Minireview of progress in the structural study of SARS-CoV-2 proteins

Guoliang Zhu\textsuperscript{a,b,1}, Chunmei Zhu\textsuperscript{a,b,1}, Yun Zhu\textsuperscript{a,*}, Fei Sun\textsuperscript{a,*}

\textsuperscript{a} National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China
\textsuperscript{b} University of Chinese Academy of Sciences, Beijing, China

\textbf{A R T I C L E I N F O}

\textbf{Keywords:}
COVID-19
SARS-CoV-2
Genome-encoded proteins
Structure-based screening
Drug screening

\textbf{A B S T R A C T}

A severe form of pneumonia, named coronavirus disease 2019 (COVID-19) by the World Health Organization, broke out in China and rapidly developed into a global pandemic, with millions of cases and hundreds of thousands of deaths reported globally. The novel coronavirus, which was designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified as the etiological agent of COVID-19. On the basis of experience accumulated following previous SARS-CoV and MERS-CoV outbreaks and research, a series of studies have been conducted rapidly, and major progress has been achieved with regard to the understanding of the phylogeny and genomic organization of SARS-CoV-2 in addition to its molecular mechanisms of infection and replication. In the present review, we summarized crucial developments in the elucidation of the structure and function of key SARS-CoV-2 proteins, especially the main protease, RNA-dependent RNA polymerase, spike glycoprotein, and nucleocapsid protein. Results of studies on their associated inhibitors and drugs have also been highlighted.

\textbf{1. Introduction}

In December 2019, a novel pneumonia named coronavirus disease 2019 (COVID-19), emerged in Wuhan, China (Zhu et al., 2020; Wu et al., 2020a). Its symptoms included dry cough, fever, headache, dyspnea, and pneumonia (Huang et al., 2020a; Liu et al., 2020; Wang et al., 2020a). Due to the high contagiousness and transmission rate of the virus during the presymptomatic phase (Hu et al., 2020; Huang et al., 2020b), COVID-19 became a pandemic and spread to more than 212 countries and territories, and community transmission took place in many countries including the United States, Germany, France, Spain, Japan, Singapore, South Korea, Iran, and Italy. Thus far, millions of cases and hundreds of thousands of deaths have been recorded, with rapidly increasing numbers globally.

A previously unknown coronavirus, designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which belongs to the \(\beta\)-coronaviruses and has a genomic sequence closely related to that of SARS-CoV, was determined as the etiological agent of this severe infection (Zhu et al., 2020; Coronavirusidae Study Group of the International Committee on Taxonomy of V 2020; Zhou et al., 2020; Wu et al., 2020b). The Coronavirusidae family includes two subfamilies: Letovirinae and Orthocoronavirinae, and the Orthocoronavirinae family consists of the \(\alpha\)-coronavirus, \(\beta\)-coronavirus, \(\gamma\)-coronavirus, and \(\delta\)-coronavirus genera (Cui et al., 2019). Many \(\beta\)-coronaviruses are human pathogens and cause severe respiratory diseases, including SARS-CoV, the Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV), and currently, SARS-CoV-2. Coronavirus are enveloped, positive-sense, single-stranded RNA viruses with mammalian and avian hosts. The length of the SARS-CoV-2 genome is approximately 30 kb (Zhou et al., 2020), and it encodes at least 29 proteins, including 16 non-structural proteins (NSP), 4 structural proteins, and 9 accessory proteins (Fig. 1).

Upon cell entry, the genomic RNA of SARS-CoV-2 is translated to produce two overlapping polyproteins, pp1a and pp1ab, from two open reading frames (ORFs), ORF1a and ORF1b, respectively. Subsequently, pp1ab is cleaved into 16 NSPs by viral proteases NSP3 and NSP5, which harbor a papain-like protease domain and a 3C-like protease domain.
domain, respectively (Fig. 1). The NSP12 (also known as the RNA-dependent RNA polymerase, RdRp) is the central component of the replication/transcription machinery in the synthesis of viral RNA, with the assistance of NSP7 and NSP8 as cofactors (Huang et al., 2020a; Subissi et al., 2014). SARS-CoV-2, as a positive-sense, single-stranded RNA virus, has the capacity to synthesize full-length negative-sense RNA chain, which serve as templates for further generation of positive-sense genomic RNA (gRNA) and subgenomic RNAs (sgRNAs). The gRNA is enveloped by structural proteins for progeny virion assembly, whereas the sgRNAs are relatively short and encode conserved structural proteins (spike [S] glycoprotein, envelope [E] protein, membrane [M] protein, and nucleocapsid [N] protein), and several accessory proteins. The nine accessory proteins (3a, 3b, 6, 7a, 7b, 8, 9b, 9c, and 10) of SARS-CoV-2 participate in various processes, ranging from coronavirus replication to resistance against immune responses (Fig. 1) (Wu et al., 2020b; Kim et al., 2020).

Elucidating the structural and functional features of the major proteins encoded by the SARS-CoV-2 genome would facilitate the development of viral specific drug therapies and vaccines to combat the global pandemic. With the tireless efforts of researchers, the atomic resolution structure of some key SARS-CoV-2 proteins have been resolved using cryo-Electron Microscopy (cryo-EM) or X-ray methods (Table 1), and their corresponding biological functions have been reported, promoting the rapid discovery of antiviral agents with clinical potential. The present review summarizes current research findings and developments with regard to the structure and function of major proteins encoded by SARS-CoV-2, and the insights could promote drug and vaccine development for the treatment and management of COVID-19.

2. Main protease (Mpro)

The functional polyproteins of SARS-CoV-2, pp1a and pp1ab, are cleaved into 16 NSPs, mainly by a 33.8-kDa main protease (Mpro), which is also referred to as 3C-like protease (3CLpro) (Anand et al., 2003; Hilgenfeld, 2014). There are at least 11 conserved Mpro restriction sites on the pp1ab protein, starting with its autolytic cleavage site (Hegyi and Ziebuhr, 2002). Because Mpro has no closely related homologues in humans, its specific inhibitor has been demonstrated to inhibit viral life cycle with no obvious biotoxicity (Pillaiyar et al., 2016; Hayden et al., 2003; Kim et al., 2016; Yang et al., 2005). Consequently, extensive studies on the structure of SARS-CoV-2 Mpro have been conducted rapidly following the COVID-19 outbreak (Anand et al., 2002; Yang et al., 2003). Jin et al., for the first time, reported the crystal structure of full-length SARS-CoV-2 Mpro in complex with N3 based on 2.1 Å X-ray diffraction data (Jin et al., 2020a). Almost at the same time, Zhang et al., determined the X-ray structures of the wild type SARS-CoV-2 Mpro protein at 1.75 Å, in addition to the structures of several inhibitor-bound Mpro proteins (Table 1) (Zhang et al., 2020).

All the aforementioned atomic structures of SARS-CoV-2 Mpro proteins are almost identical and highly similar to other Mpro proteins resolved previously (Ren et al., 2013; Tan et al., 2005; Wang et al., 2016; Xue et al., 2008), which suggests that the conserved protein is essential for the coronavirus family. Its overall structure is a dimer with a crystallographic 2-fold symmetry axis. Each protomer consists of three domains, Domains I (residues 8–101), II (residues 102–184) and III (residues 201–303), and a long loop region (residues 185–200) connects domains II and III (Fig. 2). The chymotrypsin-like and piconavirus 3C protease-like Domains I and II both form an antiparallel β-barrel structure, and the substrate-binding site is located in a cleft between them. Domain III is composed of five α-helices arranged into a largely antiparallel globular cluster, which forms a salt-bridge interaction between Glu290 of one monomer and Arg4 of the other monomer to mediate dimerization of the Mpro (Shi and Song, 2006). In addition, the dimer interface predominantly involves interactions between domain II of one protomer and the N-terminal residues of the other, which could facilitate the shaping of the substrate-binding pocket of Mpro protein (Fig. 2).
SARS-CoV-2 Mp^0 is an attractive and key target for antiviral reagent discovery. Several compounds that inhibit Mp^0 activity have been identified through a combination of structure-based virtual screening and high-throughput screening. For example, Wu et al., discovered a series of clinical drugs and natural products with antiviral, antibacterial, and anti-inflammatory activities that showed high affinity to SARS-CoV-2 Mp^0 by artificial intelligence-assisted computer virtual screening (Wu et al., 2020a). As reported previously, N3 could specifically inhibit multiple CoV Mp^0s, including those from SARS-CoV and MERS-CoV (Yang et al., 2005; Ren et al., 2013; Wang et al., 2016; Xue et al., 2008). Jin et al., resolved the crystal structure of SARS-CoV-2 Mp^0 in complex with N3 and demonstrated that N3 was a potent irreversible inhibitor of SARS-CoV-2 Mp^0. Subsequently, using structure-based virtual and high-throughput screening, they identified six promising antiviral compounds out of over 10,000 compounds, including approved drugs, drug candidates in clinical trials, and other pharmacologically active compounds as Mp^0 inhibitors (Jin et al., 2020b). Zhang et al., presented the crystal structure of SARS-CoV-2 Mp^0 in complex with an α-ketoamide inhibitor and developed the lead compound into a potent inhibitor based on the structure, which displayed obvious lung tropism and suitability for inhalation (Zhang et al., 2020). In addition, Dai et al., designed and synthesized two peptide-like compounds (11a and 11b) that exhibited high inhibitory activity against SARS-CoV-2 Mp^0 and viral infection. More importantly, both compounds showed better PK properties and lower toxicity in vivo, suggesting that they are promising drug candidates. Furthermore 1.5 Å high-resolution structures of SARS-CoV-2 Mp^0 in complex with 11a or 11b, revealed the precise interaction and molecular mechanism between Mp^0 and compounds (Dai et al., 2020).

### 3. The RNA-dependent RNA polymerase (RdRp, or NSP12)

Coronavirus employs a multi-subunit replication/transcription machinery of viral NSPs to synthesize viral RNA (Ziebuhr, 2005). The RNA-dependent RNA polymerase (RdRp, or NSP12) is the core component and plays a central role in SARS-CoV-2 replication and transcription cycle (Ahn et al., 2012; Velthuis et al., 2010). NSP7 and NSP8 increase RdRp activity in template-prime binding and processivity as accessory factors (Subissi et al., 2014; Kirchdoerfer and Ward, 2019). Therefore, RdRp is considered as the primary target for nucleotide analog antiviral inhibitors such as remdesivir (Holshue et al., 2020; Wang et al., 2020b; Warren et al., 2016), whose triphosphate form (RTP) is the active drug within cells (Siegel et al., 2017), and shows the potential for use in COVID-19 therapy. Therefore, determining the three-dimensional structures of SARS-CoV-2 RdRp with RNA template or inhibitors can provide a basis for the design of novel antiviral therapeutics (Table 1).

Gao et al., for the first time, reported the high-resolution cryo-EM structure of SARS-CoV-2 full-length NSP12 (residues 1–932) in complex with NSP7 (residues 1–83) and NSP8 (residues 1–198) cofactors at 2.9 Å (Gao et al., 2020a). It comprises one NSP12 monomer, one NSP8 monomer and one NSP7-NSP8 pair, similar to SARS-CoV RdRp complex (Kirchdoerfer and Ward, 2019). The structure of the SARS-CoV-2 NSP12 has three domains: a “right hand” RdRp domain (residues 367–920), a nidovirus-unique N-terminal extension domain (residues 4–28 and 69–249, also named as NiRAN domain) and an interface domain (residues 250–365) (Fig. 1). The RdRp domain displays the canonical arrangement of the viral polymerase family (McDonald, 2013) and consists of three subdomains: the finger subdomain (residues 366–581 and 621–679), the palm subdomain (residues 582–620 and 680–815), and the thumb subdomain (residues 816–920). The NiRAN domain adopts a nidovirus RdRp-associated nucleotidyl-transferase (NiRAN) configuration (Lehmann et al., 2015) with eight helices and a five-stranded β-sheet. The interface domain connects the other two domains with three helices and five β-strands (Fig. 2). Different from SARS-CoV’s, the RdRp structure of SARS-CoV-2 has an additional N-terminal β-hairpin (residues 29–50) which locates in the groove clamped by the NiRAN domain and the palm subdomain to stabilize the overall structure. The overall structure is further stabilized by the binding of NSP7 and NSP8, with one NSP8 monomer sitting on top of the finger subdomain and the NSP7-NSP8 pair interacting with the thumb-finger subdomains interface (Fig. 2).

In a subsequent study, Yin et al. reported another two cryo-EM structures of the SARS-CoV-2 RdRp complex (Table 1). One is in apo form, at a resolution of 2.8 Å, and the other is in binding of a 50-base template-primer RNA and the monophosphate form of remdesivir (RMP) at a resolution of 2.5 Å (Yin et al., 2020). The overall structures of the two RdRp structures are highly similar to NSP12-NSP7-NSP8 complex solved by Gao et al. (2020a). In template-RMP RdRp complex, extensive protein-RNA interactions between the template-primer RNA and NSP12 were revealed, which explicates how the RdRp complex recognizes RNA rather than DNA, and suggests no sequence-specific RNA binding. RMP is covalently incorporated at the 3’ end of the primer strand and harbors the center of the catalytic active site, with extensive hydrogen bond networks. Two magnesium ions near the bound RMP form an active catalytic site. A pyrophosphate occupies the nucleotide entrance, which could make the active site inaccessible to nucleotide triphosphate (NTP) and in turn terminate replication process.
In addition, Wang et al., reported near atomic-resolution structures of stalled pre-/post- translocated polymerase complexes, which revealed the molecular mechanism of SARS-CoV-2 RNA replication and its inhibition by remdesivir (Wang et al., 2020c).

4. Pre-fusion state of spike (S) protein

SARS-CoV-2 enters host cells mainly through the envelope-located trimeric spike (S) glycoprotein. S protein is translated as a large polypeptide, which is subsequently cleaved by host proteases to produce two functional subunits responsible for binding to the host cell receptor (N-terminal S1 subunit) and fusion of the viral and cellular membranes (C-terminal S2 subunit) (Bosch et al., 2003; Li, 2016; Walls et al., 2017). The receptor-binding domain (RBD) is located in the S1 subunit and can recognize human ACE2 receptor on target cell membranes (Figs. 1 and 2). In the pre-fusion state, the S1 subunit stabilizes the entire S protein trimer and protects the S2 protein from undergoing conformational change in advance. The S2 subunit contains a
hydrophobic fusion peptide and two heptad repeat regions, which could form a six-helix bundle to mediate viral entry through fusion pores (Fig. 2).

The S protein is a key target for vaccines, inhibitors, neutralizing antibodies, and diagnostics, considering its indispensable function. In the infection process of CoVs, S is further cleaved by proteases (Madu et al., 2009; Millet and Whitaker, 2015); however, the cleaved S proteins always remain non-covalently associated in the metastable pre-fusion conformation (Belouzard et al., 2009; Burkard et al., 2014; Kirchdoerfer et al., 2016; Millet and Whitaker, 2014; Park et al., 2016; Tortorici and Veersel, 2019; Walls et al., 2016). Such cleavage has been proposed to activate the protein for membrane fusion via extensive irreversible conformational changes (Walls et al., 2017; Belouzard et al., 2009; Heald-Sargall and Gallagher, 2012). The entry into host cells of CoVs is very complicated that requires the concerted action of receptor-binding and proteolytic processing of the S protein to promote virus-cell fusion.

The single-particle cryo-EM structure of the SARS-CoV-2 spike trimer has recently been reported in two independent studies (Table 1) (Walls et al., 2020; Wrapp et al., 2020), both of which have demonstrated that the RBD can undergo a hinge-like conformational movement to shift between “up” or “down” states, as in other coronaviruses (Gui et al., 2017; Pallesen et al., 2017; Wrapp and McLellan, 2019; Yuan et al., 2017). The “down” conformation corresponds to the receptor-inaccessible state (hide), and the “up” conformation corresponds to a receptor accessible state (expose). Wrapp et al., for the first time, reported a 3.5 Å resolution cryo-EM structure of the SARS-CoV-2 S asymmetrical trimer, in which a single RBD is observed in the “up” conformation and two other RBDS are observed in the “down” conformation (Walls et al., 2020). The major difference between the two states is the distinct organization of the RBD within the S1 apex. Using the 3D variability feature in cryoSPARC v2 (Punjani et al., 2017), it is observed that RBD underwent a hinge-like conformational movement that transiently hide or expose the determinants of receptor binding, much like the breathing of S1 subunits.

Walls et al., also presented an asymmetric reconstruction of the trimer with a single S domain opened (the “up” conformation) at 3.2 Å resolution, and determined the closed (all “down” conformations) SARS-CoV-2 S ectodomain trimer at 2.8 Å resolution (applying 3-fold symmetry; Table 1) (Walls et al., 2020). Overall, in the “down” conformation, the SARS-CoV-2 S ectodomain is a 160-Å-long trimer with a triangular cross-section and the S1 subunit has a V-shaped architecture, resembling those of closely related β-coronaviruses, such as SARS-CoV S (Fig. 2).

5. Receptor recognition of S protein

During SARS-CoV infection, the RBD domain in its S protein has a core and a receptor-binding motif (RBM), which could directly bind to the peptide domain (PD) of the angiotensin-converting enzyme 2 (ACE2), a type I membrane protein expressed in the lungs, heart, kidney, and intestine (Fig. 2) (Kirchdoerfer et al., 2018; Li et al., 2005; Song et al., 2018; Towler et al., 2004). Afterward, the S2 subunit is exposed and undergoes a conformational change from a pre-fusion to a post-fusion state. Therefore, binding to the ACE2 receptor is essential for viral infection (Belouzard et al., 2009; Simons et al., 2005). Recent studies have demonstrated that SARS-CoV-2 S and SARS-CoV S share the same functional host cell receptor ACE2 (Li et al., 2003; Wan et al., 2020; Hoffmann et al., 2020a); however, the SARS-CoV-2 RBD has a significantly higher ACE2-binding affinity than SARS-CoV RBD, based on biochemical data (Wrapp et al., 2020; Wang et al., 2020d).

Yan et al., for the first time, reported cryo-EM structures of full-length human ACE2 in complex with a neutral amino acid transporter B0AT1 at 2.9 Å as well as the complex structure of SARS-CoV-2 RBD bound ACE2 at an overall resolution of 2.9 Å (local resolution of the ACE2-RBD interface was 3.5 Å) (Table 1) (Yan et al., 2020). The ACE2-B0AT1 complex is assembled as a dimer of heterodimers, in which the collectrin-like domain of ACE2 is the key mediator of homodimerization, and the extracellular PD contributes a minor interface. According to the rotation state of the PD, the ACE2-B0AT1 complex displays open or closed conformations, whereas the rest of the complex is largely unaltered. However, in the RBD-ACE2-B0AT1 ternary complex, only the closed state of ACE2 is observed.

In a separate study, Lan et al., determined the crystal structure of the wild-type SARS-CoV-2 RBD in complex with ACE2 at 2.45 Å resolution (Lan et al., 2020). Similarly, Shang et al., resolved the crystal structure of the SARS-CoV-2 RBD (a chimera protein contains the SARS-CoV RBD core and the SARS-CoV RBM) bound to the ACE2 at 2.68 Å resolution (Table 1) (Shang et al., 2020). The overall structures of the SARS-CoV-2 RBD/ACE2 complexes are all similar to that of SARS-CoV, in which the RBM motif forms a gently concave surface with a ridge on one side and contacts the exposed outer surface of the claw-like (or arch shaped) structure of ACE2. Such atomic-level structural information provides important insights into the molecular basis for coronavirus recognition and infection, and could facilitate ongoing vaccine design and inhibitor screening efforts.

The S proteins of SARS-CoV-2 and SARS-CoV share a high sequence similarity of approximately 77% (Zhou et al., 2020); therefore, cross-reactive epitopes and neutralizing antibodies may exist. A previously characterized SARS-CoV monoclonal antibody CR3022 could bind to the SARS-CoV-2 RBD with a KD of 6.2 nM. Yuan et al., determined the crystal structure of CR3022 in complex with the SARS-CoV-2 RBD at 3.1 Å, and they observed that CR3022 targets a highly conserved epitope away from the ACE2-binding site (Fig. 2). Structural modeling further demonstrates that the epitope of CR3022 is only accessible when at least two RBD on the trimeric S protein are in the “up” conformation and slightly rotated so that all clashes can be avoided (Yuan et al., 2020).

Proteins that interact with SARS-CoV-2 S are also important targets for the design of novel antiviral therapies. Endosomal cysteine proteases-cathepsin B and L (CatB/L) (Simmons et al., 2005) and serine protease TMPRSS2 (Glowacka et al., 2011; Matsuyama et al., 2010; Shulla et al., 2011) can be used for SARS-CoV S priming in cell lines. However, only TMPRSS2 activity is essential for viral entry into primary target cells and for viral spread in infected hosts, whereas CatB/L activity is dispensable (Iwata-Yoshikawa et al., 2019; Shirato et al., 2017; Shirato et al., 2018; Zhou et al., 2015). Hoffmann et al., demonstrated that the SARS-CoV-2 S protein is primed by TMPRSS2, and catomost melyate, a TPARSS2 inhibitor, could block the infection of lung cells with SARS-CoV-2 and is a potential treatment option (Fig. 2) (Hoffmann et al., 2020b). Recently, it has been demonstrated that recombinant soluble human ACE2 (hrsACE2) can block the early stages of SARS-CoV-2 infection significantly. Clinical-grade hrsACE2 can reduce SARS-CoV-2 recovery from Vero cells by a factor of 1000–5000, and block SARS-CoV-2 infections in engineered human blood vessel organoids and human kidney organoids (Monteil et al., 2020).

6. Post-fusion state of spike (S) protein

After the binding of the S1 subunit to the ACE2 receptor, the S2 subunit undergoes conformational changes to insert the hydrophobic fusion peptide into the host target cell membrane. The heptad repeat 1 (HR1) domain in the S2 subunit forms a homotrimer, which exposes three highly conserved hydrophobic grooves on the surface that binds the heptad repeat 2 (HR2) domain in the S2 subunit; therefore, HR1 and HR2 domains interact to form a stable six-helix bundle (6-HB) fusion core. The 6-HB fusion core structure helps bring the viral and cellular membranes into close proximity for membrane fusion and viral infection (Xia et al., 2020a). SARS-CoV-2 and SARS-CoV S2 subunits share high sequence similarities, with 92.6% and 100% overall identities in HR1 and HR2 domains, respectively. Xia et al., found that HR1 and HR2 domains in SARS-CoV-2 can interact to form 6-HB (Xia et al., 2020a), and they resolved the X-ray crystal structure of SARS-CoV-2’s 6-HB (Xia et al., 2020b). The overall 6-HB structure of SARS-CoV-2 is similar to that of
SARS-CoV (Xia et al., 2019). Although the amino acid sequences of the HR2 domain from SARS-CoV-2 and SARS-CoV are identical, the HR1 domain has eight different residues, which may enhance HR1 and HR2 interactions and stabilize 6-HB conformation.

Over the past few decades, the viral HR1 domain in the S2 subunit has been proved to be another important target for the development of viral fusion and entry inhibitors. The pan-coronavirus entry inhibitor EKI peptide has been demonstrated to target the HR1s of SARS-CoV and MERS-CoV and to inhibit viral infection (Xia et al., 2020a; Liu et al., 2004; Lu et al., 2014). Xia et al., generated several lipopolypeptides by conjugating the cholesterol molecule to the EKI peptide. One of the lipopolypeptides, EK1C4, possesses the most potent inhibitory activity and is approximately 24-1 and 214 times more potent than the original EKI peptide against SARS-CoV-2 S protein-mediated membrane fusion and pseudovirus infection with IC_{50}S of 1.3 and 15.8 nM, respectively (Fig. 2). EK1C4 has also showed broad-spectrum inhibitory activity against infection of SARS-CoV, MERS-CoV and other HCoVs, suggesting EK1C4 can be used for the prevention and treatment of SARS-CoVs infection (Xia et al., 2020b).

7. Nucleocapsid (N) Protein

The nucleocapsid (N) protein of coronavirus is a multifunctional RNA-binding protein and forms a helical filament structure that is required for the assembly of viral genomic RNA into a ribonucleoprotein complex. The packaging of the viral genomic RNA regulates viral replication/transcription and modulates infectious cell metabolism (Cong et al., 2020; Masters and Sturman, 1990; McBride et al., 2014; Nelson et al., 2000; Stohlman et al., 1988). The N protein of SARS-CoV-2 is highly expressed and induces early antibody responses in infected patients, similar to in the case of SARS-CoV N protein (Ahmed et al., 2020; Liu et al., 2006; Shang et al., 2005; Hachim et al., 2020). Consequently, an enhanced understanding of the structural and functional features of SARS-CoV-2 N protein could facilitate the development of sensitive and specific immunological tests.

The common domain architecture of beta-coronavirus N protein encompasses several conserved parts: an ordered RNA-binding (N1b) domain, a dimerization (N2b) domain, and several short regions with high predicted disorder (N1a, N2a, and spacer B/N3; Fig. 1). The structures of CoVs N1b have been extensively reported, and some key residues involved in RNA binding and influencing virus infectivity have been identified (Grosselme et al., 2009; Keane et al., 2012; Saikatendu et al., 2007; Tan et al., 2006). Kang et al., determined the crystal structure of the SARS-CoV-2 N1b domain at 2.7 Å resolution and revealed distinct surface electrostatic potential characteristics between them (Kang et al., 2020) (Table 1). The overall structure is a ‘right-hand shape’ composed of a β-hairpin, a β-sheet core and a long-loop region, which resembles a protruded basic finger, a basic palm, and an acidic wrist. The β-sheet core contains five antiparallel β-strands with a short α-helix and the long β-hairpin is a large protruding loop between β2 and β5 (Fig. 2).

The N2b domain mediates self-association into a dimer, which subsequently assembles into a higher-order helical filament, potentially involving cooperative interactions. Ye et al., reported two crystal structures of the SARS-CoV-2 N2b domain, exhibiting a tightly intertwined dimer that is similar to that of SARS-CoV (Table 1) (Chen et al., 2007; Takeda et al., 2008; Yu et al., 2006; Ye et al., 2020). The dimer interface consists of two β-strands and a short α-helix from each protomer that extend toward the opposite protomer and pack against its hydrophobic core (Fig. 2). In addition, a combination of multi-angle light scattering and hydrogen-deuterium exchange mass spectrometry techniques has been used to explore the self-assembly mechanism of the SARS-CoV-2 nucleocapsid. Such studies enable researchers to identify small molecules that could inhibit nucleocapsid self-assembly processes, and in turn minimize the infection severity and pathogenic infectivity.

8. Conclusion

Following the COVID-19 outbreak, SARS-CoV-2 was rapidly identified as the etiological agent. Although our understanding of coronaviruses has improved substantially following previous SARS-CoV and MERS-CoV outbreaks, in 2003 and 2012, respectively, effective antiviral drugs for SARS-CoV-2 treatment and vaccines for its prevention remain unavailable. Extensive studies on the structures and functions of the major proteins involved in SARS-CoV-2 has been conducted, which have enhanced our general understanding of the mechanisms of its infection and replication. Furthermore, combined with structure-based predictions, experiments at cellular and animal levels in addition to high-throughput screening activities have revealed numerous compounds, including clinical drugs, drug candidates, and other pharmacologically active compounds, to be inhibitors of viral key proteins; these are potential antiviral drugs for the treatment of COVID-19 disease.

The atomic structures of SARS-CoV-2 proteins not only help to research and develop antiviral drugs, but also provide important basis for vaccine design. Among four major structural proteins, the S glycoprotein plays crucial roles in viral entry and pathogenesis, which is a key target for vaccines, but it is highly glycosylated and antigenically variable. 20 N-linked glycosylation sequons per protomer are conserved among SARS-CoV-2 S and SARS-CoV S (Walls et al., 2020), and these oligosaccharides might modulate antibody recognition in SARS-CoV (Pallese et al., 2017; Walls et al., 2019). Consequently, the precise structure of the S glycoprotein and the organization of these N-Linked Glycans will provide atomic-level information to help the design and development of vaccines. At present, multiple SARS-CoV-2 vaccine types are under development, including inactivated or weakened vaccines, viral vectored vaccines, DNA- or RNA-based vaccines, adenovirus-based vaccines and protein-based vaccines, which are based on the S antigen (Callaway, 2020; Gao et al., 2020b; Wu, 2020).

This review is a brief summary of recent developments toward the understanding of the structural and functional features of major proteins involved in SARS-CoV-2 infection and replication and corresponding drug development activities. Such molecular insights on SARS-CoV-2 infection and replication mechanisms could facilitate the effective clinical treatment and epidemiological control of the COVID-19 disease.

Declaration of Competing Interest

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

Acknowledgments

We thank Ping Shan and Ruigang Su (F.S. lab) for their assistance in lab management. This work was supported by grants from National Natural Science Foundation of China (31830020) and National Key Research and Development Program of China (2017YFA0504700 and 2018YFA0901102).

References

Ahmed, S.F., Quadeer, A.A., McKay, M.R., 2020. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses 12. doi:10.3390/v12090254.

Ahn, D.G., Choi, J.K., Taylor, D.R., Oh, J.W., 2012. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. Arch Virol. 157, 2095–2104. doi:10.1007/s00705-012-1404-x.

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., Hilgenfeld, R., 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. 21, 3213–3224. doi:10.1093/emboj/cdf272.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., Hilgenfeld, R., 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 300, 1763–1767. doi:10.1126/science.1085658.
Belouzard, S., Chu, V.C., Whitaker, G.R., 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl. Acad. Sci. USA 106, 5871–5876. doi:10.1073/pnas.0809524106.

Bolls, R.R., Han, H., Cullen, R.J., Peters, P.J., 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. 77, 8801–8811. doi:10.1128/JVI.77.17.8801-8811.2003.

Burdack, J., Deslauriers, J., Schepers, N., 2019. Structural insights into SARS-CoV-2 protease: a new target for drug discovery. J. Med. Chem. 62, 7673–7678. doi:10.1021/acs.jmedchem.9b00952.

Chang, L., Haid, P., Seidler, C., Hahn, C.A., Rolfs, P.J., 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. 77, 8801–8811. doi:10.1128/JVI.77.17.8801-8811.2003.

Chen, C.Y., Chang, C.K., Wang, Y.W., Huo, X., Cao, H.L., Li, L., 2018. Structural and functional characterization of the SARS-CoV-2 main protease. Biophys. J. 115, 2621–2630. doi:10.1016/j.bpj.2018.04.012.

Chen, Y., Liu, C., Wang, Y., Hu, H., Yan, L., Huang, W., Wang, X., Liu, C., 2020. Structural and functional characterization of the SARS-CoV-2 main protease. Nature 586, 437–443. doi:10.1038/s41586-020-2832-9.

Chen, Y., Liu, C., Wang, Y., Hu, H., Yan, L., Huang, W., Wang, X., Liu, C., 2020a. Structural and functional characterization of the SARS-CoV-2 main protease. Nature 586, 437–443. doi:10.1038/s41586-020-2832-9.
Yu, I.M., Oldham, M.L., Zhang, J., Chen, J., 2006. Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae. J. Biol. Chem. 281, 17134-17139. doi:10.1074/jbc.M602107200.

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., et al., 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat. Commun. 8, 15092. doi:10.1038/ncomms15092.

Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion Jr., R., Nuneley, J.W., Barnard, D., Pohlmann, S., McKerrow, J.H., Renslo, A.R., et al., 2015. Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res. 116, 76-84. doi:10.1016/j.antiviral.2015.01.011.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270-273. doi:10.1038/s41586-020-2012-7.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727-733. doi:10.1056/NEJMoa2001017.

Ziebuhr, J., 2005. The coronavirus replicase. Curr. Top. Microbiol. Immunol. 287, 57–94. doi:10.1007/3-540-26765-4_3.