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.
A review on structural, non-structural, and accessory proteins of SARS-CoV-2: Highlighting drug target sites

Md. Jahirul Islam a, Nafisa Nawal Islam b, Md. Siddik Alom c, Mahmuda Kabir d, Mohammad A. Halim e,*

a Division of Infectious Diseases and Division of Computer Aided Drug Design, The Red-Green Research Centre, BICCB, 16 Tejkunipara, Tejgaon, Dhaka 1215, Bangladesh
b Department of Biotechnology and Genetic Engineering, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh
c Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA
d Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka 1000, Bangladesh
e Department of Chemistry and Biochemistry, Kennesaw State University, 370 Paulding Avenue NW, Kennesaw, GA 30144, USA

ABSTRACT
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, is a highly transmittable and pathogenic human coronavirus that first emerged in China in December 2019. The unprecedented outbreak of SARS-CoV-2 devastated human health within a short time leading to a global public health emergency. A detailed understanding of the viral proteins including their structural characteristics and virulence mechanism on human health is very crucial for developing vaccines and therapeutics. To date, over 1800 structures of non-structural, structural, and accessory proteins of SARS-CoV-2 are determined by cryo-electron microscopy, X-ray crystallography, and NMR spectroscopy. Designing therapeutics to target the viral proteins has several benefits since they could be highly specific against the virus while maintaining minimal detrimental effects on humans. However, for ongoing and future research on SARS-CoV-2, summarizing all the viral proteins and their detailed structural information is crucial. In this review, we compile comprehensive information on viral structural, non-structural, and accessory proteins structures with their binding and catalytic sites, different domain and motifs, and potential drug target sites to assist chemists, biologists, and clinicians finding necessary details for fundamental and therapeutic research.

1. Introduction
Infections with the newly emerged beta coronavirus (CoV) named severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) are now widespread, affecting more than 210 countries and regions on earth (Tiwari et al., 2020). As of June 8, 2022, 537 million cases have been confirmed, with over 6.32 million deaths worldwide (World Health Organization, 2021) (WHO Coronavirus (COVID-19) Dashboard with Vaccination Data,” 2021). During the covid-19 pandemic, biomedical research received unprecedented attention. Chemists, biologists, and clinicians from all fields of biomedical science worked together to accelerate diagnostic testing and enable the development of vaccines and therapeutics for covid. Initial milestones were the publication of the viral genome sequence, solving the structure of SARS-CoV-2 main protease, and spike (S) glycoprotein. Preliminary analyses were suggested that SARS-CoV-2 has a close evolutionary association (96 % nucleotide sequence identity) with the SARS-like bat coronaviruses (Lam et al., 2020a). The early information that the SARS-CoV-2 receptor binding domain (RBD) has a higher hACE2 binding affinity than SARS-CoV RBD, and the interactions between RBD and hACE2 were the key step for the viral life cycle (Fan et al., 2020; Wrapp et al., 2020). These were crucial for understanding the molecular basis of the viral life cycle, and atomic-scale resolution of structures provided the target for designing structure-based vaccines or drugs.

Up to now, over 1800 structures of various viral proteins of SARS-CoV-2 have been resolved and reported (https://www.rcsb.org). It is important to compile the structural information together to design and develop new structure-based therapeutics. This review summarizes the current knowledge on the genome constitution and structure–function relationships of different viral proteins, and a short overview of the repurposed drugs and vaccines with therapeutic potential and ongoing trials.

2. Genome organization and overview of SARS-CoV-2 proteins
SARS-CoV-2 is a positive-sense single-stranded RNA genome of 29.8–29.9 kb nucleotides that encodes a lengthy polyprotein of 9860 amino acids (Saxena et al., 2020). The NCBI viral database has about 485,141 completed nucleotide sequences for SARS-CoV-2 as of...
the 5′ (pp1a and pp1ab), found in the first two-thirds of the viral genome from ribosomes directly translate ORF1a and ORF1ab into two polyproteins frames (ORFs) in its genome, separated into two portions. Cellular re coronavirus (Wu et al., 2020a). SARS-CoV-2 has fourteen open reading

Membrane (M) protein, and Nucleocapsid (N) protein, along with 14 structural proteins such as spike (S) protein, Envelope (E) protein, November 5, 2021 (Hatcher et al., 2017). The genome contains four structural proteins such as spike (S) protein, Envelope (E) protein, Membrane (M) protein, and Nucleocapsid (N) protein, along with 14 ORFs that encode 27 proteins and is about 80% identical to the human coronavirus (Wu et al., 2020a). SARS-CoV-2 has fourteen open reading frames (ORFs) in its genome, separated into two portions. Cellular ribosomes directly translate ORF1a and ORF1ab into two polyproteins (pp1a and pp1ab), found in the first two-thirds of the viral genome from the 5′ end. Two viral proteases, papain-like protease (PLpro) and main protease (Mpro or CLpro), then process the polyproteins and generate sixteen nonstructural proteins, nsp1–nsp16 (Y. Chen et al., 2020a), (Lu et al., 2020a). Furthermore, eight accessory proteins such as ORF3a, ORF3b, p6, ORF7a, ORF7b, ORF8b, ORF9b, and ORF14 are located at the 3′ end (Wu et al., 2020a). The genome organization and corresponding proteins have been depicted in Fig. 1. A summarized function and structural characteristics of structural, non-structural, and accessory proteins have been listed in Table 1. More than 1500 structures of SARS-CoV-2 proteins have been submitted to the Protein Data Bank (PDB) database (https://.rcsb.org/covid19) since January 2020, as presented in Fig. 2 and Supplementary Table S1. In the case of SARS-CoV-2 structural proteins, ~747 spike proteins, an envelope protein, and ~25 nucleocapsid protein structures have been deposited into the PDB database. However, the membrane protein structure of SARS-CoV-2 has not been reported yet. Among the non-structural proteins of SARS-CoV-2, a higher number of structures have been resolved for nsp5 (443), followed by nsp3, nsp13, nsp15, nsp7 & 8, nsp10, nsp12, nsp16, nsp9, and nsp14, respectively. In addition, several structures of SARS-CoV-2 accessory proteins were deposited into the PDB database, whereas only four structures for ORF8 and two structures for each of the ORF3a, ORF7a, and ORF9b proteins.

3. Structural proteins

3.1. Spike protein (S)

Among the structural proteins, the spike glycoprotein (S) is the most important. The S protein mediates the virus entry to the host cell, enhances virulence, and determines the life cycle of the virus (Walls et al., 2020). Structurally S protein is a homotrimer; three polypeptide chains of S protein are assembled to form a functional protein (Benton et al., 2020). Each monomer of the spike protein has two subunits S1 and S2, which mediate receptor-binding and membrane fusion (Fig. 3a) (Wrapp et al., 2020). Furthermore, S1 is subdivided into an N-terminal domain (NTD) and the receptor-binding domain (RBD), which directly binds to the extracellular peptidase domain (PD) of host cell surface receptor called hACE2 of human respiratory epithelial cells (Lam et al., 2020b; Walls et al., 2020). The highly mutable RBD domain consists of a five-stranded antiparallel β-sheet (β1-β3-β5-β4-β2) core, which is flanked by a short helix. The receptor-binding motif, RBM (437–508), forms a cradle-like conformation for receptor binding (Xia et al., 2020).

Basically, the interaction with the hACE2 occurs through the α1 helix of RBD which is augmented by the engagement of two polar residues of the middle segment of the α1 helix as well as the linker between β3 and β4 loops and the α2 helix (Wrapp et al., 2020). The interacting residues of RBD with the host cell receptors have been reported in several studies, shedding light on the structural basis of receptor recognition (Othman et al., 2020; Yan et al., 2020). The RBM remains buried inside the protein in the closed state, and in the open conformation, it interacts with ACE2. In the S-ACE2 complex, four disulfide bonds (C336-C361, C379-C432, C391-C525, and C480-C488) stabilize the RBD structure, and the RBM forms a concave outer surface to accommodate the N-terminal helix of ACE2. Ten H-bonds, a salt bridge, and several hydrophobic interactions contribute to ACE2 engagement (Fig. 3a) (Lan et al., 2020). The S2 subunit comprises an N-terminal fusion peptide (FP); two heptad repeats (HR1 and HR2) separated by a central helix (CH) and a connector domain (CD); a transmembrane domain (TM), and a cytoplasmic tail (CT) (Wrapp et al., 2020; Xia et al., 2020). The S protein contains three conformational states during the membrane fusion process: a native state (prefusion), an intermediate state (pre-hairpin), and a post-fusion hairpin state (stable). During the fusion process, the FP is inserted into the host cell membrane, which leads S2 to the pre-hairpin intermediate state, forming an α-helical anti-parallel complex between
| Protein name | Length (aa) | Function | Binding site/ catalytic residues | Different domain and motifs | Drug binding sites | Ref. |
|--------------|-------------|----------|----------------------------------|-----------------------------|--------------------|------|
| S (Spike)    | 1273        | Mediates binding to ACE2 | K417, E484, N487, F486, N501 | NTD (14–306), RBD (331–528), CTD1 (529–591), CTD2 (592–668), HR1 (910–985), HR2 (1163–1220), TM (1221–1234), CTD (1235–1273) | CTD of S1: V382, L390, C391, T393, T430, L517, A520, A522, L527, N544, L564, F565, F782, A1056 | (Chowdhury et al., 2020; Zhang et al., 2021) |
| E (Envelope) | 75          | Involved in virus morphogenesis and assembly | E8, N15, L18, L21, V25, L28, A32, T35 | NTD (1–8), TM (9–38), CTD (39–75) | T9, G10, T11, H13, A36, L37, S16, N15, I33, E8, N15 | (Bhowmik et al., 2020; Mandala et al., 2020) |
| M (Membrane) | 222         | Important for the budding process of coronaviruses | A50, T57, H59, R89, R92, F94, S105, R107, R149, Y172 | NTD (1–50), RBD (51–174), Linker (175–246), Dimerization domain (247–365), CTD (366–419) | N48, N49, T50, A51, R89, Y112, Y110 | (Bhowmik et al., 2020; Cubuk et al., 2021; Khan et al., 2021b) |
| N (Nucleocapsid) | 419       | Promotes genome packaging, RNA chaperoning, intracellular protein transport, DNA degradation, interference in host translation | A50, T57, H59, R89, R92, F94, S105, R107, R149, Y172 | NTD (1–50), RBD (51–174), Linker (175–246), Dimerization domain (247–365), CTD (366–419) | Y50, L51, L54, L93, A98 | (Bhowmik et al., 2020; Yan et al., 2022) |
| NSP1 | 180 | Recommended as leader protein which inhibit host translation and degrade host mRNAs | P153-N160, S166-N178 | NTD (1–128), CTD (148–180), Motif KH (164–165) | V35, E36, L39, V89, Y97, F143, F157, Q158 | (Schubert et al., 2020; Singh et al., 2021) |
| NSP2 | 638 | Binds to prohibitin 1 (PHB1) and 2 (PHB2) | V126, A127, C132, V157, L169, C240, Y242, W243, T256, G257 | NTD (1–345), CTD (438–638) | P15, D16, N94, V96, A227 | (Ma et al., 2021; Maiti et al., 2020) |
| NSP3 (PLpro) | 1945 | Responsible for cleaving of NSP1, NSP2, and NSP3 from the N-terminal region of pp1a and lab | C111, H272, D286 | Ubl (1–108), HVR (109–206), Macl or X (203–296), SU (387–745), UbII (746–805), PLPro (806–1058), NBD (1059–1200), MD (1201–1340), TM (1341–1567), Y domain (1568–1945) | L162, G163, D164, E167, P247, P248, Y264, Y268, Q256, Y273 | (Fu et al., 2021; Osipiuk et al., 2021; Yan et al., 2022) |
| NSP4 | 500 | Potential transmembrane scaffold protein which helps modify ER membranes | NA | TM1 (10–30), TM2 (280–300), TM3 (305–330), TM4 (355–380), CTD (381–500) | NA | (Santerre et al., 2021; Yan et al., 2022) |
| NSP5 (3CLpro) | 306 | Cleaves viral polyprotein | C145, H41 | N-finger (1–9), Domain-I (10–99), Domain-II (100–182), Domain-III (198–303) | T24, T25, T26, H41, F140, L141, N142, C143, C145, H163, E166, P168, H172, Q189, T190, A191, Q192 | (Jin et al., 2020; Khan et al., 2021b; Rahman et al., 2020) |

(continued on next page)
| Protein name | Length (aa) | Function | Binding site/ catalytic residues | Different domain and motifs | Drug binding sites | Ref. |
|--------------|-------------|----------|----------------------------------|----------------------------|--------------------|------|
| NSP6         | 290         | Induction of autophagosomes from host ER (Cottam et al., 2014a, Cottam et al., 2014b) | NA | NA | NA | NA |
| NSP7         | 83          | Forms hexadecameric complex with nsp8 for viral replication and participate as a cofactor for nsp12 | S4, D5, K7, C8, H36, L40, N37, V33 S15, L14, V11, A30, W29, E23 | Replicase domain (1–83) | R21, K43, D44 | (Wilamowski et al., 2021) |
| NSP8         | 198         | Makes heterodimer with nsp7 and nsp12 | P183, Y149, V131, M129, P133, A125, K127, V130, P121, L122, A110, L128, N118, I119, T123, K79, L117, I106, N109, P116, V115, M94, D112, C114, D99, L95, N104, L91, V83, I98, P92, A162, T84, R80, Q88, M90, I185, M87, A86 | Shaft domain (6–104), Head domain (105–196) | A102, A150, R190, A194 | (Wilamowski et al., 2020) |
| NSP9         | 198         | RNA-binding protein which may participate in viral replication | N33, G100, M101, V102, L103, G104, S105 | Single domain protein (1–109) Motif GxxxG (100–104) | M12, S13, N33, T35, F40, L42, L94, N98 | (Khan et al., 2021b; Littler et al., 2021a) |
| NSP10        | 139         | Forms heterodimer complex with nsp14 and nsp16, acting as a cofactor for both and stimulates ExoN (viral exoribonuclease) and 2-O-methyltransferase activity | N3, V4, T5, F8, K9, D10, P20, T21, Q22, P24, T25, H26, L27, L38, C39, D41, F60, K61, M62, N63, Y64, V66, Y69, T127, N129, N130, T131, K196, K200, I201 | Single domain protein (1–139) | V21, D22, A26, G35, Q36, P37, I38, GLY52, Q65, R78, P107, V108 | (Halder, 2021; Lin et al., 2021) |
| NSP11        | 13          | Unknown | NA | NA | NA | NA |
| NSP12 (RdRp) | 932         | Replication and methylation | Y420, F415, F441, F440, F442, N552, A443, P412, G413, D445, Q444, T409, N447, R392, D390, L391, L389, N403, V405, L388, T402, L387, N386, A379, P323, L270, F396, P326, V398, L271, S468, T532, D534 | NRAN (1–250), C-terminal RdRp (398–932), Motif G (499–511), Motif F (544–560), Motif A (612–626), Motif B (678–710), Motif C (753–767), Motif D (771–796), Motif E (810–820) | M542, K545, S549, K551, R553, R555, V557, D618, C622, ASP623, S626, S759, D760, D761, R836 | (Ahmed et al., 2020; Khan et al., 2021b; Zhang et al., 2020c) |
| NSP13 (Helicase) | 596  | A helicase core domain participates in binding interaction with ATP. | R178, H230, N361, S468, T532, D534 | ZBD (1–100), SD (101–150), 1B domain (151–261), | V45, Y70, F90, P283, G285, T286, G287, K288, H290, R443, E540 | (Chen et al., 2020; Malone et al., 2021; Yan et al., 2021) |

(continued on next page)
Table 1 (continued)

| Protein name | Length (aa) | Function | Binding site/ catalytic residues | Different domain and motifs | Drug binding sites | Ref. |
|--------------|-------------|----------|----------------------------------|-----------------------------|-------------------|------|
| Zn-binding domain is involving in replication and transcription | 527 | NSP14 (ExoN) | D90, E92, E191, H268, D273 | 1A domain (262–442), 2A domain (443–601) | W385, N386, Y420, P426, F506 | (Devkota et al., 2021; Tahir, 2021) |
| Acting on both ssRNA and dsRNA in a 5' to 3' direction and a N7-guanine methyltransferase activity | 346 | NSP15 | H235, H250, K290, T341, Y343, S294 | N-domain (1–64), Middle domain (65–182), endoU (207–347) | F44, E45, D92, H250, Y290, V292, C293, S294, Y343 | (Khan et al., 2021b; Kim et al., 2021) |
| Uridine-specific endoribonuclease activity | 298 | NSP16 | K46, D130, K170, E203 | NTD (1–29), Mtsase domain (30–210), CTD (211–298) | A80, T83, A84, L86, T94, L95, L96, V97, D98, S99, D100 | (Rosas-Lemus et al., 2020; Vithani et al., 2021) |
| ORF3a | 275 | ORF3a | NA | NTD (1–34), TM1 (35–56), TM2 (76–99), TM3 (103–129), CR domain (127–133), CTD (208–264), TRAF3-binding motif (36–40), CBM (141–149), Motif YXXΦ (160–163), Motif EXD (171–173) | Y61, I62, I63, T64, I118, V121, R122, Y206 | (Kern et al., 2021) |
| ORF6 | 61 | ORF6 | D53, E55, M58, E59, D61 | NA | NA | (Gordon et al., 2020; Li et al., 2022) |
| Type 1 IFN antagonist | 121 | ORF7a | NA | Signal peptide (1–15), Ig-like ectodomain (16–96), TM region (97–116), ER retention motif (117–121) | E33, C35, S36, S37, T39, V40, E41, G42, S44, P45, F46, F65 | (Gorgulla et al., 2021) |
| Triggers an immune response in host cells | 121 | ORF8 | P65, F86, T87, I88, N89, D90, Q91, E92 | P65, F86, T87, I88, N89, D90, Q91, E92 | I47, L60, V62, D63, V73–75, Y79, T80, Q91, K94, L95 | (Casasotto et al., 2021; Hassan et al., 2021) |

NTD = N-terminal Domain, RBD = Receptor Binding Domain, CTD = C-terminal Domain, TM = Transmembrane Domain, UbI1 = ubiquitin-like domain 1, SUD = SARS-unique domain, HVR = hypervariable region, PL2pro = papain-like protease, MD = Marker domain, NBD = nucleic acid-binding domain, ZBD = zinc-binding domain, SD = stalk domain.
HR1 and HR2, where the loop region acts as a hinge and builds a six-helix bundle (6HB) that brings the cellular and viral lipid bilayers in close proximity (Ling et al., 2020; Wang et al., 2021), as shown in Fig. 2b. In post-fusion state, HR1 and HR2 of the S2 subunit are associated with each other forming a six-helix bundle (6HB) fusion core where three HR2 helices surround the HR1 helices in an anti-parallel arrangement (Fig. 3c). This complex structure is highly stable and plays a significant role in membrane fusion (Schütz et al., 2020).

Protein-protein interaction between HR1 and HR2 revealed that multiple H-bond contacts were established between amino acids from HR1 (N925, Q935, Q949, N953, N960) and amino acids from HR2 (A1174, V1177, I1179, Q1180, A1190, N1194, I1198) (Fig. 3d). HR hairpin trimerization is mediated by a collection of tandemly ordered seven-residue repeats, with the repeatedly presented hydrophobic amino acids forming a strong hydrophobic face (Fig. 3e). Notably, the HR2 domains of SARS-CoV-2 and SARS-CoV-1 are identical, while the HR1 domains show variations; considering HR2 a good target site for developing potential fusion inhibitor (Schütz et al., 2020; Xia et al., 2020; Yan and Gao, 2021). A range of antiviral agents has been developed to target the S protein, including antibodies, inhibitors, and vaccines. In general, inhibitors of S glycoproteins prevent virus-membrane fusion by competitively blocking RBD-ACE2 interaction (Chowdhury et al., 2020). Such inhibitors include arbidol (umifenovir) (Padhi et al., 2021) and ivermectin (Caly et al., 2020). Haste et al. reported that RBD-ACE2 interactions are blocked by three kinds of antibodies (RBD1, RBD2, RBD-3 mAbs) through both steric hindrance and direct competition for interface residues. Here, RBD-1mAbs largely overlap with the RBM; the RBD-2mAbs move to the “Peak” of the RBM from the center of ACE binding; and the RMD-3mAbs bind to the “Mesa” of the RBM from the center of ACE2 binding. (Hastie et al., 2021). Furthermore, a pan-CoV fusion inhibitor, such as EK1 has been used to target HR1 of the S2 domain to inhibit membrane fusion (Efaz et al., 2021; Wang et al., 2021).

3.2. Membrane protein (M)

The M protein is the most abundant structural protein and the major component of the viral envelope (Tseng et al., 2013). Moreover, M protein is crucial for viral assembly, morphogenesis (Hu et al., 2003), budding (Vojet et al., 2009), and recruitment of S protein to the virus assembly and budding site. The M protein is also required for genome packing (Hu et al., 2003), and nucleocapsid inclusion into the virion (Vojet et al., 2009). The M protein consists of three primary domains: an ectodomain at the N-terminus, three transmembrane helices (TMH1-TMH3), and an endodomain at the C-terminus (Fig. 4a) (Mahtarin et al., 2020) TM1-TM2 intersestion is thought to be in the interior, while the TM2-TM3 intersestion is in the exterior (Hu et al., 2003). The C-terminal region is predicted to have at least two casein kinase II phosphorylation sites (T59 at codon 171, SQK at codon 183) related to S, E, and N protein interaction (Hu et al., 2003). These interactions are required for membrane bending (budding) and operate as a checkpoint to form new virions (Ujike and Taguchi, 2015).

Furthermore, SARS-CoV M protein residues L218 and L219 are needed for nucleocapsid packing (Liu et al., 2010; Tseng et al., 2013). Antigenic epitopes have been found in the TM1 and TM2 regions of the SARS-CoV M protein, leading to the designing of peptide inhibitors or vaccines against this protein. Along with the experimental study, researchers have recently employed computational techniques such as molecular dynamics (MD) simulations to uncover several potential drugs (such as remdesivir) that have a higher affinity for M protein (Khan et al., 2021a), but further studies are required to confirm these interactions.

3.3. Envelope protein (E)

The SARS-CoV-2 envelope (E) protein is the smallest of all structural proteins and is found primarily in the host cell’s endoplasmic reticulum (ER) and Golgi complex, where it is involved in viral assembly, pathogenesis, and release (Westerbeck and Machamer, 2019). The topology of E protein consists of a five-helix bundle, surrounded by a dehydrated narrow pore and bipartite channel. In terms of amino acid composition, E proteins are highly divergent but structurally highly conserved in various genera of β-coronaviruses with a hydrophobic ectodomain at the N-terminus, a hydrophobic transmembrane domain (TMD), and a lengthy hydrophilic C-terminal endodomain (Fig. 4b) (Schoeman and Fielding, 2019). Amantadine (AMT) and hexamethylene amiloride (HMA) are two ion-channel drugs having guanidinium group to interact with polar residues at the entrance of ion channel and engage the amine-terminal lumen, blocking ion channel activity of the E protein (Mandala et al., 2020; Pervushin et al., 2009). Moreover, peptide inhibitors derived from Ec18 can be employed to inhibit the interactions between the E protein and PLAS1 (the human cell junction protein) (Chai et al., 2021). Furthermore, Bacillus Calmette-Guerin (BCG) vaccination is used as an alternative approach for treating COVID-19 that induces specific immune responses that can be employed to inhibit the interactions between the E protein and PLAS1 (the human cell junction protein) (Chai et al., 2021).

3.4. Nucleocapsid protein (N)

SARS-CoV-2 N protein is encoded in the structural ORF situated at the 3’ end. In the domain architecture of SARS-CoV-2 N protein, it possesses three highly conserved domains: an N-terminal domain (NTD), a linker region or an RNA-binding domain, and a C-terminal domain (CTD) (Fig. 4c). The core region of the NTD constitutes β-sheets situated in a five-stranded anti-parallel position. Two α-helices occur inside the β-sheet core, forming an overall conformation of β1-α1-β2-β2-α3-β3-β4-α2-β5 in which β2 and β4 forms a long β-hairpin structure (Dimesh et al., 2020). The N-CTD monomer comprises a structure of β1-α1-β2-α3-α4-β1-β2-α5-β3 orientation which contains five α-helices, three 310 (η) helices, and two anti-parallel β-strands that forms a β-hairpin structure. The CTD forms a compact homodimeric structure, and the monomer of CTD is assumed to be unstable (Zhou et al., 2020). The N-CTD dimeric structure is stabilized by 40H-bond and 389 hydrophobic interactions (Zinnulza et al., 2021). The CTD region of SARS-CoV is essential for RNA binding (residue 248–280) (Chen et al., 2007; Takeda et al., 2008), and it remains almost conserved in SARS-CoV-2 (corresponding residue 247–279) with one amino acid replacement (SARS-CoV Gln268 → Ala267 SARS-CoV-2) (Zhou et al., 2020). The C-terminal
Fig. 3. (a) Schematic representation of SARS-CoV-2 spike protein primary structure and hACE2-RBD complex. Here, the top panel shows different color domains: SS, single sequence; NTD, N-terminal domain; RBD, receptor-binding domain; S1, subdomain 1; S2, subdomain 2; S1/S2, S1/S2 protease cleavage site; S2', S2' protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Arrows indicate the protease cleavage site. The lower-left panel shows the cartoon and surface representations of the overall structure of the SARS-CoV-2 RBD bound to hACE2 (PDB: 6M0J) and protein–protein interactions are presented in the lower-right panel. (b) The mechanism of human ACE2 and SARS-CoV-2 S protein-mediated virus attachment and fusion. In the native state, the S2 subunit is encapsulated by the S1 subunit. Several conformational changes occur in the S2 subunit after viral RBD engagement with the receptor. The HR1-trimer core structure is formed from three HR1 molecules, and three HR2 molecules bind to the HR1-trimer to form 6-HB, which mediates membrane fusion. A neutralizing antibody that targets RBD blocks viral infection by blocking RBD’s interaction with cellular receptors. The membrane fusion process is inhibited by the fusion inhibitor that blocks 6-HB formation. (c) Six-helix bundle fusion core is comprised of three HR2-helices packed in the HR1 side grooves (PDB: 6M1V). Structures in the top view (left panel) and side view (right panel) are presented respectively. Here, three HR1/HR2 chains are colored light green, magenta, and cyan. (d) The detailed interactions between HR1 and HR2, and residues involved in the H-bond interactions are labelled. (e) Residues involved in the hydrophobic interactions are labeled.
8

Immunobiology 228 (2023) 152302

8

nCoV396 monoclonal antibody, isolated from the blood of convalescent COVID-19 patients, forms H-bonds and hydrophobic interactions with several residues (Q163, L167, and K169) of the N-NTD to stabilize the protein complexes. These interactions work together to help the antibody neutralize the N protein’s antigenicity (Kang et al., 2021). The N protein has also been inhibited by small molecules (such as PJ34 and rapamycin) that interfere with the RNA binding of N-NTD and dimerization of N-CTD (Mat-
suo, 2021; Peng et al., 2020).

4. Non-Structural proteins (NSP)

4.1. nsp1

Nsp1 is the N-terminal cleavage product released from polyprotein precursors pp1a and pp1ab by viral papain-like protease (PLpro) through proteolysis (Clark et al., 2021). The structure of the nsp1-40S ribosomal subunit complex was resolved by the cryo-EM which revealed the mechanism of translation inhibition (Fig. 5a) (Schubert et al., 2020). The hydrophobic core of the β-barrel comprises three layers, where the first layer is formed by side chains of residues L16, L18, V69, L88, L107, and L123, but the opening of the β-barrel at this layer is obstructed by α1 helix side-chain residue L46. The middle layer consists of residues V20, L53, I71, V86, and V121, while the bottom layer features residues V84 and L104. Moreover, the two 310 helices in the globular domain form an H-bond interaction between R24 and Q63 which stabilize the position of two of the largest loops in the globular domain, the β1-α1 (L21–S34) loop and the β2-β3 (E54–P67) loop (Semper et al., 2021). Several surface residues, such as E36, E37, E41, K47, K58, R124, and K125 are significant for mRNA binding; these are highly conserved in SARS-CoV-2 (Almeida et al., 2007). Guardoño et al., reported that two C-terminal regions (aa 122–130 and aa 155–165) of nsp1 are important to inhibit IFN responses and/or antiviral signaling (Jimenez-Guardoño et al., 2015). Recently, Vankadari et al. reported that several natural product molecules including garinolic acid, glycyrrhizin acid, tirilazad, and lobaric acid were considered as potential nsp1 inhibitors. These molecules were also screened by preliminary computational studies, and it is reported that they can possibly block the nsp1/SL1 complex formation (Vankadari et al., 2020).
Md. Jahirul Islam et al.
4.2. nsp2

SARS-CoV-2 nsp2 is the second protein of pp1, which contains two domains, i.e., the N-terminal domain (1–345) and the C-terminal domain (438–638) (Heo and Feig, 2020). Although SARS-CoV-2 nsp2 has been involved in viral processes, its exact functions and the structural basis remain unknown (Y. Chen et al., 2020b). Gupta et al. reported that the macrodomain has binding impact each other.

4.3. nsp3

SARS-CoV-2 nsp3 is the largest membrane-bound protein (1945 aa) with several domains (Baez-Santos et al., 2015; Wu et al., 2020b). It acts as a membrane-anchored scaffold that associates with the host proteins and other nsp to form the viral replication-transcription complex (Angelini et al., 2013). Nsp3 consists of the N-terminal Nsp3a domain (includes ubiquitin-like domain 1 (Ub1) and acidic domain (Ac) or DNA/RNA mimicry, and protein–protein interactions (Chou and Wang, 2015)). A conserved X domain or macrodomain (also called Nsp3b) follows the hypervariable region in all CoVs (Gorbunova et al., 1991; Neuman, 2016; Neuman et al., 2008). Recently, several studies reported that the macrodomain plays a role in evoking the innate immune response of host cells (Eriksson et al., 2008; Fehr et al., 2016, 2015; Kuri et al., 2011). Imbert et al. reported that the macrodomain has binding interaction with the RNA-dependent RNA polymerase (Imbert et al., 2009). If this interaction exists in the virus life cycle, two proteins can impact each other’s enzymatic activity (Lei et al., 2018). The exact functional role of the Ub1 domain is not clear. Frieman et al. reported that the Ub1 domain is essential for antagonizing the host innate immune response by blocking IRF3 or the NF-κB pathway (Frieman et al., 2009). Recently, a study reported that SARS-CoV-2 PLpro recognizes LxGx tetrapeptide motif between nsp1 and nsp2, nsp2 and nsp3, and nsp3 and nsp (Rut et al., 2020a). The catalytically active PLpro domain of SARS-CoV cleaves PpI at three cleavage sites at the N-terminus (176ELNGG|AV182|814RLKGG|AP820|2759SLKGG|K12742) to release nsp1, nsp2, and nsp3 through proteolytical processes (Lei et al., 2018). According to the UniProtKB Database (UniProtKB: P0DDT1), these three cleavage sites have found for SARS-CoV-2 at the N-terminus (176ELNGG|AV182|814RLKGG|AP820|2759SLKGG|K12742) of ppla; this process is essential for viral replication (Harcourt et al., 2004). The PLpro monomer comprises four distinct domains, three of which adopt an extended right-hand fold with a distinct thumb, finger, and palm subdomains (Lei et al., 2018). The first 62 residues fold into a ubiquitin-like domain (Ub1). This domain is well separated from the other three domains that interact with each other and form a compact globular conformation (Alliawares et al., 2017). The Ub1 domain adopts a β-grasp fold similar to ubiquitin and is highly conserved in most β-CoVs, including SARS-CoV and MERS-CoV (Lei et al., 2014; Yang et al., 2014). The central thumb subdomain of SARS-CoV-2 PLpro is predominantly helical, comprised of six α-helices and one β-strand, and a catalytic Cys111 residue contributes to the active site. Amino acids from 189 to 314 form into the finger and palm domains. The C-terminal region of the PLpro domain is made up of mostly β-strands. The finger domain consists of one α-helices (α8), one long (β7) and two short (β8 and β9) β-strands, and the palm domain is made up of eight β-strands. A Zn-ion is coordinated by four cysteine residues (Cys189, Cys192, Cys224, and Cys226) inducing from two β-hairpins is located between β7 and β9 of the finger domain. Although the conformations of the zinc finger are variable between different CoV PLpro (Lei et al., 2014), the motif is significant for proteolytic activity and structural stability (Barretto et al., 2005). The active sites are located at the interface of the thumb domain and middle of the palm domain and comprise the typical Cys111, His272, and Asp286 triad, adjacent to the flexible “locking loop” (BL2) containing Thr106 (Bagherzadeh et al., 2020). This loop comprises six amino acid residues (GNYQCG) in the enzyme of SARS-CoV PLpro (Lei et al., 2018). The enzyme has four substrate recognition subunits (S1-S4) (Arya et al., 2020). Residues Gly271, Thr106, Cys111, and Tyr112 are included in the S1-subsite, whereas residues Leu162, Asp164, Gly271, and Tyr273 are engaged in shaping the conserved S2 subsite. S3 subsite hid residue Gly271 that is partially solvent-exposed, and S4 subsite includes residues Asp302, Pro228, Thr264, Tyr268, Tyr273, and Thr301 that are buried and structured (Kong et al., 2015). The competitive inhibitors of SARS-CoV PLpro are bound at the S2 and S4 subsites (Baez-Santos et al., 2014; Kong et al., 2015). The deubiquitinase and desglycylating activity of CoV PLpro are well installed, but the elaborated mechanism of the PLpro antagonism of the host innate immune response is still uncertain (Lei and Hilgenfeld, 2017). Various cytokines such as TNF-α and IFN-α are induced to inhibit virus replication by the IRF3 and the NF-κB pathway.
Fig. 6. (a) Domain organization of nsp7 and nsp8. Cartoon representation of SARS-CoV-2 nsp7 bound to the C-terminal of nsp8 in the lower-left panel (PDB: 6M5I). The lower-right panel shows protein–protein interactions between SARS-CoV-2 nsp7 and nsp8. (b) The overall topology of SARS-CoV-2 nsp12 with different colors domains. Cartoon representation in the below (left panel) shows SARS-CoV-2 nsp12-nsp7-nsp8 complex (PDB: 7BV2). The conserved zinc-binding motifs are highlighted in the SARS-CoV-2 nsp12 structure. The coordinate details of the zinc-binding residues are shown in stick representation. Protein-protein interactions between SARS-CoV-2 nsp12, nsp7, and nsp8 are shown in the left and right panels.
Fig. 7. (a) Overall topology of SARS-CoV-2 nsp10 and nsp14. The lower-left panel shows the cartoon and surface representations of the nsp10 and nsp14 complex (PDB: 7DIY). Zinc and magnesium ions are shown as spheres and are colored magenta and lemon. Magnified views on the two zinc-finger motifs in nsp10 and nsp14, and one magnesium-finger motif in nsp14. Residues involved in zinc and magnesium-coordination are shown in sticks. The lower-right panel shows protein–protein interactions between nsp10 and nsp14 proteins. (b) Domain organization of SARS-CoV-2 nsp16 with different colors. The lower-left panel shows the cartoon and surface representations of the nsp16 and nsp10 complex (PDB: 7L6R). The lower right-panel shows protein–protein interactions between nsp16 and nsp10.
Fig. 8. Structures of SARS-CoV-2 accessory proteins. (a) The upper panel shows a schematic representation of the domains of SARS-CoV-2 ORF3a protein; N-terminal ectodomain, three transmembrane regions, a cysteine-rich domain, YxxΦ domain, diacidic domain, and C-terminal endodomain. The lower-left panel shows a cartoon and surface representation of cryo-EM structure of dimeric ORF3a (PDB: 6XDC). The lower-right panel shows protein–protein interactions between the A-chain and B-chain of the dimeric structure. (b) The upper panel shows the overall topology of SARS-CoV-2 ORF8. The lower-left panel shows the cartoon and surface representation of the crystal structure of dimeric ORF8 (PDB: 7JTL). The lower-right panel shows detailed protein–protein interactions between the A-chain and B-chain of the dimeric structure.
Unfortunately, the page contains a table and some text which seems to be cut off or incomplete. I will attempt to read and interpret what I can see:

### Table 2

| Candidate     | Vaccine type         | Manufacturer                  | Mechanism of action                                                                 | Efficacy  |
|---------------|----------------------|-------------------------------|-------------------------------------------------------------------------------------|-----------|
| AZD122        | Viral Vector         | Oxford/AstraZeneca            | The ChAdOx1 vector has been engineered to contain genetic information that encodes the S protein of the wild-type SARS-CoV-2 | 62-90 %   |
| mRNA-1273     | mRNA                 | Moderna/US NIAID              | Activates T cells to assist in the development of B cells that produce antibodies; initiating an adaptive immune response to the virus’ S protein | 95 %      |
| BNT162b2      | mRNA                 | Pfizer/BioNTech               | Lipid nanoparticles encapsulate the mRNA of SARS-CoV-2 S protein. Immunological responses are triggered when these proteins are released from the injected cells | 95 %      |
| Convidicena Ad5-nCoV | Viral vector | CanSino               | Enables to recognize the SARS-CoV-2 S protein and triggers an immune response | 65.7 % in moderate cases, and 90.98 % in severe cases |
| Spunvik V/Gam-COVID Vac | Viral vector | Gamaleya Research Institute of Epidemiology and Microbiology | Enables to recognize the SARS-CoV-2 S protein and triggers an immune response | 92 %      |
| CoronaVac     | Inactivated          | Sinovac Biotech               | Recognizes the antigen of S protein without the virus to propagate once inside human cells | 66 %      |
| BBBIp-CovV    | Inactivated          | Sinopharm (Beijing)           | Well tolerated and β-propiolactone-activation virus provides an immune response in the host cell | 50 %      |
| Covaxin       | Inactivated          | Bharat Biotech                | It induced robust humoral responses in a short period of time | 79 %      |
| NVX-CoV2373   | Recombinant nanoparticle | Novavax                   | Well tolerated and β-propiolactone-activation virus provides an immune response in the host cell | 100 % in severe cases and 70 % in asymptomatic cases |
| ZF/2001       | Protein subunit      | Anhui Zhifei Longcom Bio      | S protein-containing nanoparticles are injected into the arm muscle to activate the antigen-presenting cells | 89 %      |
| Unknown       | Inactivated          | Sinopharm (Wuhan)             | Full immunity still under investigation | NA        |
| EpiVacCorona  | Protein subunit      | Federal Budgetary Research   | Induces virus-specific and neutralizing antibodies | 82 %      |

---

**Data source:** (Covid-19: China approves Sinopharm vaccine for general use - BBC News, n.d.; Kaur and Gupta, 2020; Kyriakidis et al., 2021; Sanyasoju et al., 2022)
4.6. nsp6

SARS-CoV-2 nsp6 is a multiple-spanning transmembrane protein located to the endoplasmic reticulum (ER) (Benvenuto et al., 2020; Cottam et al., 2011). It is associated with the generation of autophagosomes to release viral components to lysosomes for degradation (Cottam et al., 2014a, Cottam et al., 2014b). Oostra et al. reported that SARS-CoV nsp6 and mouse hepatitis virus (MHV) nsp6 contain six TM domains (Oostra et al., 2008). In TM2 and TM3 domains, the highly conserved lysine and histidine residues are designated as KH loop, but the function of this cytosolic loop is unknown. Another point is that both N- and C-terminus are exposed to the cytosol (Baliji et al., 2009; Oostra et al., 2008). Notably, the C-terminal domain possesses a palmitoylation site (Hagemeijer et al., 2012) and predicted it to be cysteine residue within the conserved G(XC)XG motif (Baliji et al., 2009).

4.7. nsp7, nsp8, and nsp12

The SARS-CoV-2 nsp7 is composed of α-helical structure with three helical bundle folds, whereas nsp8 has two subdomains: An N-terminal ‘shaft’ domain (6–104 residues) and C-terminal ‘head’ domain comprised of four anti-parallel β-strands11β (Fig. 6a) (Konkolova et al., 2020). The crystal structure of SARS-CoV-2 nsp7 with nsp8 is a hollow cylindrical hexadecameric complex that forms a dimer conformation with negatively charged outer skin and positive charged inner core channel. The channel is mainly formed by attaching the four N-terminal helices of nsp8, of which the structure resembles the “shaft” of a “golf-club” (Zhai et al., 2005). This charge distribution helps the phosphate backbone of nucleic acid to pass the cylindrical channel without any electrostatic repulsion (Krishna et al., 1994). The cylindrical nsp7-nsp8 complex is stabilized by a salt bridge, four H-bond, and 90 hydrophobic interactions. The structural details of the nsp7-nsp8 complex suggest that the development of a potential allosteric inhibitor can block RdRp activity. Distinct from nucleotide analogs that directly target RdRp, allosteric inhibitors disrupt the assembly of nsp7-nsp8-nsp12, which inhibits the activity of RdRp machinery (Biswal et al., 2021).

Nsp12 is a multi-subunit RNAdependent RNA polymerase (RdRp) (van Hemert et al., 2008). Structurally, nsp12 contains two main functional domains, i.e., N-terminal (1–379) and a polymerase domain (398–919) (Yin et al., 2020). The polymerase domain at the C-terminus part resembles a “right hand” cupped shaped conformation with finger subdomain (398–581, 628–687), palm subdomain (582–627, 688–815), and a thumb subdomain (816–919) (Fig. 6b) (de Clercq, 2006). The N-terminus of RdRp contains a norovirus RdRp-associated nucleotideN transferase domain or NiRAN domain (115–250), and an extended N-terminal β-hairpin domain (31–50) (Yin et al., 2020). The NiRAN domain is followed by an interface domain (251–365), connected to the RdRp domain, and the β-hairpin domain inserts into a groove clamped by the palm domain and NiRAN domain (Gao et al., 2020). However, the active site of RdRp is situated at the finger and thumb sub-domain interface, which are the center of the substrate domain where RNA synthesis takes place (Cheng et al., 2005). In the NiRAN domain and fingers domain, two Zn2+ ions are situated distally from the RdRp catalytic site, coordinated with highly conserved residues that may be essential for structural stability. Two Mg2+ ions are coordinated to D618, D760, D761, and NTP in the palm subdomain, which forms the polymerase catalytic core (Yin et al., 2020). The complex of SARS-CoV-2 nsp12 is bound to nsp7 and nsp8 (PDB: 7BV2), forming two salt bridges, 11H-bonds, and 166 hydrophobic interactions with nsp8 and three H-bond, and 40 hydrophobic interactions with nsp7. The nsp7-nsp8 heterodimer’s attachment to the finger loop stabilizes the polymerase domain, allowing for greater affinity for template RNA. The second subunit of nsp8 is thought to play a key role in polymerase activity, potentially by binding to the template RNA and providing an expanded interaction surface, which keeps the RNA strand in place. In the presence of both nsp7 and nsp8, nsp12’s binding affinity is increased significantly to template-primer RNA and the polymerase activity is also enhanced (Gao et al., 2020). Currently, numerous nucleoside analog drugs such as remdesivir, and galidesivir (adenosine analogs) (Elffy, 2020), ribavirin and favipiravir (guanine analogs) (de Clercq, 2019), molnupiravir (Sheahan et al., 2020), sofosbuvir (uridine analog) (Gane et al., 2013) have been used to block the catalytic active site of RdRp. Moreover, RdRp is also targeted by the drug Suramin, a poly-sulfonated trypan blue derivative that effectively suppresses a range of viruses, including SARS-CoV-2 (Zeltner et al., 2020).

4.8. nsp9

SARS-CoV-2 nsp9 is an ssRNA-binding protein involved in viral replication (Littler et al., 2020). Structurally, nsp9 is homologous to a subdomain of serine protease, particularly the first domain of picornaviral 3CLpro (PDB ID: 1L1N) and the second domain of the SARS-CoV 3CLpro (PDB ID: 1P9U, 1I2W, and 1P9S). Like other nsp9 homologs it exhibits an unusual fold that is yet to be observed outside coronaviruses (Littler et al., 2020; Sutton et al., 2004). The folding core consists of a 6-stranded enclosed β-barrel, from which most of the extended loops are observed outward (Fig. 5d) (Biswas et al., 2021). Two glycine-rich loops, such as β2-β3 and β3-β4 are positively charged and involved in RNA-binding. The N-terminal β-strand in SARS-CoV-2 nsp9 forms a dimer interface with the C-terminal α1-helix. The GXXG motif is highly conserved at the dimer interface, allowing the helices in a close pack at the residues G100 and G104 (Miknis et al., 2009). Littler et al. reported that hydrophobic residues of the α1-helix form a funnel-like hydrophobic cavity on either side of the dimer interface. In the structure of SARS-CoV-2 nsp9, the N-terminal tag is incorporated together with a rhinoviral 3C protease sequence (LEV1). This 3C sequence entered into the other side cavities of the dimer interface and proximal to the GXXG motif. Furthermore, additional β-sheet interactions are formed by the 3C sequence with the N-terminus from the neighboring protomer (Biswas et al., 2021). The uracil-analog FR6, which has a weak backbone affinity with nsp9 (Littler et al., 2021b). This compound induces a hexameric form and modifies the oligomerization state, altering RNA entry channels and thus affect RNA binding. Furthermore, FR6 disrupts the dimer interface of the nsp9 GXXG, impacting RNA binding and viral proliferation (Hu et al., 2017).

4.9. nsp10, nsp14, and nsp16

The nsp10 protein is a small, single-domain protein with 139 amino acids that acts as a scaffold to bind with the nsp14 (exonuclease and N7-methyltransferase) and nsp16 (20-O-methyltransferase) for forming the mRNA cap methylation complex (Chen et al., 2013). The crystal structure of nsp10 (PDB: 7DIY) shows that it comprises a helical domain, an anti-parallel β-sheet, and two zinc-binding sites (Fig. 7a). A short peptide, K29 is found in SARS-CoV nsp10 (resides 68–96) that can inhibit the 2’-O-methyltransferase activity of nsp16 (Ke et al., 2012).

The nsp14 is a bifunctional protein, which contains the N-terminal exonuclease (ExoN) domain and N7 methyltransferase (MTase) domain for RNA cap formation (Fig. 7a) (Konkolova et al., 2020), and is also critical for proofreading, repair, and safeguarding the viral genome as a part of the replication-transcription complex (RTC) throughout the viral life cycle (Bouvet et al., 2012). The ExoN domain contains two zinc-finger motifs (C207-C210-C226-H229 and H257-C261-H264-C279) are sterically located close to the MTase domain, connecting the β11/β12, β13/β14, and α4/α5 intervening loops to helix α5. In the nsp10-nsp14 complex structure, nsp10 helices α1, α2, α3, q1, strands β1, β2 and most of their intervening loops are involved in the binding including nsp14 helix 1, strands β2, β8, α11, a long N-terminal loop and multiple intervening loops (β2/α3, β3/α1, β7/β8, α3/β11, and β11/β12) (Lin et al., 2021). Several studies revealed that the relationship between nsp14-nsp10 is explicitly responsible for establishing the ExoN activity, i.e., 35-fold enhanced activity (Bouvet et al., 2012; Ma et al., 2015). For
SARS-CoV-2, nsp10 in complex with nsp14 stabilizes the ExonN activity by binding of a salt bridge, 20H-bonds, and 272 hydrophobic interactions. Therefore, the ExoN domain’s catalytic pocket visibility collapses in the absence of nsp10 (Ziebuhr et al., 2004). Nsp14 is an exoribonuclease that removes both nucleotide analogs and mis-incorporated nucleotides from the nascent RNA, which can lead to the development of nucleotide analog-based antiviral resistance (Ferron et al., 2017; Robson et al., 2020). This issue can be resolved using a combination of nsp14 inhibitors and nucleotide analogs (such as sofosbuvir, remdesivir, and ribavirin) (Eastman et al., 2020; Jockusch et al., 2020; Khater et al., 2021). For example, the activity of ExoN is inhibited by 3′-deoxy nucleotide analogs (Lin et al., 2021); Ebselen and Disulfiram, zinc-ejecting agents, inhibit the activity of nsp14 via their three Zn-binding sites (Chen et al., 2021; Sargysan et al., 2020); Several SAM analogs and competitive inhibitors such as sinefungin (SFG), aurantricarboxylic acid (ATA), and S-adenosylhomocysteine (SAH) impede N7-Mase activity (Ahmed-Belkacem et al., 2020; He et al., 2004) ultimately preventing 5′-end cap formation (Devkota et al., 2021). Additionally, to these inhibitors, mutations in nsp14 can impair virus replication and induce a higher level of the interferon response. Therefore, the development of live attenuated virus vaccines and antibodies are alternative options for this target (Graham et al., 2012; Lu et al., 2020b).

The nsp16 is an S-adenosyl methionine (SAM) dependent and m7GpppA (Cap-1) specific protein which catalyzes the 5′-methyl capping of viral mRNA (Bouvet et al., 2010). It contains a canonical SAM-MTase fold with eight stranded cores twisted β-sheet flanked by three helices on one side and two α-helices on the other side (Fig. 7b) (Viswanathan et al., 2020). Nsp16 is inactive by itself and requires nsp10 for the activity. Nsp10 interacts with nsp16 through seven H-bonds and 90 hydrophobic interactions to stabilize the SAM binding site (Lin et al., 2020; Rosas-Lemus et al., 2020). The genetic disruption of SARS-CoV nsp16 causes a tenfold reduction in the viral RNA synthesis. Disrupting the nsp10/nsp14 or nsp10/nsp16 interface might be an excellent therapeutic target method since the interaction of nsp10 with nsp14 and nsp16 is required for their optimum activity (Bouvet et al., 2014). RNA-cap methyltransferase (nsp16) might be a crucial target in developing antiviral drugs against SARS-CoV-2, although no effective inhibitors or licensed drugs are currently available (Rohaim et al., 2021).

4.10. nsp13 (Helicase)

SARS-CoV-2 nsp13 is an RNA helicase that unwinds double-stranded RNA (dsRNA) or DNA (dsDNA) into single strands in an ATP-dependent manner, and its helicase activity can be enhanced by binding of polymerase protein (Jia et al., 2019). Guo et al. reported that nsp13 has an inhibitory role in regulating type I interferon production, and over-expression of nsp13 suppressed IFN-β levels in the host cell (Guo et al., 2021). The structural studies of SARS-CoV-2 nsp13 contain an N-terminal zinc-binding domain (ZBD), two helicase subdomains RecA1 (1A) and RecA2 (2A), a β-barrel 1B domain being connected to ZBD via a helical “stalk” region (Fig. 5e) (Yan et al., 2020). Two copies of nsp13 can interact with the core nsp7-nsp10-nsp12 complex on opposing sides of the RNA binding cleft. In both cases, the interactions are mediated from the ZBD (residues V45, A46, M68, I79, S80, F81, P90, G91, L92, Y93, K94, N95) and 1B domain (residues V193, Q194, H230, R248, Y253, L256), which interacts with the “shaft” and “head” domains of nsp8 and the thumb domain of nsp12. The structure of nsp13 in a complex with nsp8 and nsp12, might have potential implications for helicase activity and regulation (Chen et al., 2020c). Zeng et al. identified novel inhibitors such as PFA-124 and suramin-like compounds that can inhibit viral helicase activities (Zeng et al., 2021).

4.11. nsp15

SARS-CoV-2 nsp15 is a uridine-specific endoribonuclease and is highly conserved across coronaviruses. It is responsible for interference with the innate immune response (Pillon et al., 2021). An earlier study reported that overexpression of nsp15 inhibits the interferon (IFN) response and the mitochondrial antiviral signaling protein (MAVS) mediated apoptosis in SARS-CoV (Friedman et al., 2009). Nsp15 contains three discrete domains: N-terminal “wing” domain, middle “body” domain and C-terminal catalytic NendoU “wing” domain (Fig. 5f) (Gordon et al., 2020). The N-terminal domain consists of two α-helices with three anti-parallel β-sheets: uridine-specific endoribonuclease (Ricagno et al., 2006). The middle domain comprises ten β-strands, a major mixed β-sheet, three small and short two α-helices, which form a hexamer resulting in the concave surfaces that may act as an interaction hub (Joseph et al., 2007). The C-terminal catalytic NendoU domain contains two anti-parallel β-sheets that play a catalytic function in SARS-CoV-2 and likely many others coronaviruses family (Ulferts and Ziebuhr, 2011). However, the active site in the C-terminal is formed by six major residues (His235, His250, Lys290, Thr341, Tyr343, and Ser294), which is significant for replication in the host cell (Xu et al., 2006). The inhibitors that target nsp15, are modified oligonucleotides containing derivatives or uracil derivatives, such as 2′-fluoroine-modified RNA and tipiracil on the uridine ribose (Guo et al., 2017; Kim et al., 2021). As part of the nsp15′ tipiracil complex, tipiracil inhibits the activity of EndoU and analogously binds to uridine. EndoU is inhibited by compounds that perturb the stability of the hexamer conformation (Kim et al., 2021; Kish and Uppal, 2016). Additionally, SARS-CoV-2 with nsp15 defects has been proposed as a live attenuated virus vaccine (Hackbart et al., 2020).

5. Accessory factors

5.1. ORF3a

ORF3a is a membrane-associated protein located between the gene of the spike and envelope protein (Yu et al., 2004), which is associated with apoptosis, pathogenicity, and virus release (Ren et al., 2020). The recent cryo-EM structure of SARS-CoV-2 ORF3a (PDB: 6XDC) has N-terminal ectodomain, three transmembrane regions, a cysteine-rich domain, YxxΦ domain, diacidic domain, and C-terminal endodomain (Fig. 8a) (Kern et al., 2021). Cysteine-rich domain forms homo and hetero-tetramers, which cluster together with S protein by forming interchain disulfide bonds and a potassium-permeable ion channel (McBrride and Fielding, 2012). Potassium–permeable ion channel is significant for integrating viral particles (Issa et al., 2020). It also interacts with the M and E proteins and helps viral assembly. It downregulates the type 1 interferon receptor by inducing serine phosphorylation within the IFN alpha-receptor subunit 1 (IFNAR1) degradation motif. By promoting TRAF3-dependent ubiquitination of caspase recruitment domain (ASC), ORF3a protein activates the NLR family pyrin domain containing 3 (NLRP3) inflammasome and inflammation plays an important role in viral infection (Siu et al., 2019). Apoptosis is a cell-death program that occurs in response to several extrinsic and intrinsic signals such as cellular stress, including virus infection. The ORF3a protein also triggers the mitochondrial death pathway by activating p38 MAP kinase (Padhan et al., 2008). The tyrosine-based sorting motif, YXXΦ, is essential for its trafficking to the plasma membrane from Golgi. ORF3a interacts with Caveolin-1, a protein that is part of lipid-rich regions of the membrane and plays an important role in cell signaling, cell cycle, and virus uptake (Minakshi and Padhan, 2014). Many signaling pathways are regulated by caveolin-1, including the extracellular regulated kinase (ERK) and inducible nitric oxide synthase (iNOS) pathways. These two critical pathways are involved in cell survival, proliferation, and response to viral infection (Padhan et al., 2007). ORF3a-HMOX1 (Gordon et al., 2020) complex plays an important role in anti-inflammatory effects through the NLRPS pathway (Lee and Chau, 2002; Lv et al., 2018).
Treatment of COVID-19 with drugs that block the interaction between ORF3a and HMOX1 is an effective approach. In addition, anti-ORF3a antibodies have been detected in the plasma of patients convalescing from COVID-19 (Grifoni et al., 2020; Oja et al., 2020). Suppressing the expression of ORF3a in SARS-CoV resulted in less virus release and less morbidity (Castano-Rodriguez et al., 2018; Lu et al., 2006).

5.2. ORF6

The ORF6 of SARS-CoV-2 is a 61 amino acid long protein found in the endoplasmic reticulum and membranes of vesicles such as lysosomes and autophagosomes (Lee et al., 2021). It is a potent IFN antagonist, a role reported in SARS-CoV previously reported by Kopecky-Bromberg and coworkers (Kopecky-Bromberg et al., 2007). In SARS-CoV, ORF6 was found to co-localize and interact with nsP8. It is still unknown whether ORF6 has a role in RNA synthesis or whether its association with nsP8 influences the replication-transcription complex’s activity (Mariano et al., 2020).

5.3. ORF7a

SARS-CoV-2 ORF7a is a type-1 transmembrane protein that contains a signal peptide, luminal domain or ectodomain, transmembrane segment, and cytoplasmic tail (Liu et al., 2014). Huang et al. reported that SARS-CoV ORF7a physically interacts with SARS-CoV S protein and ORF3a protein for incorporation into mature virions (Huang et al., 2006). An earlier study reported no significant impact on viral RNA replication and synthesis in cell culture due to the elimination of ORF7a from the SARS-CoV genome (Yount et al., 2005). Schaecher et al. reported that both ORF7a and ORF7b deletion do not significantly impact on SARS-CoV replication (Schaecher et al., 2007).

5.4. ORF8

The ORF8 is a unique accessory protein of SARS-CoV-2 having an N-terminal signal peptide (1–17) for transport to the endoplasmic reticulum, and an Ig-like domain (18–121) consisting of a beta-strand core (Fig. 8b) (Hassan et al., 2021). A study reported a novel immunoglobulin (Ig) domain that may play a role as a potential immune modulator to reduce the host immune response against the virus and it contains highly conserved cysteine residues that formed disulfide-bond might enhance to dimerization (Grifoni et al., 2020). Although ORF8 does not participate in viral replication, it possibly plays a direct role in viral pathogenesis by interaction with host molecules and possibly played a significant role in viral trafficking into the cells (Gordon et al., 2020). However, the protein is fast-evolving in SARS-CoV-2 viruses that might be considered a mutational hotspot of the recent pandemic (Su et al., 2020). Drugs can be developed based on the crystal structure of SARS-CoV-2 ORF8 to perturb the interaction between ORF8 and a variety of host proteins, including IL17RA (interleukin 17 receptor), LOX (lysyl oxidase), and GDF15 (growth/differentiation factor 15) (Zinzula, 2021). A substantial antibody response is also generated in the host by ORF8 protein, which has become the major serological marker for SARS-CoV-2 infection (Gordon et al., 2020; Hachim et al., 2020).

6. Potential vaccines for COVID-19

There has been an increased effort around the world to develop vaccines in response to the SARS-CoV-2 outbreak and emerging variants. In order to fight infection, vaccines are administered to people of all ages to build and strengthen humoral and cellular immunity (Fahmi et al., 2021). Globally, 57 authorized/approved vaccines have been developed and 97 vaccines are in various phases of trial (https://www.raps.org/news-and-articles/news-articles/2020/3/covid-19-vaccine-tracker; accessed on July 24, 2022). The majority of these are aimed at inducing neutralizing antibodies against the spike protein (S). As a result, human ACE-2 receptors are prevented from being occupied by these antibodies, preventing viral entry (Thanh Le et al., 2020).

Currently, several vaccine candidates are being tested in Phase 3 clinical trials (Table 2) The mRNA-1273 and mRNA-BNT162b2 vaccines both are encapsulated in a lipid nanoparticle (LNP) that encodes full-length sequence of SARS-CoV-2 S protein and stabilized prefusion conformation, including a transmembrane anchor and an entire S1-S2 cleavage site (Jackson et al., 2020; Sahin et al., 2021). A similar mechanism of action is expressed by both Pfizer and Moderna vaccines. The purpose of these vaccines is to induce both B- and T-cell responses against the spike protein (Patel et al., 2022). The Ad5-nCoV is a replication-defective adenovirus type-5 vectored vaccine that encodes the full-length of SARS-CoV-2 S protein. Although this vaccine’s efficacy is relatively high, but a drawback is that it may not be effective for people with recessive infectious diseases (Han et al., 2021). The AZD1222 is a recombinant adenovirus vaccine that was developed using codon-optimized S glycoprotein. In the shuttle (plasmid) vector, amino acid (2 to 1273) and tissue plasminogen activator (tPA) leader sequences at 5’ end are encapsulated (Kaur and Gupta, 2020). In the human body, these modified adenoviruses cannot replicate as the gene which facilitates virion assembly has been removed and introducing the SARS-CoV-2 antigenic component into the host cell in a safe way (Li et al., 2021). This approach results in the expression of the S protein by host cells, which activates a potent humoral and cell-mediated immune response (Ewer et al., 2020; van Doremalen et al., 2020).

Sputnik V is an adenoviral vaccine containing two vectors which carry the SARS-CoV-2 gene for S protein (Chugh et al., 2021). The SARS-CoV-2 S protein is transferred into the cells through the gene-containing vector (rAd26) during the first vaccination, which then triggers an immune response. During a second vaccination, an additional vector (rAd5) is introduced to the body for boosting immunity, and ensuring long-term protection (Zahid et al., 2021). The JNJ-78436735 is a recombinant vector vaccine that expresses the SARS-CoV-2 spike protein within cells using a human adenoviral vector. By introducing a segment of DNA from SARS-CoV-2 into the adenovirus, which has been genetically modified so that it cannot replicate in the body, CoronaVac is an inactivated vaccine. A dead version of SARS-CoV-2 is used to prevent replication, however, the surface spike protein is preserved to trigger the immune system to form antibodies against the live virus (Wong et al., 2022).

7. Conclusion

In response to the SARS-CoV-2 pandemic, considerable progress has been made in revealing the structural features of the non-structural, structural, and accessory proteins of SARS-CoV-2 and their functionalities in the virus’s life cycle. The viral protein’s structures have assisted the development of various antiviral drugs and antibody treatments. Particularly, antiviral drug Veklury (also known as remdesivir) was designed targeting the RNA-dependent polymerase. In addition, Paxlovid and Lagevrio (molnupiravir) were developed inhibiting the main protease, and several monoclonal antibody treatments are offered for treatment targeting the spike protein. However, for emerging variants, continuous research and development initiatives are expected to improve the strategy of the structural-based rational drug design. Drugs and biological products targeting the various conserved binding sites or multiple key sites of viral and host proteins in the life cycle of SARS-CoV-2 can assist developing more potent therapeutics for covid treatment. In summary, this review provided details structural insights of the representative viral proteins of SARS-CoV-2 to advance the development of antiviral drugs, peptides, antibody, and vaccine preventing and treating the COVID-19.

Declaration of Competing Interest

The authors declare that they have no known competing financial
Yan, F.F., Gao, F., 2021. Comparison of the binding characteristics of SARS-CoV and SARS-CoV-2 RdRPs to ACE2 at different temperatures by MD simulations. Brief Bioinform 22, 1132–1136. https://doi.org/10.1093/BIB/BRAB044.

Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M., Liu, Z., Wang, H., Li, Y., Chen, Y., Guddat, L.W., Wang, Q., Rao, Z., Lou, Z., 2021. Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. Cell 184, 184. https://doi.org/10.1016/J.CELL.2020.11.016.

Yang, X., Chen, X., Bian, G., Tu, J., Xing, Y., Wang, Y., Chen, Z., 2014. Proteolytic processing, deubiquitinase and interferon antagonist activities of Middle East respiratory syndrome coronavirus papain-like protease. J. Gen. Virol. 95, 614–626. https://doi.org/10.1099/vir.0.059014-0.

Yin, W., Mao, C., Luan, X., Shen, D.D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., Chang, S., Xie, Y.C., Tian, G., Jiang, H.W., Tao, 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 1979 (368), 1499–1504. https://doi.org/10.1126/science.abc1560.

Yount, B., Roberts, R.S., Sims, A.C., Deming, D., Frieman, M.B., Sparks, J., Denison, M.R., Zhai, Y., Sun, F., Li, X., et al., 2005. Insights into SARS-CoV transcription and replication stages of clinical trials, mode of actions and efficacy. 28, 225–242. https://doi.org/10.1074/JBC.RA120.012355.

Zhang, L., Lin, D., Huang, X., Sun, C., Caruth, U., Drosten, C., Sauererling, L., Becker, S., Rox, K., Hilgenfeld, R., 2020b. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science 1979 (368), 409–412. https://doi.org/10.1126/science.abb3405.

Zhang, W.F., Stephen, P., Stephen, P., Theriault, J.F., Wang, R., Lin, S.X., 2020c. Novel Coronavirus Polymerase and Nucleotidyl-Transferase Structures: Potential to Target New Outbreaks. J. Phys. Chem. Lett. 11, 4430–4435. https://doi.org/10.1021/ACS.JPCL.9B00571.

Zhang, J., Xiao, T., Cai, Y., Chen, B., 2021. Structure of SARS-CoV-2 spike protein. Curr. Opin Virol 50, 173–182. https://doi.org/10.1016/J.COVIRO.2021.08.010.

Zhang, C., Xu, D., Duan, Y., Pan, X., Sun, Y., You, T., Han, L., Jin, Z., Zhang, W., Ju, J., Guo, H., Liu, Q., Wu, Y., Peng, C., Wang, J., Zhu, C., Yang, X., Yang, K., Lei, Y., Guddat, L.W., Xu, W., Xiao, G., Sun, L., Zhang, L., Rao, Z., Yang, H., 2021. High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. Protein Cell 12, 877. https://doi.org/10.1007/s13238-021-00836-9.

Zhou, R., Zeng, R., von Brunn, A., Lei, J., 2020. Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. Molecular Biomedicine 1, 2. https://doi.org/10.1186/s43556-020-00001-4.

Zebuyr, J., 2004. Molecular biology of severe acute respiratory syndrome coronavirus. Curr. Opin. Microbiol. https://doi.org/10.1016/j.mib.2004.06.007.

Zebuyr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. https://doi.org/10.1099/vir.0.059014-0.

Ziebuhr, J., 2021. Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. Cell 184, 184. https://doi.org/10.1016/J.CELL.2020.11.016.

Yount, B., Roberts, R.S., Sims, A.C., Deming, D., Frieman, M.B., Sparks, J., Denison, M.R., Zhai, Y., Sun, F., Li, X., et al., 2005. Insights into SARS-CoV transcription and replication stages of clinical trials, mode of actions and efficacy. 28, 225–242. https://doi.org/10.1074/JBC.RA120.012355.