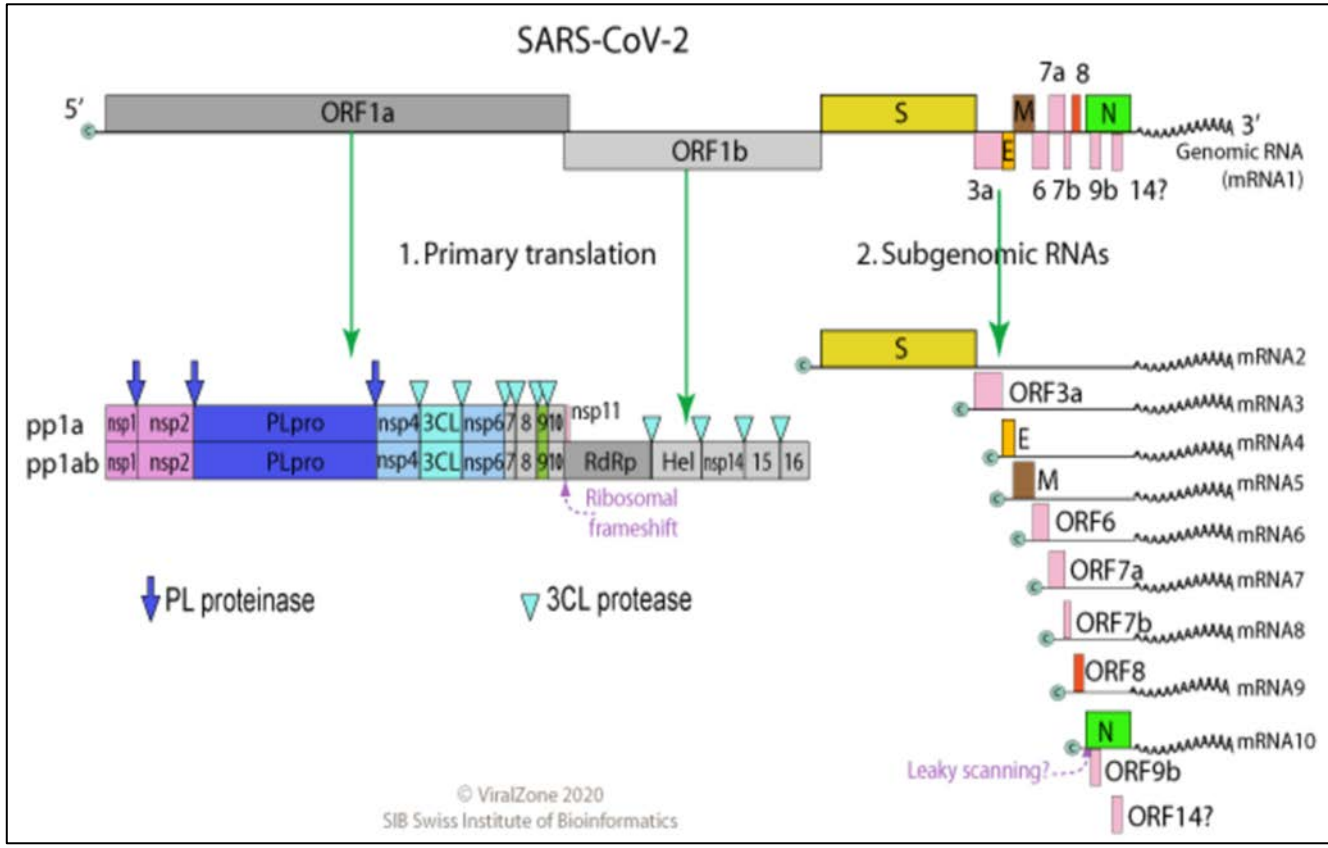
- Biocuration of COVID-19 structures prioritized, including post-release revisions (e.g., citation updates)
- PDB depositors strongly encouraged to release COVID-19 structures immediately
- Consistent taxonomy name/ID
  - Severe acute respiratory syndrome coronavirus 2; 2697049
- Consistent UniProt referencing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





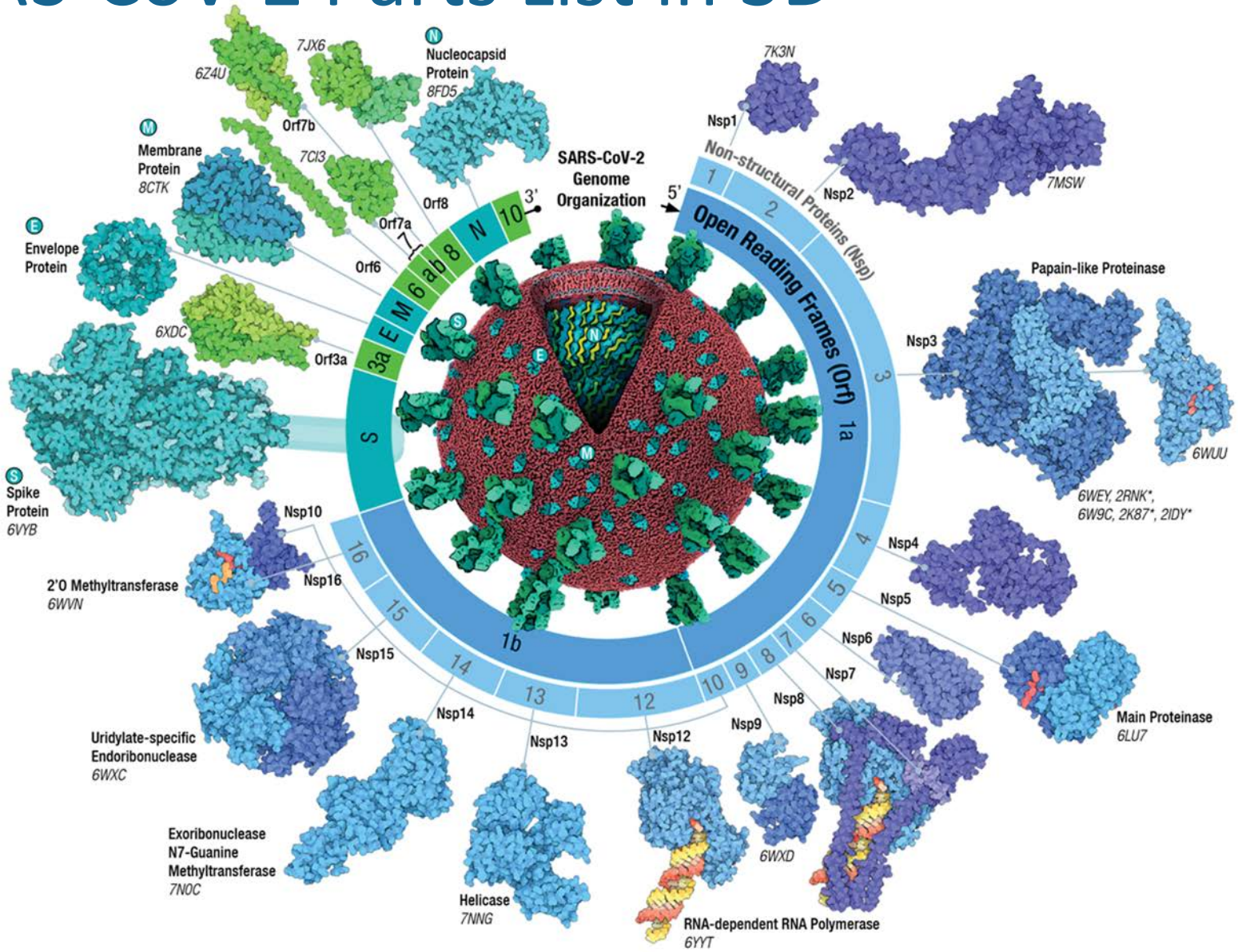
Coronavirus (SARS-CoV-2) Genome Organization

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



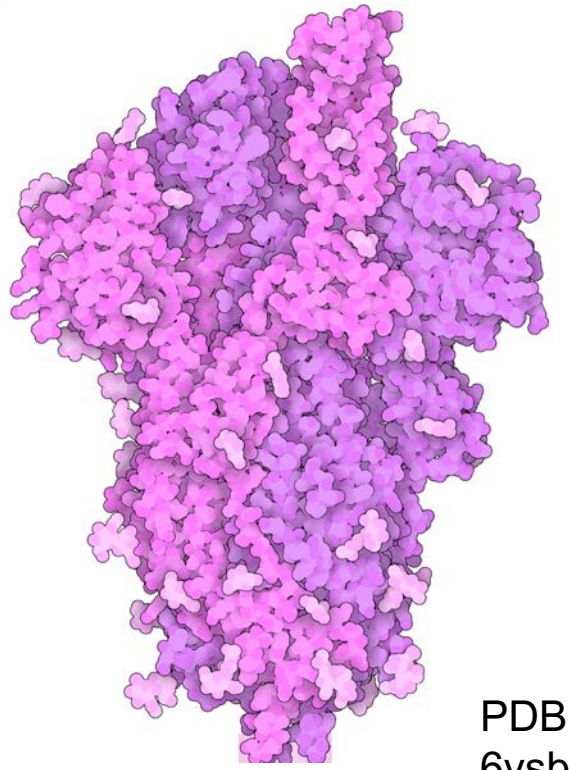
Near Complete SARS-CoV-2 Parts List in 3D

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



Structure-Based Vaccine Design: Spike Protein

- Spike Protein
  - ~1,600 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
- ~5.5 billion vaccinated!
- Tens of millions of lives were saved!



PDB ID
6vsb

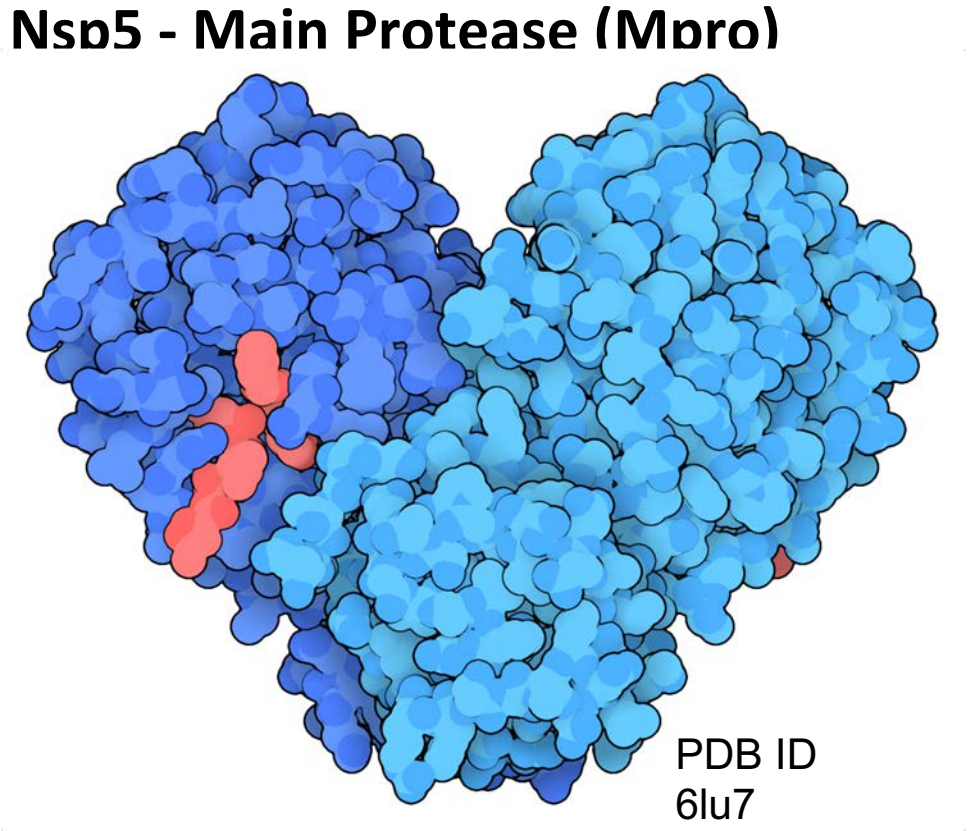
Vaccine Discovery Target
Antibody Discovery Target

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

Main Protease: Achilles Heel of SARS-CoV-2

PDB Structures

- Nsp5/Main Protease (Mpro)
  - >750 Apo/Co-crystal structures
  - Target of Pfizer's nirmatrelvir (+ritonavir=Paxlovid)
- Paxlovid is approved for outpatient treatment of individuals infected with SARS-CoV-2



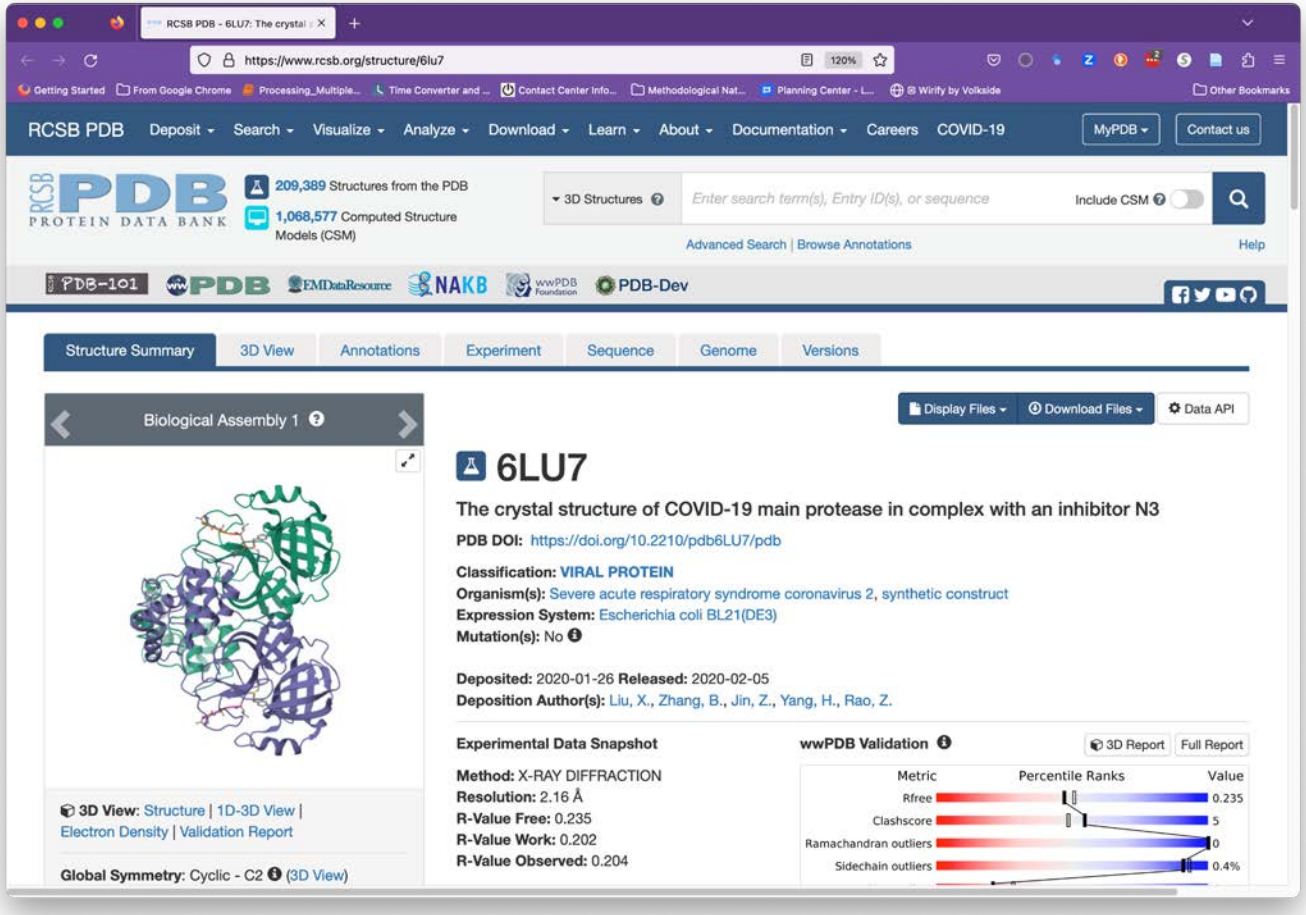
Drug Discovery Target
Symmetric Homodimer; Two Active Sites

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
- Introduction to 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
- Discussion

The Structure Summary Page for COVID-19 Main Protease



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

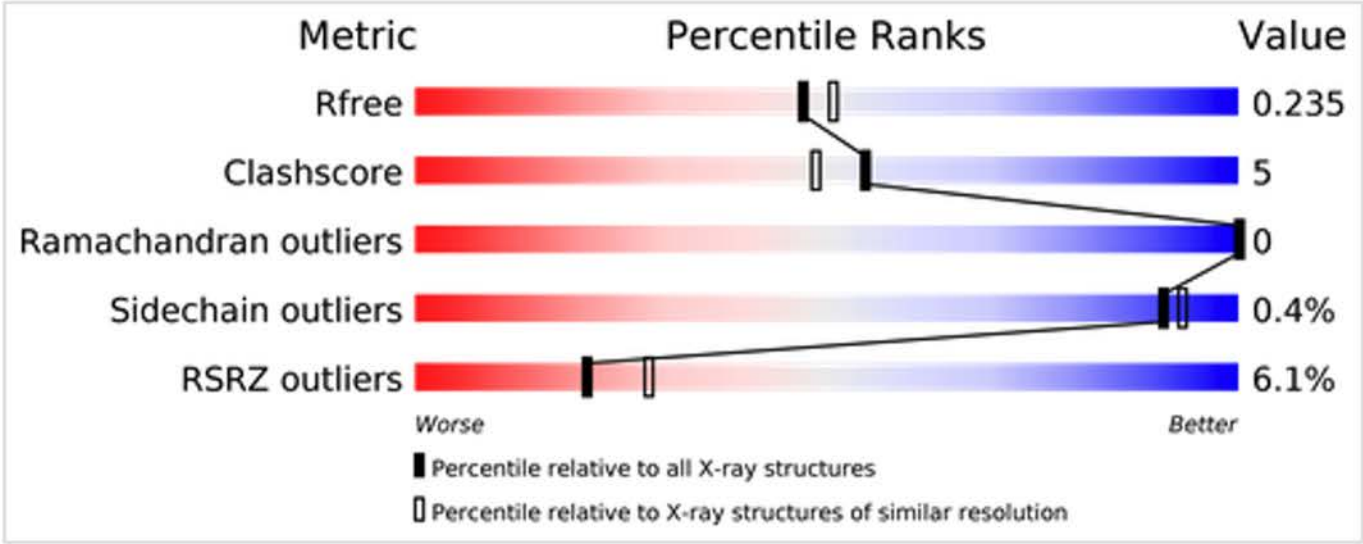
To follow along in this section, go to [Exploring the SARS-CoV-2 Main Protease]

Structure Validation for 6LU7

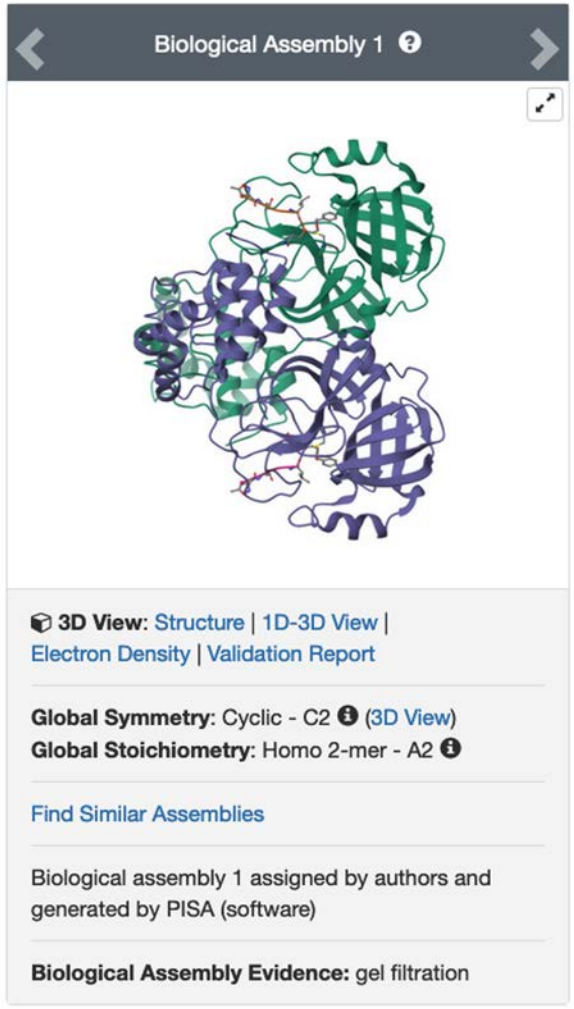
Experimental Data Snapshot

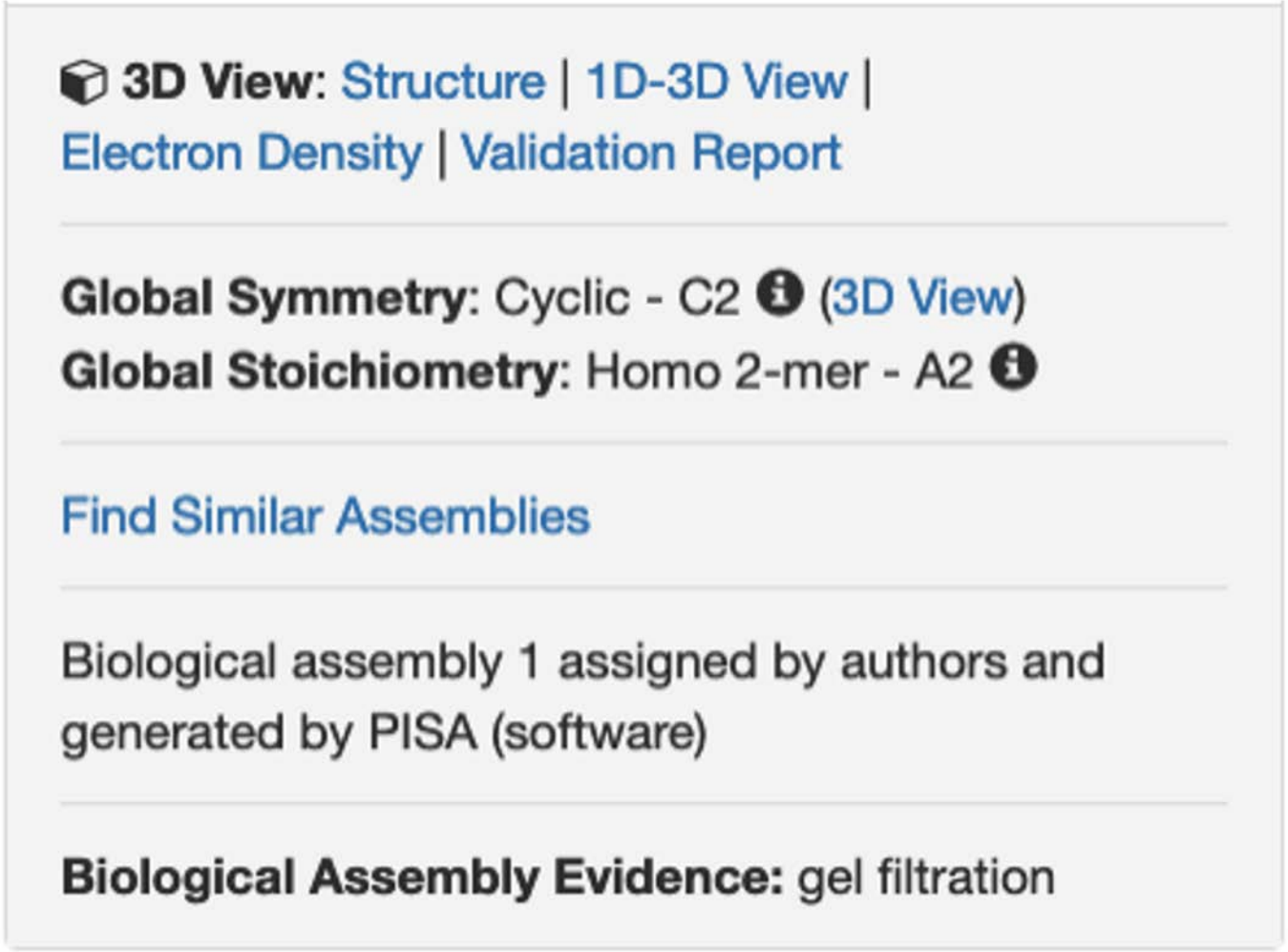
Method: X-RAY DIFFRACTION
Resolution: 2.16 Å
R-Value Free: 0.235
R-Value Work: 0.202
R-Value Observed: 0.204

wwPDB Validation 3D Report | Full Report



Exploring COVID-19 Main Protease with Mol*

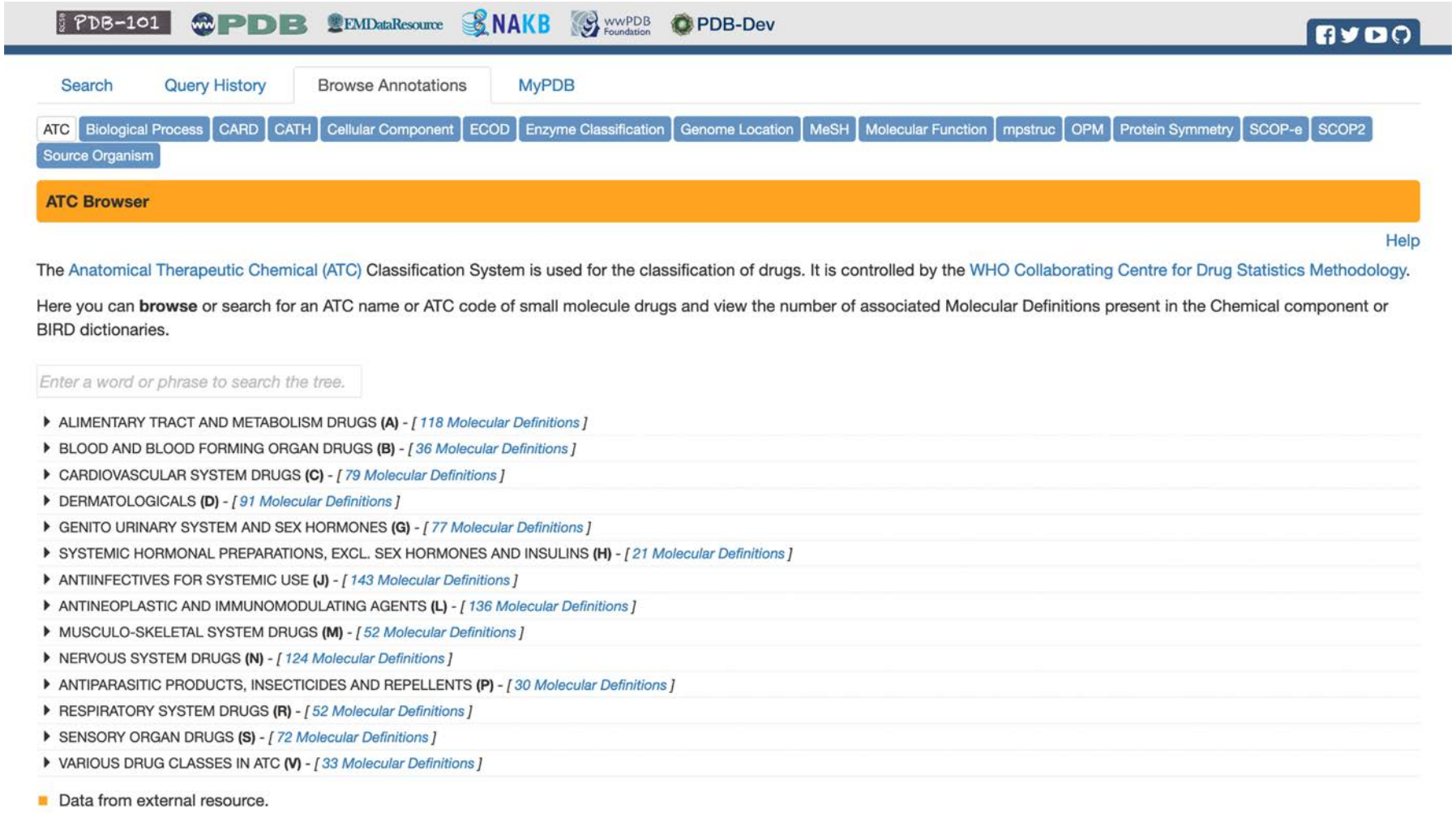




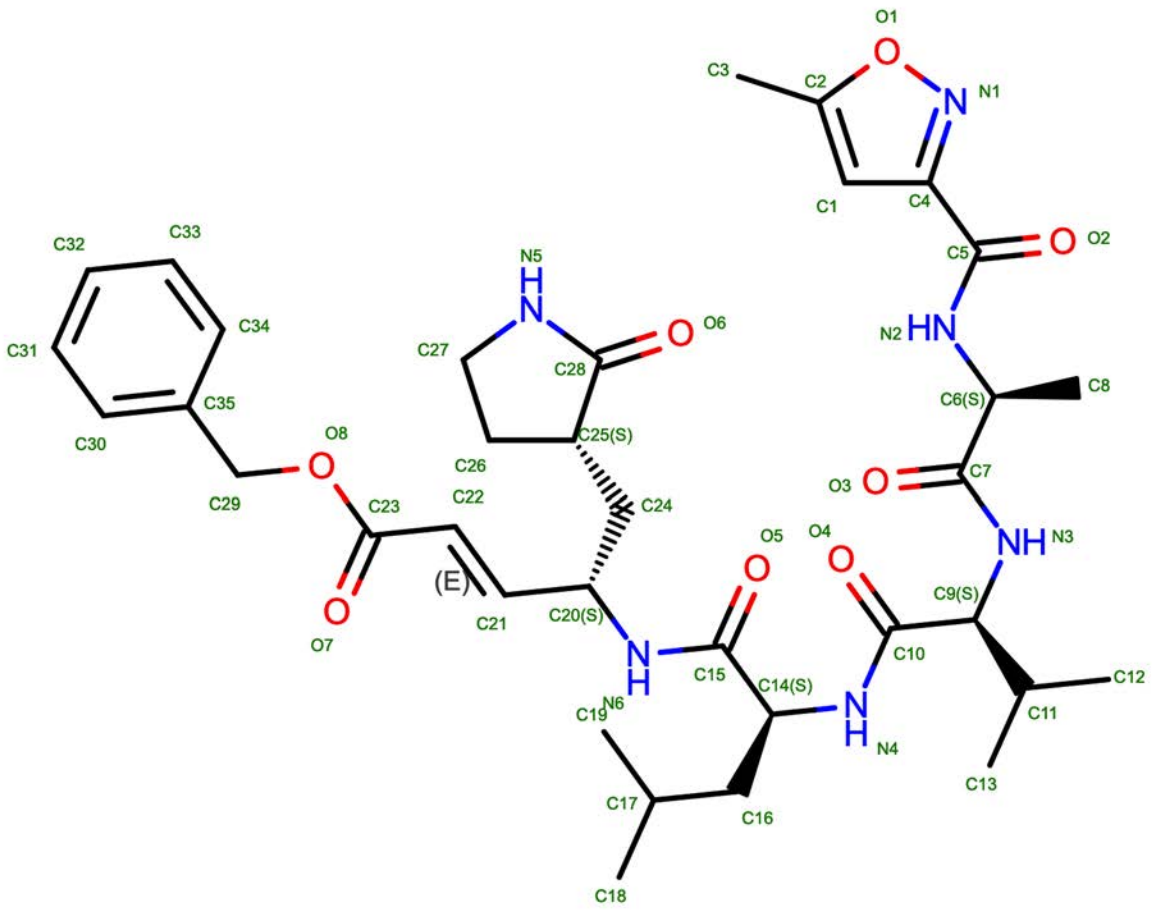
Browse Annotations

- ECOD - Evolutionary Classification of protein Domains
- Enzyme Classification
- MeSH - Medical Subject Headings
- Molecular Function
- Protein Symmetry

Browse Annotations



Biologically Interesting Molecule



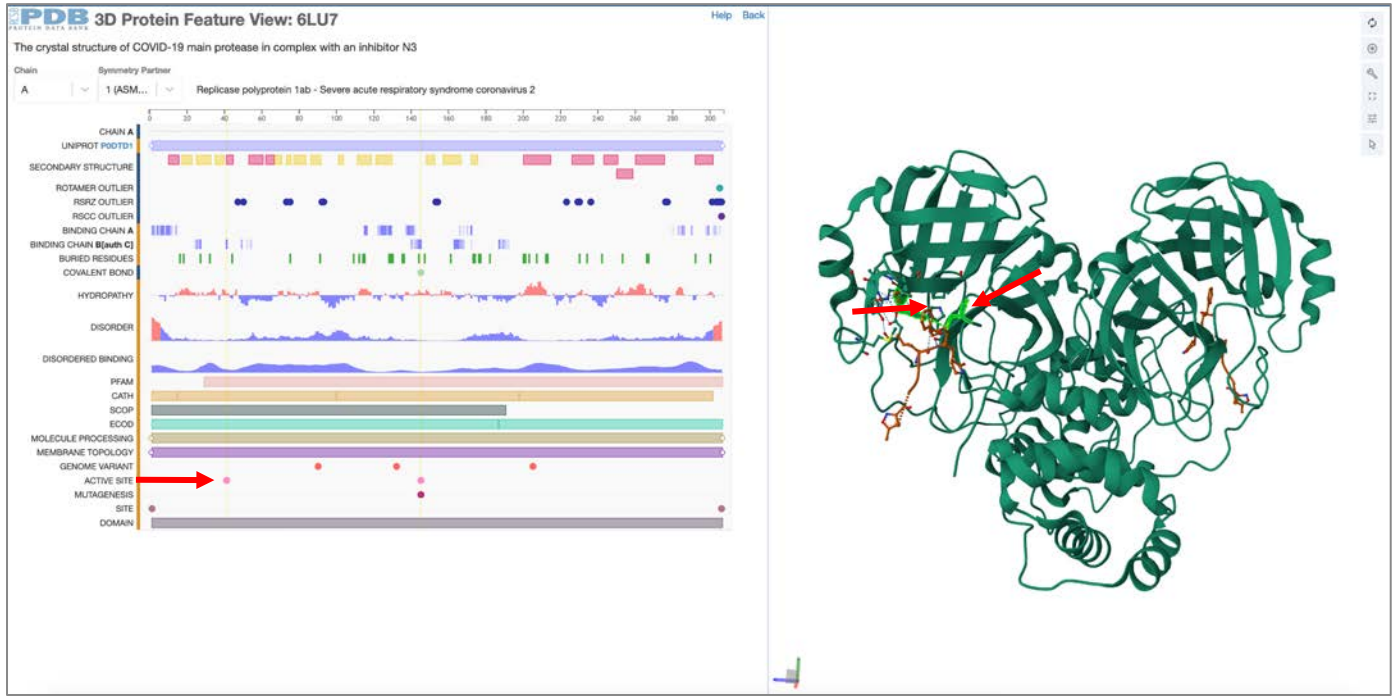
1. 2D diagram

2. Ligand Interaction

    a. Limit the image to one chain
    b. Explore the non-covalent interactions

1D-3D View

- Access from the Structure Summary Page
- Identify the active site residues
- Explore interactions
- Find the covalent ligand
- Name the ligand

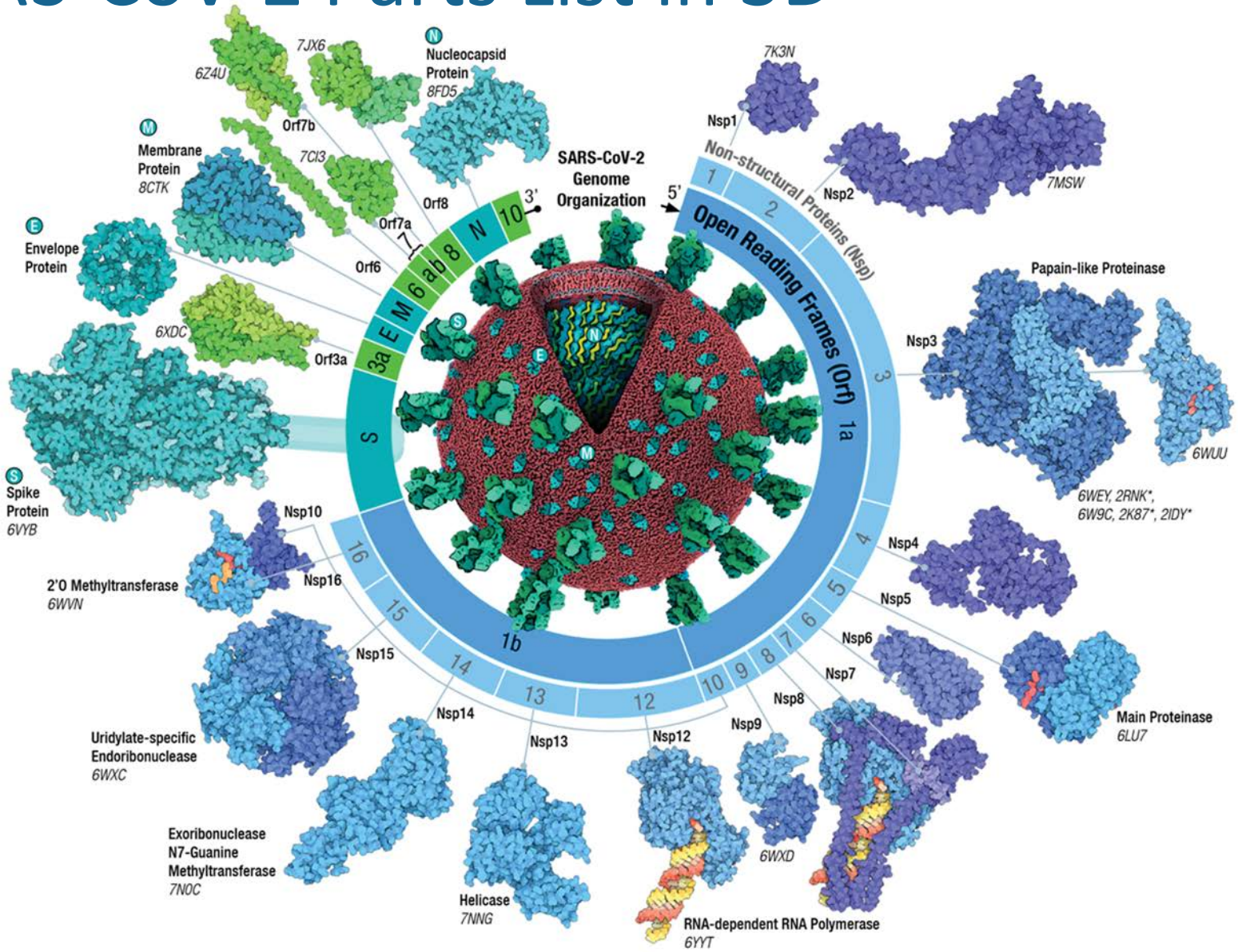


Today's Agenda

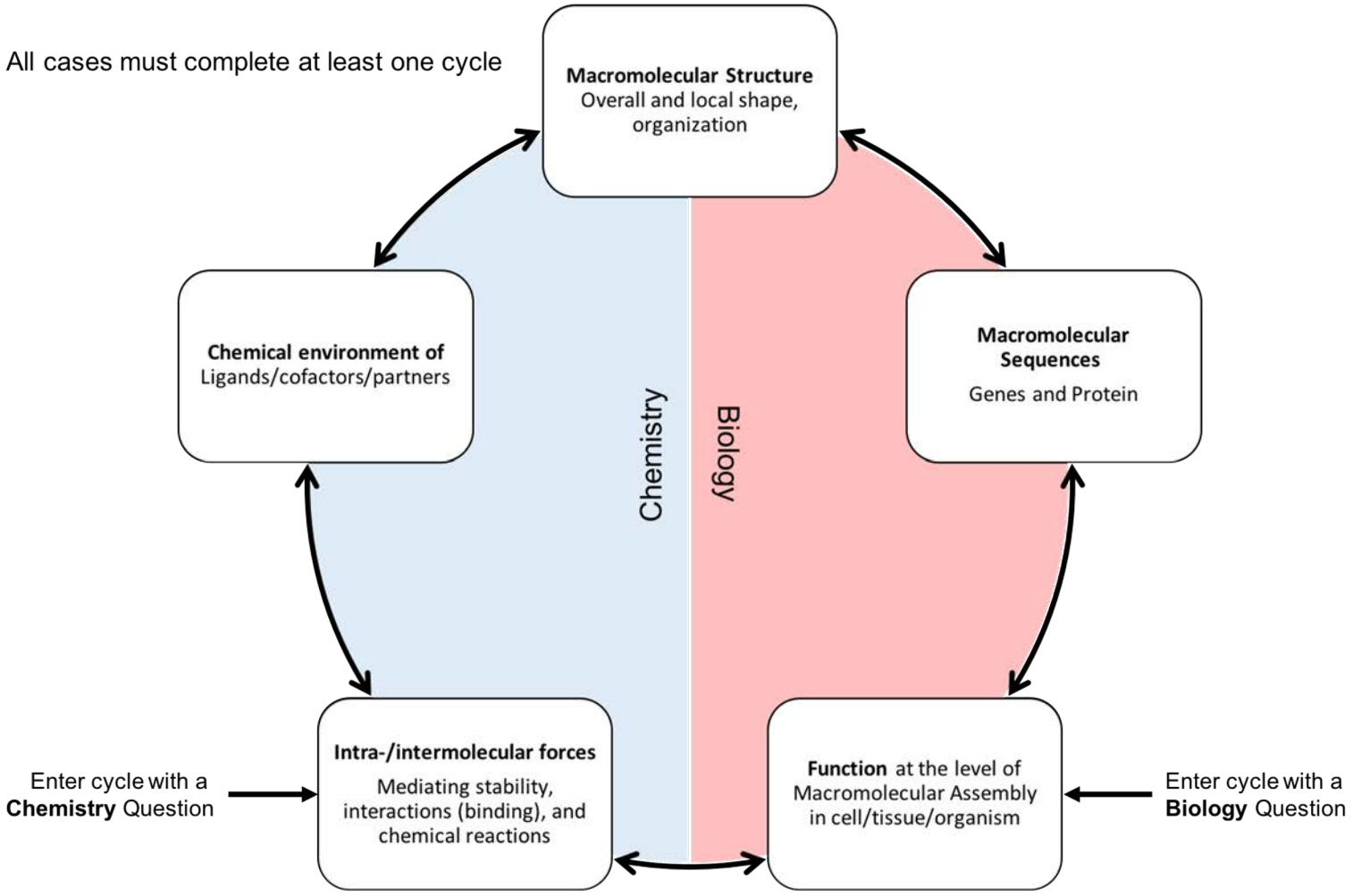
- Introduction to the Protein Data Bank and SARS-CoV-2
  - Stephen K. Burley, M.D., D.Phil. - Director, RCSB Protein Data Bank
- Exploring the SARS-CoV-2 Main Protease using RCSB.org
  - Paul Craig, Ph.D. - Rochester Institute of Technology
- Making connections using RCSB.org
  - Shuchismita Dutta, Ph.D.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

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

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

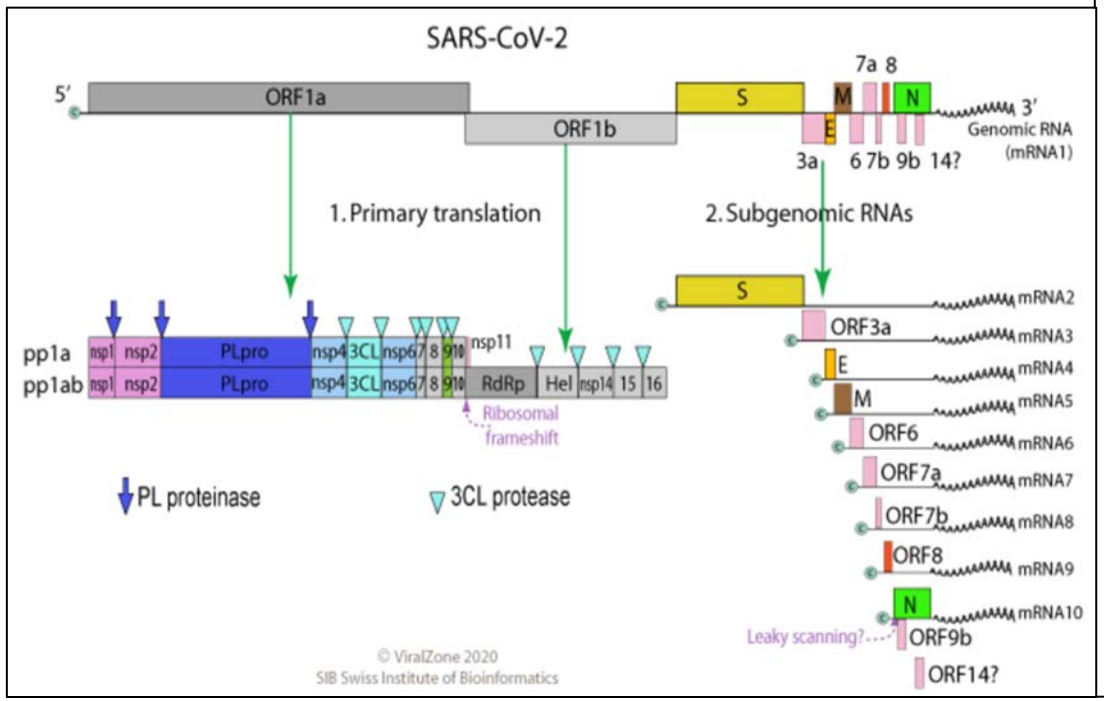


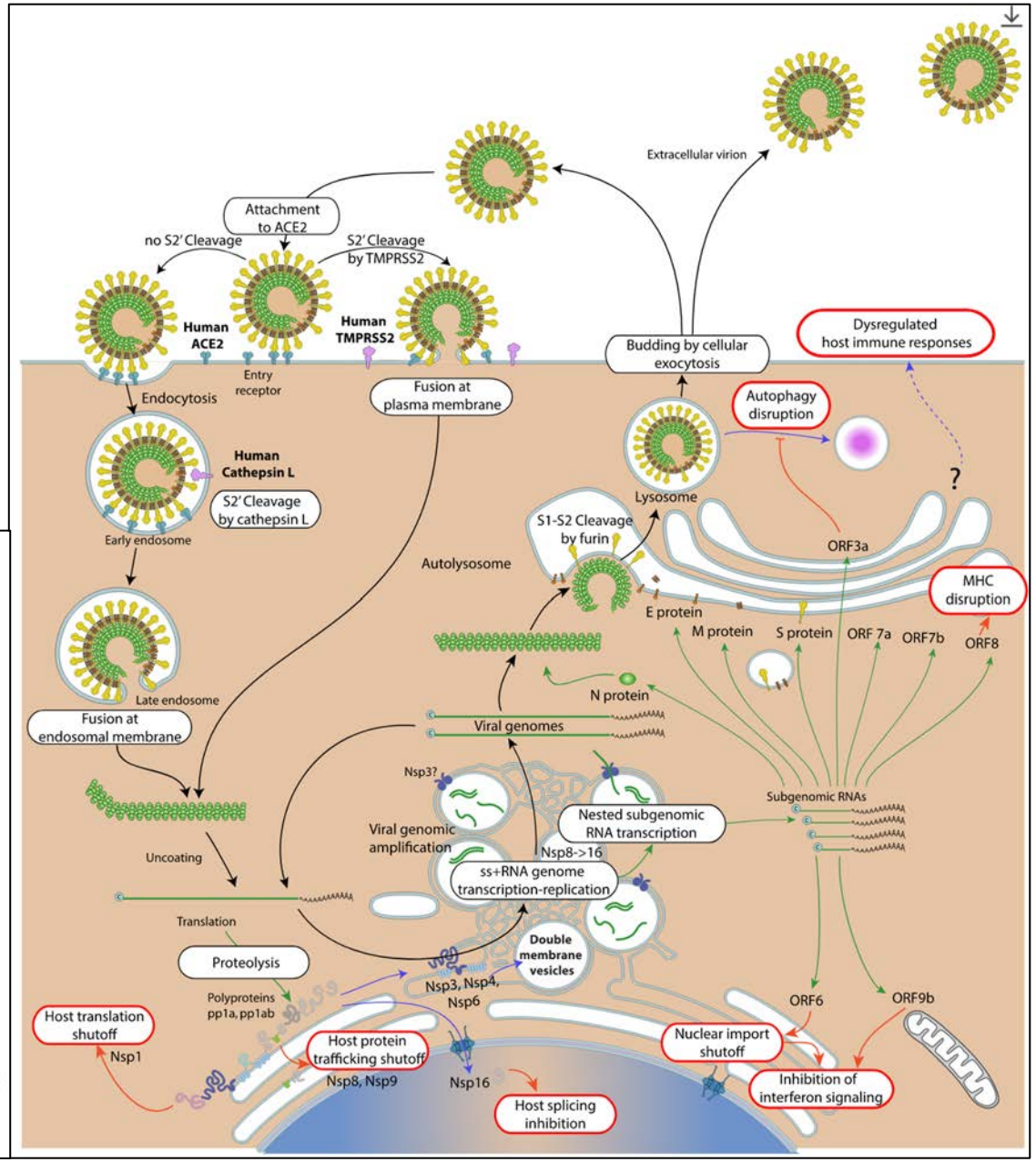
The Molecular Case Study Cycle



This project is supported by the National Science Foundation - DBI 1827011; DBI 2018884

Setting the Stage





From Pandemic to Paxlovid

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.

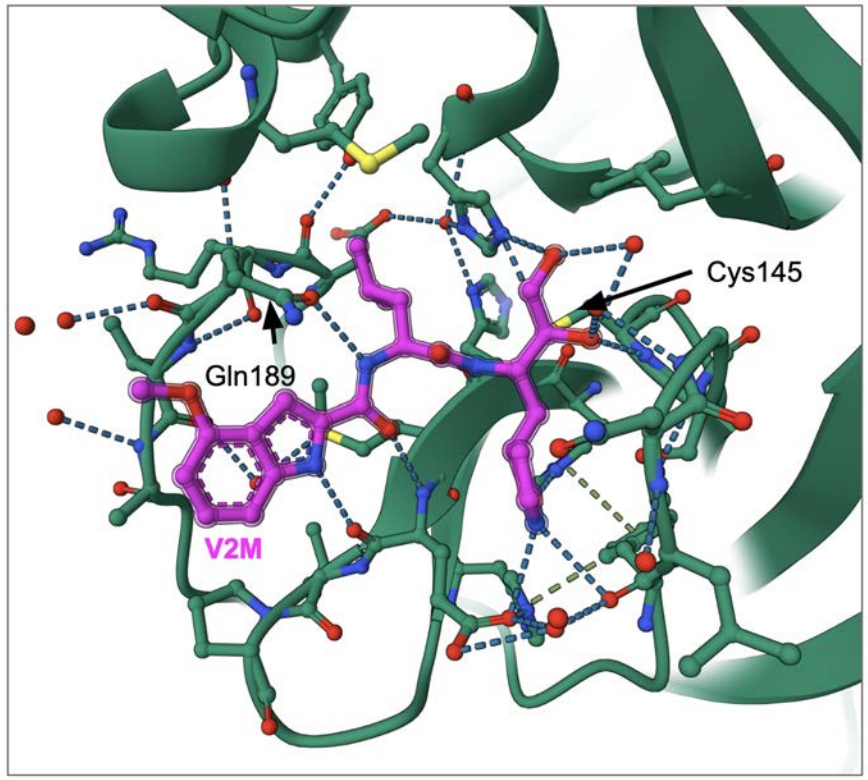
On Friday March 13th 2020, our CEO published a five-point plan for our response to Covid-19. One of those was 'marshalling our people'. That was the day I was asked to plan some specifics for the potential program. I was asked '**What would you need and how would you prosecute an oral protease inhibitor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

Why was PF-00835231 a good starting point?

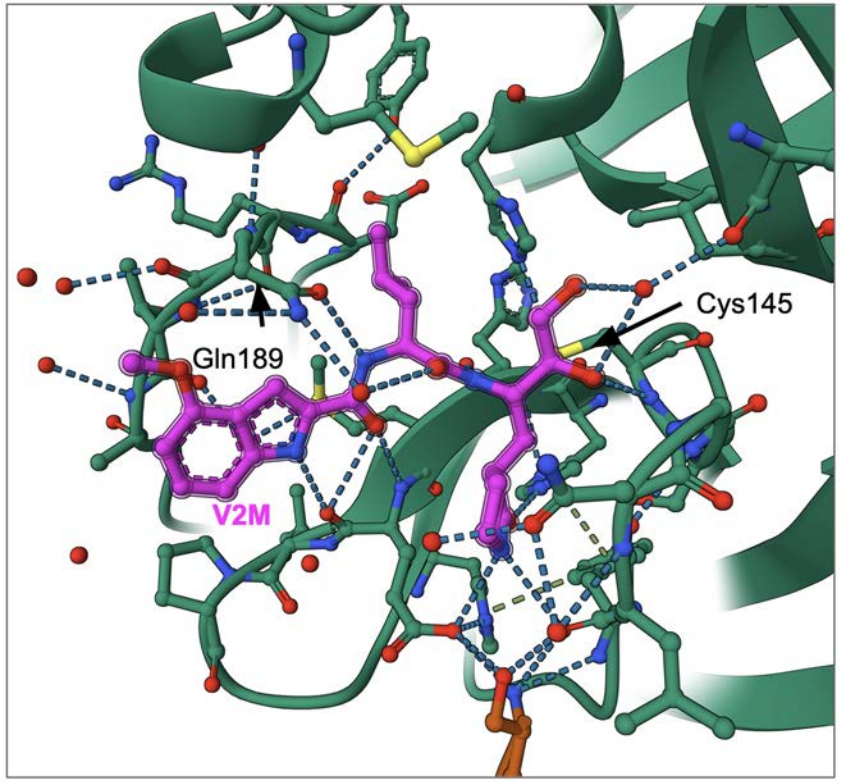
- What does PF-00835231 look like - can we find it in the PDB?
- What does it bind to:
  - Does it bind the SARS-CoV-2 Main Protease and where?
  - What else does this molecule bind and where?
- Compare key binding targets - What does this mean?

To following along in this section go to From Pandemic to Paxlovid

Paxlovid Binding to the Main Protease



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



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

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

Review Learning

Knowledge

- Structure of drug PF-00835231
- Identify enzymes it binds.
- Nature of inhibitor binding.

Skills

- Search drug by name
- Visualize and learn about drug
- Identify structures in the PDB which have this drug bound
- Visualize and analyze inhibitor bound structures
- Compare structures

Assess Learning and More

***Q**: What did it feel like when you first saw the PF-07321332 data? Was the team together when the data were revealed?*
Dr. Owen's reply: There were certainly lots of 'data moments' for PF-7321332 (*PF-07321332*). Ultimately it was the Ph2/3 high risk patient readout in early November 2021 that many people will remember, but in a preclinical sense our team got pretty excited about 50% oral bioavailability in an animal model – we were after an oral therapeutic after all. We also remember the day we saw the safety margins established in the mandatory regulatory toxicology studies. This was a critical milestone in taking a molecule into human trials for the first time. People worked through their 2020 holiday season to generate that data. The team was never together for any of these moments in person. Having worked from home since March 2020, I didn't see a team member face to face until April 2021. By that point we'd built a program, invented a clinical candidate and were already a month into human clinical trials.

Q. Is PF-07321332 (the key component of Paxlovid) an effective therapeutic agent?

- Does it bind to the many known mutants of SARS-CoV-2?
- Will it bind to newer types of coronaviruses that may evolve in future?

Is Paxlovid an effective drug?

- What does PF-07321332 look like - can we find it in the PDB?
- What does it bind to
  - Does it bind to different coronaviral main proteases?
  - Does it bind to mutants of SARS-CoV-2?
- Can you identify at least one protease mutant where the Paxlovid mutant may not be effective as a therapeutic agent?

Review Learning

Knowledge

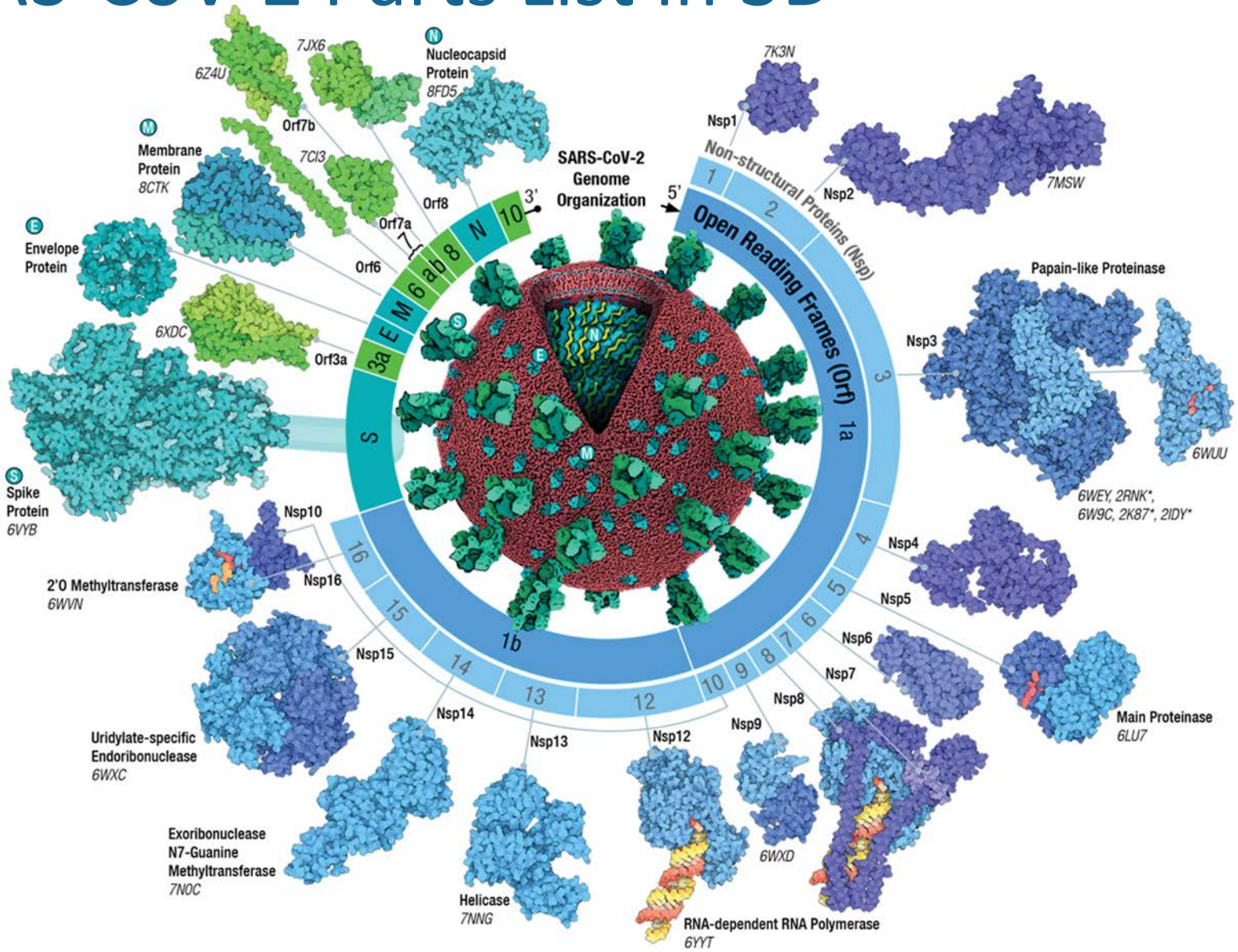
- Structure of drug PF-07321332
- Identify enzymes it binds
- Nature of inhibitor binding to
  - Enzyme from other species
  - Impact of mutations
- *Predict future effectiveness of drug*

Skills

- Search drug, visualize and learn about it
- Identify structures in the PDB which have this drug bound
- Visualize, analyze, and compare inhibitor bound structures
- *Explore Groups of Structures*

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

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



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
- Discussion

Provide Feedback

Please complete the Exit Survey to

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

https://rutgers.ca1.qualtrics.com/jfe/form/SV_2fsfdttRAmTBRHg

RCSB PDB Team

RCSB.ORG
info@rcsb.org

Core Operations Funding

National Science Foundation (DBI-1832184),
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
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

OPPORTUNITIES for SCIENTIFIC SOFTWARE DEVELOPERS, POSTDOCTORALS, and STUDENTS

Develop innovative analysis, integration, query, and visualization tools for 3D biomolecular structures at **RCSB.org** to help accelerate research and training in biology, medicine, and related disciplines. Design, develop, and deploy modern web and data applications and complex interactive graphical user interfaces.

Visit www.rcsb.org/pages/jobs for more information

- Database Administrator (Rutgers)
- Postdoctoral Researchers
  - Metalloproteins (Rutgers)
  - Bioinformatics (UCSD)
- Gap Year Interns (Rutgers)
- Undergraduate Summer Research Interns (Rutgers)

RUTGERS UC San Diego SDSC San Diego Supercomputer Center UCSF University of California San Francisco

Summer Scholars Emma and Jordi beta testing the Bound! card game

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