

# Ligand-centered assessment of SARS-CoV-2 drug target models

A. Wlodawer<sup>1</sup>, Z. Dauter<sup>2</sup>, I. Shabalin<sup>3,4</sup>, M. Gilski<sup>5,6</sup>, D. Brzezinski<sup>3,6,7</sup>, M. Kowiel<sup>6</sup>,  
W. Minor<sup>3,4</sup>, B. Rupp<sup>8,9</sup>, M. Jaskolski<sup>5,6</sup>

<sup>1</sup>*Protein Structure Section, Macromolecular Crystallography Laboratory, NCI, Frederick, MD, USA*

<sup>2</sup>*Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, NCI, Argonne National Laboratory, IL, USA*

<sup>3</sup>*Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, USA*

<sup>4</sup>*Center for Structural Genomics of Infectious Diseases (CSGID), Charlottesville, VA, USA*

<sup>5</sup>*Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland*

<sup>6</sup>*Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland*

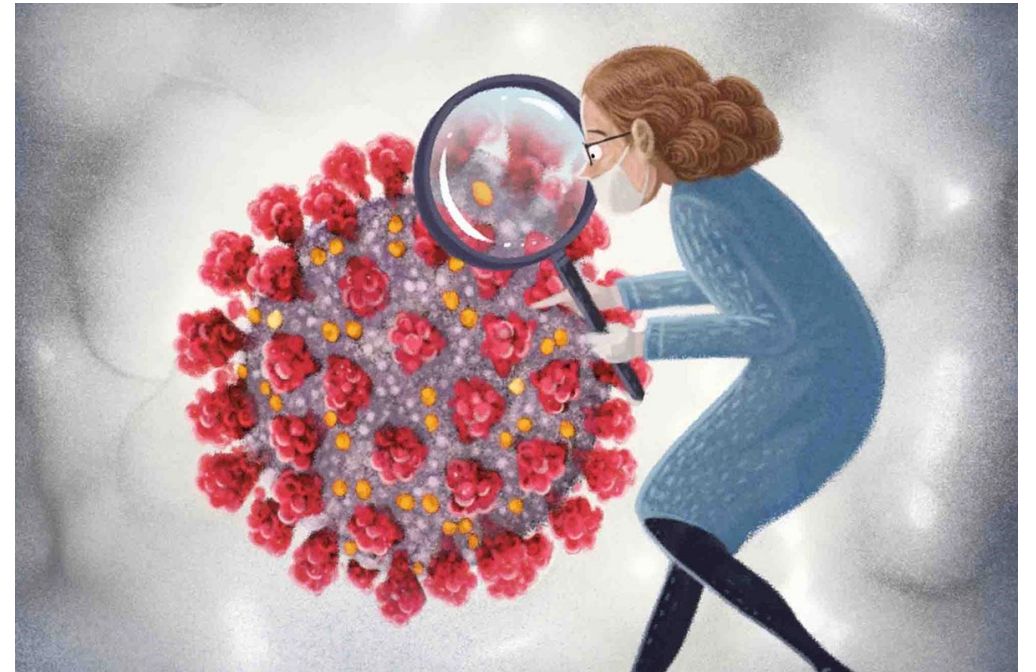
<sup>7</sup>*Institute of Computing Science, Poznan University of Technology, Poznan, Poland*

<sup>8</sup>*k.-k. Hofkristallamt, San Diego, CA, USA*

<sup>9</sup>*Institute of Genetic Epidemiology, Medical University Innsbruck, Austria*

# Outline

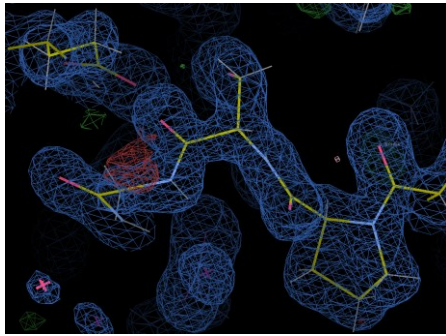
1. Atomic structure determination and drug discovery
2. SARS-CoV-2 drug targets
3. Assessment protocol of SARS-CoV-2 structure models
4. Examples of detected problems
5. Future plans



*Illustration by Marcin Minor*

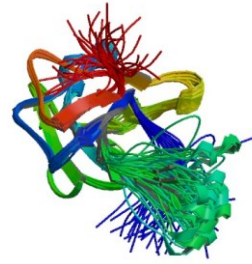
# Atomic structure determination

- The goal of structure determination is to experimentally reveal the 3D atomic architecture of a chemical compound (e.g. a protein)
- Main methods used for this purpose:



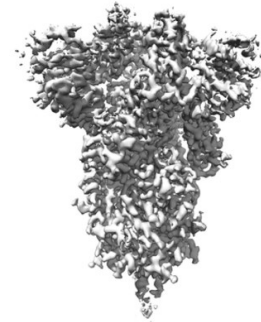
**X-ray crystallography**

(**6WNP**: SARS-CoV-2 Main Protease)



**NMR spectroscopy**

(**6YI3**: SARS-CoV-2 RBD)



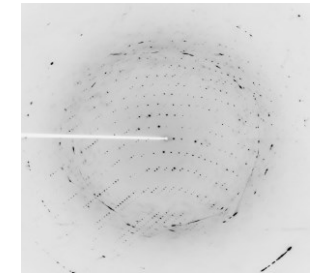
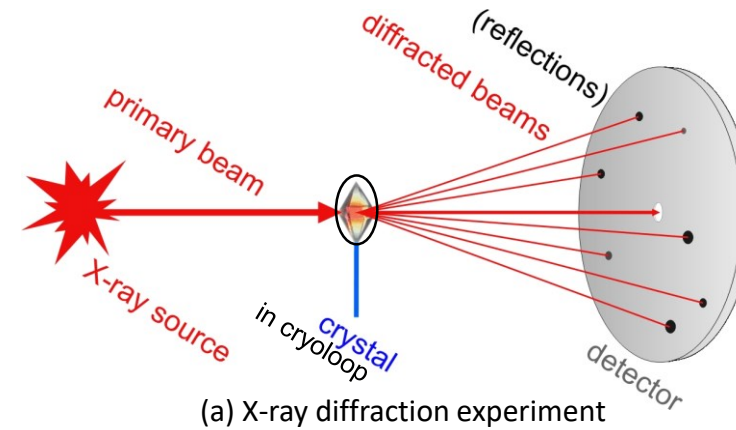
**Cryo electron microscopy**

(**6X29**: SARS-CoV-2 Spike)

- The resulting 3D models are made publicly available through databases
- Protein structures are deposited in the Protein Data Bank (PDB)

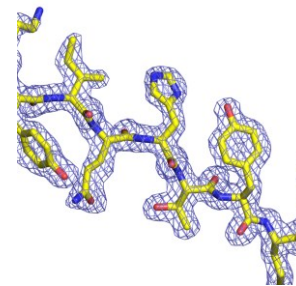
# X-ray crystallography

- Most popular structure determination method (89% PDB, 90% SARS-CoV-2)
- Offers highest resolution
- Best choice for drug design and fragment screening
- Like each structure determination method, requires a degree of human interpretation



(b) Diffraction image

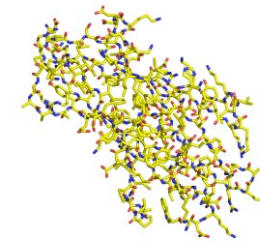
FT  
→  
phase problem



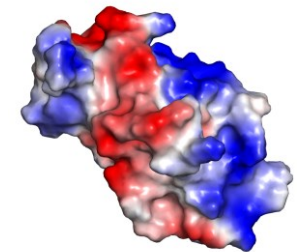
(c) Electron density



(d) Cartoon model



(e) All-atom model



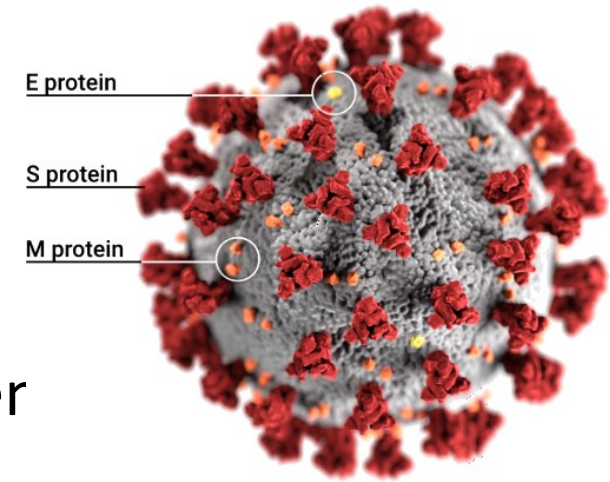
(f) Surface

(g) deposit in PDB

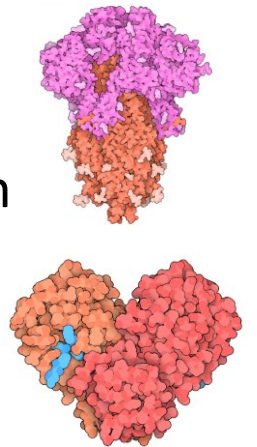
# Structure-based drug design

- Knowledge of the atomic structure of biological macromolecules is necessary to understand the mechanisms of life processes
- In the case of viruses, such knowledge is the basis for the design of drugs (*bullets*) that target certain parts of the virus and block their function
- Usually this requires:
  - finding a suitable binding site (*pocket*) in one of the virus's proteins
  - designing a small-molecule with tight & specific binding in that site
- With iteration cycles, this is the most rational way to develop efficient drugs targeting specific diseases
- HIV treatments have been designed this way

# Drug targets for SARS-CoV-2



- SARS-CoV-2 consists of ~30 proteins and encapsulated RNA genome that codes those proteins
- The proteins can be classified as:
  - **Structural proteins:** M, E, S, N
  - Non-structural proteins (NSP): mainly **enzymes** (biocatalysts) and regulatory proteins
- The main proteins that can be used for drug design:
  - **Spike protein (S):** structural protein that recognizes the ACE2 receptor on human cell; if this protein (or ACE2) is blocked by a drug, the virus will not be able to enter the host cell
  - **Main protease (Mpro):** an enzyme whose function is to cut the viral polyproteins produced in the infected cell to their active form; if this enzyme is blocked by a drug, the virus will not be able to mature and will be non-infectious



# Project goal

Critically evaluate the experimentally determined SARS-CoV-2 protein structures, with special focus on potential drug targets

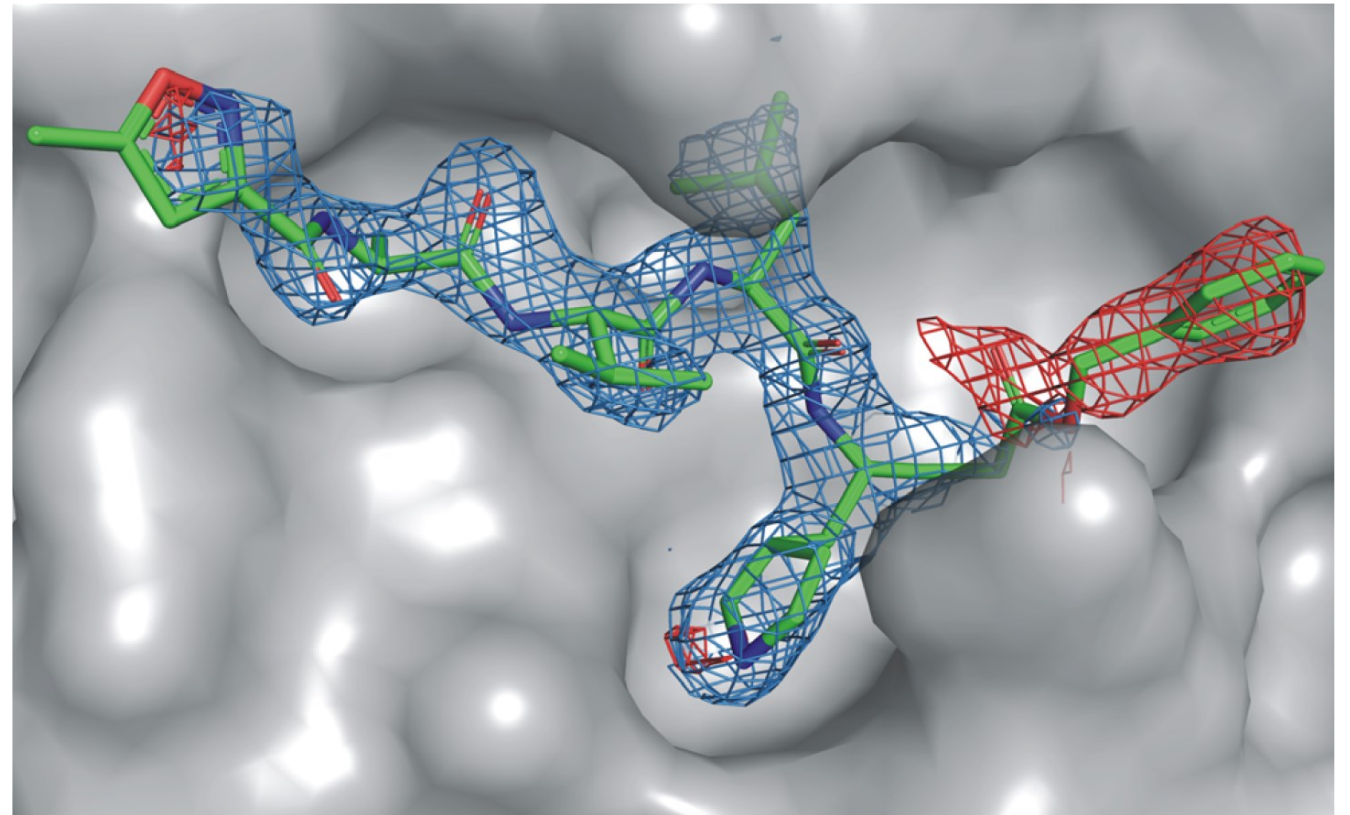
# Proposed assessment protocol

- Extract data from the PDB
- Look for raw diffraction data (IRRC or Zenodo)
- Run validation tools:
  - MolProbity (geometry checking, assessment of the entire model)
  - Twilight (real space correlation coefficient, assessment of ligands)
- Pass data to expert structural biologists
- Determine protein type and ligand status
- If needed, re-refine the structure
  - Run ACHESYM (standardization of model placement in the unit cell)
  - [If interesting case] Prepare Molstack visualization for comparison



# Example problems – incorrect ligand model

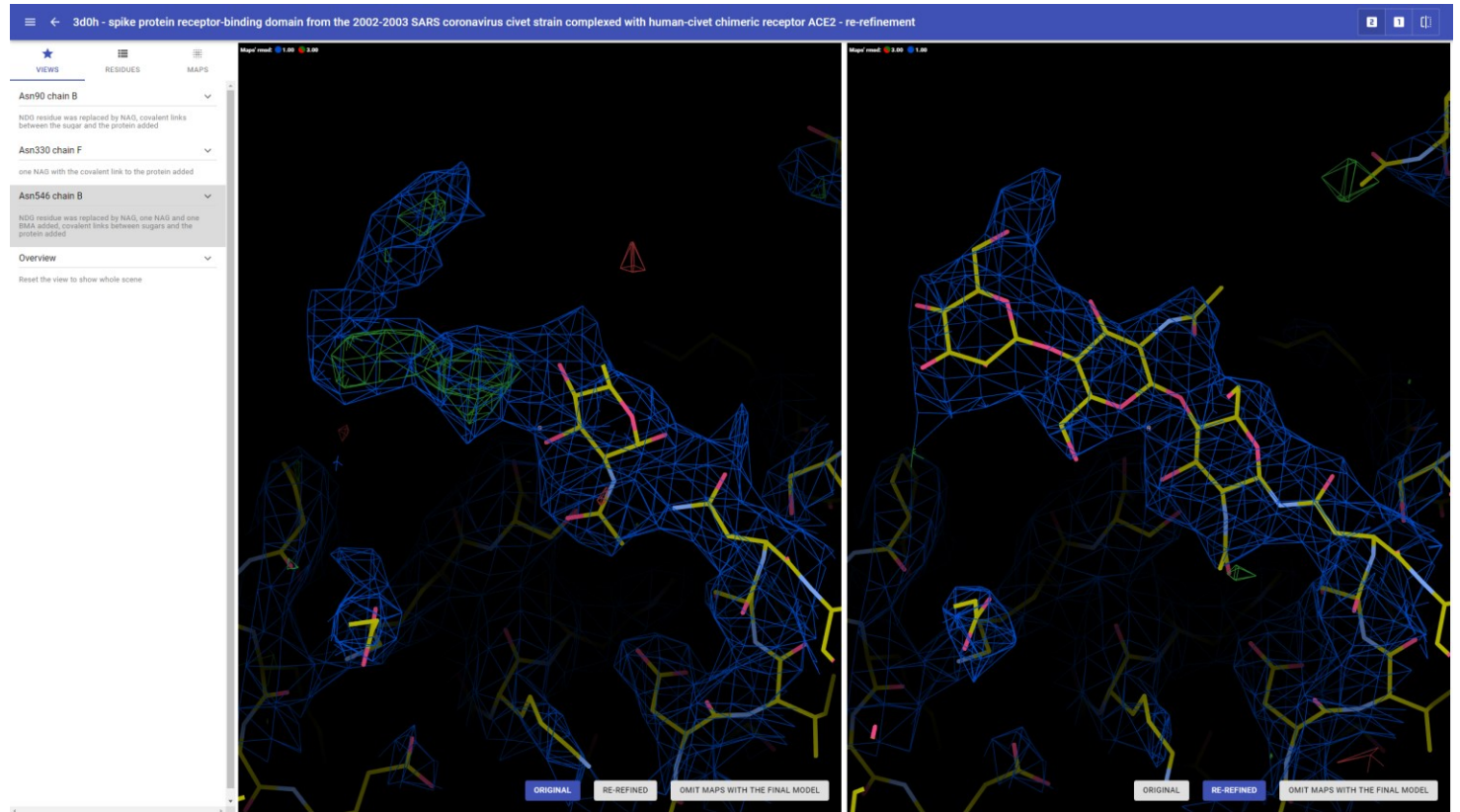
- Peptidic inhibitor in the substrate-binding site of the structure with PDB ID **6LU7**
- The presence of negative difference electron density (red contour) for the terminal benzyl group indicates that this group has been eliminated by hydrolysis and is not there



*Incorrectly modeled inhibitor molecule in the protein binding site*

# Example problems – missing chain fragment

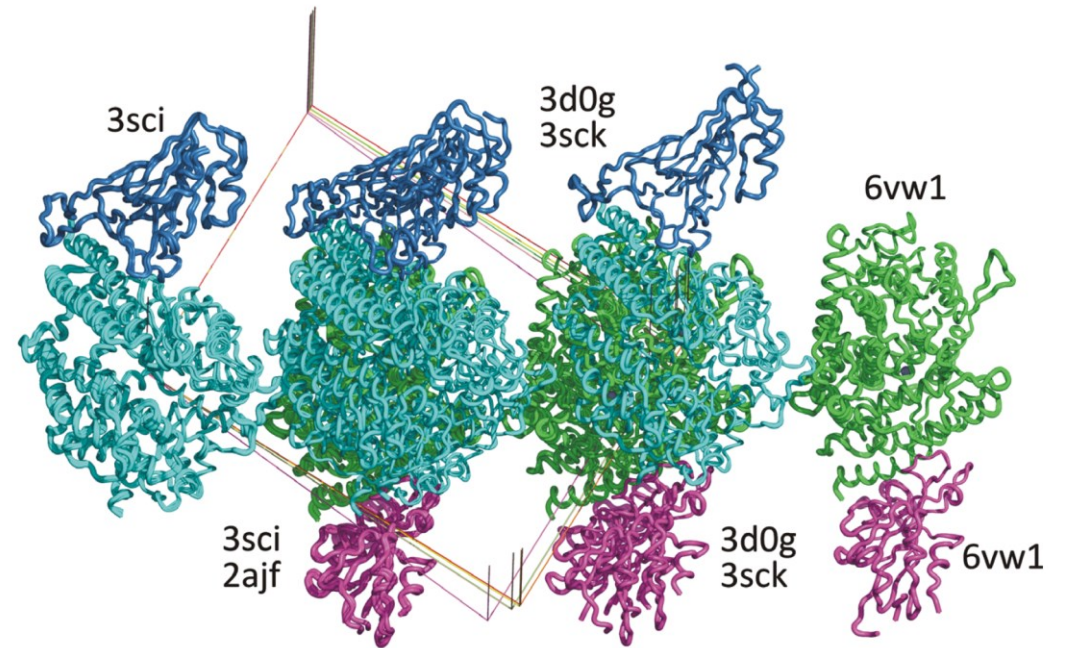
- Structure **3D0H**
- Three chemically linked carbohydrate molecules (NAG-NAG-BMA) should be connected to residue Asn546B
- Left panel shows the original (wrong) model
- Right panel after corrections



<https://molstack.bioreproducibility.org/project/view/Wr12XsIE978LiF95PQYo/>

# Example problems – unit cell placement

- Structures of the same protein although crystallized isomorphously are often presented inconsistently
- This means that different versions of the same protein are hard to compare
- To alleviate this issue we used our ACHESYM server in each re-refinement to unify model placement in the unit cell



*Protein structures after placing in isomorphous unit cells*



# Web resource

## covid-19.bioreproducibility.org

- Aggregates all the mined SARS-CoV-2 data
- Provides info about original model problems & links to re-refinements
- Classifies proteins according to:
  - experimental method
  - virus type
  - protein type
  - ligand status
- Allowing flexible and versatile selection of cases

The screenshot displays the 'Structures' page of the covid-19.bioreproducibility.org website. It features a search bar and several filter panels. The 'Filters Active - 1' section shows the following filters:

- Method: Cryo-EM (23/23), NMR (1/1), X-ray (118/231)
- Virus: HCoV-229E (6/6), MERS (1/1), SARS-CoV (18/18), SARS-CoV-2 (117/232)
- Protein: Main protease (36/151), NSP10/NSP16 (12/12), NSP12 (1/1), NSP15 (6/6)
- Ligand status: Functional ligand (54/54), Possible functional ligand (9/9), No functional ligands (50/50), Pathogen-host interaction (17/17)
- Presets: Non-PanDDA structures (149/149)

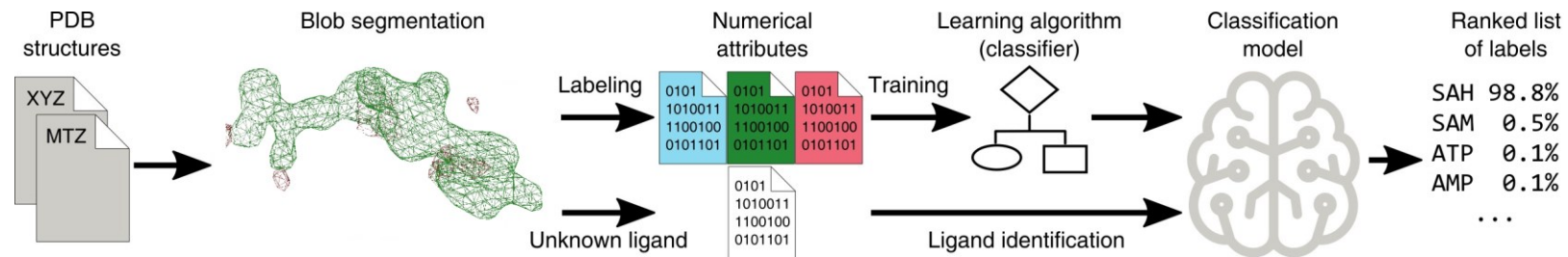
The search results table is as follows:

PDB	Resol.	Released	Title	Method	Ligand IDs	Virus	Protein	Ligand status	R-work	R-free	Issues	Re-refined?	Raw data	Ref.
6WX4	1.66 Å	2020-05-20	Crystal structure of the SARS C...	X-ray	-	SARS-CoV-2	Papain-like protease	Functional ligand	0.17%	0.20%	moderate	Yes	-	-
6YZ6	1.70 Å	2020-05-20	Structure of the hemiacetal com...	X-ray	IMD	SARS-CoV-2	Main protease	Functional ligand	0.18%	0.22%	moderate	Yes	-	-
7BR0	2.00 Å	2020-05-13	Crystal structure of the 2019-nC...	X-ray	-	SARS-CoV-2	Main protease	Protein-protein complex	0.23%	0.26%	-	Yes	-	-
7BRP	1.80 Å	2020-05-13	Crystal structure of the 2019-nC...	X-ray	HU5	SARS-CoV-2	Main protease	Functional ligand	0.22%	0.24%	-	Yes	-	-
7BRR	1.40 Å	2020-05-13	Crystal structure of the 2019-nC...	X-ray	K36	SARS-CoV-2	Main protease	Functional ligand	0.18%	0.20%	-	Yes	-	-
6YT8	2.05 Å	2020-05-06	Structure of SARS-CoV-2 Main P...	X-ray	PK6, IMD	SARS-CoV-2	Main protease	Functional ligand	0.20%	0.24%	moderate	Yes	-	-
6YNQ	1.80 Å	2020-04-29	Structure of SARS-CoV-2 Main P...	X-ray	-	-	Main protease	Functional ligand	18.21%	22.85%	moderate	Yes	-	-
6YLA	2.42 Å	2020-04-15	Crystal structure of the SARS-C...	X-ray	-	-	NSP3	Protein-protein complex	21.27%	23.67%	minimal	Yes	-	-
6W9C	2.70 Å	2020-04-01	The crystal structure of papain-li...	X-ray	-	-	Papain-like protease	No functional ligands	23.51%	30.88%	moderate	Yes	-	-
6W41	3.08 Å	2020-03-25	Crystal structure of SARS-CoV-2...	X-ray	-	-	Spike protein	Possible functional ligand	22.26%	24.33%	minimal	Yes	-	-
6YB7	1.25 Å	2020-03-25	SARS-CoV-2 main protease with...	X-ray	-	-	Main protease	No functional ligands	16.90%	19.20%	minimal	Yes	-	-
6Y7M	1.90 Å	2020-03-18	Crystal structure of the complex...	X-ray	-	-	Main protease	Functional ligand	20.62%	25.76%	minimal	Yes	-	-
6M83	2.00 Å	2020-03-11	The crystal structure of COVID-1...	X-ray	-	SARS-CoV-2	Main protease	No functional ligands	19.62%	24.58%	minimal	Yes	-	-

<https://covid-19.bioreproducibility.org>

# Future plans

- Use Machine Learning validation as an addition to correlation-based validation metrics (<https://checkmyblob.bioreproducibility.org/>)



- Work on combining genetic/structural visualizations with our quality assessment data (<https://coronavirus3d.org/>)
- Evaluate PanDDA fragment screening procedure to prevent flooding of the PDB with low-quality ligand complexes

# Conclusions

- New structures of SARS-CoV-2 proteins with ligands appear every week
- Due to the accelerated pace of COVID-related science, these structures have to be double-checked for correctness as drug design targets
- We use bioinformatic tools and expert knowledge to review, validate & rectify these structures
- Through our [covid-19.bioreproducibility.org](https://covid-19.bioreproducibility.org) server we want to pass our results on to the biomedical community
- We plan to expand it with new validation metrics and categorizations
- **Tools that combine knowledge and translate it to other fields are as important as tools that generate new knowledge within one field**