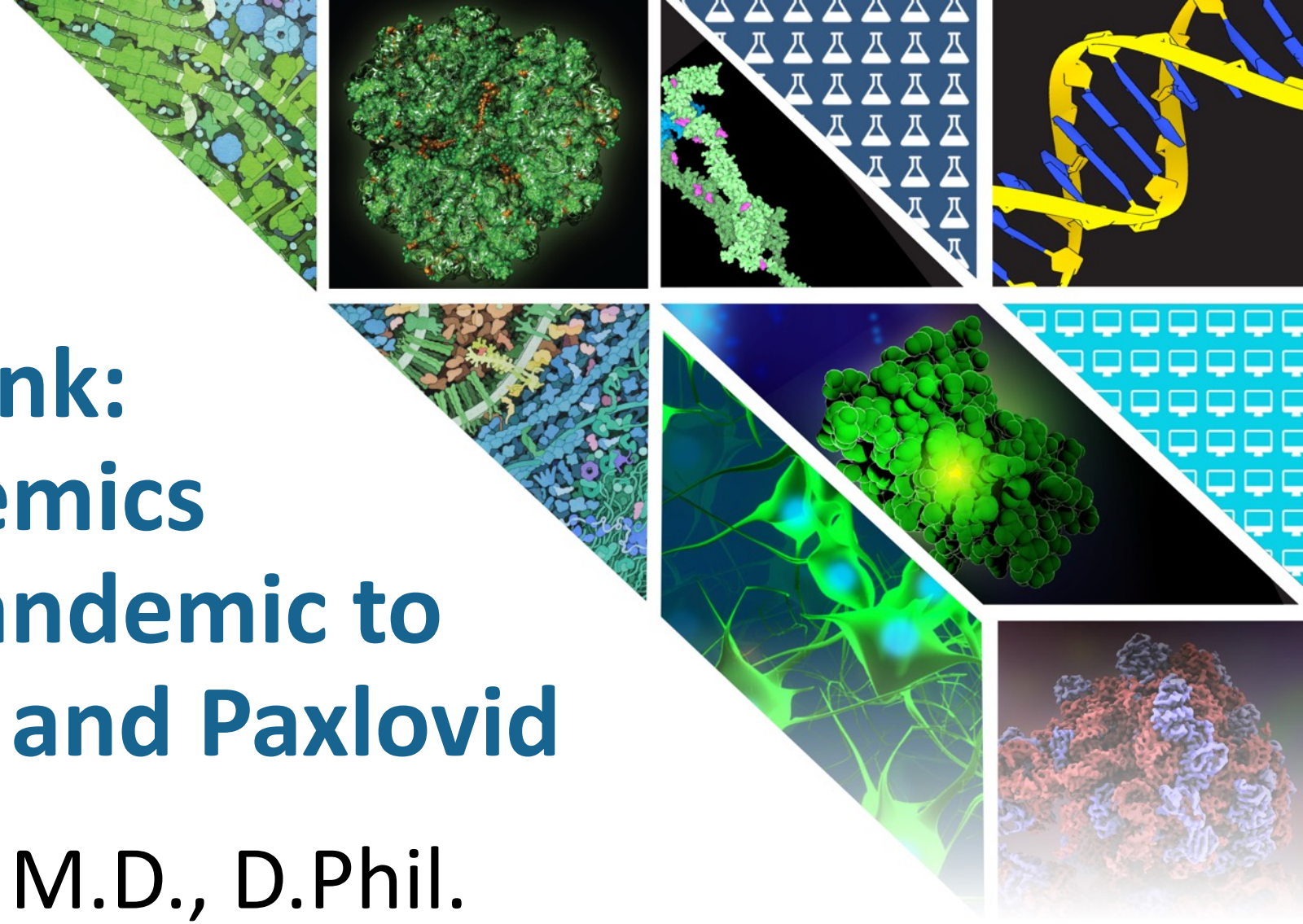
RCSB PDB
PROTEIN DATA BANK

rcsb.org

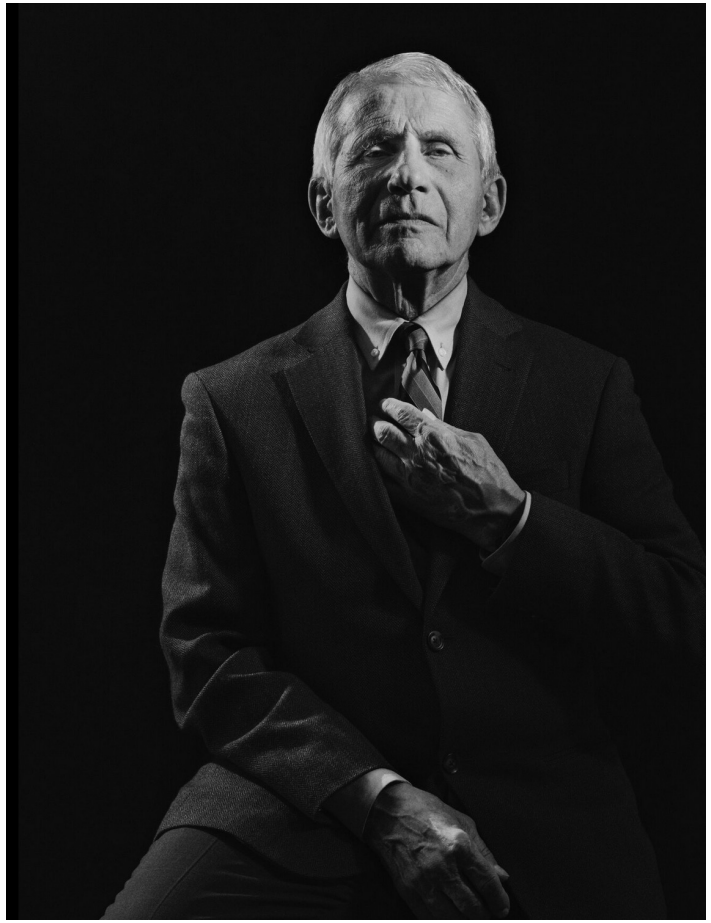


Protein Data Bank: From Two Epidemics to the Global Pandemic to mRNA Vaccines and Paxlovid

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

National Academy of Sciences, Engineering, and Medicine Advancing Drug Discovery Webinar
June 4th 2024 Noon-1:00pm Eastern Daylight Time

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.

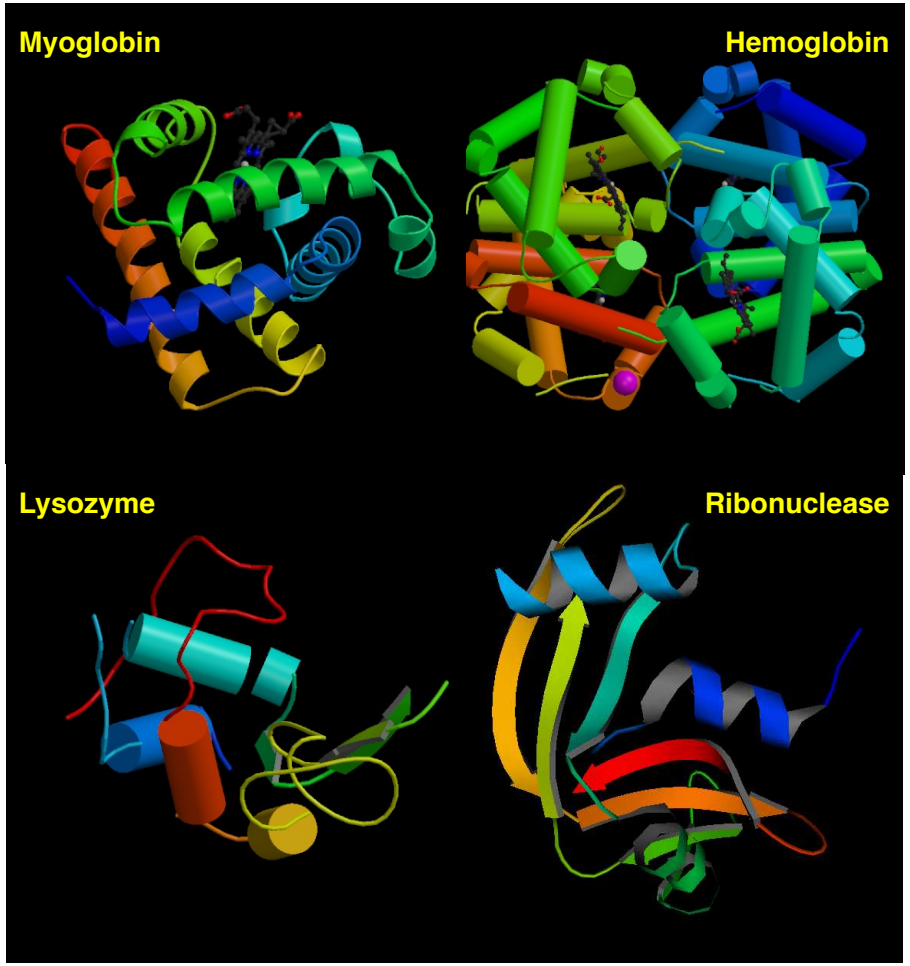
https://www.nytimes.com/interactive/2023/04/24/magazine/dr-fauci-pandemic.html

Outline

- History of the Protein Data Bank and the Worldwide PDB partnership
- RCSB Protein Data Bank: US Data Center for the wwPDB
- RCSB.org: One-Stop-Shop for public 3D Biostructure Data
- How Structural Biologists and the PDB contributed to the Fight Against the SARS-CoV-2 and Taming of the COVID-19 Pandemic
  - Open Access to >4,200 SARS Related Structures
    - Guiding design of mRNA Vaccines encoding the SARS-CoV-2 Spike Protein;
    - Explaining the success of drugs purposed for use during the pandemic (molnupiravir); and
    - Facilitating Structure-Guided Discovery of Paxlovid (active ingredient: nirmatrelvir).
- Sobering Postscript
- Acknowledgements

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 >220,000 structures!**
- wwPDB Partnership founded in 2003
- Members: RCSB PDB (US), PDBe (EMBL-EBI), PDBj (Japan), and PDBc (China); plus EMDB (3DEM) and BMRB (NMR)

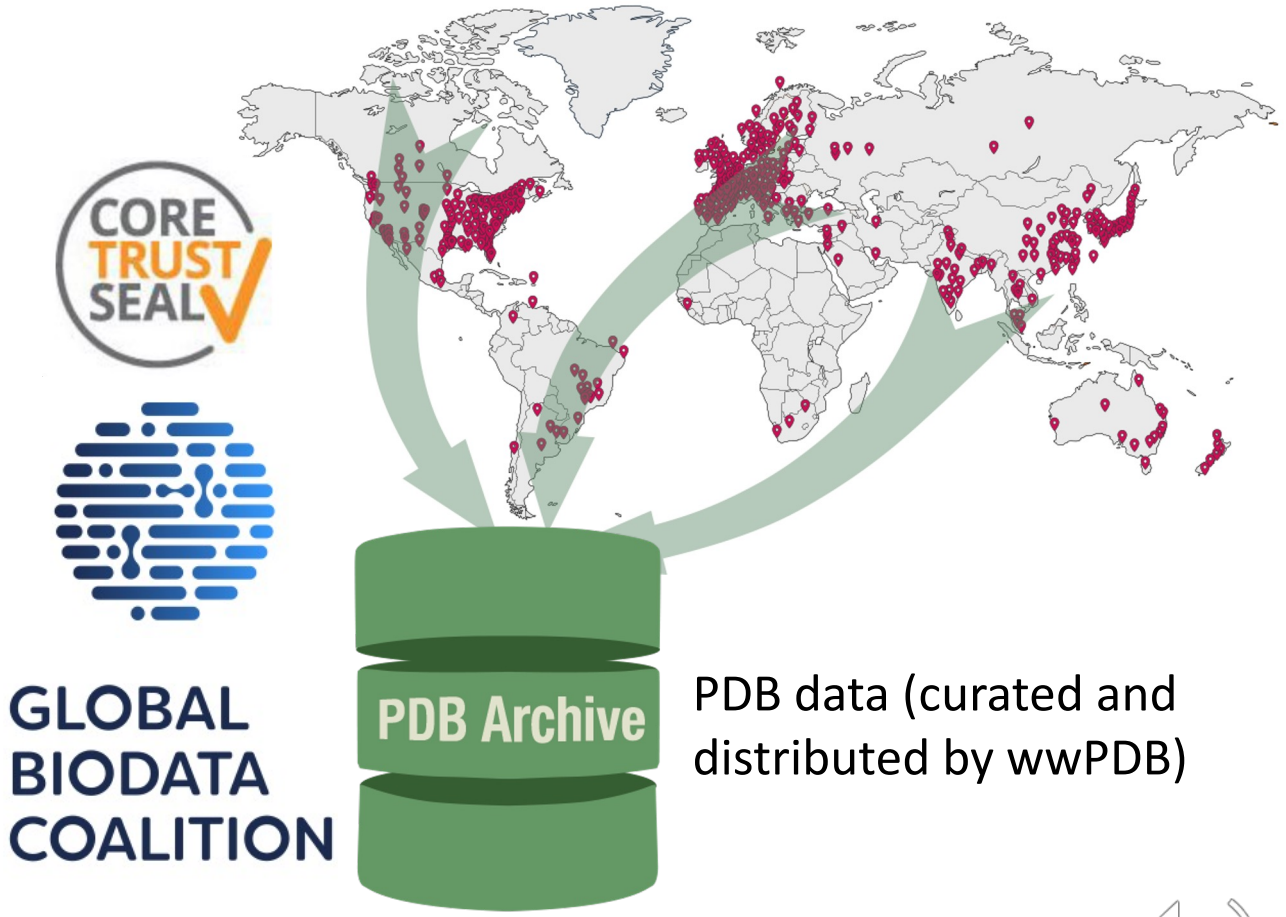


Structures that Inspired Launch of the PDB

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

Protein Data Bank is “Crucial for Sustaining the Broader Biodata Infrastructure”

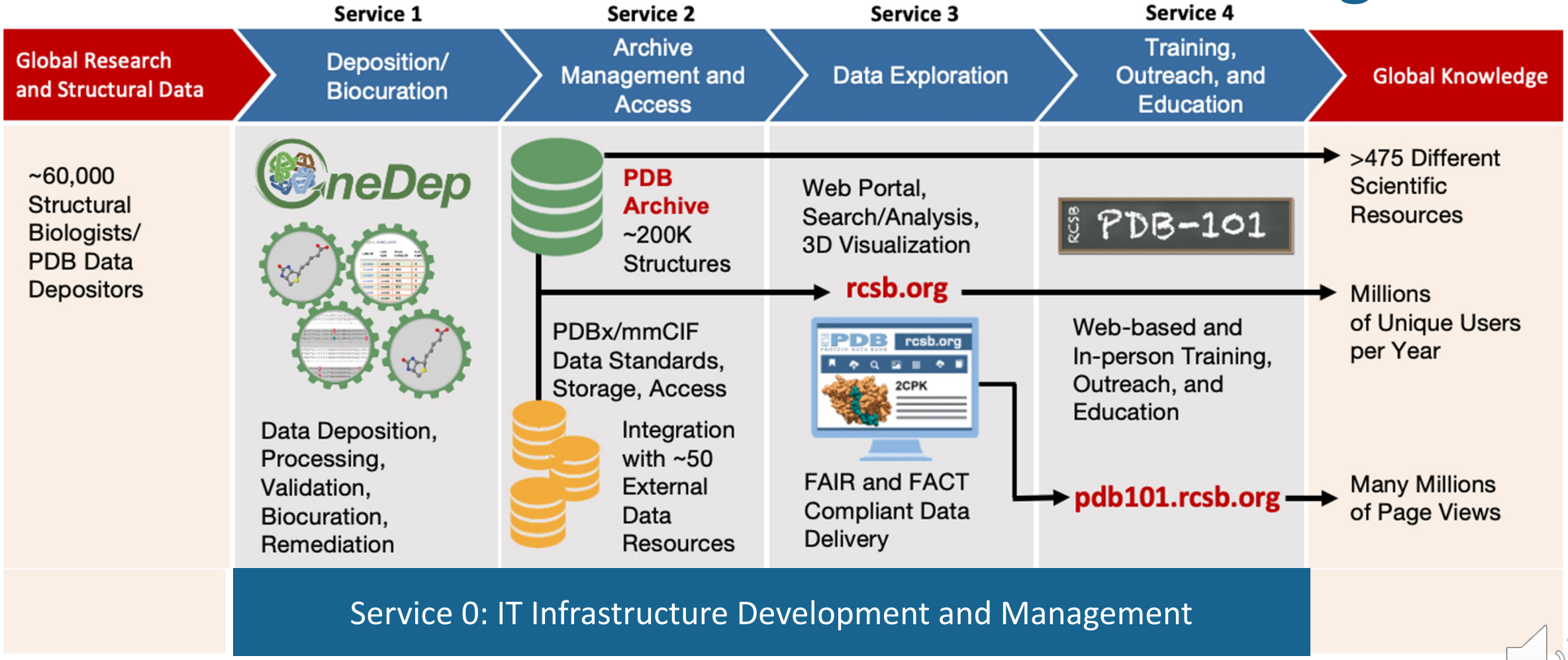
- Single FAIR/FACT compliant global archive providing Open Access to public domain experimental 3D biostructures
- PDB data distributed under the Creative Commons CC0 License
- RCSB Protein Data Bank is the US wwPDB Data Center jointly supported by NSF, NIH, and DOE



PDB data (curated and distributed by wwPDB)

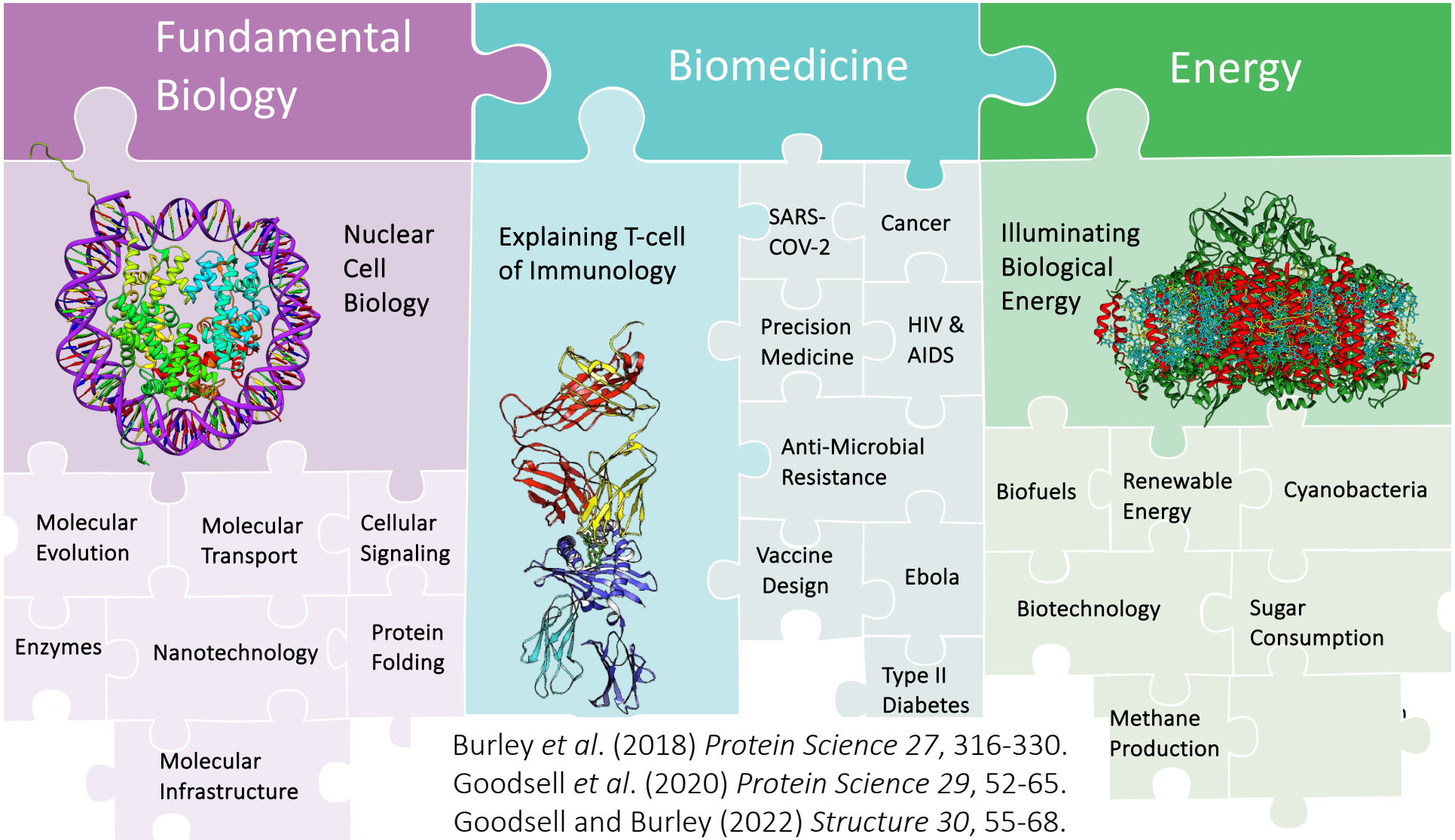
Protein Data Bank (1971) *Nature New Biology 233*, 223.
Worldwide Protein Data Bank (2019) *Nucleic Acids Research 47*, D520–D528.

RCSB Protein Data Bank (RCSB PDB) Services Convert Global Data into Global Knowledge



Burley *et al*. (2022) *Protein Science 31*, E4482.
Burley *et al*. (2023) *Nucleic Acids Research 51*, D488–D508.

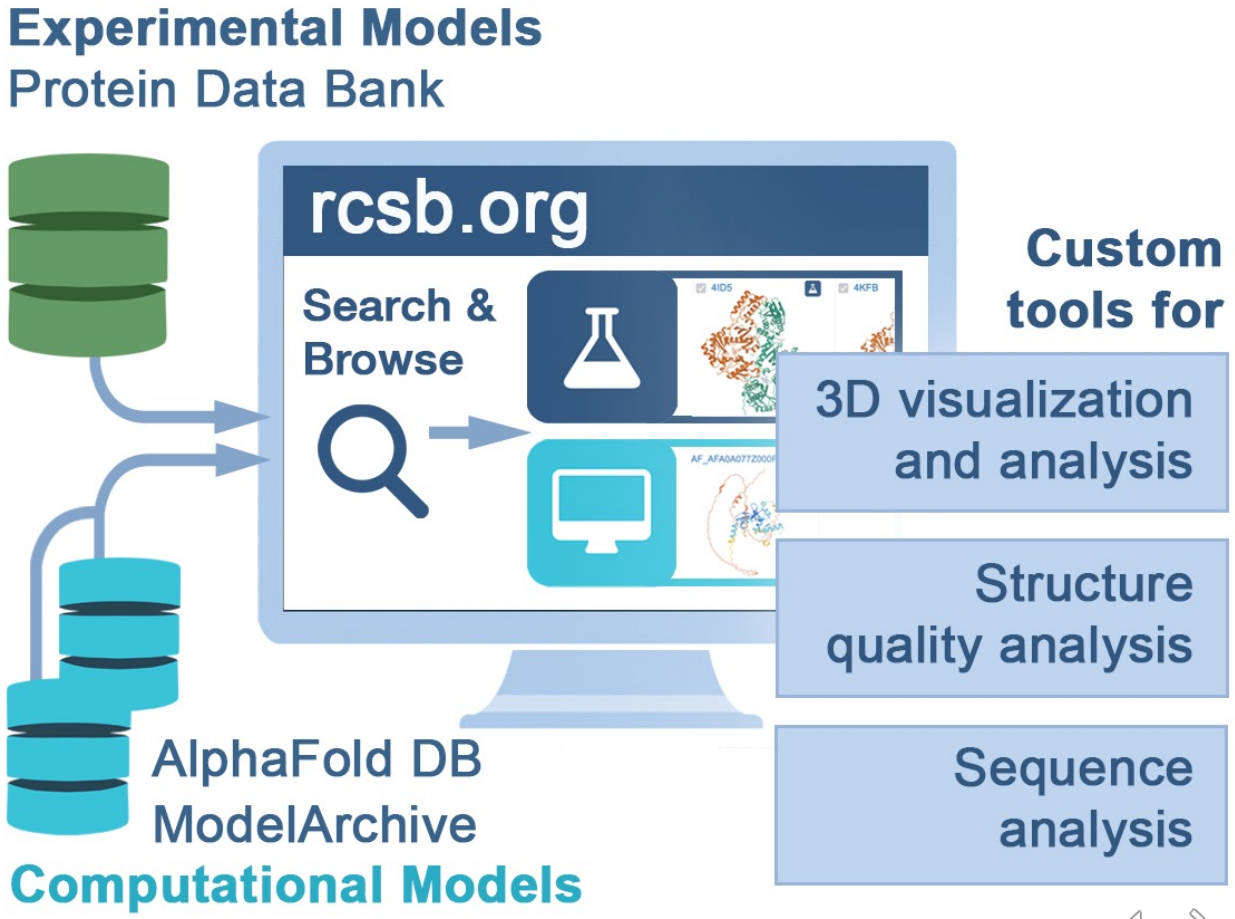
RCSB PDB Impact Across the Biosciences



Burley *et al*. (2018) *Protein Science 27*, 316-330.
Goodsell *et al*. (2020) *Protein Science 29*, 52-65.
Goodsell and Burley (2022) *Structure 30*, 55-68.
Burley *et al*. (2022) *Protein Science 31*, 187-202.

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

- RCSB.org delivers
  - >220,000 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



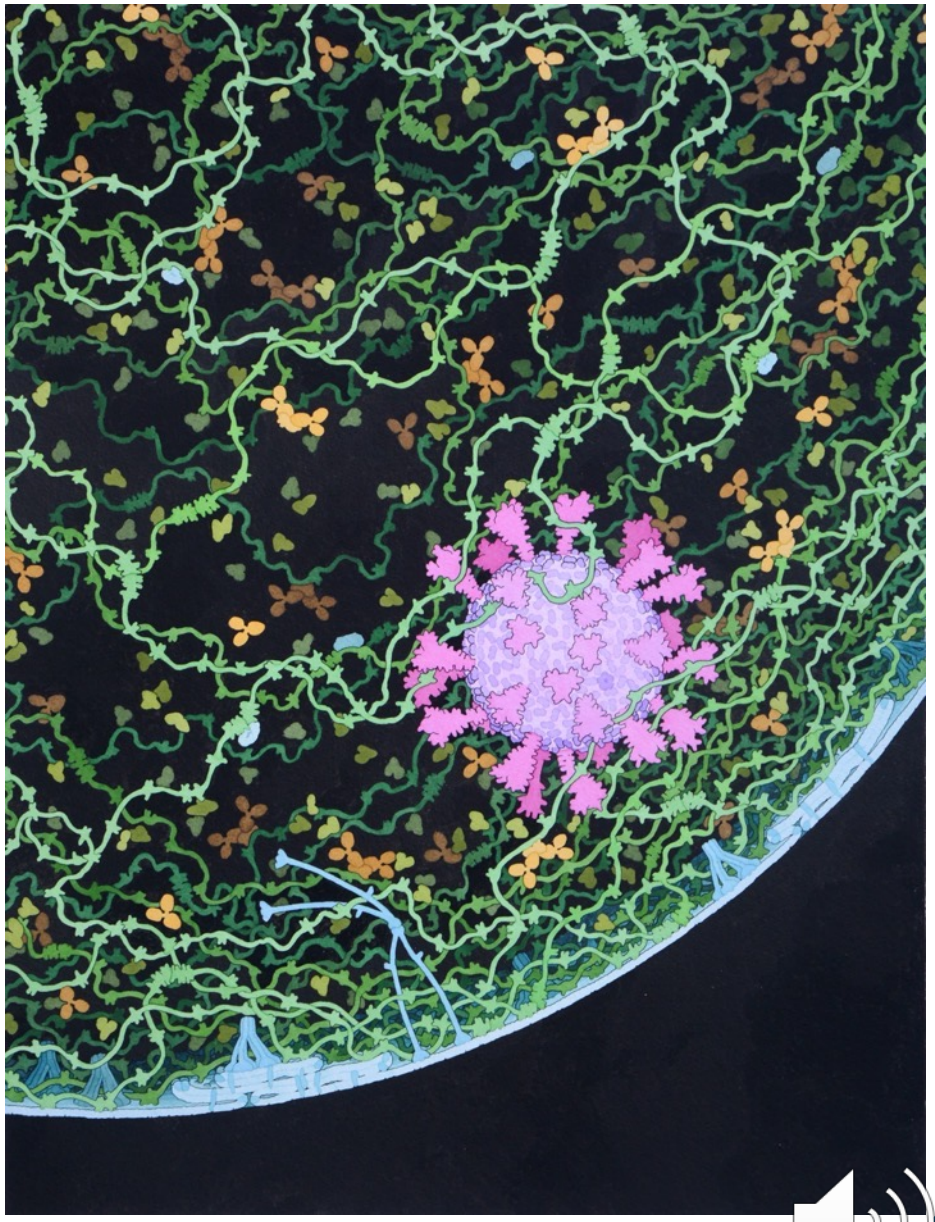
Burley *et al*. (2022) *Protein Science 31*, E4482.
Burley *et al*. (2023) *Nucleic Acids Research 51*, D488–D508.

RCSB.org Opt In: Computed Structure Models

Burley *et al*. (2022) *Protein Science 31*, E4482.
Burley *et al*. (2023) *Nucleic Acids Research 51*, D488–D508.

PDB Essential for Responding to Emerging Viruses

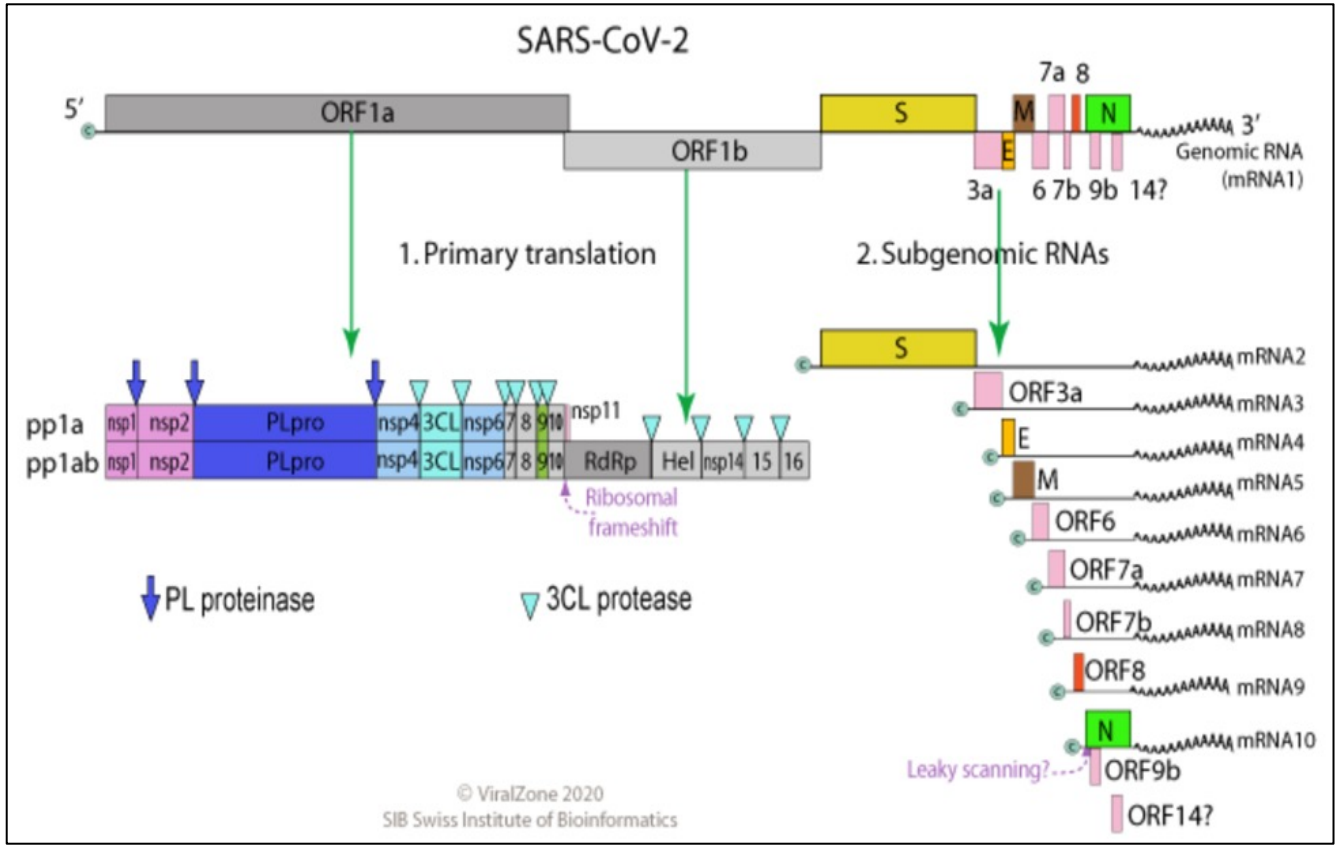
- SARS-CoV Epidemic 2002
  >170 SARS-CoV structures→PDB
- MERS-CoV Epidemic 2012
  >100 MERS-CoV structures→PDB
- COVID-19 Pandemic 2019
  ~4,000 SARS-CoV-2 structures→PDB
- Effective mRNA vaccines designed and Paxlovid discovered/developed using PDB structures of SARS-CoV, MERS-CoV, and SARS-CoV-2 proteins



Respiratory Droplet, 2020; David S. Goodsell

Coronavirus (SARS-CoV-2) Genome Organization

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

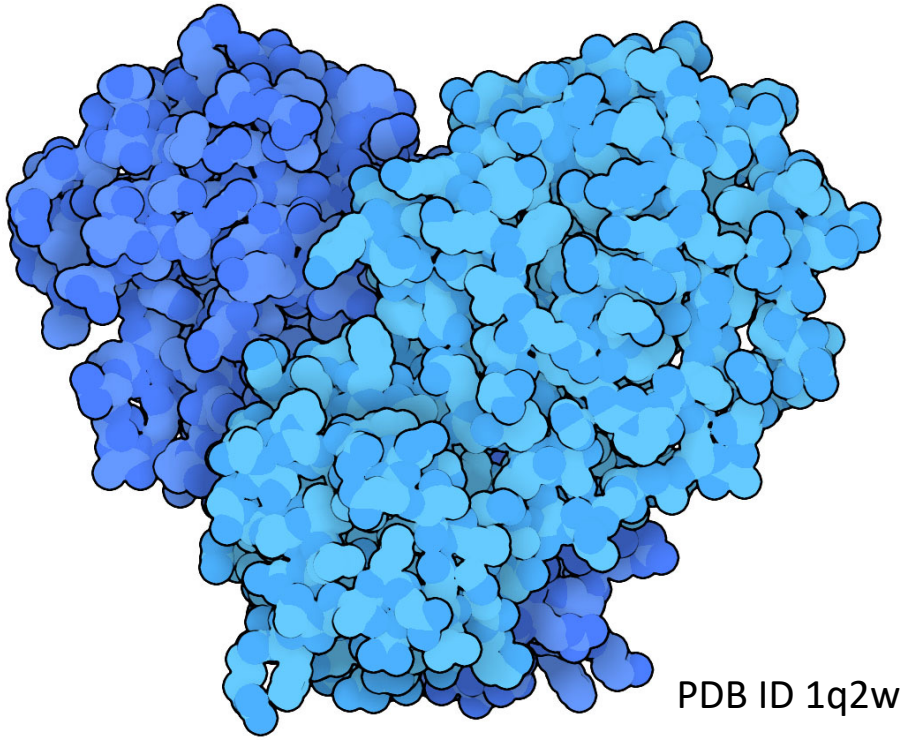


SARS-CoV: PDB Structures I

PDB Structures (>170)

- Main Protease (Mpro or Nsp5)
  - >90 Apo/Co-crystal structures
- Nsp3/PL Proteinase (PLpro)
  - >10 Apo/Co-crystal structures
- Spike Protein
  - >70 3DEM/Crystal structures
  - All Down and 1 Up/2 Down Trimers
  - Post-fusion Trimers
  - Complexes with ACE2 Receptor, *etc*.

Main Protease (Mpro or Nsp5)



PDB ID 1q2w

Drug Discovery Target
Symmetric Homodimer; Two Active Sites

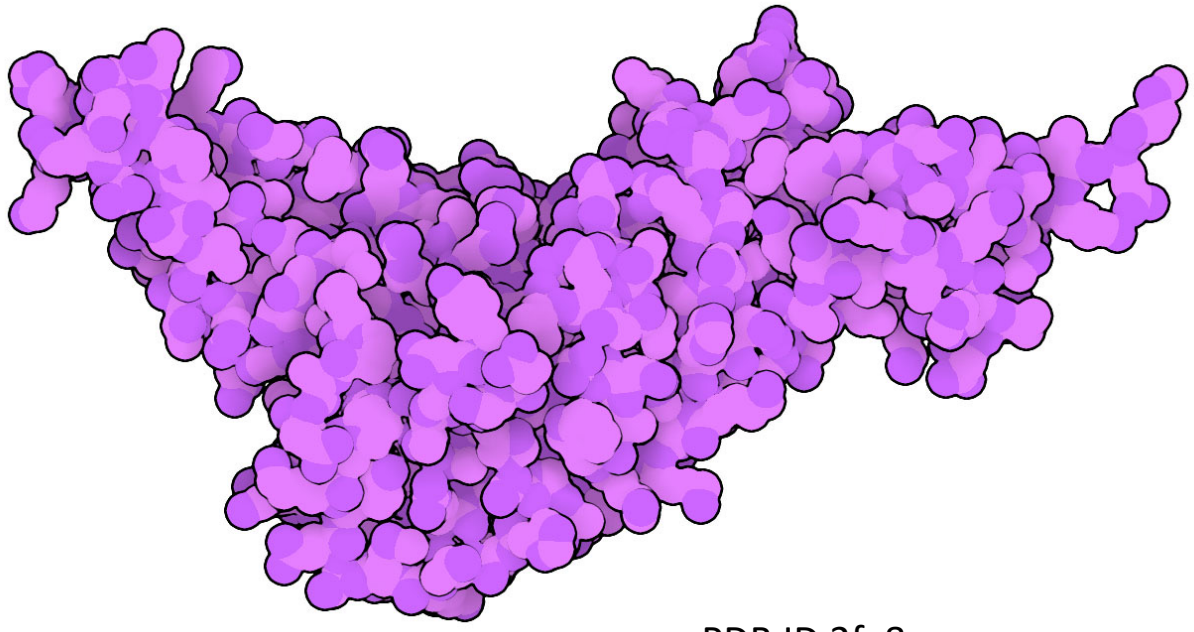
New York Times 30 July 2003, C2.

SARS-CoV: PDB Structures II

PDB Structures (>170)

- Main Protease (Mpro or Nsp5)
  - >90 Apo/Co-crystal structures
- Nsp3/PL Proteinase (PLpro)
  - >10 Apo/Co-crystal structures
- Spike Protein
  - >70 3DEM/Crystal structures
  - All Down and 1 Up/2 Down Trimers
  - Post-fusion Trimers
  - Complexes with ACE2 Receptor, *etc*.

Nsp3 – Papain-Like Proteinase (PLpro)



PDB ID 2fe8

Drug Discovery Target
Monomer; One Active Site

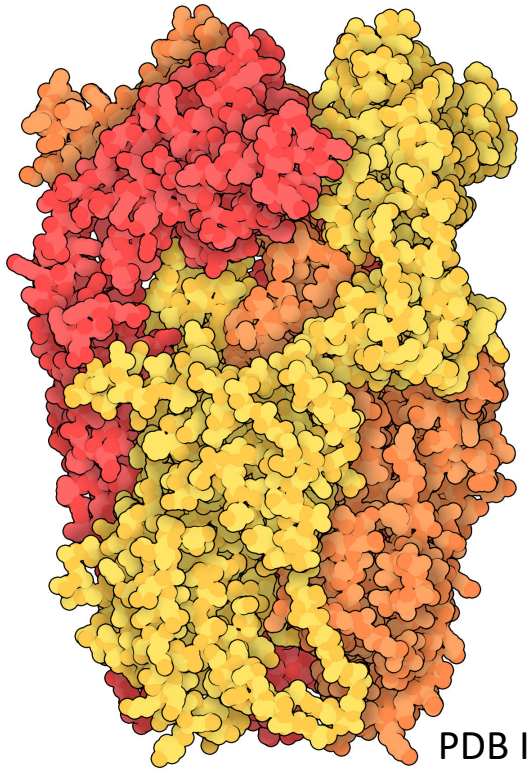
Ratia *et al*. (2006) *Proceedings of the National Academy of Sciences 103*, 5712-5722.

SARS-CoV: PDB Structures III

PDB Structures (>170)

- Main Protease (Mpro or Nsp5)
  - >90 Apo/Co-crystal structures
- Nsp3/PL Proteinase (PLpro)
  - >10 Apo/Co-crystal structures
- Spike Protein
  - >70 3DEM/Crystal structures
  - All Down and 1 Up/2 Down Trimers
  - Post-fusion Trimers
  - Complexes with ACE2 Receptor, *etc*.

Spike Protein



PDB ID 5wrg

Vaccine/Antibody Discovery Target
Homotrimer

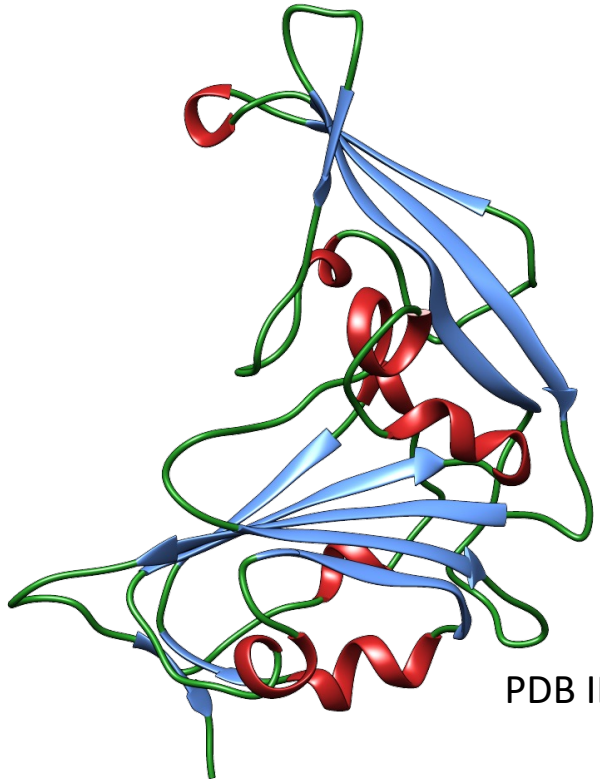
Gui *et al*. (2017) *Cell Research 27*, 119-129.

MERS-CoV: PDB Structures

PDB Structures (>120)

- Spike Protein
  - ~70 3DEM/Crystal structures
  - All Down and 1 Up/2 Down Trimers
  - Post-fusion Trimers
  - Complexes with ACE2 Receptors, *etc*.
- Nsp5/Main Protease (Mpro)
  - ~40 Apo/Co-crystal structures
- Nsp3/PL Proteinase (PLpro)
  - >10 Apo/Co-crystal structures

Spike Protein-Receptor Binding Domain



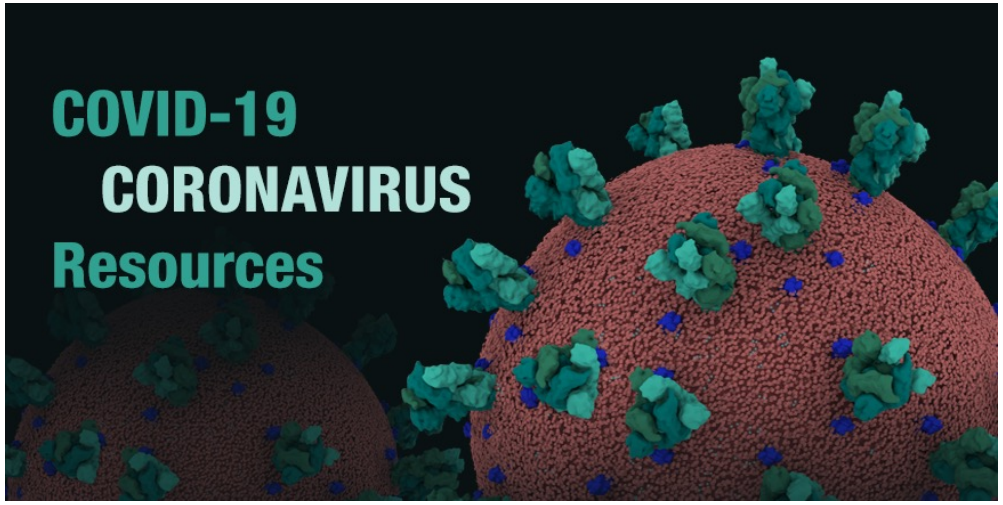
PDB ID 4kqz

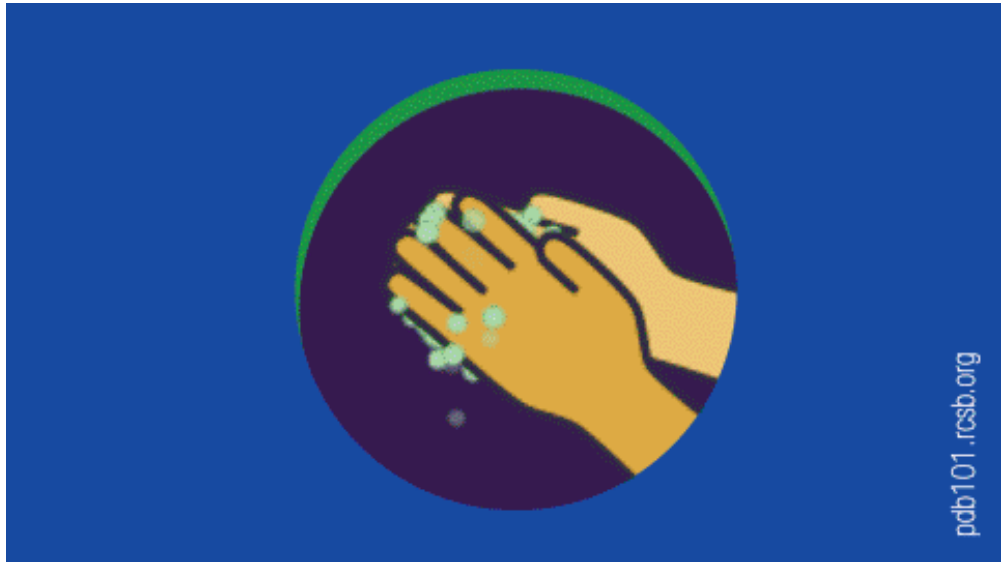
Vaccine/Antibody Design Target

Lu *et al*. (2013) *Nature 500*, 227-231.

RCSB PDB Response to COVID-19

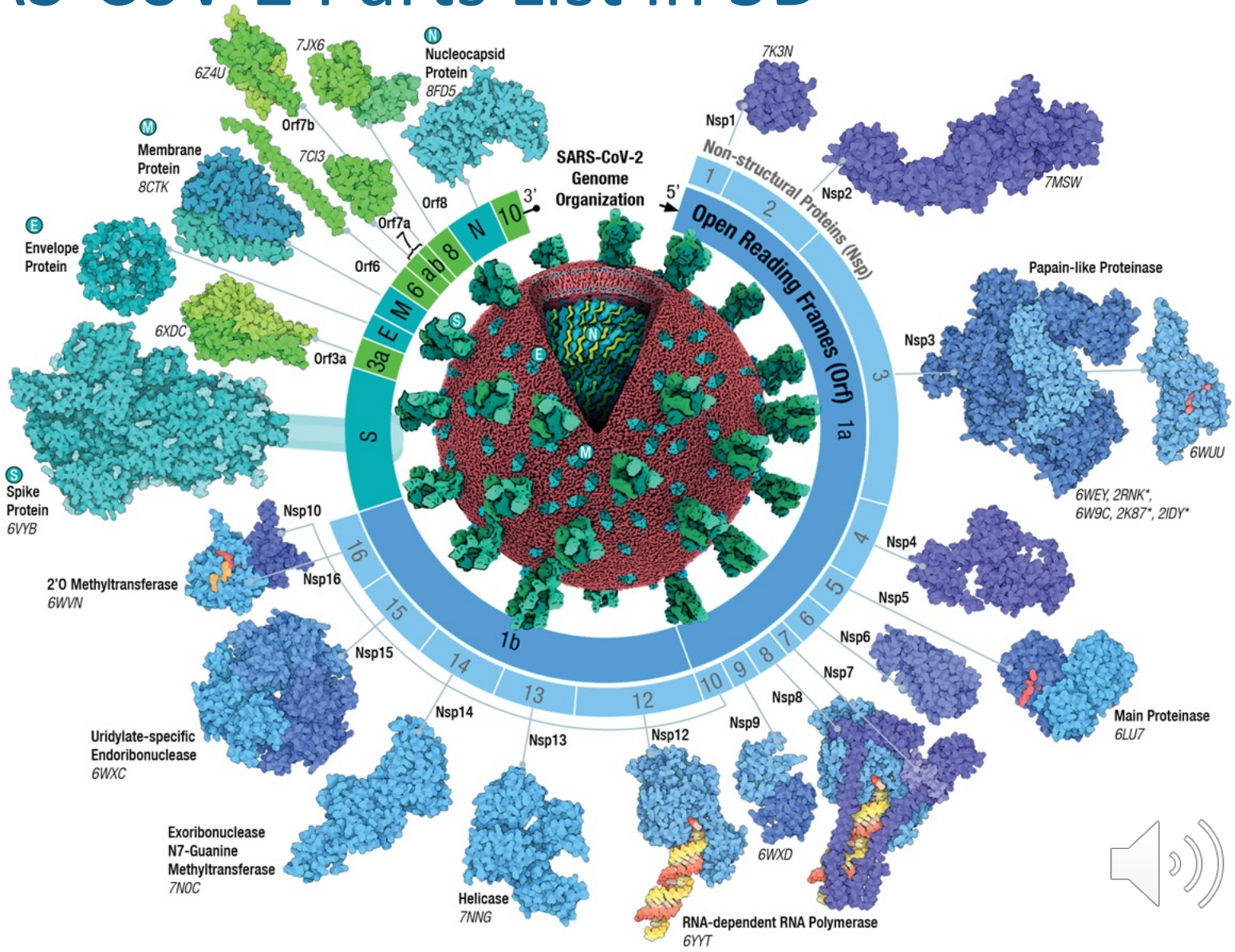
- Deposition, Validation, and Biocuration of COVID-19 structures prioritized, including post-release revisions (*e.g.*, citation updates)
- PDB depositors strongly encouraged to release COVID-19 structures immediately
- Consistent taxonomy name/ID
  - Severe acute respiratory syndrome coronavirus 2;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





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

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



SARS-CoV-2 Proteome Evolution: Study Design

Qualitative Findings

- ~49,000 Viral Proteomes Analyzed
  - All 29 study proteins were mutated
  - Non-synonymous substitutions only
  - No insertions or deletions
- Structural Modeling
  - Structural models publicly available
- Rosetta Energetics
  - Each Unique Sequence Variant (USV) for each study protein was analyzed for change in free energy of stabilization ($\Delta\Delta G^{App}$) Rosetta Energy Units (REUs)

Quantitative Findings

- Analyzed 1,248,712 clean viral protein sequences across 29 study proteins
- 1,094,893 (~88%) proteins unchanged
- Least Conservation: ~58%
- Greatest Conservation: ~98%
- ~1 Unique Sequence Variant (USV)/residue in each study protein (range: ~0.6-2.5)

Burley *et al.* (2020) *Biochemistry and Molecular Biology Education 48*, 511-513.
Lubin *et al*. (2022) *Proteins 90*, 1054-1080.

SARS-CoV-2 Proteome Evolution: Key Findings

- Amino acid changes are relatively rare (due to Nsp14 proofreading)
- Many of the substitutions result from single base changes
- Most substitutions occurred on protein surfaces, where they were predicted to have little or no effect on structure stability (**N.B.: Spike Protein changes contributed to immune evasion!)**
- Minority occurred within hydrophobic cores
- Key enzyme active sites were highly conserved
- Main Protease, Papain-like Proteinase, RNA-dependent RNA Polymerase, *etc*. represent attractive targets for drug discovery

Burley *et al.* (2020) *Biochemistry and Molecular Biology Education 48*, 511-513.
Lubin *et al.* (2022) *Proteins 90*, 1054-1080.

Why a Pandemic? What did it cost the World?

Differences in Virus Behavior:
SARS *versus* MERS *versus* SARS-CoV-2

- SARS-CoV and MERS-CoV:
  Not infectious until patients became seriously ill (*i.e.*, hospitalized)
- Health care workers were at very high risk during both epidemics
- SARS-CoV-2 pandemic driven by pre-symptomatic/asymptomatic transmission around the world facilitated by commercial air travel

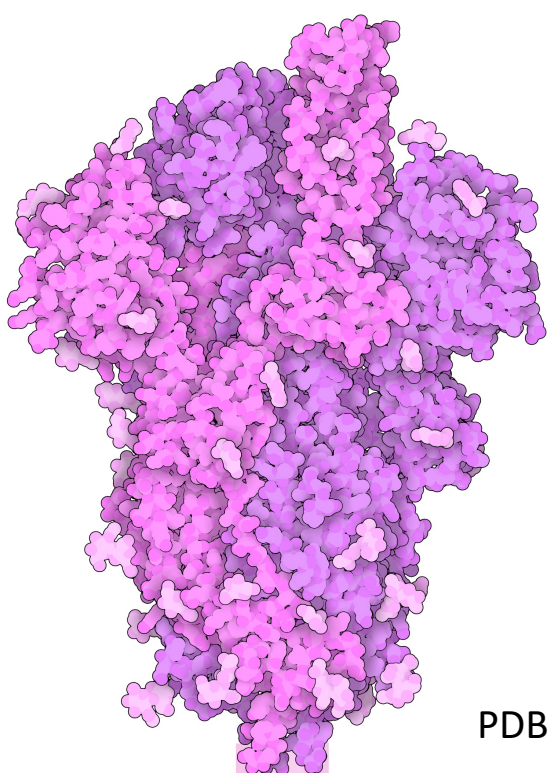
Global Morbidity, Mortality, and Economic Impact Statistics

- Cases: ~770 million (WHO)
- Deaths: ~7 million (WHO)
- Vaccine Doses: ~13.5 billion (WHO)
- Coronavirus Case Fatality Rates (CFRs)
  - SARS-CoV: ~10%
  - MERS-CoV: ~34%
  - SARS-CoV-2: 1-2%
- Economic Losses in 2020:
  ~3.5 trillion US$ (Financial Times)

Financial Times 23 September 2020.

Structure-Based Vaccine Design: Spike Protein

- Spike Protein
  - >1,700 3DEM/Crystal structures
  - All Down and 1 Up/2 Down Trimers
  - Post-fusion Trimers
  - Complexes with ACE2, Fabs, etc.
- mRNA vaccine designs relied on PDB structures of SARS-CoV and MERS-CoV spike proteins
- ~5.5 billion vaccinated!
- Tens of millions of lives saved!



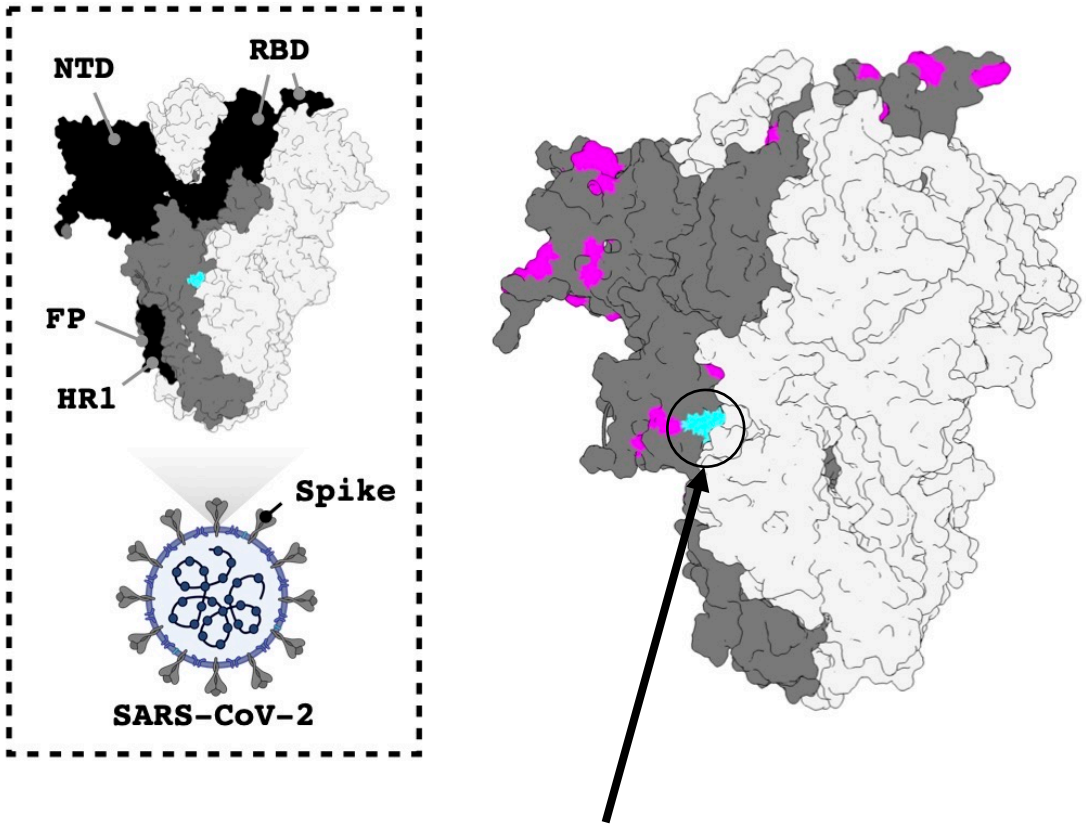
PDB ID 6vsb

Vaccine Discovery Target
Antibody Discovery Target

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

SARS-CoV-2 Spike Protein Universal Epitope

- PDB structures and Computed Structure Models of the Spike Protein were used to identify a universal epitope (C662–C671)
- Amino acid sequence CDIPIGAGIC is fully conserved across all known major variants of SARS-CoV-2 (and SARS-CoV)
- Conformation and antibody accessibility were predicted to be similar across all SARS-CoV-2 spike proteins

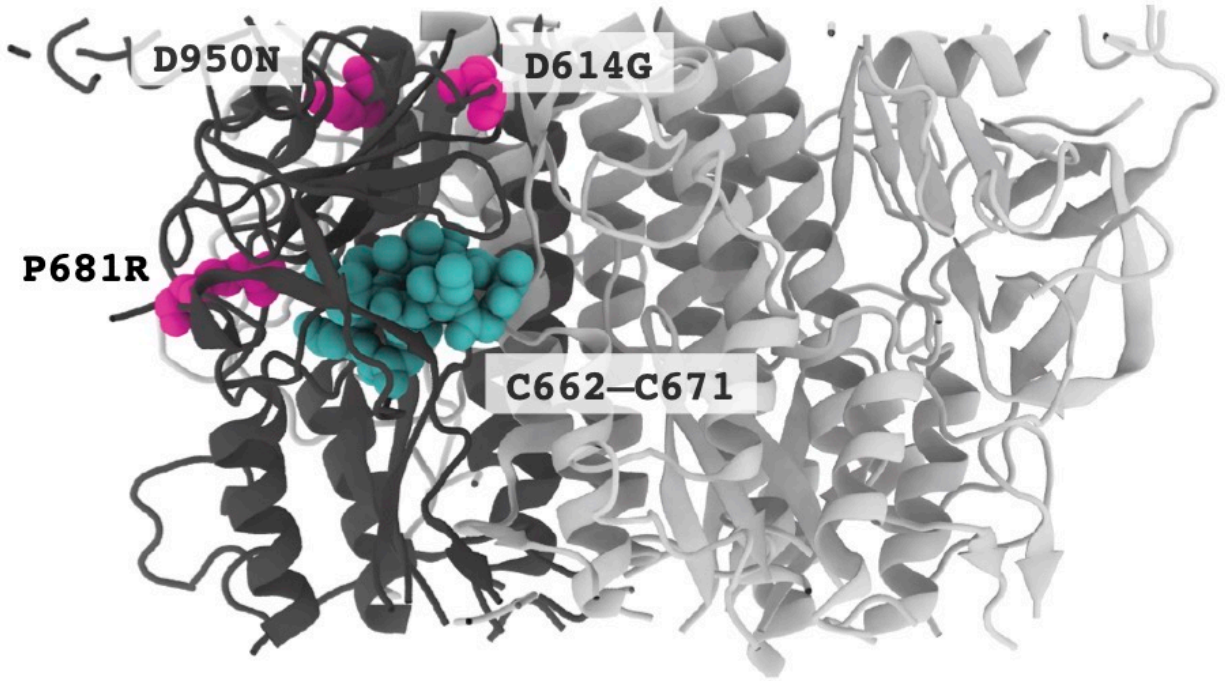


Location of the Universal Epitope

Markosian *et al*. (2022) *Molecular Biology and Evolution 39*, msac091.
Staquicini *et al.* (2021) *Proceedings of the National Academy of Sciences 118*, e2105739118.

SARS-CoV-2 Spike Protein Universal Epitope

- PDB structures and Computed Structure Models of the Spike Protein were used to identify a universal epitope C662–C671
- Amino acid sequence CDIPIGAGIC is fully conserved across all known major variants of SARS-CoV-2 (and SARS-CoV)
- Conformation and antibody accessibility were predicted to be similar across all SARS-CoV-2 spike proteins

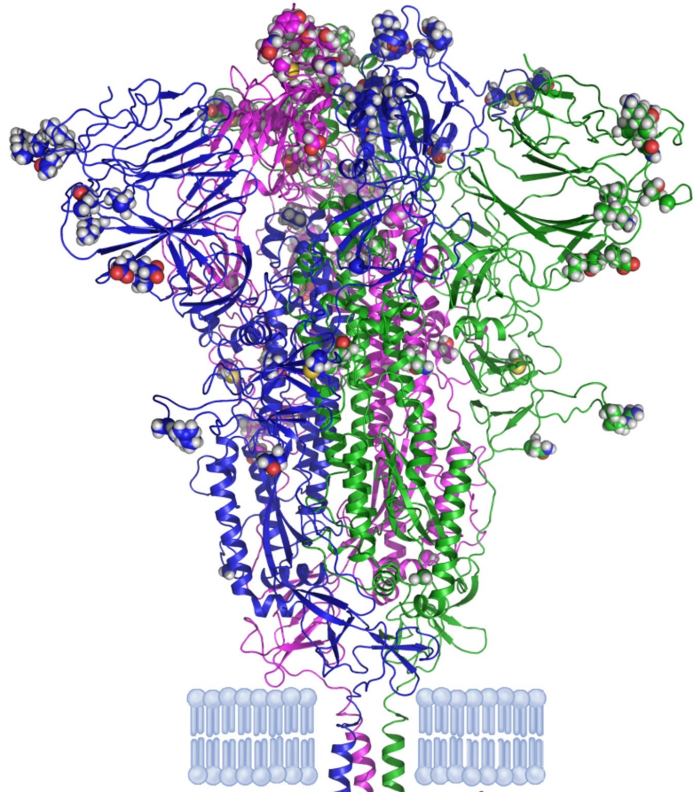


Location of the Universal Epitope

Markosian *et al*. (2022) *Molecular Biology and Evolution 39*, msac091.
Staquicini *et al.* (2021) *Proceedings of the National Academy of Sciences 118*, e2105739118.

SARS-CoV-2: Omicron Variant Spike Protein

- ~1,500 Spike Protein related PDB structures were used to analyze the Omicron Variant of Concern
- Computed Structure Models were used to assess the likely impact of amino acid changes on Omicron Spike Protein binding to the ACE2 cellular receptor and its recognition by Rx antibodies



SARS-CoV-2 Omicron Spike Protein
Computed Structure Model
Amino Acid Changes Shown in CPK

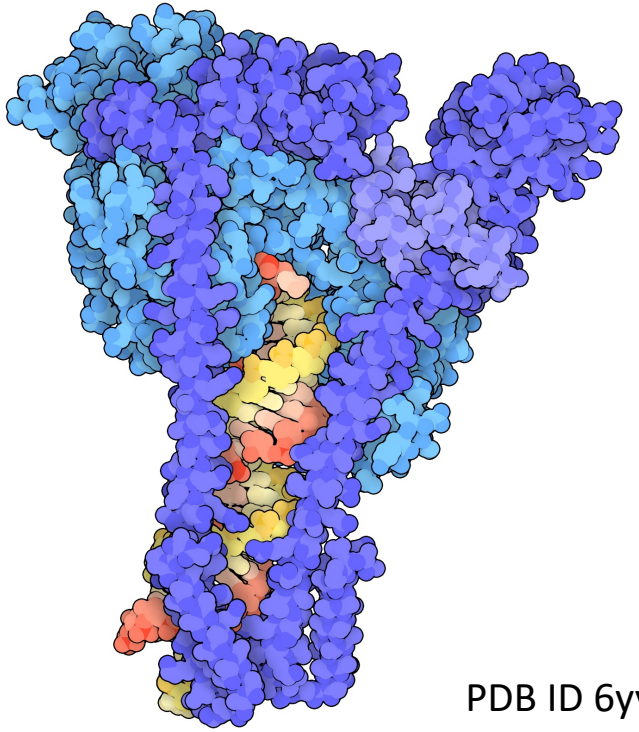
Lubin *et al*. (2023) *Journal of Clinical Investigation Insight 8*, 168296.

Drug Repurposing: Viral RNA Polymerase

PDB Structures

- Nsp7/Nsp8_2/Nsp12 RDRP (RNA-dependent RNA polymerase)
  - >60 3DEM structures
  - Target of Gilead's remdesivir (IV)
  - Target of Merck's molnupiravir (Oral)
- Both drugs were intended to treat other viral infections then repurposed for SARS-CoV-2

Nsp7/8_2/12 – RDRP/dsRNA



PDB ID 6yyt

Validated Drug Target
Heterotetramer; Two Active Sites

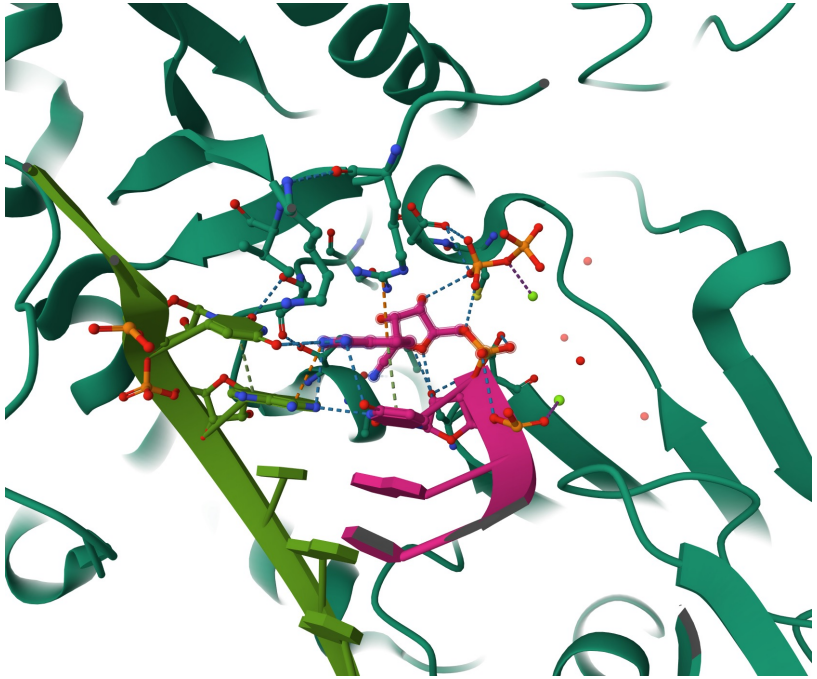
Hillen *et al*. (2020) *Nature 588*, 154-156.

Drug Repurposing: Viral RNA Polymerase

PDB Structures

- Nsp7/Nsp8_2/Nsp12 RDPR (RNA-dependent RNA polymerase)
  - ~60 3DEM structures
  - **Target of Gilead's remdesivir (IV)**
  - Target of Merck's molnupiravir (Oral)
- Both drugs were intended to treat other viral infections then repurposed for SARS-CoV-2

Nsp7/8_2/12 – RDPB/Remdesivir



PDB ID 7bv2

Remdesivir incorporated into the RNA product of the holoenzyme

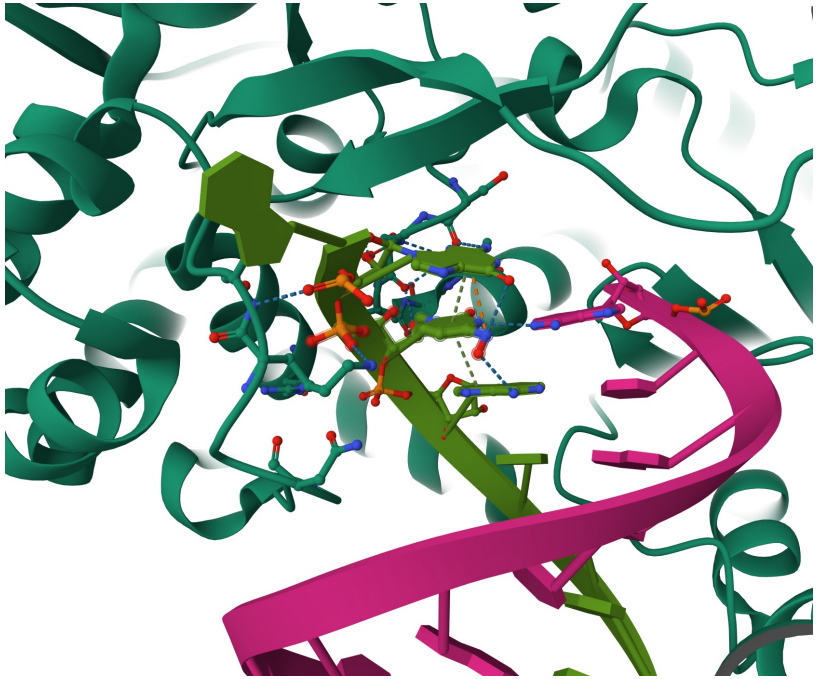
Yin *et al*. (2020) *Science 368*, 1499-1504.

Drug Repurposing: Viral RNA Polymerase

PDB Structures

- Nsp7/Nsp8_2/Nsp12 RDRP (RNA-dependent RNA polymerase)
  - ~60 3DEM structures
  - Target of Gilead's remdesivir (IV)
  - **Target of Merck's molnupiravir (Oral)**
- Both drugs were intended to treat other viral infections then repurposed for SARS-CoV-2

Nsp7/8_2/12 – RDRP/Molnupiravir



PDB ID 7ozu

Molnupiravir incorporated into the RNA product of the holoenzyme

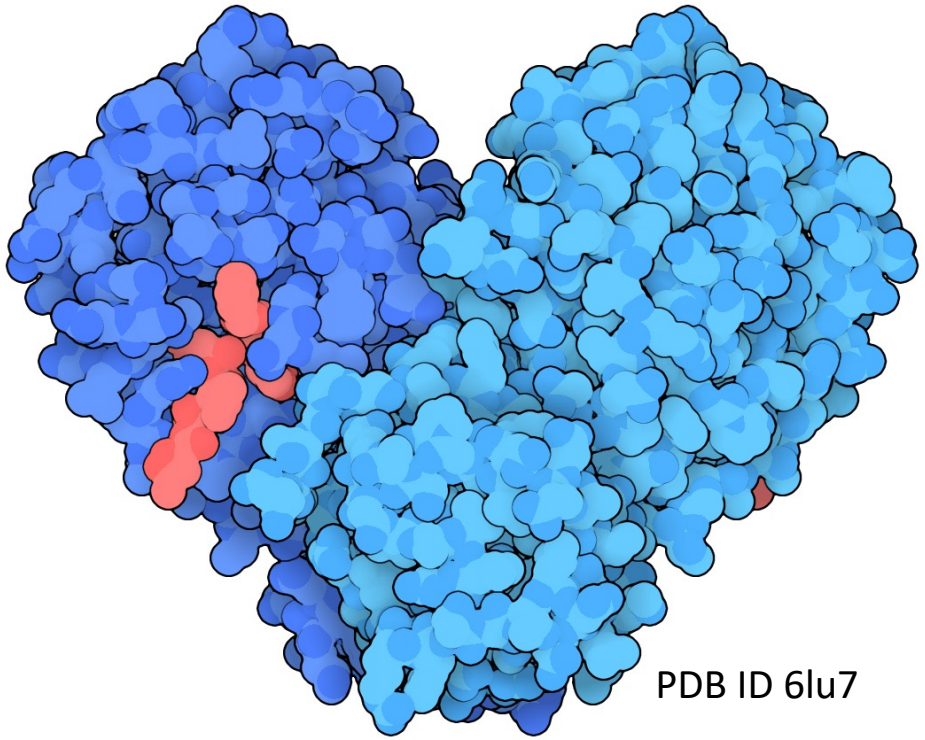
Kabinger *et al*. (2021) *Nature Structural Biology 28*, 740-746.

SARS-CoV-2: PDB Structures of Drug Targets

PDB Structures

- Main Protease (Mpro)
  - >1400 Apo/Co-crystal structures
  - Target of Pfizer's nirmatrelvir (+ritonavir=Paxlovid)
- Paxlovid is the more effective than the repurposed RNA polymerase inhibitors from Merck and Gilead
- N.B.: Paxlovid cannot be prescribed safely to all infected individuals because of drug-drug interactions with ritonavir

Main Protease (Mpro)



PDB ID 6lu7

Drug Discovery Target

Symmetric Homodimer; Two Active Sites

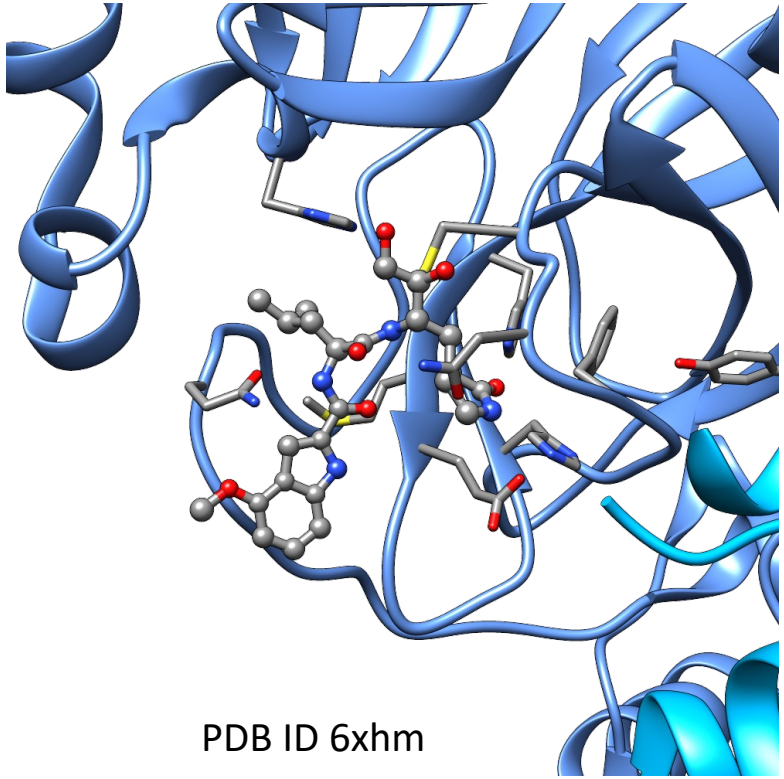
Jin *et al*. (2020) *Nature 582*, 299-293.

Structure-Guided Drug Discovery: PF-00835231

PDB Structures

- Main Protease (Mpro)
  - >1400 Apo/Co-crystal structures
  - Target of Pfizer's nirmatrelvir (+ritonavir=Paxlovid)
- In the 2000s, Pfizer used SGDD to discover PF-00835231 as a potential SARS-CoV antiviral
- Project halted in the mid-2000s (no commercial market!)

PF-00835231 Mechanism of Action



PDB ID 6xhm

PF-00835231 bound to the Mpro Active Site

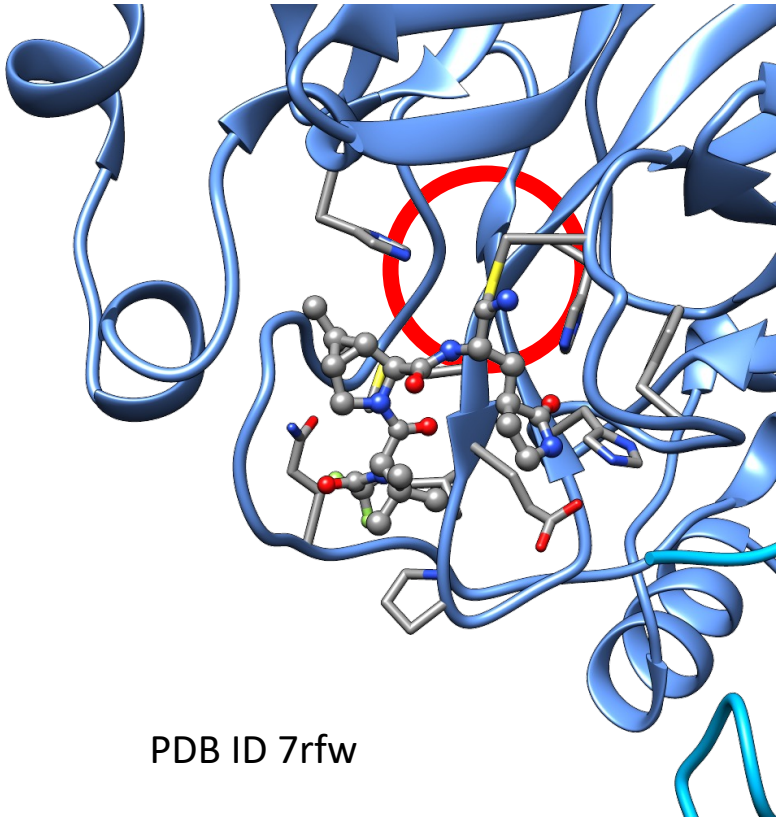
Hoffman *et al*. (2020) *Journal of Medicinal Chemistry 63*, 12725-12747.

Structure-Guided Drug Discovery: Nirmatrelvir

PDB Structures

- Main Protease (Mpro)
  - >1400 Apo/Co-crystal structures
  - Target of Pfizer's nirmatrelvir (+ritonavir=Paxlovid)
- In 2020, Pfizer used SGDD to turn PF-00835231 into nirmatrelvir as a SARS-CoV-2 antiviral
- N.B.: Also active against SARS-CoV and MERS-CoV Main Proteases

Nirmatrelvir Mechanism of Action



PDB ID 7rfw

Nirmatrelvir bound to the Mpro Active Site

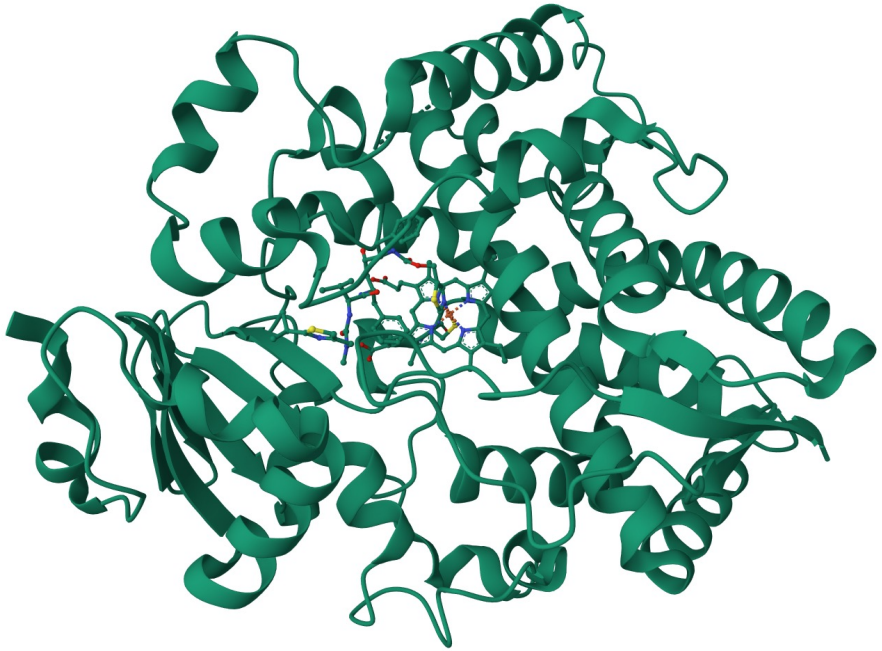
Owen *et al*. (2021) *Science 374*, 1586-1593.

Mechanism of Action: Ritonavir

PDB Structures

- Cytochrome P450 isozyme Cyp 3A4
  - ~100 Apo/Co-crystal structures
  - Target of Ritonavir
- Ritonavir was discovered and developed by Abbott (now AbbVie) as an HIV protease inhibitor
- Co-administered with other anti-viral agents for Hepatitis-C, SARS-CoV-2

Ritonavir Mechanism of Action



PDB ID 5vc0

Ritonavir bound to the Cyp 3A4 Active Site

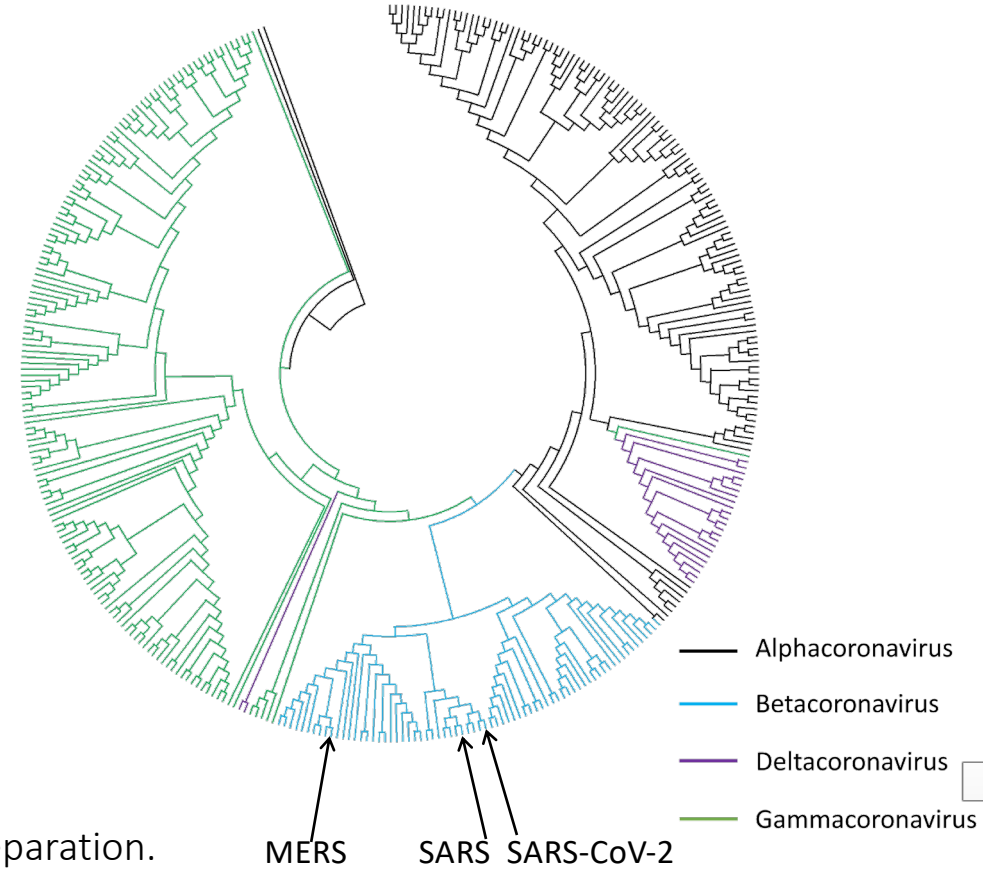
Sevrioukova (2017) *Biochemistry 56*, 3058-3067.

Understanding Coronavirus Mpro Evolution

Mpro Conservation

- Mpro highly conserved across all known coronaviruses
- Mpro active sites are even more highly conserved
- Mpro polyprotein cut sites are also highly conserved
- Double selection pressure!

Mpro Full-Length Amino Acid Sequences



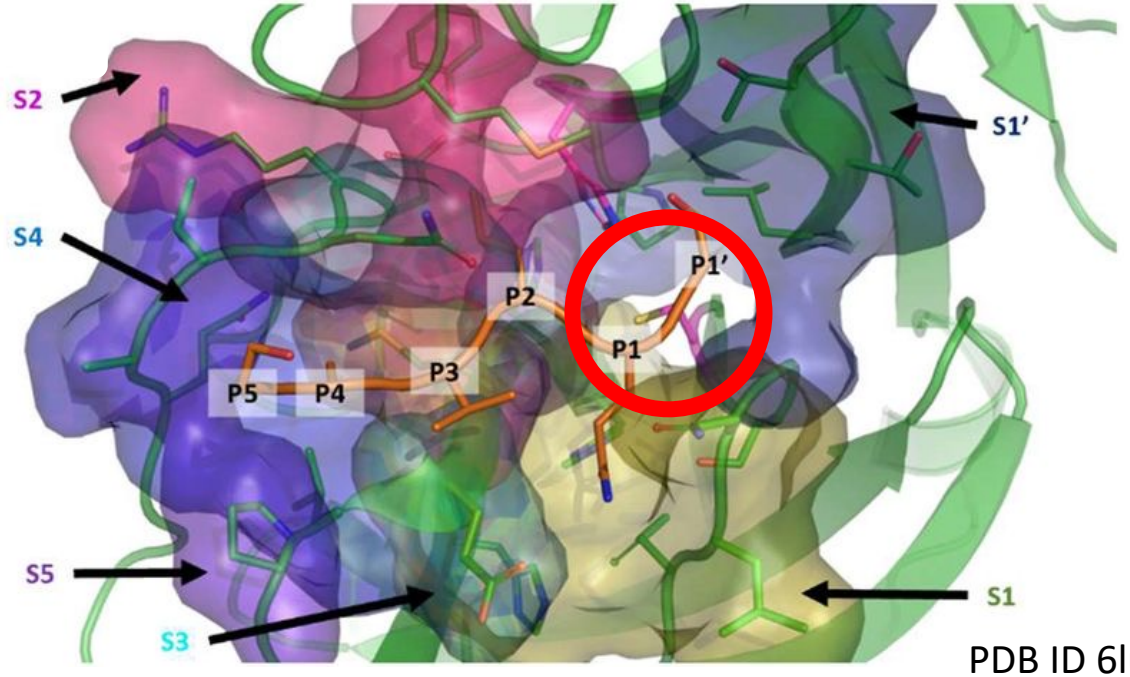
Lubin *et al*. (2024) In Preparation.

Understanding Coronavirus Mpro Evolution

Mpro Conservation

- Mpro highly conserved across all known coronaviruses
- Mpro active sites are even more highly conserved
- Mpro polyprotein cut sites are also highly conserved
- Double selection pressure!

Mpro Active Site/Polyprotein Cut Site



PDB ID 6lu7

Active Site encompasses 6 subsites S1'-S5

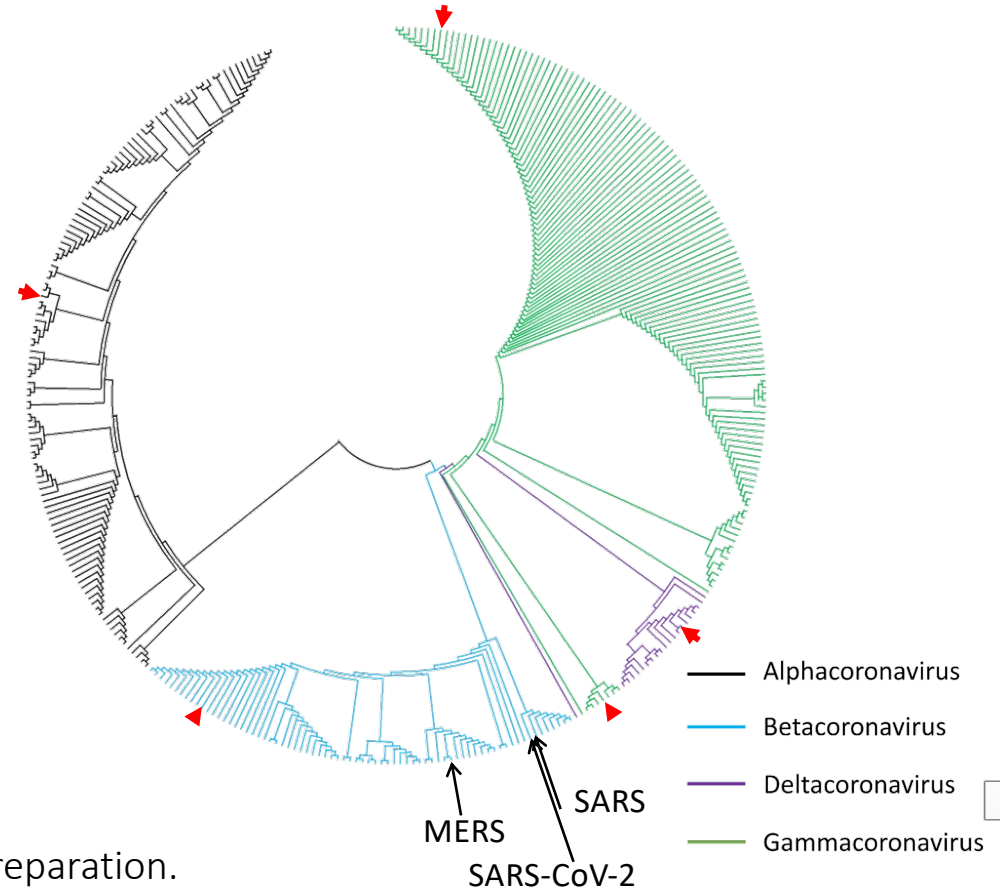
Lubin *et al*. (2024) In Preparation.

Understanding Coronavirus Mpro Evolution

Mpro Conservation

- Mpro highly conserved across all known coronaviruses
- Mpro active sites are even more highly conserved
- Mpro polyprotein cut sites are also highly conserved
- Double selection pressure!

Mpro Active Site Residues



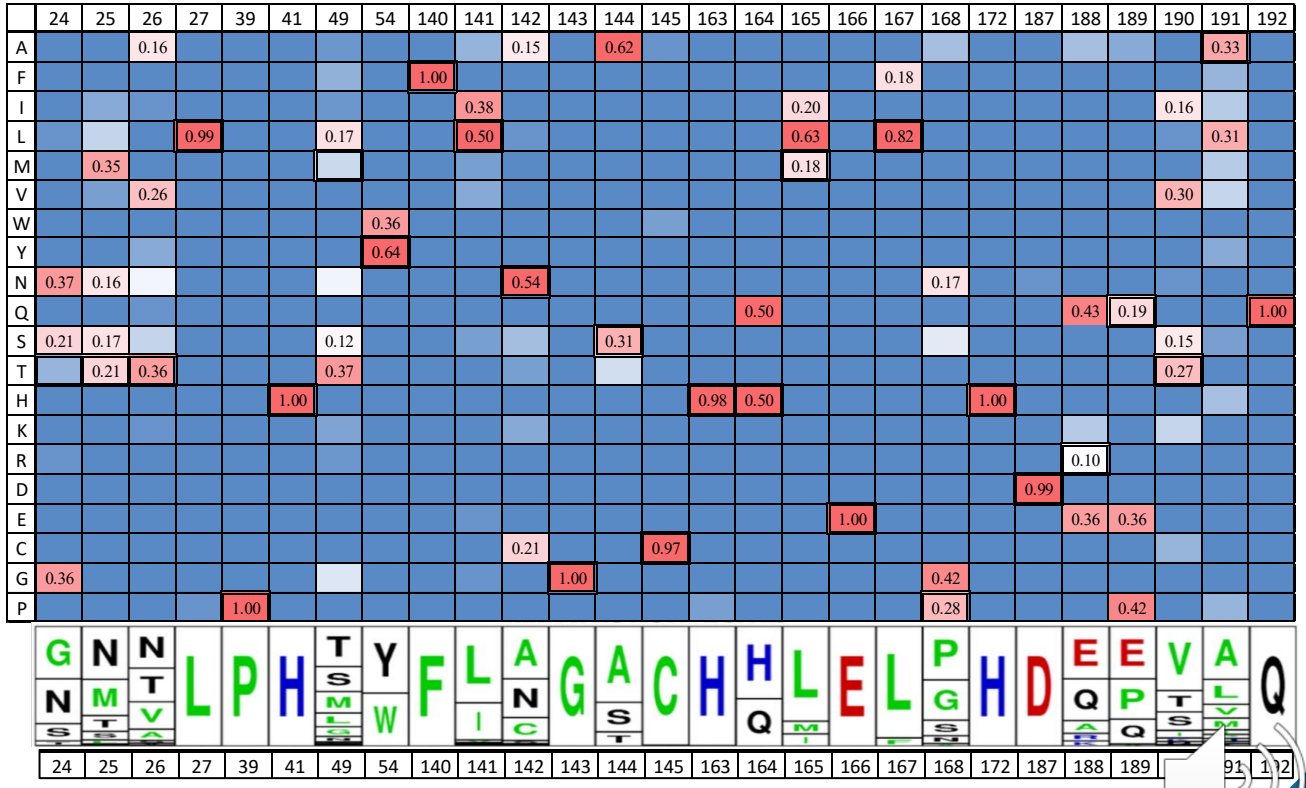
Lubin *et al*. (2024) In Preparation.

Understanding Coronavirus Mpro Evolution

Mpro Conservation

- Mpro highly conserved across all known coronaviruses
- Mpro active sites are even more highly conserved
- Mpro polyprotein cut sites are also highly conserved
- Double selection pressure!

Mpro Active Site Residues



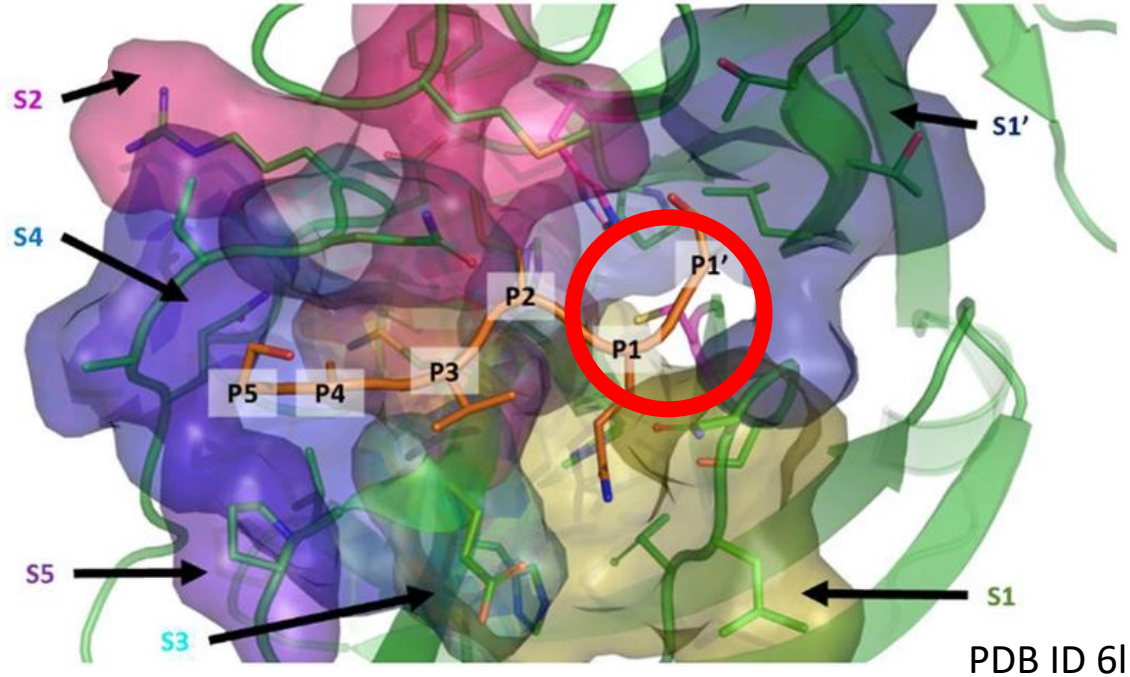
Lubin *et al*. (2024) In Preparation.

Understanding Coronavirus Mpro Evolution

Mpro Conservation

- Mpro highly conserved across all known coronaviruses
- Mpro active sites are even more highly conserved
- Mpro polyprotein cut sites are also highly conserved
- Double selection pressure!

Mpro Active Site/Polyprotein Cut Site



PDB ID 6lu7

Polyprotein cut sit spans six residues P1'-P5

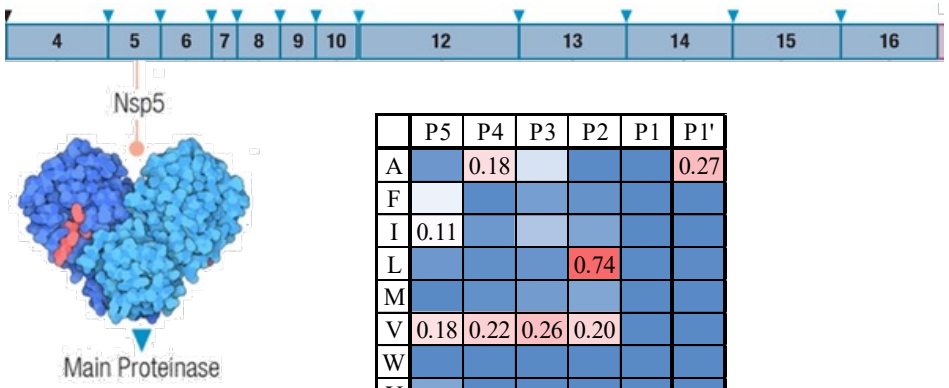
Lubin *et al*. (2024) In Preparation.

Understanding Coronavirus Mpro Evolution

Mpro Conservation

- Mpro highly conserved across all known coronaviruses
- Mpro active sites are even more highly conserved
- Mpro polyprotein cut sites are also highly conserved
- Double selection pressure!

Mpro Polyprotein Cut Sites



	P5	P4	P3	P2	P1	P1'
A		0.18				0.27
F						
I	0.11					
L				0.74		
M						
V	0.18	0.22	0.26	0.20		
W						
Y						
N						0.10
Q					0.99	
S	0.16	0.26				0.59
T	0.13	0.22	0.18			
H						
K						
R			0.17			
D						
E						
C						
G						
P						

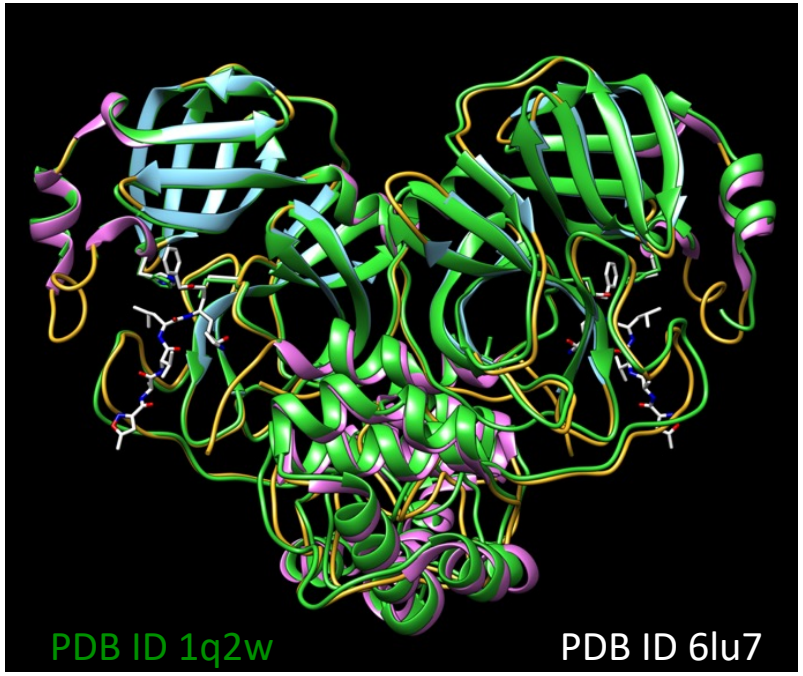
Lubin *et al*. (2024) In Preparation.

Failure of the Free Market: Sobering Postscript

Mpro Conservation

- Mpro highly conserved
- SARS-CoV-2 and SARS-CoV Mpro structurally similar (Active Site R.M.S.D.<0.5Å)
- **If Pfizer had been incentivized to continue in the mid-2000s we would already have had a drug to combat SARS-CoV-2!**

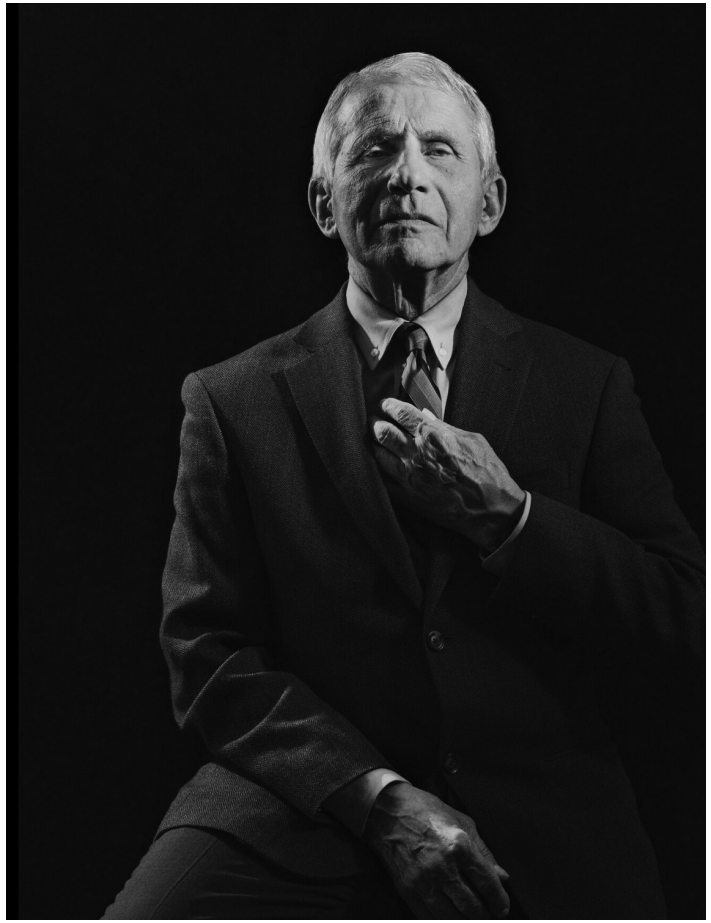
SARS-CoV-2 and SARS-CoV Mpro



Investment of US$250M in a SARS-CoV Mpro inhibitor could have saved tens of millions of lives and trillions of US$ in economic losses

Burley (2020) *Nature 580*, 167.

No Wonder 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

RCSB PDB Team

RCSB PDB PROTEIN DATA BANK

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

Management

US Department of Energy (DE-SC0019749)

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

UC San Diego

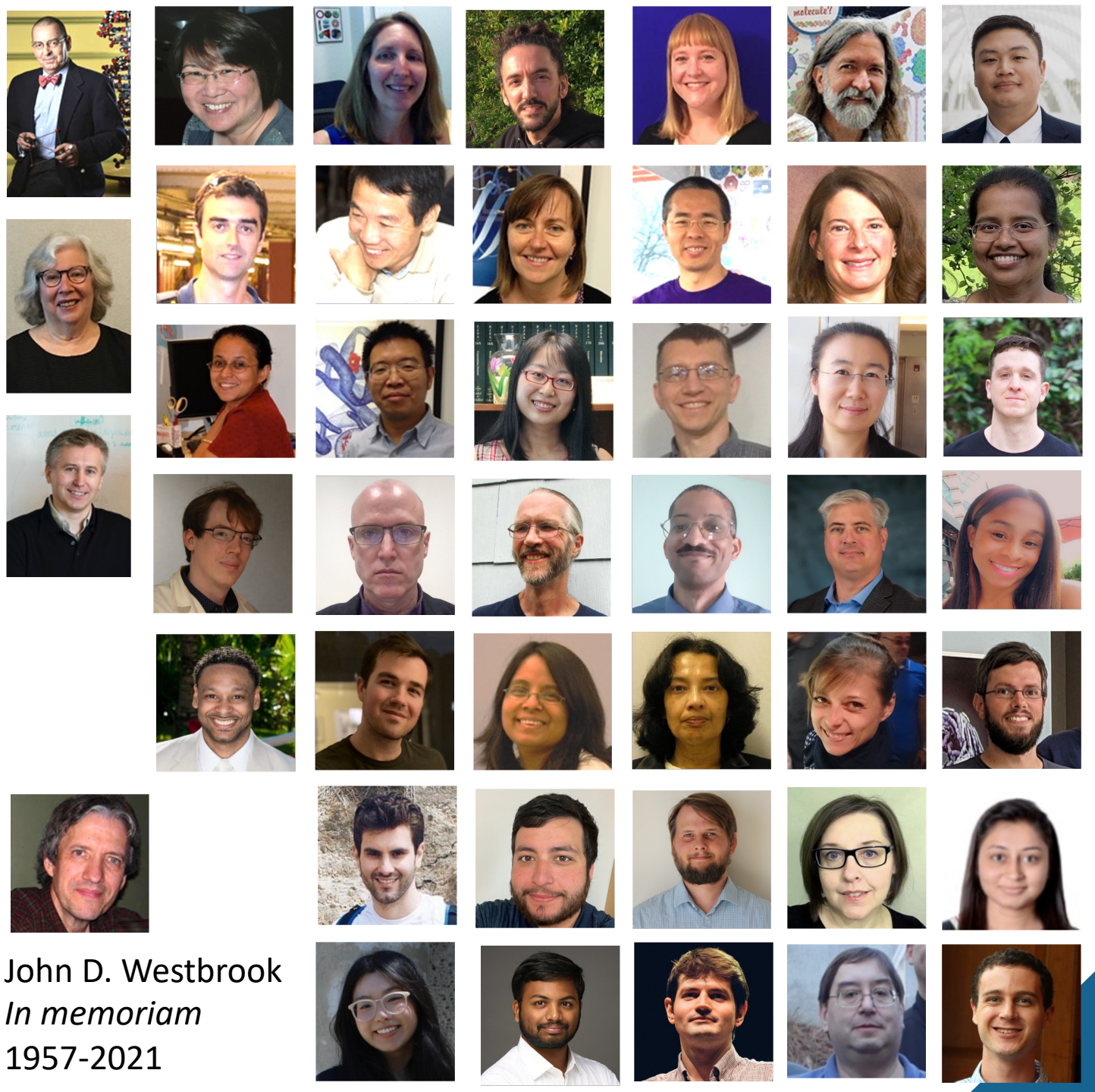
SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF University of California San Francisco

WORLDWIDE PDB PROTEIN DATA BANK

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

Follow us



John D. Westbrook
In memoriam
1957-2021