Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Structural insights into SARS-CoV-2 infection and therapeutics development

Gangyu Sun, Lulu Xue, Qingjing He, Yue Zhao, Wenqing Xu, Zhizhi Wang*

School of Life Science and Technology, ShanghaiTech University, Shanghai, China

ARTICLE INFO

Keywords: Structure SARS-CoV-2 Spike protein Main protease RdRp Nsp

ABSTRACT

The current COVID-19 pandemic is caused by the severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). By late October 2020, more than 43 million cases of infections, including over 1.15 million deaths, have been confirmed worldwide. This review focuses on our current understanding of SARS-CoV-2 from the perspective of the three-dimensional (3D) structures of SARS-CoV-2 viral proteins and their implications on therapeutics development against COVID-19.

1. Introduction

SARS-CoV-2 is a member of the Betacoronavirus genus that also includes severe acute respiratory syndrome coronavirus (SARS-CoV), murine hepatitis coronavirus (MHV) and Middle East respiratory syndrome coronavirus (MERS-CoV), etc. The single-stranded positive-sense RNA genome of SARS-CoV-2 is about 30 kb and consists of six major open reading frames (ORFs) and several other accessory genes (Fig. 1 A). Polyprotein (pp1a/1ab, translated from ORF1a and ORF1b, are cleaved into 16 non-structural proteins (nsps) through the collaboration of papain-like protease (PLpro) and main protease (Mpro) (Fig. 1 B). The remaining of the genome are translated to four major structural proteins: Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) proteins and several accessory proteins (Fig. 1 A) (Lu et al., 2020; Zhou et al., 2020b; Zhu et al., 2020).

Since the SARS-CoV-2 genome was sequenced, scientists all over the world have raced to solve the structures of SARS-CoV-2 proteins in order to reveal the structural basis of the SARS-CoV-2 viral infection, replication and assembly. We summarized here our current knowledge of their structures, especially the key drug targets, including S protein, Mpro, RNA-dependent RNA polymerase (RdRp), and their implications in therapeutics development.

2. Structures of SARS-CoV-2 proteins and their implications in therapeutics development

2.1. S protein

The infection mechanisms of the SARS-CoV-2 have been extensively studied. For all enveloped viruses, membrane fusion is considered to play a key role in the viral infection through a dramatic conformational change of the S protein. The S protein belongs to the class I fusion glycoproteins which mediate the attachment of virus to a variety of cell-surface receptors and subsequent fusion between viral and host cell surface (Belouzard et al., 2012). The SARS-CoV-2 S protein is a homotrimer, and each protomer consists of S1 and S2 subunits. S1 folds into 4 subdomains which are N-terminal domain (NTD), receptor binding domain (RBD), C-terminal domain 1 (CTD1) and C-terminal domain 2 (CTD2) to protect the trimerized pre-fusion S2 core. S2 includes fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM) and C-terminal domain (CT) (Fig. 2 A) (Wrapp et al., 2020b).

2.1.1. Mechanism of membrane fusion through the S protein

SARS-CoV-2 uses human angiotensin converting enzyme 2 (ACE2) as the key cellular receptor to enter cells, the same as SARS-CoV (Letko et al., 2020; Zhou et al., 2020b). Structures of the RBD of SARS-CoV-2 S protein in complex with ACE2 were determined in the early stage of the epidemic (Lan et al., 2020; Shang et al., 2020b; Wang et al., 2020c; Yan...
structures are in grey.

Fig. 1. Schematic diagram of the SARS-CoV-2 genome. (A) SARS-CoV-2 genome. (B) Pp1a, polyprotein 1a; pp1ab, polyprotein 1ab. Proteins with no reported 3D structures are in grey.
also shifts the S structure to opening conformation and facilitate the ACE2 binding (Clausen et al., 2020). Furthermore, free fatty acid linoleic acid (LA) can also change the conformational distribution of SARS-CoV-2 S protein. In the presence of LA, the probability of closed conformation of SARS-CoV-2 S protein increases to 70%, which indicates that LA binding stabilizes the closed conformation and may reduce the S-ACE2 interaction and infection (Toelzer et al., 2020).

2.1.3. Antibodies and therapeutics targeting the S protein

Discovering the structural details of antibodies in complex with S protein of SARS-CoV-2 is critical for vaccine design and immunotherapeutic development. CR3022 is the first structurally characterized antibody, previously isolated from the plasma of convalescent SARS individuals, and interacts with the RBD domain of SARS-CoV-2 S protein (Yuan et al., 2020b). Unfortunately, although it could interact with SARS-CoV S, CR3022 could not effectively neutralize SARS-CoV-2. The crystal structure shows that CR3022 targets a highly conserved epitope from RBD, and enables cross-reactive binding between SARS-CoV and SARS-CoV-2. However, the epitope can only be accessed by CR3022 when at least 2 RBDs on the trimeric S protein are in the erect conformation, which is rare for SARS-CoV-2 (Yuan et al., 2020b).

Numerous antibodies have been structurally characterized that contains many distinct epitopes, including various regions in the RBD at different S conformations, as well as many different non-RBD epitopes such like NTD (Barnes et al., 2020a, 2020b; Brouwer et al., 2020; Chi et al., 2020). These antibodies can be classified into four categories by their neutralizing mechanisms: (I) those target the RBD and neutralize SARS-CoV-2 by competing with ACE2; (II) those target the RBD and lock the S protein in the closed conformation; (III) those bind to epitopes outside the RBD; (IV) those presumably neutralize virus through the unclear mechanism. Class I is the major category including most of the neutralizing antibodies, all of which recognize the epitopes around the top of the RBD and occludes ACE2 binding, even though their epitopes are slightly different (Fig. 3A, 3B) (Brouwer et al., 2020; Cao et al., 2020b; Du et al., 2020; Hansen et al., 2020; Ju et al., 2020; Lv et al., 2020; Shi et al., 2020; Wec et al., 2020; Yuan et al., 2020a; Zhou et al., 2020).

Fig. 2. SARS-CoV-2 S protein in different conformations. (A) Domain structure of the full-length SARS-CoV-2 S protein. (B) Structural model of SARS-CoV-2 S trimer. Three protomers are colored in green, blue and purple, with A in “RBD-erect” and B, C in “RBD-down” conformations, respectively (PDB code: 6VSX). (C) Structural components of SARS-CoV-2 S trimer in the same color scheme as Fig. 2A. The S1/S2 cleavage site and the S2' cleavage site are also indicated (PDB code: 6VSX). (D) Structural rearrangement in the process of membrane fusion. From the left to right are the S trimer in closed conformation (PDB code: 6VXX), the 1 RBD “erect” conformation (6VSX), and ACE2 bounded S protein. After the S1 dissociation, S2 components were exposed and the post-fusion intermediate was sequentially formed. Finally, folding back of HR2 formed the post-fusion conformation (PBD code: 6XRA) and brings the viral envelope close to the host cell membrane. The ACE2 bounded S obtained by superimposed with ACE2/RBD complex. The conformations masked in gray were modelled to illustrate the proposed membrane fusion process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Antibodies targeting the S protein of SARS-CoV-2 (Pinto et al., 2020). Class II antibodies behave like a lock that fixes the S trimers in the closed conformation and blocks ACE2 interaction (Fig. 3C) (Liu et al., 2020; Tortorici et al., 2020). Class III antibodies bind to regions outside RBD, such as NTD of the S protein. They may restrain conformational changes of the S protein needed for membrane fusion (Fig. 3D) (Chi et al., 2020; Liu et al., 2020). Class IV antibodies recognize epitopes also on the RBD of SARS-CoV-2 S but far away from the ACE2 binding site and generate no collision with ACE2. The underlying neutralizing mechanism of Class IV antibodies awaits further investigation (Fig. 3E) (Pinto et al., 2020).

Antibodies targeting the S protein of SARS-CoV-2 especially the RBD are broadly expected to play a key role in combating the COVID-19 pandemic. However, mutations in RBD rapidly appeared following the viral passageing, and resulted in the loss of antibody neutralization activity. Consequently, researchers reckon that a therapeutic antibody cocktail would have the potential to decrease the influence by virus mutations. The ideal antibody cocktail partners should bind to non-overlapping epitopes on the S protein which can prevent the virus escape mutants (Baum et al., 2020). For example, REGN10933 and REGN10987 have great potential to form an antibody cocktail (Hansen et al., 2020) as well as BD-368-2/BD-629 which can also successfully rescue mutation-induced neutralization escapes (Fig. 3F) (Du et al., 2020).

Besides the antibodies, researchers discovered many nanobodies that potently disrupt the cell entry of SARS-CoV-2. There are five structurally characterized nanobodies: three of them bind to the top of RBD and block the interaction with ACE2, function like the Class I antibodies (Hanke et al., 2020; Huo et al., 2020); another one locks the S protein in the closed conformation like the Class II antibodies (Schoof et al., 2020); the last one is similar with Class IV antibodies, the epitope of which is on the side of RBD and far away from ACE2 binding site (Wrapp et al., 2020a).

In addition to antibodies and nanobodies, de novo protein design, a rapidly developing interdisciplinary field (Huang et al., 2016), has also brought new hope to the therapeutics. Based on the structures of ACE2 and SARS-CoV-2 S protein, researchers used two major strategies to de novo design of RBD binders which can block the interaction with ACE2: (1) Incorporating the major elements of ACE2 which form the interactions with RBD into small proteins and redesigning the small proteins to attain higher neutralizing ability to the RBD (Cao et al., 2020a; Huang et al., 2020; Linsky et al., 2020); (2) Docking the RBD against a previously generated mini-protein libraries (Chevalier et al., 2017) to find new binding patterns, subsequently the sequences of the target mini-proteins are designed to optimize the binding affinity, neutralizing ability and stability (Cao et al., 2020a). These de novo designed mini-proteins have provided many great starting points for the SARS-CoV-2 therapeutics.

2.2. Mpro

Mpro (also known as 3CLpro) with 3C-like protease activity from coronavirus is a cysteine protease. The genomic Mpro of SARS-CoV-2 is located at nsp5 region in pp1a/pp1ab. Its prominent function is to cleave pp1a and pp1ab at 11 or more sites for viral replication, transcription and immune regulation (Kiemer et al., 2004). It cleaves the polyproteins mostly at the L-Q↓-(S/A/G) site. The lack of a similar protease in human ensures relatively low toxicity of the Mpro inhibitors (Zhang et al., 2020c), making the Mpro of SARS-CoV-2 an attractive drug target against SARS-CoV-2. The Mpro of SARS-CoV-2 shares over 96% sequence homology with SARS-CoV (Jin et al., 2020a), despite the fact that their associated viral behavior and pathological performance vary significantly (Lu et al., 2020).

Structural analyses have been performed with SARS-CoV-2 Mpro and its complexes with various ligands for drug discovery. SARS-CoV-2 Mpro is composed of three domains. Domain I and II contain six-strand anti-parallel β barrels, and domain III contains five helices (Fig. 4A) (Dai et al., 2020; Jin et al., 2020a; Zhang et al., 2020c). In the absence of ligands, active SARS-CoV-2 Mpro forms homodimers mainly through interactions between the domain III of the two protomers, which is critical for its catalytic activity by forming substrate-binding pocket (Fig. 4B) (Dai et al., 2020; Kneller et al., 2020; Zhang et al., 2020c). The Cys145 and His41 catalytic dyad and the substrate binding site are located in the cleft between domains I and II (Fig. 4A) (Dai et al., 2020; Jin et al., 2020a; Zhang et al., 2020c). Structural comparison revealed that...
residues Thr$^{285}$ and Ile$^{286}$ in SARS-CoV Mpro were replaced by Ala and Leu in SARS-CoV-2 respectively, which resulted in a more closed conformation of the two domain III in the Mpro dimer and may explain the increased catalytic activity in SARS-CoV-2 Mpro (Zhang et al., 2020c).

To discover lead compounds, the crystal structure of SARS-CoV-2 Mpro with the broad-spectrum antiviral inhibitor N3 were determined and elucidated their binding details (Jin et al., 2020a), and the binding mode was similar to these of SARS-CoV-2 Mpro with other ligands (Dai et al., 2020). In the symmetric complex composed of two Mpro protomers and one N3 inhibitor, N3 binds covalently with Cys$^{145}$ in the substrate binding pocket, and its kinetic analysis confirmed an irreversible inhibitory mechanism (Jin et al., 2020a). Furthermore, researchers turned attention to designing and modifying specific compounds based on structures and worked on repurposing approved drugs or operating high-throughput screening from natural compounds or databases. For example, the covalent interactions between Mpro and antineoplastic carmofur were revealed with crystal structure (Jin et al., 2020b).

Since the studies mentioned above, structural studies of SARS-CoV-2 Mpro interacting with different types of inhibitors such as covalent, non-covalent or non-peptide inhibitors, continue to emerge, according to the PDB database (https://www.rcsb.org/). These studies have provided deep insights into structure-based drug design targeting Mpro.

### 2.3. RdRp

The genome of SARS-CoV-2 consists of a single-stranded positivesense RNA, which is directly translated by host cell factors (Faheem et al., 2020). In contrast, the viral nsp12 protein, also knowns as RdRp, acts as the core component in RNA synthesis instead of host polymerases (Eskier et al., 2020). Cofactors nsp7/nsp8 are required for full function to promote the binding between nsp12 and RNA template (Yin et al., 2020).

The complex structures of nsp12/nsp7/nsp8 of SARS-CoV-2 have been determined (Fig. 4C) (Gao et al., 2020b; Yin et al., 2020). The first successfully determined cryo-EM structure of SARS-CoV-2 nsp12/nsp7/nsp8 complex (PDB: 6M71; 7BTF), showed the complex structure of nsp12 with one nsp7-nsp8 pair and one nsp8 monomer. Compared with the nsp8 in the nsp7-nsp8 pair, the conformation of the nsp8 monomer was shifted. The nsp7-nsp8 pair was similar to the SARS-CoV nsp7-nsp8 pair, which showed the conservation between them. In addition, a unique “hairpin” domain at N terminal of RdRp was discovered, which stabilized the overall structure by forming a set of close contacts (Gao et al., 2020b). To investigate the specific function of cofactors, conformations of the complex in the stable extension stage were captured (PDB code: 6YYT; 7C2K). N-terminal regions of the two nsp8 subunits formed long α-helical extensions flanking exiting RNA duplex, which offered a place on upstream exist pathway for product RNA. The processivity of RdRp may be explained by the extension conformation of nsp8 and its interactions with RNA (Hillen et al., 2020; Wang et al., 2020b).

Since RdRp plays a vital role in the life cycle of SARS-CoV-2, it is proposed to be the target of a class of antiviral drugs that are nucleotide analogues, including Remdesivir. Remdesivir is an investigational nucleotide produrg of a 1′-cyano-substituted adenosine analog, which was initially developed for the treatment of Ebola virus. Although Remdesivir failed in a clinical trial during an Ebola outbreak, it has been recognized as a promising drug against RNA viruses including SARS-
CoV, MERS-CoV and SARS-CoV-2 (Pardo et al., 2020; Wang et al., 2020a). Structural studies were performed with RdRp in complex with RNA and Remdesivir, because Remdesivir in triphosphate (RDV-TP) format can compete with ATP and be incorporated into synthesis complex in Remdesivir monophosphate (RDV-MP) format. Structure and sequence alignments suggest that the mechanism of template RNA recognition and Remdesivir inhibition is highly conserved in multiple RdRp complexes from various RNA viruses. To gain insight into the mechanism of SARS-CoV-2 RNA replication and its inhibition by Remdesivir, 3D structures of SARS-CoV-2 RdRp-RNA complex with incorporated RDV-MP (PDB:7C2K; 7BV2) were determined (Wang et al., 2020b; Yin et al., 2020). Combined with biochemical evidence, the mechanism of Remdesivir-mediated RNA synthesis inhibition was proposed. Despite different structure resolutions and different RDV-MP insertion states, both studies agreed with previously proposed delayed chain termination hypothesis. The incorporation of RDV-MP into the product RNA chain at position “i” would cause the termination of the RNA synthesis at position “i + 3”, because the 1’–CN substituent of the incorporated RDV-MP would encounter a steric clash with the sidechain of Ser861 in RdRp and hamper further translocation, probably leading to a significant distortion of the positioning of the RNA and intercepting RNA replication (Fig. 4D) (Gordon et al., 2020a; Wang et al., 2020b; Yin et al., 2020).

The structural studies suggest the therapeutic potential of nucleotide analogues. The efficiency of incorporation of RDV-TP with its natural counterpart ATP and other RdRp inhibitors were compared, showing Favipiravir triphosphate and Ribavirin triphosphate were less well-incorporated by SARS-CoV-2 RdRp (Gordon et al., 2020a). Another nucleotide analog with broad-spectrum antiviral activity against SARS-CoV-2, MERS-CoV and SARS-CoV, β-D-N4-hydroxycytidine (NHC; EIDD-1931), showed increased potency against coronaviruses bearing resistance mutations to Remdesivir (Sheahan et al., 2020).

In addition to screening therapeutic drugs through experimental methods, computational approaches targeting RdRp complex catalytic cavity (including Asp760, and Asp761) were also wildly used in potential antiviral drug screening and design (Elfiky, 2020; Iftikhar et al., 2020).

2.4. Other potential drug targets with known structures

Although S, Mpro, and RdRp proteins are the most essential drug targets against SARS-CoV-2, extending the range of drug targets may expedite the process of therapeutic discovery and expand the diversity of available drugs, which may alleviate the probable antiviral drug-resistance problem. Here we summarize the current structural knowledge of other potential drug targets of SARS-CoV-2 as follows. They were grouped based on their functions (Fig. 5).

2.4.1. Host response suppression

Nsp1 proteins are the major virulence factors of alpha- and beta-coronavirus that show similar biological functions in suppressing host gene expression, facilitating efficient viral replication and immune evasion. SARS-CoV-2 nsp1 shows 84% amino acid sequence identity with SARS-CoV. Structural analysis of nsp1-40S ribosomal subunit complex revealed that the C terminus of nsp1, comprising two α-helices and two short loops, locates close to the “latch” between the head and body domains of the 40S ribosomal subunit and obstructs the mRNA entry channel, effectively blocking RIG-I-dependent innate immune responses (Fig. 5A) (Schubert et al., 2020; Thoms et al., 2020; Yuan et al., 2020c). Therefore, the structural characterization of the inhibitory mechanism of nsp1, provided a valuable insight into viral protein synthesis, and may aid structure-based drug design against SARS-CoV-2.

Nsp15 is a uridylate-specific endoribonuclease that preferentially...
cleaves both double- and single-stranded RNA substrates 3’ of uridines with different efficiencies, regulating the viral RNA to make it escape from the host’s immune recognition to evade antiviral defense of the host cell (Deng et al., 2019; Hackbart et al., 2020). The UTP bound structure of SARS-CoV-2 nsp15 showed it as a hexamer consisting of a dimer of trimers with key residues in the active site, which can recognize uridine and promote the catalysis of phosphodiester bonds, making the uridine derivatives like Tipiracil plausible therapeutic candidates (Fig. 5A) (Kim et al., 2020a, 2020b).

2.4.2. Replication, transcription, and RNA capping

The RdRp complex works with a number of accessory factors during the viral life cycle including nsp9 and key factors including nsp13. Nsp9 is a single-stranded RNA-binding subunit that assists the RNA-dependent RNA replicase machinery of coronaviruses to replicate and transcribe viral RNA. SARS-CoV-2 nsp9 shares a 97% amino acid sequence identity with SARS-CoV (Zhang et al., 2020a). The crystal structure of SARS-CoV-2 nsp9 is a horseshoe-shaped tetramer with two stable dimeric interfaces, providing a structural basis for understanding the assembly mechanism of RNA binding proteins (Fig. 5B) (Zhang et al., 2020a). However, it remains to be investigated how nsp9 dimerization may affect interactions with other replicase machinery subunits. In addition, there is a peptide binding site close to the dimer interface, but the identity and binding affinity of the peptide need to be determined (Littler et al., 2020; Zhang et al., 2020a). Several molecules, such as convivaptan and telmisartan, were identified as potent inhibitors using a computational approach of structure-based drug repurposing, and the efficacy of these compounds needs further verification (Chandel et al., 2020).

Nsp13 is a member of superfamily 1B helicase that has RNA and DNA duplex-unwinding activities with 5’-to-3’ polarity using the energy of nucleotide hydrolysis. In addition to helicase activity, nsp13 also has RNA 5’-triphosphatase activity, which may play an important role in formation of the 5’ cap structure of viral RNA. The SARS-CoV2 nsp13 shares a 99.8% sequence identity with SARS-CoV with only one different residue (White et al., 2020). Biochemical and biophysical studies suggested that nsp13 and nsp12 can work together to improve the efficiency of viral replication. Structural studies showed that nsp13 binds to RdRp complex with NTPase domain sitting in front of the replication-transcription complex (RTC), making the RTC dynamically retrospective for proofreading or changing template during the process of RNA transcription by backtracking. In this model, RdRp complex and nsp13 translocate on the same template RNA strand but in the opposite direction. When the translocation of nsp13 is dominant, which might push the RdRp backward on the template RNA, but the model needs to be verified experimentally (Fig. 5B) (Chen et al., 2020). Simeprevir, paritaprevir and grazoprevir were identified as potential inhibitors by a comparative homology modeling approach, and more biological assays needs to further confirm the finding (Gurung, 2020).

After the replication and transcription process, the RNA transcripts are further capped at the 5’ end. The capping of mRNA involves nsp13, nsp14, and nsp16/nsp10, which promotes initiation of translation, protects mRNAs and helps virus hijack the host’s translation machinery to produce viral proteins. The cap is added to the 5’ hydroxyl group of first nucleotide of the growing mRNA strand in four sequential steps (Souvet et al., 2010). Firstly, the 5’ triphosphate group is hydrolyzed by the triphosphatase nsp13 to make the diphosphate-RNA. Secondly, guanosine 5’-triphosphate added by unknown guanylyltransferase to produce the guanosine cap. Thirdly, a methyl group is transferred by the methyltransferase nsp14 to the guanosine cap to yield 7-methylguanosine cap. Lastly, nsp16/nsp10 complex, a cap ribose 2’-O-RNA methyltransferase, uses S-adenosyl methionine (SAM) as substrate to methylate the first nucleotide of the nascent mRNA at the ribose 2’-O position, causing the host cell fail to distinguish the non-self mRNA. By transforming viral mRNA from the Cap-0 state to the Cap-1 state, nsp16/nsp10 complex enhances the efficiency of translation and helps coronavirus destroy host innate immune responses by escaping the type I interferon suppression. The structures of SARS-CoV-2 nsp16/nsp10 in complex with the pan-methyltransferase inhibitor sinefungin or SAM demonstrate the catalytic mechanism of mRNA capping, making nsp16 a potential target to limit pathogenesis (Fig. 5B) (Krafcikova et al., 2020; Lin et al., 2020b; Rosas-Lemus et al., 2020; Viswanathan et al., 2020).

Additionally, the N protein, which is highly conserved among all coronaviruses, is a multifunctional RNA-binding protein. It binds the viral RNA and forms helical ribonucleoprotein, which is vital for viral RNA replication and packaging (Zhou et al., 2020c). In addition, it can interact with numerous host cell proteins to modulate the metabolism of infected cells (Gordon et al., 2020b). The N- and C-terminal structures of SARS-CoV-2N protein reveals the surface electrostatic potential characteristics and self-assembly properties, explaining the mechanism of its binding ability to single-stranded RNA (Fig. 5B) (Kang et al., 2020; Peng et al., 2020; Ye et al., 2020; Zhou et al., 2020c). Furthermore, compound PJ34 targeted to HCoV-OC43 and 5-benzoxoygamine targeted to MERS-CoV may be potential drugs to block SARS-CoV-2 replication, transcription and virus assembly based on conserved key residues in structures (Lin et al., 2020c; Peng et al., 2020).

2.4.3. Double-membrane vesicles formation

Coronavirus genome replication is related with virus-induced cytosolic double-membrane vesicles (DMVs), which can offer an isolated micro-environment in the infected cell for viral RNA synthesis (Snijder et al., 2020). Similar to SARS-CoV infected cells, DMVs were observed in cells when co-expressing nsp3, nsp4 and nsp6. However, the DMVs were not observed when nsp3 lacking its C-terminal (residues 1319–1922) was co-expressed with nsp4 and nsp6 (Angelini et al., 2013). Additionally, the co-expression of nsp4 and C-terminal of nsp3 (residues 1256–1922) can drive the formation of the ER membrane rearrangements where the coronavirus RTC assemble in MHV or SARS-CoV infected cells (Hagemeijer et al., 2014). The structure of SARS-CoV-2 DMV is still elusive. The low-resolution structure of DMV in MHV resolved by Cryo-ET showed that a molecular pore complex that spans both membranes of the DMV, and nsp3 is the main component of the pore structure (Wolf et al., 2020).

Except forming the DMV, the multi-domain and multifunctional nsp3, which is highly conserved between SARS-CoV and SARS-CoV-2, plays various roles in coronavirus life. Nsp3 can process viral polyproteins, which is essential for viral replication. It can also remove post-translational modifications on host proteins to evade host antiviral innate immune response, and interact with host proteins to support virus survival. Nsp3 consists several domains including Ubl1, HVR, Mac1, Mac2, Mac3, DPUP, Ub1b, PLpro, NAB, jsm, TM, Ecto, AH1, Y1 and CoV-Y, in which not every domain has clear structure features and functions (Lei et al., 2018). A few domain structures of SARS-CoV-2 nsp3 have been solved including Mac1 domain and PLpro domain (Fig. 5C). Mac1 domain can remove ADP-ribose from ADP-ribosylated proteins and the structures in the apo form and in complexes with ligands including 2-(N-morpholino)-ethanesulfonic acid, AMP and ADP-ribose have been solved, respectively (Alhammadi et al., 2020; Frick et al., 2020; Lin et al., 2020a; Michalska et al., 2020). PLpro domain recognizes the consensus cleavage sequence LXXG to process three cleavage sites at the boundaries of nsp1/2, nsp2/3 and nsp3/4, and to remove the ubiquitin and ISG15 from host proteins, showing deubiquitinating and delS-Glyating activities to antagonize innate immune response (Lei et al., 2018). The inhibitors of PLpro may provide a dual therapeutic effect – suppressing virus replication and enhancing antiviral immune response in the host. Small molecule inhibitors have been developed targeting SARS-CoV PLpro, including thiopurol compounds, tashinones, diallylephantanoids, gananylated flavonoids, zinc ion/zinc conjugate compounds, and naphthlene-based compounds (Baez-Santos et al., 2015). Since the SARS-CoV-2 PLpro and SARS-CoV PLpro share 83% sequence identity, these inhibitors may accelerate the drug development against SARS-CoV-2. However, PLpro is a cysteine protease which belongs to
ubiquitin specific protease family (USP) class with 56 members in human, and the catalytic domains among the USP class are highly conserved (Baez-Santos et al., 2015). As a result, it would be difficult to design the inhibitors selectively against PLpro but not human USPs, and the structure-based drug design approaches may help solve this problem. The structures of the SARS-CoV-2 PLpro domain in apo form, and its C111S mutant, in complex with ubiquitin, ISG15 or inhibitors have been solved, respectively (Gao et al., 2020a; Klemm et al., 2020; Rut et al., 2020; Shin et al., 2020), providing the structural basis of their substrate specificities. These inhibitors including VIR250, VIR251 and GRL0617, and rac5c, were identified by a combination of substrate library and activity analysis or in a repurposing approach, and the efficacy of these inhibitors needs to be further verified (Gao et al., 2020a; Klemm et al., 2020; Rut et al., 2020). As a key component for coronavirus replication and survival, other structure features and functions of nsp3 worth further investigation.

It should be noted that the structures and functions of nsp4 and nsp6 remain poorly understood. Currently, there is not any known 3D structure for any coronavirus nsp6 protein, and only one 3D structure for the C-terminal cytoplasmic domain of nsp4 from MHV (Xu et al., 2009).

2.4.4. Putative ion channels

The SARS-CoV-2 genome encodes two putative ion channels, 3a and E, which are highly conserved within the beta-coronavirus subgenus Sarbecovirus including bat coronaviruses (Kern et al., 2020; Srinivasan et al., 2020), suggesting that they are potential therapeutic targets to cure COVID-19.

SARS-CoV 3a forms an ion channel involved in viral release, and necrotic cell death, which is required for SARS-CoV replication and virulence (Castano-Rodriguez et al., 2018). SARS-CoV-2 3a forms large conductance cation channels in dimeric or tetrameric state and the high resolution structure may provide the structural basis for antiviral drugs design (Fig. 5D) (Kern et al., 2020).

SARS-CoV E protein forms cation-selective ion channels in endoplasmic reticulum - Golgi intermediate compartment and/or Golgi membranes that are permeable to calcium ions, which mediates the process of budding and releasing of progeny viruses and may activate the host immune response (Schoeman and Fielding, 2019). SARS-CoV-2 E forms a five-helix bundle surrounding a narrow central pore which can interact with ions, drugs and other host and viral proteins. The amantadine, known to inhibit the proton channels of influenza A virus and HIV-1, and hexamethylene amiloride both bind to polar residues in the N-terminal lumen, establishing the structural basis for designing E inhibitors against SARS-CoV-2 (Fig. 5D) (Mandala et al., 2020).

2.4.5. Orf8

ORF8 is one of the most rapidly evolving accessory proteins in all beta-coronaviruses and the SARS-CoV-2 ORF8 has less than 20% sequence identity to SARS-CoV ORF8 (Mohammad et al., 2020), which may be vital for understanding SARS-CoV-2 as a spread pathogen. SARS-CoV-2 ORF8 is presumably a secreted protein (Hachim et al., 2020) that interacts with a variety of host proteins, which inhibits type I interferon signaling pathway (Li et al., 2020) and mediates immune evasion through potently downregulating MHC-I (Zhang et al., 2020d). SARS-CoV-2 ORF8 has two novel dimer interfaces that can form unique large-scale assemblies which is impossible for SARS-CoV ORF8, potentially providing essential immune evasion activities (Fig. 5E) (Flower et al., 2020). The understanding of structure of SARS-CoV-2 ORF8 is very helpful to the development of novel therapies.

3. Perspectives

Although many structures of SARS-CoV-2 proteins have been solved by X-ray crystallography, Cryo-EM and NMR methods, and have provided much-needed structural basis for developing effective therapeutic approaches against COVID19, including inhibitors, antibodies and vaccines, the pathogenesis of SARS-CoV-2 infection and the molecular mechanism of virus-host interaction are still unclear. It is urgent to further study the host immune response to SARS-CoV-2 and the processes of SARS-CoV-2 infection, replication and assembly, which will lay a foundation for formulating prevention and treatment strategies against COVID-19 and future pandemics.

Funding

The work in the authors’ laboratory is currently supported by the ShanghaiTech University Startup grant to WX and the Shanghai Pujiang program (Grant # 20PJ1410500 to ZW) from Science and Technology Commission of Shanghai Municipality and Shanghai Municipal Human Resources and Social Security Bureau.

Author contributions

ZW conceived the review. GS, LX, QH, YZ, WX & ZW wrote the manuscript.

Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The manuscript was prepared in late October. During the period of manuscript submission and revision, new studies on SARS-CoV-2 continued to emerge. The authors regret that they cannot cite those works. The authors thank Ms. Zhizhuo Dai for her critical reading and comments on the manuscript.

References

Alhamad, Y.M.O., Kashipathy, M.M., Roy, A., Johnson, D.K., McDonald, P., Battaile, K.P., Gao, P., Lovell, S., Fehr, A.R., 2020. The SARS-CoV-2 conserved macrolactam is a highly efficient ADP-ribosehydrolase enzyme. bioRxiv.
Angelin, M.M., Abkhazpour, M., Neuman, B.W., Buchmeier, M.J., Mosesta, A., 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. mBio 4 (4).
Baez-Santos, Y.M., St John, S.E., Menacar, A.D., 2015. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. Antiviral Res. 115, 31–38.
Barnes, C.O., Jette, C.A., Abernathy, M.E., Dam, K.A., Enwein, S.R., Gristick, H.B., Malyutin, A.G., Sharaf, N.G., Huey-Tubman, K.E., Lee, Y.E., Robbiani, D.F., Nussenzweig, M.C., West, A.P., Bjorkman, P.J., 2020a. Structural classification of neutralizing antibodies against the SARS-CoV-2 spike receptor-binding domain suggests vaccine and therapeutic strategies. bioRxiv.
Barnes, C.O., West, J.R., Huey-Tubman, K.E., Hoffmann, M.A.G., Sharaf, N.G., Hoffman, P.R., Koranda, N., Gristick, H.B., Goebl, C., Muecksch, F., Lorenzi, J.C.C., Finkin, S., Hagglof, T., Hurley, A., Millard, K.G., Weisblum, Y., Schmidt, F., Hatziioannou, T., Bieniasz, P.D., Caskey, M., Robbiani, D.F., Nussenzweig, M.C., Bjorkman, P.J., 2020b. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. Cell 182 (4), 828–842.e814.
Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., Wei, Y., Arwal, G.S., Murphy, A.J., Stahl, N., Yacopoulos, G.D., Kyrtoussis, C.A., 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science 369 (6506), 1014–1018.
Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4 (6), 1011–1033.
Benton, D.J., Wrobel, A.G., Xu, P., Routtan, C., Martin, S.R., Rosenthal, P.B., Skelhel, J.J., Gamblin, S.J., 2020. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature.
Bouvet, M., Derbent, C., Imbert, L., Selisko, B., Sujjler, E.J., Canard, B., Decroly, E., 2010. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS Pathog. 6 (4), e1000863.
Brouwer, P.J.M., Caneel, T.G., van der Straten, K., Sluiterlaar, J.J., Aldon, Y., Bangaru, S., Torres, J.L., Okba, N.M.A., Claireaux, M., Kerster, G., Bentlage, A.E.H., van
Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Teng, Z., Zhang, W., Wu, L., Gao, M., Chang, S., Xie, Y.-C., Tian, G., Jiang, H.-W., Luo, S.C., Shen, J., Jiang, Y., Jiang, H., Xu, Y., Zhang, S., Zhang, Y., Xu, H.-E., 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science 368 (6486), 1499–1504.

Yuan, M., Liu, H., Wu, N.C., Lee, C.D., Zhu, X., Zhao, F., Huang, D., Yu, W., Hua, Y., Tien, H., Rogers, T.F., Landais, E., Sok, D., Jardine, J.G., Burton, D.R., Wilson, I.A., 2020a. Structural basis of a shared antibody response to SARS-CoV-2. Science 369 (6507), 1119–1123.

Yuan, M., Wu, N.C., Zhu, X., Lee, C.D., So, R.T.Y., Iiv, H., Mok, C.K.P., Wilson, I.A., 2020b. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV-1. Science 369 (6509), 620–623.

Yuan, S., Peng, L., Park, J.J., Hu, Y., Devzarkar, S.C., Dong, M.B., Wu, S., Chen, S., Lomakina, I., Xiong, Y., 2020c. Nonstructural protein 1 of SARS-CoV-2 is a potent pathogenicity factor redirecting host protein synthesis machinery toward viral RNA. bioRxiv.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T., Wang, Y., Baun, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S.F., Lemieux, J.E., Munro, J.B., Rafique, A., Barve, A., Saberi, P.C., Kyritis, C.A., Gestrichsa, N.V., Shen, K., Luban, J., 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell.

Zhang, C., Chen, Y., Li, L., Yang, Y., He, J., Chen, C., Su, D., 2020a. Structural basis for the multimerization of nonstructural protein nsq9 from SARS-CoV-2. Molecular Biomedicine 1 (1).

Zhang, L., Jackson, C.B., Mo, H., Ojha, A., Randolph, J., Izard, T., Farzan, M., Choe, H., 2020b. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv.

Zhang, L., Liu, D., Sun, X., Curbt, D., Drott, C., Sauermann, L., Becker, S., T regulations are required to control the SARS-CoV-2 Spike protein. bioRxiv.