

Protein Data Bank in the Time of Coronavirus: How open-access 3D structures are enabling research and development in fundamental biology, biomedicine, biotechnology, and drug discovery as we confront the global SARS-CoV-2 pandemic.

S Burley¹

¹*RCSB Protein Data Bank, Rutgers University, Piscataway, NJ*
Stephen.Burley@rcsb.org

The Protein Data Bank (PDB) archive currently holds >165,000 atomic level three-dimensional (3D) structures of biomolecules experimentally determined using macromolecular crystallography (MX), NMR spectroscopy, electron microscopy (3DEM), and electron diffraction. The archive was established in 1971 as the first open-access, digital-data resource in biology, and is now managed by the Worldwide Protein Data Bank partnership (wwPDB; wwPDB.org). US PDB operations are the responsibility of the RCSB Protein Data Bank (RCSB PDB; RCSB.org; based at Rutgers, The State University of New Jersey, UC San Diego, and UC San Francisco). The RCSB PDB serves many millions of users worldwide by delivering PDB data integrated with ~40 external biodata resources via RCSB.org, APIs, and FTP download, providing rich structural views of fundamental biology, biomedicine, biotechnology, and energy sciences. In addition, the RCSB PDB outreach/education portal serves more than half a million PDB101.RCSB.org users worldwide, who are primarily university educators and their undergraduate students. Not counted in these usage metrics are the many PDB users working in biopharmaceutical companies, wherein copies of the PDB archive are retained within firewalls for interoperation with proprietary structures. Discovery/development of nearly 90% of the 210 new drugs approved by the US Food and Drug Administration 2010-2016 was facilitated by open access to target protein structures in the PDB [1]. With public release of the SARS-CoV-2 genome sequence on January 10th 2020, structural biologists joined the fight against the global COVID-19 pandemic. The first atomic-level 3D structure of a COVID-19 protein (Nsp5 or Main Protease, a promising drug target; PDB ID 6lu7) [2] was released by the PDB on February 5th 2020. Since then, more than 200 additional COVID-19 protein structures have been released by PDB, including those of the Spike Protein (illuminating receptor binding to Angiotensin Converting Enzyme 2 and neutralization by monoclonal antibodies), Papain-like Protease (part of Nsp3, and another promising drug target), RNA-dependent RNA polymerase (Nsp7, Nsp8, and Nsp12 heterotrimer; including complexes with RNA and the US FDA approved small-molecule inhibitor remdesivir), ADP Ribose Phosphatase (part of Nsp3), and Endoribonuclease (Nsp15). Of particular significance for global structure-guided drug discovery efforts are more than 100 PDB structures of the Main Protease bound to small-chemical fragments (or scaffolds) generated using the XChem Fragment Screening Pipeline at the UK Diamond Light Source. Select examples drawn from PDB holdings will be presented, highlighting the impact of MX and 3DEM on our understanding of SARS-CoV-2 biology, evolution, infection, and pathogenesis, and discovery of vaccines and small- and large-molecule therapeutics. Released COVID-19 PDB structures and related resources are regularly updated at <http://rcsb.org/covid19>. Acknowledgements: RCSB Protein Data Bank is funded by the National Science Foundation (DBI-1832184), the US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198. References: 1. Westbrook, J.W., and Burley, S.K. (2019) Structure. <https://doi.org/10.1016/j.str.2018.11.007> 2. Jin, Z. et al. (2020) Nature. <https://doi.org/10.1038/s41586-020-2223-y> Coronavirus 2020. Portrait painted by David S. Goodsell, RCSB Protein Data Bank.



Figure 1