



Fighting SARS-CoV-2 with structural biology methods

High-resolution structural information is critical for rapid development of vaccines and therapeutics against emerging human pathogens. Structural biology methods have been at the forefront of research on SARS-CoV-2 since the beginning of the COVID-19 pandemic. These technologies will continue to be powerful tools to fend off future public health threats.

Jun Zhang and Bing Chen

Forty years have passed since the publication of the first atomic structures of virus particles and viral glycoproteins^{1–6}. The influenza hemagglutinin structure has since become a key reference for structural studies of viral glycoproteins, including the human immunodeficiency virus 1 (HIV-1) envelope glycoprotein and coronavirus spike proteins. It enabled immunologists to visualize antigenicity of a human pathogen in three dimensions³, and epitope mapping was the first major application of the atomic model⁴. Within a year of determining the structure, the outlines of a membrane fusion mechanism had also begun to take shape⁷. The influenza neuraminidase structure⁵ also facilitated development of drugs still in use today. Nonetheless, structural biology has generally lagged behind other approaches — genetics, cell biology, molecular biology and biochemistry — for understanding viruses and viral infectivity, as defining the three-dimensional (3D) coordinates of thousands or even hundreds of thousands of atoms in a purified sample of a biological assembly experimentally has been a time-consuming and challenging task. Rapid advances in our structural understanding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) during the ongoing COVID-19 pandemic show that the balance has changed.

Structure determination of biological macromolecules at atomic resolution mainly relies on three principal methods in structural biology — X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryogenic electron microscopy (cryo-EM). The need for well-ordered crystals in X-ray crystallography and the upper limits of molecular mass in NMR exclude many macromolecules and their complexes, which are critical for mechanistic understanding, from application of these two methods. Since 2013, developments in cryo-EM single-particle analysis have enabled electron

microscopy to reach high resolution, dramatically altering the structural biology landscape^{8,9}. Owing to new technologies and software, cryo-EM, which requires far less sample (~0.1 mg) than crystallography or NMR (several milligrams or more), has become a powerful tool for determining structures of biological macromolecules at or near the atomic resolution. First, introduction of direct electron detectors has significantly increased the detective quantum efficiency (and thus enhanced contrast) of recorded images over the levels afforded by conventional charge-coupled device (CDD) detectors and photographic film¹⁰, effectively preserving high-resolution information^{11–13}. Second, data collection in a movie mode, made possible by the new detectors, has enabled corrections for beam-induced sample motion, allowing computational de-blurring of images^{14,15}. Third, maximum-likelihood algorithms have enabled robust 3D classification that separates a compositionally or conformationally heterogeneous sample into structurally homogenous subsets, which can yield density maps at a much higher resolution than previously possible^{16,17}.

Removing a biological macromolecule from its cellular context, however, can preclude full understanding of its function(s) in its native environment. Imaging a structure in situ at a sufficient resolution by cryogenic electron tomography (cryo-ET) promises to bridge the gap between a high-resolution structure derived from an isolated sample and the same molecules in a complex and heterogeneous assembly, such as a SARS-CoV-2 particle, or even in an intact cell. Cryo-ET reconstructs molecular 3D images of a target of interest together with its surrounding context at a resolution in the nanometer range by recording a tilt series (for example, $\pm 60^\circ$) of projection images and computationally reconstructing them into a 3D tomogram¹⁸. Electron absorption limits the sample thickness to ~500 nm — larger than the

dimensions of most virus particles but thinner than almost all parts of a eukaryotic cell. Focused-ion-beam (FIB) milling¹⁹, combined with correlative light–electron microscopy (CLEM)²⁰, helps solve this problem. Milling a thin (for example, 150–200 nm) ‘lamella’ in a cell provides a window into the part of the cell included in that thin layer. Reconstructed tomograms can be used directly for structural interpretation, but the complex cellular background and intrinsically low signal-to-noise ratio in tomograms make it hard to extract high-resolution information from individual intracellular particles. If many copies of the particle are present in a cell, averaging the images of those particles extracted from the tomogram (subtomogram averaging) can greatly extend resolution, even to near-atomic detail in favorable cases^{21,22}.

Since its initial outbreak in 2019, SARS-CoV-2 has captured the attention of the scientific community. After the release of the first viral genomic sequence²³, structural studies of various viral components, as well as the whole virus, have advanced at an astonishing pace. The crystal structure of the main protease (ProM) — a key enzyme in viral replication — was deposited in the Protein Data Bank (PDB) just two weeks later, followed shortly by publication of several structures of spike (S)-protein complexes. These included cryo-EM structures of the ectodomain stabilized in the prefusion conformation^{24–26} and X-ray crystal structures of the receptor-binding domain (RBD) in complex with the receptor angiotensin-converting enzyme 2 (ACE2)^{27–30}. The S protein mediates the fusion of viral and target cell membranes to allow the virus to enter a host cell. It is also a major surface antigen that induces neutralizing antibody responses and thus the key component of the first-generation vaccines³¹. The strategy, built on previous studies in both the coronavirus and HIV fields^{32,33}, to stabilize the S-prefusion

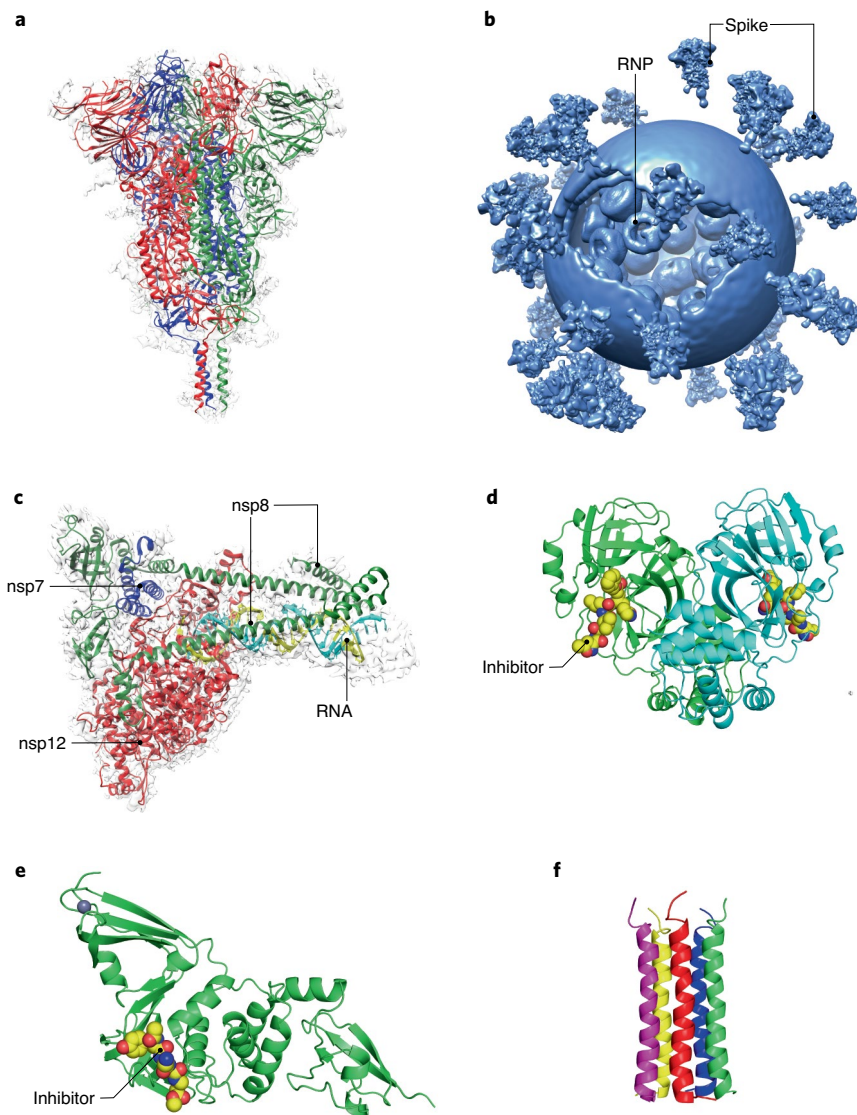


Fig. 1 | Key structures of SARS-CoV-2. **a**, Cryo-EM structure of the full-length S protein derived from the early variant, carrying a single-residue substitution (D614G) in the closed prefusion conformation with all three RBDs in the down conformation (EMD-23010; PDB ID: 7KRQ). Three protomers in the S trimer are colored in red, green and blue, respectively. **b**, Cryo-ET reconstruction of the intact SARS-CoV-2 (EMD-30430). The spikes and RNP assemblies are indicated. **c**, Cryo-EM structure of the SARS-CoV-2 RdRp, including nsp12 (red), nsp8 (green) and nsp7 (blue), as well as two turns of RNA template-product duplex (yellow and cyan) (EMD-11007; PDB ID: 6YYT). **d**, Crystal structure of the SARS-CoV-2 ProM dimer in complex with an inhibitor (N3) with the two protomers shown in a ribbon diagram (green and cyan, respectively) and the inhibitor in a sphere model (PDB ID: 7BQY). **e**, Crystal structure of the SARS-CoV-2 PL^{pro} in complex with a peptide inhibitor (VIR251). The enzyme is shown in a ribbon diagram (green) and the inhibitor in a sphere model (PDB ID: 6WX4). **f**, Structure of the SARS-CoV-2 E protein transmembrane domain in a pentameric assembly determined by solid-state NMR (PDB ID: 7K3G). Five protomers are colored in red, magenta, yellow, blue and green.

conformation with two proline residues to prevent formation of a long helix in the postfusion structure, has been used in several successful COVID-19 vaccines³¹. Cryo-EM structures of the purified full-length S protein of the original Wuhan-Hu-1 strain, in both its prefusion

and postfusion conformations, were also reported not long after the stabilized ectodomain structures³⁴, as was the structure of the full-length S protein from an early variant carrying a single-residue substitution: D614G (Fig. 1a; ref. 35). These structures, consistent with the

results of studies on chemically inactivated SARS-CoV-2 virions with combined cryo-ET subtomogram averaging and cryo-EM single-particle analysis^{36–38}, suggested that greater stability might make it preferable to use an immunogen with the D614G substitution instead of the unmodified Wuhan-Hu-1 S sequence³⁹. Numerous structures of monoclonal neutralizing antibodies in complex with the S protein have also been determined^{40,41}. These structures have not only led to mechanistic understanding of antibody neutralization mechanisms, but can also guide development and optimization of therapeutic antibodies by rational strategies⁴².

The overall molecular architecture of SARS-CoV-2 was subsequently analyzed by cryo-ET and subtomogram averaging, showing the structures of the S protein in both prefusion and postfusion states at medium resolution at the virion surface, as well as the assemblies of the ribonucleoproteins (RNPs) inside of the viral membrane (Fig. 1b; ref. 43). The key component of the viral replication machinery is an RNA-dependent RNA polymerase (RdRp), which carries out both replication and transcription. A cryo-EM structure of the SARS-CoV-2 RdRp, including non-structural protein 12 (nsp12), nsp8 and nsp7, and two turns of RNA template-product duplex (Fig. 1c; ref. 44), showed how the enzyme interacts with the double-stranded RNA and how antiviral compounds target the enzyme. Crystal structures of ProM (Fig. 1d) and the papain-like cysteine protease (PL^{pro}, nsp3; Fig. 1e) of SARS-CoV-2 in complex with various inhibitors have illustrated the inhibition mechanisms and guided efforts to repurpose approved drugs and develop novel therapeutics^{45,46}. NMR spectroscopy has also contributed significantly to our structural understanding of certain viral components, in particular, those buried in membrane, which are difficult to access by other approaches. For example, the envelope (E) protein forms a homopentameric cation channel that is critical for viral pathogenicity. A structure of the E transmembrane domain, determined by solid-state NMR spectroscopy (Fig. 1f; ref. 47), may facilitate development of E inhibitors as antiviral drugs. Similarly, because it has been either removed or is invisible in the S structures determined by cryo-EM or cryo-ET, the transmembrane domain of the S protein has been isolated and reconstituted into bicelles that mimic a lipid bilayer for structure determination by solution NMR. The structure shows a novel trimeric leucine-isoleucine zipper

with tetrad repeats⁴⁸. As of February 2022, 1,832 structures containing SARS-CoV-2 components have been deposited in the PDB, with 1,195 determined by X-ray crystallography, 628 by electron microscopy, 8 by solution NMR, 3 by neutron diffraction and 1 by solid-state NMR (<http://www.wwpdb.org>). Additionally, powerful molecular simulation tools have been applied to extrapolate dynamic mechanisms of the steps of SARS-CoV-2 infection, such as the large conformational changes of S protein required for membrane fusion, well beyond what the static atomic structures can offer^{49,50}.

We strongly believe that structural biology will continue to make critical contributions — not only in efforts to control the COVID-19 pandemic, but also to fight against any future outbreaks of human pathogens — by revealing atomic details of infection, pathogenesis and host immune responses, as well as by facilitating rapid development of vaccine and therapeutic strategies. In particular, the advances in cryo-EM and cryo-ET technology in the past decade have greatly broadened the variety of biological samples accessible by high-resolution structural biology methods, as it is no longer necessary to truncate a full-length protein, remove physiologically relevant posttranslational modifications (for example, glycosylation) or select one homogenous conformation to produce well-diffracting crystals or interpretable NMR spectra. Several years ago, it would have been unthinkable that an atomic structure of a full-length and heavily glycosylated protein with more than a thousand amino-acid residues, such as the SARS-CoV-2 spike, could be determined in less than three months³⁴.

Other technological advances have also made major impacts on the structural biology field. For example, live-cell optical imaging, such as lattice light-sheet microscopy combined with deep-learning-based data analyses, has already achieved unprecedented spatial and temporal resolutions at the single-molecule level in the context of living cells⁵¹. Protein structures predicted by the deep-learning-based programs AlphaFold and RoseTTAFold have now become remarkably accurate^{52,53}, representing a further important advance in structural biology as they can rapidly generate atomic models (including some SARS-CoV-2 proteins^{54,55}) on the proteomic scale that not only can guide functional studies, but can also serve as initial models for structure determination by experimental methods. These developments are revolutionizing the way in which we study molecular events of

biological processes such as viral infection, marking a new era of structural biology that directly connects static atomic models with dynamic events in a crowded cellular environment.

What are the next methodological challenges in structural biology? In cryo-EM, most of the initially selected particles from a cryo grid are rejected as ‘junk’, even from a biochemically ‘clean’ preparation, and are thus removed from subsequent data processing to filter for a few ‘good’ classes that are homogenous enough for sufficient alignment to reach high resolutions. It is unclear whether those ‘bad’ particles are indeed misfolded or damaged molecules (for example, due to well-documented denaturation at the air–water interface⁵⁶), or simply physiologically meaningful species in low abundance or with intrinsic conformational dynamics that are difficult to align. It is possible that certain important structural information is discarded along with the genuinely damaged particles. In cryo-ET, which represents a fast-evolving frontier in electron microscopy, there is still much to be desired for the resolution that can be achieved. Moreover, small regions embedded in membranes are invisible from many cryo-ET reconstructions even though there should be enough contrast to reveal them if enough tomograms are collected, at least in theory. Finally, as powerful as artificial-intelligence-based programs, such as AlphaFold, have already been for structure prediction, high levels of accuracy are still heavily biased to those that have set a homology model determined by experimental methods during training⁵⁷. Predicting the structure of a large and complex protein with ligands or posttranslational modifications continues to be a formidable challenge.

Structural biology has become an indispensable tool for understanding biology and medicine, and continues to evolve rapidly. We should not be surprised if an atomic image of a complete virus particle is one of the first results released at the onset of any future pandemic comparable to the one caused by SARS-CoV-2. □

Jun Zhang ^{1,2} and Bing Chen ^{1,2}

¹Division of Molecular Medicine, Boston Children's Hospital, Boston, MA, USA. ²Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

✉e-mail: juzhang@crystal.harvard.edu; bchen@crystal.harvard.edu

Published online: 8 April 2022
<https://doi.org/10.1038/s41592-022-01448-9>

References

- Harrison, S. C. et al. *Nature* **276**, 368–373 (1978).
- Abad-Zapatero, C. et al. *Nature* **286**, 33–39 (1980).
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
- Wiley, D. C., Wilson, I. A. & Skehel, J. J. *Nature* **289**, 373–378 (1981).
- Colman, P. M., Varghese, J. N. & Laver, W. G. *Nature* **303**, 41–44 (1983).
- Varghese, J. N., Laver, W. G. & Colman, P. M. *Nature* **303**, 35–40 (1983).
- Skehel, J. J. et al. *Proc. Natl Acad. Sci. USA* **79**, 968–972 (1982).
- Cheng, Y. *Cell* **161**, 450–457 (2015).
- Cheng, Y. et al. *Cell* **161**, 438–449 (2015).
- Faruqi, A. R. & McMullan, G. Q. *Rev. Biophys.* **44**, 357–390 (2011).
- McMullan, G. et al. *Ultramicroscopy* **147**, 156–163 (2014).
- Sander, B., Golas, M. M. & Stark, H. J. *Struct. Biol.* **151**, 92–105 (2005).
- McMullan, G. et al. *Ultramicroscopy* **109**, 1126–1143 (2009).
- Brilot, A. F. et al. *J. Struct. Biol.* **177**, 630–637 (2012).
- Campbell, M. G. et al. *Structure* **20**, 1823–1828 (2012).
- Sigworth, F. J. *J. Struct. Biol.* **122**, 328–339 (1998).
- Scheres, S. H. et al. *Nat. Methods* **4**, 27–29 (2007).
- Turk, M. & Baumeister, W. *FEBS Lett.* **594**, 3243–3261 (2020).
- Marko, M. et al. *Nat. Methods* **4**, 215–217 (2007).
- Sartori, A. et al. *J. Struct. Biol.* **160**, 135–145 (2007).
- Schur, F. K. et al. *Science* **353**, 506–508 (2016).
- von Kugelgen, A. et al. *Cell* **180**, 348–358 (2020).
- Wu, F. et al. *Nature* **579**, 265–269 (2020).
- Wrapp, D. et al. *Science* **367**, 1260–1263 (2020).
- Walls, A. C. et al. *Cell* **181**, 281–292 (2020).
- Henderson, R. et al. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020).
- Lan, J. et al. *Nature* **581**, 215–220 (2020).
- Yan, R. et al. *Science* **367**, 1444–1448 (2020).
- Shang, J. et al. *Nature* **581**, 221–224 (2020).
- Wang, Q. et al. *Cell* **181**, 894–904 (2020).
- Dai, L. & Gao, G. F. *Nat. Rev. Immunol.* **21**, 73–82 (2021).
- Sanders, R. W. et al. *PLoS Pathog.* **9**, e1003618 (2013).
- Kirchdoerfer, R. N. et al. *Sci. Rep.* **8**, 15701 (2018).
- Cai, Y. et al. *Science* **369**, 1586–1592 (2020).
- Zhang, J. et al. *Science* **372**, 525–530 (2021).
- Ke, Z. et al. *Nature* **588**, 498–502 (2020).
- Turoňová, B. et al. *Science* **370**, 203–208 (2020).
- Liu, C. et al. *Structure* **28**, 1218–1224 (2020).
- Koenig, P. A. & Schmidt, F. I. N. *Engl. J. Med.* **384**, 2349–2351 (2021).
- Barnes, C. O. et al. *Nature* **588**, 682–687 (2020).
- Tong, P. et al. *Cell* **184**, 4969–4980 (2021).
- Chen, J. et al. *Annu. Rev. Biophys.* **50**, 1–30 (2021).
- Yao, H. et al. *Cell* **183**, 730–738 (2020).
- Hillen, H. S. et al. *Nature* **584**, 154–156 (2020).
- Jin, Z. et al. *Nature* **582**, 289–293 (2020).
- Rut, W. et al. *Sci. Adv.* <https://doi.org/10.1126/sciadv.abd4596> (2020).
- Mandala, V. S. et al. *Nat. Struct. Mol. Biol.* **27**, 1202–1208 (2020).
- Fu, Q. & Chou, J. J. *J. Am. Chem. Soc.* **143**, 8543–8546 (2021).
- Dodero-Rojas, E., Onuchic, J. N. & Whitford, P. C. *eLife* **10**, e70362 (2021).
- Pak, A. J. et al. *Nat. Commun.* **13**, 1002 (2022).
- Liu, T. L. et al. *Science* **360**, eaaq1392 (2018).
- Jumper, J. et al. *Nature* **596**, 583–589 (2021).
- Baek, M. et al. *Science* **373**, 871–876 (2021).
- Jumper, J. et al. *Computational Predictions of Protein Structures Associated with COVID-19 v.3* (DeepMind, 2020); <https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>
- Heo, L. & Feig, M. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.03.25.008904> (2020).
- Noble, A. J. et al. *eLife* **7**, e34257 (2018).
- Jones, D. T. & Thornton, J. M. *Nat. Methods* **19**, 15–20 (2022).

Acknowledgements

We thank S. Harrison for critical reading of the manuscript and other members of the Chen laboratory for their contributions to our research. Our work was supported by National Institutes of Health grants, a Fast Grant from Emergent Ventures and a COVID-19 award from the Massachusetts Consortium on Pathogen Readiness (MassCPR).

Competing interests

The authors declare no competing interests.