Structural biology of SARS-CoV-2 and implications for therapeutic development

Haitao Yang and Zihe Rao

Abstract | The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is an unprecedented global health crisis. However, therapeutic options for treatment are still very limited. The development of drugs that target vital proteins in the viral life cycle is a feasible approach for treating COVID-19. Belonging to the subfamily Orthocoronavirinae with the largest RNA genome, SARS-CoV-2 encodes a total of 29 proteins. These non-structural, structural and accessory proteins participate in entry into host cells, genome replication and transcription, and viral assembly and release. SARS-CoV-2 proteins can individually perform essential physiological roles, be components of the viral replication machinery or interact with numerous host cellular factors. In this Review, we delineate the structural features of SARS-CoV-2 from the whole viral particle to the individual viral proteins and discuss their functions as well as their potential as targets for therapeutic interventions.

Coronaviruses are enveloped viruses that possess a positive-sense single-stranded RNA genome 26–32 kb in length. Coronaviruses belong to the Coronavirusidae subfamily Orthocoronavirinae. According to variations in the genome sequence and serological reactions, coronavirus members in the subfamily are classified into four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus. Among them, Betacoronavirus is classified into five subgenera. Although infectious bronchitis virus was the first coronavirus isolated in chicken embryos in 1937 (REF 1), it was not until the 1960s that these viruses, particularly the human respiratory coronaviruses2, were characterized by electron microscopy. This subfamily of viruses has a unique structural feature on their surfaces which resembles a solar corona. This feature arises due to the presence of spike proteins on the virion surface.

Coronaviruses are characterized by high genetic recombination and mutation rates, which result in their ecological diversity. They are able to infect and readily adapt to a wide range of hosts, from birds to whales. Seven coronaviruses have been found to infect humans. Human coronaviruses 229E, OC43, NL63 and HKU1 are responsible for 10–30% of upper respiratory tract infections annually, characterized by mild respiratory illnesses, such as the common cold3. By contrast, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are able to cause severe human respiratory diseases, potentially resulting in high mortality. In 2002–2003, SARS-CoV resulted in 8,096 reported cases and 774 deaths (case-fatality rate of ~10%)4. By the end of January 2020, 2,500 cases of Middle East respiratory syndrome and more than 800 associated deaths (case-fatality rate ~34%) were reported worldwide5. In late December 2019, clustered cases of a severe pneumonia were reported, and the aetiological agent was isolated and identified as a novel betacoronavirus, named SARS-CoV-2, that shares ~80% similarity in genome sequence with SARS-CoV6. SARS-CoV-2 causes COVID-19, with symptoms including fever, cough, fatigue, nausea and shortness of breath7. To date, there have been more than 160 million confirmed COVID-19 cases and more than 3 million related deaths worldwide8.

To date, there has been a lack of effective therapies to treat COVID-19. Due to the rampant and continuous spread of COVID-19, it is a matter of urgency to identify and characterize drug and vaccine targets for SARS-CoV-2. The genome of SARS-CoV-2 is close to 30 kb on size, contains 14 open reading frames (ORFs) and encodes 29 viral proteins. Approximately two thirds of the 5’ end of the SARS-CoV-2 genome encodes two overlapping polyproteins: pp1a and pp1ab. These two polyproteins are digested by two viral proteases into 16 non-structural proteins (NSPs), which are essential for viral replication and transcription (FIG. 1a). Four ORFs at the 3’ terminus of the viral genome encode a canonical set of structural proteins that include the nucleocapsid (N), spike (S) protein, membrane (M)
The two polyproteins are cleaved by viral proteases, a papain-like protease (PLpro) and a 3C-like protease (3CLpro), to generate 16 NSPs and to form the replication and transcription machinery. The domains that have an important function within each NSP are shown in the genome structure. The first three peptide cleavages are performed by PLpro. The remainders are cleaved by 3CLpro (also known as the main protease). The second part of the RNA genome encodes mainly four structural proteins: the spike (S) protein, the membrane (M) protein, the envelope (E) protein and the nucleocapsid (N) protein. In addition to these structural proteins, several accessory proteins are also encoded.

The life cycle of SARS-CoV-2, including viral entry, replication and transcription, assembly and release, is outlined in Fig. 1. Briefly, in the first step of the SARS-CoV-2 life cycle, the S protein on the outer surface of the virion is responsible for binding to the host receptor or receptors for attachment to the cell membrane, which is followed by viral and host cellular membrane fusion and the release of viral genomic RNA into the cells. Subsequently, host ribosomes are hijacked to produce the two viral replication polyproteins, which can further be processed into 16 mature NSPs through two virus-encoding proteases: main protease (Mpro) and papain-like protease (PLpro). These NSPs are able to assemble into the replication and transcription complex (RTC) to initiate viral RNA replication and transcription. The genomic RNA and structural proteins then assemble into mature progeny virions, which are then released by exocytosis into the extracellular space. The newly released virions are then able to infect new host cells and complete the viral life cycle.

Structural proteins in the viral life cycle

S protein in viral entry

The S protein is a homotrimer, which protrudes from the virion and extensively decorates the viral surface like a crown. It is heavily glycosylated, belongs to the type I membrane–protein family and is anchored in the viral membrane, where it mediates fusion of the viral membrane with the host cell membrane. In the native state, prefusion and postfusion conformations of S proteins can be traced simultaneously on the reconstructed virions. The SARS-CoV-2 S protein comprises ~1,200 residues and can be cleaved by a furin-like protease into two functional subunits, S1 and S2, which are responsible for mediating attachment to host cells and membrane fusion, respectively. After cleavage during viral entry into the host cells, S1 and S2 remain associated with each other through non-covalent interactions. As shown by cryo-electron microscopy (cryo-EM) (Fig. 2a), the S1 subunit of the SARS-CoV-2 S protein wraps around a threefold axis, covering the S2 subunit underneath. The S1 subunit contains a receptor-binding domain (RBD) and an amino-terminal (N-terminal) domain (NTD). The RBD has a five-stranded antiparallel β-sheet core, flanked on either side by a short helix. The receptor-binding motif (RBM) extends out of the core (connecting β4 and β5), taking on a cradle-like structure for receptor binding. The RBM, which is stabilized by a disulfide bond, does not possess a regular secondary structure but has a variable conformation. The RBD can adopt two distinct conformational states: the closed ‘down’ state and the open ‘up’ state. In the ‘down’ state, RBD angles are close to the central cavity of the trimer to shield the receptor-binding regions, while in the ‘up’ state, the RBD undergoes hinge-like conformational movement, exposing its determinant regions to recognize the human angiotensin-converting enzyme 2 (hACE2) receptor on the host cellular membrane. The state of the NTD of the S protein adopts a galectin-like fold with a sugar-binding pocket and contains a ceiling-like structure on top. The NTD may recognize sugar moieties upon initial attachment and play a significant role in the transition of the conformation of the S protein. The S2 subunit comprises four conserved structural regions: a fusion peptide, two heptad repeats (HR1 and HR2) and a transmembrane region. The HR1 region constitutes the main helical stalk of S2, whereas the HR2 region is temporarily flexible in the prefusion state. The fusion peptide forms a short hydrophobic segment.

Undergoing a substantial structural rearrangement, from the metastable prefusion conformation to the postfusion conformation, the S protein fulfils its function in regulating the fusion of viral membrane with the host cell membrane. Fusion is triggered when the S1 subunit binds to hACE2 (Fig. 2b). As observed in the complex structure, the N-terminal helix of hACE2 interacts with the outer surface of the RBM in the S1 subunit. The interaction involves 16 residues in the
Viral entry

Spike (6VXX)

Spike antibody

Spike + 4A8 (7C2L)
NTD neutralizing antibody

Spike + CB6 (7C01)
Class I RBD neutralizing antibody

Spike + 4A8 (7C2L)
NTD neutralizing antibody

Spike RBD + ACE2 + B'AT1 (6M17)

Spike RBD + ACE2 (6LZG)

Spike + BD23-Fab (7BYR)
Class III RBD neutralizing antibody

Spike + S309 (7CWL)
Class IV RBD neutralizing antibody

Spike + H11-H4 (6ZHD)
Class IV RBD neutralizing antibody
RBD and 20 residues in hACE2, which forms a network consisting of 14 hydrogen bonds and one salt bridge. The binding of hACE2 to the RBD can lock the RBD in the ‘up’ conformation and trigger S1 shedding, which is mediated by the proteolytic cleavage of host TMPRSS2 and cathepsin B or cathepsin L. Thus, three HR1 helices of trimeric S2 interact with the pairing HR2 helices and constitute a stable six-helix bundle. In this unique helix bundle, three HR2 helices are packed into the hydrophobic grooves of the HR1-trimer core in an antiparallel manner. This conformational arrangement brings viral and host cell membranes into proximity and facilitates subsequent membrane fusion. Because of the indispensable function of the S protein, it is an attractive target for inhibition by neutralizing antibodies (nAbs), and characterization of the S protein structure provides atomic-level information for rational vaccine design.

S protein-neutralizing antibodies
nAbs targeting the SARS-CoV-2 S trimer have shown protection from viral infection in animal models and are being evaluated as therapeutics in humans. These antibodies comprise human monoclonal antibodies isolated from COVID-19 convalescent donors and single-domain antibodies (also known as nanobodies) which can bind novel epitopes, including buried cavities that are inaccessible to conventional antibodies. Determination of a number of structures of nAbs in complex with the S trimer has elucidated their modes of neutralization. Although some nAbs target the NTD or S2, most nAbs bind to the RBD, the latter of which can be further classified into four distinct classes (classes I, II, III and IV) on the basis of the nAb–RBD binding characteristics.

The nAbs in class I can bind to the RBD only in the ‘up’ state (Fig. 2e). They are expected to bind to the flat area on the top side of the cradle-like surface of the RBD, which extensively overlaps with the binding site for hACE2. Through direct competition with hACE2, nAbs in this class would produce steric hindrance when binding to RBD, blocking hACE2 binding attachment. CB6 (REF.25), C105 (REF.26), CV30 (REF.25), B38 (REF.27), CC12.1, CC12.3 (REF.26,28), PR1077 (REF.26) and P4A1 (REF.29) nAbs belong to this class. Most contain IGHV3-53- or IGHV3-66-encoded heavy chains and utilize residues in complementarity-determining regions 1, 2 and 3.

The nAbs in class II also bind to the RBD in the ‘up’ state, but exhibit no overlap with hACE2-binding sites (Fig. 2f,g). CR3022 (REF.23), EY6a (REF.30) and nanobody VHH-72 (REF.31) belong to this class. The binding region is located at the bottom of the RBD, and is spatially separated from the hACE2-binding sites. Structural analysis showed that the RBD undergoes a rotation that exposes the epitopes for these nAbs. Such a rearrangement is considered to cause a premature conversion of the S protein from the prefusion state to the postfusion state. The resulting unstable configuration of the S protein consequently inactivates SARS-CoV-2.

The nAbs in class III can bind to RBDs only in the ‘down’ conformation (Fig. 2h). They comprise Fab 2–43, Fab 2–43 (REF.24) and BD23 (REF.25). The heavy chains of the nAbs reach the RBD and interact with the cradle-like surface or the flexible ridge region. However, the binding pattern between these nAbs and the RBD is different from that for class I nAbs, according to the orientation change in the RBD, and the binding area becomes narrower. Notably, N-glycan chains are supposed to play a significant role in stabilizing the binding of class III nAbs to the ‘down’ RBD. Additionally, epitopes of some nAbs extend to the NTD, which may help to resist dynamic instability. Collectively, this binding mode would lock the RBD in the ‘down’ conformation, which also sterically hinders hACE2 access.

The nAbs in class IV can recognize both the ‘up’ RBD conformation and the ‘down’ RBD conformation (Fig. 2i). They comprise H11–D4, H11–H4 (REF.20), P2B–2F6 (REF.19), Ty1 (REF.26), S309 (REF.26), REGN10987 (REF.20) and P17 (REF.19). Structural studies show that these nAbs target different regions. P2B–2F6 and the nanobodies H11–D4 and H11–H4 can bind to the top cradle-like surface in a similar orientation as class III nAbs. Their binding can be further reinforced by a protruding loop on the RBD. These three nAb epitopes are largely located on the opposite side of the RBD compared with the epitopes of class I nAbs. By partially overlapping with the hACE2-binding site, these nAbs sterically block hACE2 binding to the RBD as well. S309 targets a region distinct from the RBD. Its epitope comprises the α1 helix, a section of the β1 strand and two loops formed by residues 358–361 and 333–335. REGN10987 is another class VI nAb that binds distal to the hACE2-binding site. The binding of this nAb would spatially hinder hACE2 attachment.

4A8 (REF.24), COV57 (REF.25), 2–17, 5–24, 4–8 (REF.19) and FC05 (REF.19) are nAbs that target other parts of the S protein. Structural analysis reveals that 4A8, which shows a high level of neutralization of SARS-CoV-2, recognizes the NTD and does not sterically hinder the binding between hACE2 and the S protein (Fig. 2d). Regarding the S2 subunit, only a few targeted monoclonal antibodies have been reported. Antibody 1A9 (REF.34) has been found to interact with the S2 subunit but fails to neutralize SARS-CoV-2. In a recent report, the nAb CC40.8 was identified and found to neutralize SARS-CoV-2 and specifically recognize the S2 subunit(ref.30). The discovery of non-RBD-targeted nAbs may benefit the strategy of nAb cocktail therapeutics.

Since SARS-CoV and SARS-CoV-2 share the same host cell receptor, hACE2, development of cross-neutralizing antibodies to both coronaviruses seems feasible. H014 (REF.35) is a recently reported humanized antibody which efficiently neutralizes both SARS-CoV-2 and SARS-CoV.
SARS-CoV-2. It can recognize and interact with open RBDS, but the binding interface is located distinct from the RBM, and exhibits no competition with hACE2 attachment. Consistently, other cross-neutralizing antibodies (for example, VHH-72, ADI-56046 (Ref. 54), COV21 (Ref. 55) and CC6.33 (Ref. 56)) also avoid the RBM and prefer to recognize the core domain of the RBD.

It is noteworthy that SARS-CoV-2 has a high mutation rate, and numerous mutant strains (variants) have been reported. Mutations in the S protein, especially the epitopes for nAbs, would attenuate the potency of nAbs. The D614G mutation is the most commonly reported mutation in the S protein, and results in increased infectivity and morbidity. The cryo-EM structure of the trimeric S protein with D614G demonstrated a conformational shift towards the hACE2-binding fusion-competent state and exhibited attenuation of efficacy in nAb binding. N501Y is a mutant variant emerging from the United Kingdom, South Africa and Brazil. The mutation site is located at the RBD–hACE2 interface and has been experimentally shown to cause an increase in hACE2 affinity. Other mutations worth noting include K417N and K417T, which appear in the epitopes of class I nAbs and are considered to affect the binding of class I antibodies. Mutations at residues in the NTD were also found in the new variants of concern, such as ΔY144 and Δ242–244. They were shown to abrogate neutralization of NTD-specific nAbs (Ref. 57) or TCE0 (Ref. 58). Combination treatment with two or more nAbs targeting distinct epitopes would be a strategy to suppress nAb escape variants.

**E protein**

After a coronavirus enters host cells, the E protein regulates viral lysis and the subsequent viral genome release. The E protein was found to be involved in viral assembly and budding by localizing to endoplasmic reticulum (ER) and Golgi body membranes. Moreover, the E protein has been shown to participate in activating the host inflammasome. The structure of the SARS-CoV-2 E protein solved by nuclear magnetic resonance spectroscopy shows that it is composed of a five-helix bundle ∼35 Å in length (Fig. 3b). As the E protein can function as an ion channel, the pore inside the transmembrane region is predominantly occupied by hydrophobic residues except for the N-terminal pore. Owing to non-specific interhelical interactions, the entrance site at the N terminus is a drug target for inhibitor binding. The E protein is recognized topologically to be N-lumen–C cytosol (N-terminal ER–Golgi intermediate compartment lumen and carboxy-terminal (C-terminal) cytoplasm) and involved in regulation of pumping Ca²⁺ out of the ER, which may lead to activation of the cellular inflammasome, thereby enhancing the host antiviral response.

**N protein**

The N protein serves as the only structural protein inside the virion. It is a crucial component that protects the viral RNA genome and packages it into a ribonucleoprotein complex. A native reconstruction of SARS-CoV-2 using electron cryotomography suggests that a significant number of ribonucleoproteins may be membrane proximal. The N protein also plays a role in antagonizing the host immune response and has been identified to counter cellular RNAi-mediated antiviral
activities through its binding with double-stranded RNA ‘strings’, and can be regarded as a viral suppressor of RNA silencing. The N protein has potential as a target for vaccine development because it induces a severe immune responses during infection.

The N protein has two conserved structural domains, the N-NTD (N-NTD) and the CTD (N-CTD), each of which is independently folded. In the crystal structures of the N protein, the N-NTD exists as a monomer, whereas the N-CTD exists as a dimer (FIG. 5a). The N-NTD has the shape of a right-handed fist and contains a four-stranded antiparallel β-sheet as a core subdomain. The loops protruding out of the core are positively charged, putatively to allow RNA binding. The N-CTD homodimer forms a rectangular shape, with each protomer displaying a crescent shape. To stabilize the dimer interface, two β-hairpin structures from each protomer can form four antiparallel β-strands by inserting themselves into each cavity. Compared with other coronaviruses, the N protein from SARS-CoV-2 displays different charge distributions in the N-terminal loop, the RNA protruding tip, the bottom of the N-NTD core and the N-CTD β-stand face. Hence, the variations in RNA binding to the N protein may further guide inhibitor optimization.

**NSPs and inhibitors**

**Host translation shutdown by nsp1**

nsp1 originates from the N-terminal cleavage of polyproteins pp1a and pp1ab by PLPro. The biological functions of nsp1 manifest themselves mainly in virus–host interactions to suppress host translation 3–11, and thus nsp1 can be regarded as a canonical virulence factor. To hinder the host translation process, nsp1 is proposed to function by two mechanisms: the first is to bind the ribosomal 40S subunit during the initiation stage 64 and the second is to induce host mRNA degradation 65. Importantly, nsp1 does not impede viral protein expression while it binds to the mRNA 5′ untranslated region, leading to efficient viral translation and replication. The structure of nsp1 and the ribosomal 40S subunit has been determined to show the interactions between them and to explain the potential inhibition mechanism 66. In this cryo-EM structure, the C-terminal domain of nsp1 possesses a short α-helix which is connected to a longer α-helix through a short loop (FIG. 4a). Thus, the host mRNA entry channel is blocked by nsp1 insertion. This hypothesis is corroborated by the loss of host translation inhibition in the K164A–H165A double mutant. The long α-helix also contributes to the interactions between nsp1 and the ribosome. Through the shutdown of host translation, especially antiviral factors, nsp1 assists in evading immune defences, which suggests that disrupting nsp1–ribosome interactions is a plausible approach for SARS-CoV-2 drug discovery.

**Multidomain protein nsp3**

nsp3 consists of 10–16 domains depending on the coronavirus genus. Eight are present in all coronaviruses, including ubiquitin-like domain 1 (Ub1), a hypervariable region, a macrodomain, ubiquitin-like domain 2 (Ub2), a PLPro, a zinc-finger domain, a Y1 domain and a CoV-Y domain7. Most of the conserved domains perform essential functions in the life cycle of the virus. The macrodomains possess highly conserved structures and similar functions. Macrodomain Mac1 can cleave the phosphate group of ADP-ribose 1-phosphate and reverse protein ADP-ribosylation by hydrolysis. The core structure of Mac1 contains seven β-strands flanked by six α-helices (FIG. 4b). ADP-ribose interacts with the Mac1 hydrophobic cleft through conserved hydrogen bonds. This indicates that compounds targeting Mac1 may have broad-spectrum antiviral activities.

The ‘SARS-unique domain’ (SUD) participates in virus–host interactions. SUD has three subdomains: SUD-N (Mac2), SUD-M (Mac3) and SUD-C (DPUP). SUD-N and SUD-M adopt a macrodomain fold, whereas SUD-C has a frataxin-like fold. Deletion of Mac2 decreases the viral replication rate to 65–70%, whereas Mac3 is indispensable for replication activity 66. PAIP1, which is a component of the eukaryotic translation machinery, has been identified to interact with SUD. The structure of the Mac2–PAIP1M (middle domain of PAIP1) complex shows that Mac2 displays a typical α/β/α macrodomain fold, whereas PAIP1M adopts a HEAT repeat fold. Strong complementarity which enhances complex stability is observed at the interface. This structure also supports the suggestion that Mac2–PAIP1M participates in regulating viral mRNA translation and is thus a good antiviral drug target.

PLPro is located in nsp3 between SUD and a nucleic acid-binding domain. It cleaves the viral polyprotein precursors pp1a and pp1ab at three sites to produce NSPs nsp1, nsp2 and nsp3 (REF. 1). Apart from viral polyproteins, PLPro can also cleave host proteins to antagonize the innate immune response 1. It preferentially recognizes and cleaves interferon-stimulated gene product 15 (ISG15) from interferon regulatory factor 3 (IRF3) and attenuates type I interferon responses, facilitating escape of the virus from the immune system 1. PLPro is a 36-kDa cysteine protease with a catalytic triad 1. It contains an N-terminal ubiquitin-like domain and a catalytic core domain 71. The catalytic core domain comprises three subdomains, the thumb, palm and fingers, which together fold like an open right hand. The thumb subdomain is composed of four α-helices, whereas the palm is formed by a six-stranded β-sheet. A four-stranded, twisted, antiparallel β-sheet makes up the finger subdomain. In the fingertip region, four cysteine residues constitute a zinc-finger motif, which coordinates a zinc ion with tetrahedral geometry. This zinc-finger is essential for structural integrity and protease activity.

The substrate-binding site is located in the solvent-exposed cleft between the thumb subdomain and the palm subdomain, which possess a catalytic triad composed of C111, H272 and D286. The substrate-binding site recognizes the consensus sequence LXG↓GX (the amino acid residues of the substrate are numbered P4–P3–P2–P1↓P1′–P2′ around the cleavage site, denoted by the downwards arrow). Subsites S1–S4 provide the binding sites for P1–P4, respectively 2. The S1 and S2 subsites are rather narrow, and can accommodate only glycine residues. The S3 subsite is partially solvent exposed but prefers positively charged and hydrophobic
residues. The S4 subsite is relatively large and accommodates only hydrophobic residues. A flexible β-hairpin BL2 loop, which contains an unusual β-turn at Y268 and Q269, is involved in controlling substrate access to the active site. Consideration of the conformation of the BL2 loop may be important for rational drug design.

Besides the catalytic site, PLpro harbours two distinct binding subsites (SUb1 and SUb2) for recognizing diubiquitin chains and ISG15. SUb1 recognizes one ubiquitin molecule of diubiquitin chains and the C-terminal ubiquitin-like domain of ISG15. SUb2 recognizes the other (K48-linked) ubiquitin molecule and the N-terminal ubiquitin-like domain of ISG15 (refs 74, 76) (fig. 4d). As shown in the complex structures of PLpro–diubiquitin and PLpro–ISG15, SUb1 of SARS-CoV-2 PLpro preferentially binds ISG15 through a different binding mode compared with ubiquitin. Moreover, PLpro SUb2 provides exquisite specificity for K48-linked diubiquitin chains, which makes diubiquitin a suitable substrate compared with monoubiquitin.

Inhibitors targeting PLpro

Owing to the substantial role in mediating viral replication and suppressing the host immune response, PLpro is an attractive target for antiviral drug development. Thousands of compounds, including approved drugs and molecules in clinical trials, have been screened against this target, but the hit rate is extremely low compared with that of drug leads that target Mpro, another viral protease encoded by SARS-CoV-2. The peptidomimetic inhibitors VIR250 and VIR251 were the first identified covalent inhibitors of PLpro (ref. 77) (fig. 4c). A catalytic residue, C111, of PLpro engages in a Michael addition reaction with the β-carbon of the vinyl group of the vinylmethyl ester warheads from VIR250 and VIR251, resulting in the formation of a covalent thioether linkage. Residues at P2, P3 and P4 participate in an extensive network of hydrogen bonds and van der Waals interactions with their corresponding subsites. Similar substrate preferences and catalytic efficiencies are observed for SARS-CoV-2 and SARS-CoV PLpro, suggesting that inhibitors of
**Napththalene group**
A chemical group which is composed of two aromatic rings sharing two adjacent carbon atoms.

**Half-maximal effective concentration (EC₅₀).** A quantitative measure that indicates how much of a substance (for example, an antiviral agent) is effective in inducing a response (for example, eliminating a virus in cultured cells) halfway between the baseline level and the maximal level.

**Michael acceptor**
The activated alkene in an α,β-unsaturated carbonyl compound, which is involved in the Michael addition reaction.

SARS-CoV PL₆₁₇ is a good starting point for lead compound optimization against SARS-CoV-2. GRL0617, an inhibitor of SARS-CoV PL₆₁₇, also inhibits SARS-CoV-2 PL₆₁₇ (REF 79). Structural studies show that GRL0617 fits in the substrate cleft which was formed between the BL2 loop and the loop connecting α3 and α4, where it occupies the S3 and S4 subsites. The aromatic ring of GRL0617 fits into the S3 subsite, while the napththalene group fills the S4 subsite. Thus, the binding of GRL0617 blocks the substrate from gaining access to the active site. Inspired by the success of GRL0617, several naphtalene-based compounds were synthesized and also show good inhibition of SARS-CoV-2 PL₆₁₇ (REF 79). YM155, an anticaner drug candidate in clinical trials, has also been shown to inhibit SARS-CoV-2 PL₆₁₇ and has potent antiviral activity (half-maximal effective concentration (EC₅₀) of 170 nM) 80. YM155 achieves such a strong inhibition by simultaneously recognizing three hotspots in PL₆₁₇. The first binding site is located at the entrance of the substrate-binding pocket and blocks substrate entry to the active site. The second is located on the thumb domain and hampers interactions between PL₆₁₇ and ISG15. The third site is located on the zinc-finger motif, and the binding perturbs the stability of the zinc-finger motif and enzyme activity.

**M₆₁₇**
M₆₁₇ is the major protease encoded by SARS-CoV-2. It cleaves replicate polyproteins at no fewer than 11 sites to release NSPs, allowing the assembly of the viral replication and transcription machinery. The pivotal role that M₆₁₇ plays in regulating viral replication and transcription makes it an attractive drug target. Crystal structures show that this 306 amino acid protease comprises three domains (domain I, residues 10–99; domain II, residues 100–182; and domain III, residues 198–303) and adopts a chymotrypsin-like fold. Due to the similar substrate specificity and presence of a cysteine as a catalytic residue, M₆₁₇ is classified as a 3C-like protease.

Since the first crystal structure of SARS-CoV-2 M₆₁₇ in complex with a Michael acceptor inhibitor N3 (Protein Data Bank accession code 6LU7) (FIG. 5a) was published 81, many structures of M₆₁₇ in complex with inhibitors have been reported. SARS-CoV-2 M₆₁₇ functions as an active homodimer, in which the two protomers are nearly perpendicular to each other. The N-terminal finger (residues 1–7) of one protomer inserts itself between domains II and III of its neighbouring protomer, and promotes the formation of the dimer and the S1 subsite in the neighbouring protomer 82. Dimerization is additionally regulated by domain III through a salt-bridge interaction between E290 of one protomer and R4 from its adjacent protomer. In each protomer, a deep cleft between domains I and II forms the substrate-binding site, with a catalytic dyad (H41 and C145) at its centre. Domain III contains five α-helices that arrange themselves into a large antiparallel globular cluster and exhibit a unique topology in coronaviruses. Domains II and III are connected by a long loop (residues 183–198).

Coronavirus M₆₁₇ recognizes the P₄–P₁ positions of the substrate 83–86 (FIG. 5b). The S1 subsite has an absolute preference for glutamine at P1. P2 is usually a bulky side chain that can be accommodated by the deep hydrophobic S2 subsite. The P3 side chain is solvent exposed, and the corresponding S3 subsite also shows tolerance to a wide range of functional groups. The hydrophobic S4 subsite is smaller than S2 and thus accommodates residues with small side chains. This binding pocket is highly conserved among coronavirus M₆₁₇s, suggesting that antiviral inhibitors targeting this pocket should have broad-spectrum activity against coronaviruses in general.

**Inhibitors of SARS-CoV-2 M₆₁₇**
Recently, numerous inhibitors of M₆₁₇ have been identified exhibiting a range of binding mechanisms (FIG. 5c). N3 is the representative peptidomimetic inhibitor, and harbours a Michael acceptor as a warhead and substituents spanning all substrate-binding subsites. The Michael acceptor forms a covalent bond with the active site residue, C145. N3 bears a lactam ring, an aliphatic isobutyl group, an isopropyl group, a methyl group and an isoxazole as the side chain for the P1–P5 sites, respectively. The lactam ring, which replaces glutamine at the P1 site, exhibits favourable binding at the S1 subsite 87,88. Studies have shown that N3 displays strong inhibition of M₆₁₇s from different coronaviruses, and it could inhibit SARS-CoV-2 with EC₅₀ of 16.77 μM in a Vero cell-based assay. This value may not be truly representative of activity as it is not clear whether the high levels of expression of the efflux transporter P-glycoprotein in Vero cells affected the evaluation of its antiviral efficacy 89.

A recent study reported a series of α-ketoamides that inhibit SARS-CoV-2 M₆₁₇ (REF 90). Distinct from the previously designed α-ketoamides, the P2–P3 amide bond is replaced with a pyridone ring, which increases the half-life in plasma. Replacement of the P2 cyclohexyl moiety with smaller cyclopropyl increases the antiviral activity against betacoronaviruses. Approved hepatitis C virus drugs, such as boceprevir, telaprevir and narlaprevir, are α-ketoamide inhibitors and also exhibit inhibition of SARS-CoV-2 M₆₁₇. The ketone group undergoes a nucleophilic attack by the C145 thiolate to form a hemithioketal. Because boceprevir, telaprevir and narlaprevir are peptidomimetic inhibitors with similar structures, they form very similar interactions with the S1–S4 subsites 80. Another ketone-based potent inhibitor was discovered in the hydroxymethylketone class 91. One of the hydroxymethylketone derivatives demonstrated inhibition of SARS-CoV-2 M₆₁₇ and also possesses antiviral activity with EC₅₀ of 4.8 μM.

Another study presented two peptidomimetic aldehydes (named ‘11a’ and ‘11b’) which bear an indole moiety at the N terminus (P3 site) and an aldehyde warhead at the C terminus 92. The complex structures show that the aldehyde groups covalently bind to C145 of the catalytic dyad to inhibit M₆₁₇ activity. Both inhibitors exhibited excellent inhibition of SARS-CoV-2 M₆₁₇ with half-maximal inhibitory concentrations of 0.043 μM and 0.040 μM, respectively. The inhibitors also exhibited strong anti-SARS-CoV-2 infection activity in Vero cell-based assays and good pharmacokinetic and toxicity properties. A recent study reported another series of aldehyde derivatives with EC₅₀ ranging from 7.6 to 748.5 nM in cell-based assays. In a transgenic mouse model of SARS-CoV-2 infection, oral or intraperitoneal...
treatment with two compounds, MI-09 or MI-30, significantly reduced lung viral loads and lung lesions. Both also displayed good pharmacokinetic properties and safety in rats. GC376, an inhibitor of feline infectious peritonitis virus in preclinical studies, has been found to efficaciously inhibit SARS-CoV-2 in Vero cells by targeting Mpro. It utilizes an aldehyde bisulfite to covalently bind to C145 (refs 94, 95). Based on 11a, 11b and GC376, a number of aldehyde-based dipeptidyl and tripeptidyl inhibitors of Mpro were designed, and the organocatalyst-mediated protein aldol ligation to C145 of the protease occurs. A series of Mpro inhibitors that possess an aldehyde group for covalent inhibition have been reported. Among them, two compounds inhibited SARS-CoV-2 replication in cultured primary human airway epithelial cells.

The repurposing of approved drugs, drug candidates and pharmacologically active compounds provides an alternative approach to identify potential drug leads that could rapidly be approved as clinical treatments for COVID-19. Through high-throughput screening, one study identified multiple drug leads that target Mpro, including ebselen, disulfiram and carmofur. Ebselen exhibited antiviral activity in a plaque-reduction assay (EC_{50} = 4.67 μM). As an organoselenium compound, ebselen was previously investigated for treatment of bipolar disorders and hearing loss. It has been shown to have low cytotoxicity in humans in clinical trials. Ebselen has been approved by the US Food and Drug Administration to enter phase II clinical trials (NCT04484025 and NCT04483973) for COVID-19 treatment. Carmofur, which also exhibited antiviral activity in vitro, is a derivative of 5-fluorouracil. It is an approved antineoplastic agent, and has been investigated as a cancer treatment. As observed in the complex structure of Mpro and carmofur, the catalytic C145 residue is covalently bound to the carbonyl reactive group of carmofur and its fatty acid tail extends into the hydrophobic S2 subsite. Such a novel inhibitory mode makes carmofur a good lead compound for rational drug design. GRL-1720 and 5h were also identified as covalent inhibitors targeting Mpro through high-throughput screening.

**Fig. 5 | Structures of SARS-CoV-2 M^{pro} and its inhibitors.**

(a) Structure of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) main protease (M^{pro}) in complex with inhibitor N3. Protomer A is shown as a cartoon, and protomer B is shown as a surface representation. The surface representation of the substrate-binding site and N3 is shown in the right panel. Subsites S1, S1, S2 and S4 are labelled. (b) Interactions between N3 and SARS-CoV-2 M^{pro}. P1–P5 are labelled in N3. The residues interacting with P1 are shown as cyan sticks, and the residues forming the S1, S2 and S4 subsites are shown as green sticks, white sticks and orange sticks, respectively. The residues interacting with P5 are shown as blue sticks. Intermolecular hydrogen bonds are shown as dashed lines. (c) Surface representation of the substrate-binding site and with various inhibitors bound. GC376, boceprevir, carmofur and 11a are coloured yellow, blue, green and orange, respectively. All inhibitors are presented in stick form. Protein Data Bank accession codes are indicated in parentheses.

**Antineoplastic agent**
A drug for cancer treatment. Such drugs interfere with the ability of a cancer cell to grow and spread.
screening. Crystal structures show that both GRL-1720 and Shikonin form extensive interactions with C145 and other residues in the Mpro active site\textsuperscript{102}.

A recent study performed large-scale fragment screening against Mpro by combining mass spectrometry and X-ray approaches\textsuperscript{103}. Seventy-one hits were identified to bind at the substrate-binding site, and three hits were found to bind near the dimer interface. These structures provide a starting point to design more elaborate and potent drug leads that target SARS-CoV-2 Mpro.

Another study performed a high-throughput X-ray crystallographic screening of two drug repurposing libraries (the Fraunhofer IME Repurposing Collection and the Safe-in-Man library from Dompé Farmaceutici) against the SARS-CoV-2 Mpro (REF\textsuperscript{104}); the study authors identified 37 compounds that bind to Mpro. In subsequent cell-based assays, one peptidomimetic compound (callpeptin) and six non-peptidic compounds showed antiviral activity at non-toxic concentrations. Additionally, two allosteric binding sites representing potential targets against SARS-CoV-2 were identified. The first allosteric site is in the immediate vicinity of the S1 pocket of the adjacent protomer within the native dimer. The second allosteric site is formed by the deep groove between the catalytic domain and the dimerization domain.

Baicalin and baicalein, which are natural products derived from the flowering plant Scutellaria baicalensis, have been shown to inhibit SARS-CoV-2 Mpro with half-maximal inhibitory concentrations of 6.41 μM and 0.94 μM, respectively\textsuperscript{105}. The structure of Mpro in complex with baicalein shows that the phenyl ring with three hydroxy groups forms π–S and π–π interactions with C145 and H41 of the catalytic dyad, while the hydroxy groups form multiple hydrogen bonds with the S1 subsite. The distal phenyl ring occupied the S2 subsite. Another example is shikonin\textsuperscript{106}. The complex structure shows that shikonin forms a hydrogen bond network with the catalytic dyad C145 and H164 located in the S1 subsite. The aromatic head groups of shikonin form a π–π interaction with H41 on the S2 subsite. The hydroxy and methyl groups of the isohexenyl side chain of the shikonin tail form hydrogen bonds with R188 and Q189, respectively, in the S3 subsite. Such a unique mode of action expands our knowledge of Mpro inhibition.

**Replication and transcription complex**

**Replication mechanism of the central RTC**

In coronavirus infection, replication and transcription is regulated through a multisubunit mechanism\textsuperscript{107}, where the RdRP nsP12 catalyses viral RNA synthesis and thus acts as the key component of the RTC\textsuperscript{108}. In addition, the primase nsP8 (REF\textsuperscript{109}) and an auxiliary factor, nsP7, contribute to the activation and continuous production of viral RNA\textsuperscript{110}. nsP12 along with nsP7 and nsP8 makes up the complete RdRP complex.

SARS-CoV-2 nsP12 is composed of three major domains, a nidovirus RdRp-associated nucleotidytransferase (NiRAN) domain, an interface domain and a right-handed RdRp domain (finger, palm and thumb)\textsuperscript{111} (FIG. 6a). The active site of SARS-CoV-2 RdRp is located in the palm subdomain, which has a shape like other RNA polymerases, such as those from hepatitis C virus nsP5\textsuperscript{112} and poliovirus 3Dpol\textsuperscript{113}. The architecture of the central cavity is shared by other conserved polymerases involving the primer-template entry, nucleoside triphosphate (NTP) entry and nascent strand exit paths. Residues D760 and D761 are involved in the coordination of two Mg\textsuperscript{2+} ions essential for polymerase activity. One Mg\textsuperscript{2+} ion coordinates motif C and binds at the 3′ end (‘i’ site) of the RNA primer, facilitating the condensation reaction in RNA chain synthesis, while the second Mg\textsuperscript{2+} positions the incoming NTP and stabilizes the charge environment. Separate from conserved motifs A–E at the active site, motif F and motif G inside the fingers subdomain are conducive to guiding the RNA template. During viral RNA synthesis, notable structural rearrangements occur in this complex to accommodate the RNA\textsuperscript{114}. Along with the product chain synthesis, the protruding RNA template–product duplex exits through the active site without steric hindrance and extends to two positively charged ‘sliding poles’ formed by two nsP8 N-terminal helices\textsuperscript{115} (FIG. 6b). Consistent with SARS-CoV nsP8 adopting variable conformations\textsuperscript{116,117}, N-terminal extensions of nsP8-2 (the second copy of nsP8) have two different orientations at the early replicating stage. In one orientation, it is adjacent to the finger subdomain, whereas in the other orientation, it interacts with the RNA duplex, suggesting that nsP8 may have regulatory functions in replication initiation. The complex consisting of nsP12, nsP7, nsP8 and RNA duplex reflects the replicating state in RdRP activity; therefore, it is referred to as the central RTC (C-RTC).

**RNA elongation, capping and backtracking**

The RTC needs to guarantee processive RNA duplex elongation without template–product dissociation so that viral genome or subgenome synthesis can be rapidly completed inside the host cell\textsuperscript{116}. For coronaviruses, which have the largest known positive-sense RNA genomes, both replication efficiency and replication fidelity are essential for maintaining genetic integrity. The former relies on the functional elongation RTC (E-RTC), whereas the latter depends on proofreading by nsP14. An E-RTC is composed of a C-RTC and two coupled copies of the nsP13 helicase: nsP13-1 and nsP13-2 (REF\textsuperscript{118}) (FIG. 6c). nsP13 is believed to be crucial in viral replication and the mRNA capping process, which includes unwinding of the RNA duplex into single strands, 5′ to 3′ polarity formation and RNA 5′-triphosphatase activity\textsuperscript{119,120,121}. The unique domains of coronavirus nsP13, such as the zinc-binding domain, the stalk and the 1B domain, are all important for helicase activity\textsuperscript{122}. In the structure of E-RTC, two nsP13 zinc-binding domains form extensive interactions with two nsP8 N-terminal helices. In particular, the zinc-binding domain from nsP13-2 forms additional interactions with the nsP8 thumb subdomain, stabilizing the overall structure during elongation\textsuperscript{119,122}. Before entering the nsP12 active site, the template RNA strand undergoes disruption of RNA secondary structure and guidance between the nsP13-2 RecA domain and the 1B domain to ensure the 5′ to 3′ translocation direction\textsuperscript{124}. Structural characterization of E-RTC not only helps
elucidate the RNA elongation mechanisms but also sug-
gests different functional roles that nsp13 may play in
this event. In nsp13-2, residues N361 in the domain 1A,
S468, T532 and D534 in the domain 2A and R178 and
H230 in the domain 1B collectively contribute to tem-
plate RNA recognition and elongation, demonstrating
that nsp13-2 is directly involved in positioning down-
stream template RNA. Interestingly, the interactions
between the nsp13-1 1B domain and the nsp13-2
1B domain have been shown to play a pivotal role in
E-RTC helicase activity, even though nsp13-1 is far from
nsp13-2 (REF 12) (FIG. 6d). Therefore, nsp13-1 is indis-
ispensable for RNA elongation in that it is cooperatively
coupled with nsp13-2 in the functioning E-RTC.

The capping modification of mRNA, which rigor-
sously follows subgenomic mRNA synthesis, is essential
Prodrug
A pharmacologically inactive substance which can be converted into a pharmacologically active drug in vivo by metabolic reactions.

for viral translation and propagation, mRNA protection and escape from host immune response. Similarly to the RNA elongation process, multiple NSPs participate in RTC assembly during sequential stages of mRNA capping, which can be divided into four main steps: (1) removal of the γ-phosphate of 5′-pppA by nsp13 with RNA 5′-triphosphatase activity; (2) transfer of GMP to 5′-pppA by the nsp12 NiRAN domain with guanylyltransferase (GTase) activity, leading to the generation of a GpppA cap structure; (3) methylation of N7-guanine by nsp14, which has N7-methyltransferase activity; and (4) methylation of the ribose 2′-O nucleotide into the final O-methyl cap structure by nsp1, which has 2′-O-methyltransferase activity. Multiple NSPs are assembled into the RTC in order according to their functional roles, a process which is accompanied by structural conformational changes. On one hand, the nsp12 NiRAN domain is involved in the second step to catalyse the ppA to GpppA transfer through its newly identified GTase activity. On the other hand, an intermediate state which has been captured by cryo-EM, shows that nsp9 can inhibit the GTase activity by tight insertion into the NiRAN catalytic centre in order to terminate the reaction. nsp9 is an RNA-binding protein, which is characterized by a positively charged groove. This groove, together with a β-hairpin at the nsp12 N terminus, provides an exit path for postcatalytic GpppA-RNA. Several hydrophobic interactions and hydrogen bonds enhance nsp9 binding to nsp12, suggesting that nsp9 plays a substantial role in the viral life cycle. Because it has been shown that disruption of the nsp9–nsp10 cleavage site is not lethal and nsp10 is able to tightly bind to nsp14 or nsp16, nsp9 may serve as a core regulator in recruiting the nsp10–nsp14 or nsp10–nsp16 complex for the following capping RTC assembly with N7-methyltransferase activity and 2′-O-methyltransferase activity.

Another important aspect relating to the RTC is its proofreading mechanism. Most RNA viruses replicate with estimated error rates between 10−3 and 10−5, which results in approximately one mutation per genome per round of replication for a typical ~10-kb genome, a much higher mutation rate than occurs in cellular DNA replication. The lower fidelity may largely be due to the lack of proofreading activity in these viruses. By contrast, SARS-CoV-2, which encodes nsp14 (an exonuclease with proofreading activity), can maintain high fidelity during replication of its large genome. Proofreading involves the backtracking of mismatched template–product RNA chains. The single-stranded 3′ segment of the product RNA generated by backtracking extrudes through the RdRP NTP entry tunnel. Then a mismatched nucleotide located at the 3′ end of product RNA enters the conserved NTP entry tunnel to initiate backtracking, and meanwhile, nsp13 stimulates RdRP backtracking. The structure of C-RTC in complex with the essential nsp13 helicase and RNA 5′-pppA suggests that the helicase can facilitate the backtracking mechanism.

RdRP inhibitor discovery
The RdRP is a prime drug target for SARS-CoV-2 (FIG. 1a). Inhibition of RdRP activity will prevent viral replication and can potentially achieve clinical efficacy. Major efforts have been devoted to identify both nucleotide and non-nucleotide inhibitors, which have also been used as probes to understand the replication cycle of SARS-CoV-2 and to provide a basis for development of broad-spectrum antiviral drugs.

The prodrug remdesivir, which was initially developed for the treatment of Ebola virus infection, shows good activity against SARS-CoV-2 in vitro assays but limited efficacy in clinical trials. In the cell, remdesivir is phosphorylated to remdesivir triphosphate, enabling it to act as an ATP analogue. The structure of pretranslocated catalytic C-RTC clearly demonstrates the incorporation mode of remdesivir and suggests its inhibition mechanism. Kinetic analysis shows remdesivir triphosphate is preferred as a substrate over ATP and terminates product chain elongation at a delayed position (i + 3). Once the inserted remdesivir monophosphate is transferred to the i + 3 position, the distance between the serine hydroxy oxygen from S861 and the 1′-cyano nitrogen from remdesivir monophosphate will be close to 2 Å, causing delayed chain termination. Further investigations indicate that an remdesivir-induced translocation barrier and RdRP stalling occur after the addition of three nucleotides upon incorporation of remdesivir into the product chain. Favipiravir is another nucleoside analogue that has been approved as an anti-influenza virus drug in Japan. Favipiravir simulates the incorporation of ATP and GTP into the product RNA, yet it inhibits its viral proliferation by increasing the mutation rate of the viral genome rather than causing product chain terminations. The structure of the RdRP–favipiravir complex delineates a precatalytic state and identifies the conserved residues for favipiravir recognition.

Although nucleotide inhibitors can be inserted into RNA chains, they can later be cleaved by proofreading activity. Thus, non-nucleotide inhibitors have been considered as an alternative approach for drug development. Suramin, a century-old drug used to treat African sleeping sickness and river blindness, can effectively inhibit SARS-CoV-2 polymerase activity with at least 20-fold
Accessory protein–host interactions

**ORF3a, ORF9b, ORF7a and ORF8**

ORF3a protein, encoded by ORF3a, is an ion channel membrane protein with 274 amino acids. It forms a potassium-sensitive channel and may promote virus release.

The cryo-EM structure of SARS-CoV-2 ORF3a is the first viroporin family structure determined in coronaviruses. The overall structure shows that ORF3a forms a dimer with the ion channel decorated with charged residues for cation conduction (Fig. 7a). It is noteworthy that ORF3a has a TRAF-binding domain at the N terminus that can activate NF-κB and the NLRP3 inflammasome, suggesting an important role in the host immune response. As ion channels are important therapeutic targets and many ion-channel drugs have already been approved for clinical trials, ORF3a is another good antiviral drug target.

ORF9b is encoded by an alternative ORF within the N protein gene. ORF9b suppresses the type I interferon immune response by interacting with the mitochondrial import receptor subunit TOM70. Targeting the interactions between ORF9b and TOM70 has been proposed as a therapeutic option for SARS-CoV-2. The structure of SARS-CoV-2 ORF9b shows that it is dimeric, with each protomer composed mainly of β-strands (Fig. 7b). The centre of the dimer has a hydrophobic environment for accommodating lipid molecules and membrane attachment.

ORF7a is a type I transmembrane protein and is also involved in virus–host interactions and protein trafficking within the ER and Golgi body. The ectodomain of ORF7a is a type I transmembrane protein and is also involved in virus–host interactions and protein trafficking within the endoplasmic reticulum and Golgi body. The ectodomain of ORF7a exhibits a seven-stranded β-sandwich fold and is shown in a cartoon representation.

ORF7a is a type I transmembrane protein and is also involved in virus–host interactions and protein trafficking within the ER and Golgi body. The ectodomain of ORF7a is a type I transmembrane protein and is also involved in virus–host interactions and protein trafficking within the endoplasmic reticulum and Golgi body. The ectodomain of ORF7a exhibits a seven-stranded β-sandwich fold and is shown in a cartoon representation.

ORF8 is an accessory protein that is composed of 121 amino acids. It has an N-terminal signal sequence and adopts an immunoglobulin-like fold (Fig. 7c). The structure of ORF8 shows that it can form a dimer, and each promotor of ORF8 contains eight antiparallel β-strands tied by three disulfide bonds. The covalently bonded dimer structure is stabilized by surface hydrophobic interactions and a series of hydrogen bonds. ORF8 is capable of assembling itself into large-scale homologous complexes; however, the oligomerization mechanism needs to be investigated further.

Conclusions

Coronaviruses have the largest genomes among all RNA viruses, encoding structural proteins and NSPs that achieve sustainability in a wide variety of ecological niches and hosts. Evolving viral proteins help coronaviruses to achieve host recognition and entry, genome replication, assembly and release of progeny viruses, and host immune surveillance evasion. In response to the COVID-19 pandemic, great efforts have been devoted to structural studies of SARS-CoV-2 proteins and viral–cellular protein complexes using X-ray crystallography and cryo-EM. Among them, the S protein, M

PLpro and RdRP are the most widely studied drug targets. A multidisciplinary combination of structural virology, omics technologies, immunology and virology will produce a more effective approach to structure-aided design of vaccines and therapeutics that have the potential for clinical use.

Published online 17 September 2021
1. Masters, P. S. The molecular biology of coronaviruses. *Adv. Virus Res.* **66**, 193–292 (2006).
2. Weiss, S. R. & Navas-Martin, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* **69**, 635–664 (2005).
3. Cavanagh, D. Coronaviruses in poultry and other birds. *Avian Dis.* **49**, 459–469 (2005).
4. Hamre, D. & Procknow, J. A. A new virus isolated from the human respiratory tract. *Proc. Soc. Exp. Biol. Med.* **171**, 190–193 (1986).
5. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of coronavirus pathogenesis. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
6. Fung, G. & Donnelly, D. X. Human coronavirus: host-pathogen interaction. *Annu. Rev. Microbiol.* **75**, 529–557 (2019).
7. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
8. World Health Organization. Middle East respiratory syndrome coronavirus. WHO https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-merscov#tab-1 (2020).
9. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
10. Harrison, A. C., Lin, T. & Wang, P. Mechanisms of SARS-CoV-2 transmission and Pathogenesis. *Trends Microbiol.* **28**, 1260–1263 (2020).
11. Johns Hopkins Coronavirus Resource Center. COVID-19 dashboard. JHU https://coronavirus.jhu.edu/ (2021).
12. Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.* **89**, 618–625 (2020).
13. Wang, M. Y. et al. SARS-CoV-2 structure, biology, and structure-based therapeutics development. *Front. Cell. Infect. Microbiol.* **10**, 587269 (2020).
14. Fauquier et al. Drug targets of SARS-CoV-2 and treatment opportunities for COVID-19. *Bioorg. Chem.* **9**, e01969-20 (2021).
15. Lu, X., Pan, J., Tao, J. & Guo, D. SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference. *Cell Res.* **31**, 1413–1416 (2021).
16. Chang, K. H., Hou, M. H., Chang, C. F., Hsiao, C. D. & Huang, T. H. The SARS-CoV-2 spike protein—functions and forms. *Antivir. Res.* **103**, 39–50 (2014).
17. Peng, Y. et al. Structures of the SARS-CoV-2 spike and Mpro for drug design. *EMBO J.* **39**, e105938 (2020).
18. Lee, L. et al. Attenuation of mouse hepatitis virus by deletion of the LRRK2G domain is restored by nsP1-I linkage. *PLoS ONE* **8**, e61166 (2013).
19. Narayanan, K. et al. Severe acute respiratory syndrome coronavirus nsP1 suppresses host gene expression, including that of type I interferon, in infected cells. *J. Virol.* **82**, 4471–4479 (2008).
20. Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K. G. & Makinson, S. A two-armed strategy to suppress host protein synthesis by SARS-CoV nucleoprotein. *Nat. Struct. Mol. Biol.* **27**, 1115–1125 (2020).
21. Lu, X. et al. Neutralizing antibodies block COVID-19 virus binding to its receptor and potentiate viral spread and innate immunity. *J. Virol.* **94**, e01969-20 (2020).
22. Harcourt, B. H. et al. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J. Virol.* **87**, 15600–15610 (2013).
23. Barretto, N. et al. The papain-like protease of severe acute respiratory syndrome coronavirus has dual enzymatic activity. *J. Virol.* **81**, 15189–15198 (2007).
24. Xiong, J. & Chen, Z. R. The role of ubiquitylation in immune defence and pathogen evasion. *Nat. Rev. Immunol.* **12**, 55–68 (2011).
25. Shn, D. et al. Papain-like protease regulates SARS-CoV viral spread and innate immunity. *Nature* **587**, 657–662 (2020).
26. Fu, Z. et al. The complex structure of GRK2L17 and SARS-CoV-2 Plpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **12**, 488–488 (2021).
27. Klemm, T. et al. Mechanism and inhibition of the papain-like protease, Plpro, of SARS-CoV-2. *EMBO J.* **39**, e106275 (2020).
28. Rut, W. et al. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 PLpro, a framework for anti-COVID-19 drug design. *Sci. Adv.* **6**, eabd4596 (2020).
29. Gao, X. et al. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm. Sinica.* **B**, 237–245 (2021).
30. Osiptu, J. et al. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **12**, 743–745 (2021).
31. Zhao, Y. et al. High-throughput screening identifies established drugs as SARS-CoV-2 Plpro inhibitors. *Protein Cell* https://doi.org/10.1007/s13238-021-00356-9 (2021).
32. Jin, Z. et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
82. Matthews, D. A. et al. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell 77, 761–771 (1994).

83. Anand, K. et al. Discovery of coronavirus main protease inhibitors reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. 21, 5213–5220 (2002).

84. Xue, X. et al. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J. Virol. 82, 2515–2527 (2008).

85. Yang, H. et al. De design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS Biol. 3, e24 (2005).

86. Yang, H. et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl Acad. Sci. USA 100, 13190–13195 (2003).

87. Yang, H., Bartlam, M. & Rao, Z. Drug design targeting the main protease, the achilles heel of coronaviruses. Curr. Pharm. Des. 15, 2876–2883 (2009).

88. de Vries, M. et al. A comparative analysis of SARS-CoV-2 antivirals characterizes SCL™ inhibitor PF-00855261 as a potential new treatment for COVID-19. J. Virol. 95, e01819-20 (2021).

89. Zhang, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved inhibitors. Science 368, 469–472 (2020).

90. Kneller, D. W. et al. Malleability of the SARS-CoV-2 SCL M™ active-site cavity facilitates binding of clinical antivirals. Structure 28, 1513–1520.e3 (2020).

91. Hoffman, R. L. et al. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases. Nat. Virol. 107, 22505–22510 (2010).

92. Wang, G. et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. Cell 182, 417–428.e3 (2020).

93. Hillen, H. S. et al. Structure of replicating SARS-CoV-2 polymerase reveals alternative mechanism. Science 368, 1003–1007 (2020).

94. Li, S. et al. New nsp13 isoform suggests mechanism for tuning viral RNA synthesis. Protein Cell 11, 198–204 (2010).

95. Zhai, Y. et al. Insights into SARS-CoV-2 transcription and replication from the structure of the nsp8-nsp12 hexamer. Nat. Struct. Mol. Biol. 12, 980–986 (2005).

96. Choi, K. H. Viral polymerases. Adv. Exp. Med. Biol. 726, 267–304 (2012).

97. Chen, J. et al. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. Cell 182, 1560–1573.e15 (2020).

98. Ivanov, K. A. et al. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J. Virol. 78, 5619–5632 (2004).

99. Ivanov, K. A. & Ziebuhr, J. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5′-triphosphatase activities. J. Virol. 78, 7835–7858 (2004).

100. Jia, Z. et al. Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis. Nucleic Acids Res. 47, 6538–6550 (2019).

101. Yan, L. et al. Architecture of a SARS-CoV-2 mini replication and transcription complex. Nat. Commun. 11, 5874 (2020).

102. Saikrishnan, K., Powell, B., Cook, N. J., Webb, M. R. & Wrigley, D. B. Mechanistic basis of SARS-CoV-2 NSP3 translocation in SF1B1 helices. Cell 157, 849–859 (2020).

103. Daffis, S. et al. 2′-O-methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature 468, 452–456 (2010).

104. Ramanathan, A., Robbins, G. B. & Chan, S. H. RNA methylation, biological functions and applications. Nucleic Acids Res. 44, 7511–7526 (2016).

105. Yan, L. et al. Cryo-EM structure of SARS-CoV-2 ORF7a as an immunomodulating factor for human CD8+ T cells. Cell 184, 184–193.e10 (2021).

106. Chen, Y. et al. Functional screen reveals SARS-CoV coronavirus nonstructural protein nsps14 as a novel cap N7 methyltransferase. PLoS Pathog. 17, e1009063 (2021).

107. Sutton, G. et al. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure 12, 541–553 (2004).

108. Deming, D. J., Graham, R. L., Denison, M. R. & Baric, R. S. Processing of open reading frame 1a replicase proteins nsps7 to nsp10 in murine hepatitis virus strain A59 replicase. J. Virol. 81, 10280–10290 (2007).

109. Ma, Y. et al. Structural basis and functional analysis of the SARS coronavirus nsp4-m10 complex. Proc. Natl Acad. Sci. USA 112, 9436–9441 (2015).

110. Rosas-Lemus, M. et al. High-resolution structures of the SARS-CoV 2′-O-methylester transferase reveal strategies for structure-based inhibitor design. Sci. Signal. 13, eabc202 (2020).

111. Smith, E. C., Sexton, N. R. & Denison, M. R. Thinking outside the triangle: replication fidelity of the largest RNA viruses. Annu. Rev. Virol. 11, 111–132 (2014).

112. Malone, B. et al. Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. Proc. Natl Acad. Sci. USA 118, e2012516118 (2021).

113. Williamson, B. N. et al. Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. Nature 565, 275–276 (2019).

114. Gordon, C. J. et al. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from several severe acute respiratory syndrome coronavirus 2 with high potency. J. Biol. Chem. 295, 6785–6797 (2020).

115. Koki, G. et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. Nat. Commun. 12, 279 (2021).

116. Peng, G. et al. Structural basis of SARS-CoV-2 polymerase inhibition by favipiravir. Innovation 2, 100800 (2020).

117. Yin, W. et al. Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by surnam. Nat. Struct. Mol. Biol. 28, 319–325 (2021).

118. Kern, D. M. et al. Cryo-EM structure of SARS-CoV-2 ORF5a in lipid nanoparticles. Nat. Struct. Mol. Biol. 28, 573–582 (2021).

119. Sui, K. L. et al. Severe acute respiratory coronavirus ORF5a protein activates the NLRP3 inflammasome by promoting TRAF5-dependent ubiquitination of ASC. FASEB J. 33, 8865–8877 (2019).

120. Zhong, X. et al. Amino terminus of the SARS coronavirus protein 3a elicits strong, potentially protective humoral responses in infected patients. J. Gen. Virol. 87, 569–573 (2006).

121. Gordon, D. E. et al. Comparative host-virus interaction protein networks reveal prion-like disease mechanisms. Science 370, eabe1403 (2020).

122. Zhou, Z. et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD8+ T cell monocytes. Sci. Transl. Med. 13, eabc2187 (2021).

123. Flower, T. G. et al. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. Proc. Natl Acad. Sci. USA 118, e201785118 (2021).