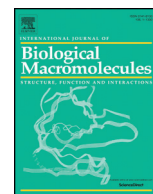




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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.



Review

SARS-CoV-2: Insights into its structural intricacies and functional aspects for drug and vaccine development

Mandeep Kaur ^{a,1}, Akanksha Sharma ^{a,c,1}, Santosh Kumar ^b, Gurpal Singh ^c, Ravi P. Barnwal ^{a,*}^a Department of Biophysics, Panjab University, Chandigarh 160014, India^b Department of Biotechnology, Panjab University, Chandigarh 160014, India^c UIPS, Panjab University, Chandigarh 160014, India

ARTICLE INFO

Article history:

Received 9 October 2020

Received in revised form 4 January 2021

Accepted 27 February 2021

Available online 1 March 2021

Keywords:

SARS-CoV-2

COVID-19

Coronaviruses

Spike protein

SARS-CoV

Vaccine

ABSTRACT

Globally, SARS-CoV-2 has emerged as threat to life and economy. Researchers are trying to find a cure against this pathogen but without much success. Several attempts have been made to understand the atomic level details of SARS-CoV-2 in the past few months. However, one review with all structural details for drug and vaccine development has been missing. Hence, this review aims to summarize key functional roles played by various domains of SARS-CoV-2 genome during its entry into the host, replication, repression of host immune response and overall viral life cycle. Additionally, various proteins of SARS-CoV-2 for finding a potent inhibitor have also been highlighted. To mitigate this deadly virus, an understanding of atomic level information, pathogenicity mechanisms and functions of different proteins in causing the infection is imperative. Thus, these structural details would finally pave the way for development of a potential drug/vaccine against the disease caused by SARS-CoV-2.

© 2021 Elsevier B.V. All rights reserved.

Contents

1.	Introduction	46
2.	The domain organization of SARS-CoV-2	48
2.1.	ORF1a/b	48
3.	Spike protein	50
3.1.	ORF3	51
3.1.1.	ORF3a	51
3.1.2.	ORF3b	52
4.	Envelope protein.	52
5.	Membrane protein.	53
5.1.	ORF6	53
5.2.	ORF7	53
5.2.1.	ORF7a	53
5.2.2.	ORF7b	53
6.	ORF8	53
7.	Nucleocapsid protein.	54
7.1.	ORF9b	55
7.2.	ORF10	55
8.	Potential targets for vaccine design	55
9.	Future perspectives	56
	CRedit authorship contribution statement.	56

Abbreviations: SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; COVID-19, Coronavirus disease; CoVs, Coronaviruses; SARS-CoV, Severe Acute Respiratory Syndrome Coronavirus; MERS, Middle East Respiratory Syndrome.

* Corresponding author.

E-mail address: barnwal@pu.ac.in (R.P. Barnwal).¹ Equal contributions.

Declaration of competing interest	56
Acknowledgements	56
Appendix A. Supplementary data	56
References	57

1. Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) caused a highly infectious viral disease referred to as coronavirus disease (COVID-19). Its first outbreak in Wuhan city of Hubei province of China in 2019 has been responsible for an unprecedented pandemic globally [1]. After spreading rapidly in the China, this deadly virus has now gripped whole world. Despite best efforts made by the science community to discover a potential cure against this disease, no specific vaccine and drug are accessible either to control the spread or for the treatment of COVID-19 till date.

Coronaviruses (CoVs) are single stranded positive sense RNA viruses belonging to the Order Nidovirales, family *Coronaviridae*, sub-family *Coronavirinae*. Further these can be classified into four genera- *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* [2]. The viruses belonging to the family *Coronaviridae* are the largest of the RNA viruses and are known to target the respiratory tract leading to mild infections in humans. These are also known to cause illness ranging from common cold to more severe diseases such as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). Both SARS-CoV and MERS-CoV have been classified under the genus *Betacoronavirus*. Usually these viruses cause mild respiratory diseases but the past episodes of SARS-CoV and MERS-CoV have proved that members of this family could be potentially fatal and have resulted in endemics [3].

For the last two decades, coronaviruses have been a constant threat for humans; for instance, in November 2002, the first case of Severe Acquired Respiratory Syndrome Coronavirus (SARS-CoV) infection in humans was reported in Foshan, China, later plaguing more than 30 countries worldwide and causing around 774 deaths out of a total 8096 reported cases. Palm civets and racoon dogs were identified as hosts for SARS-CoV [4]. In the year 2012, a severe respiratory infection caused by Middle East Respiratory Syndrome (MERS-CoV) was first

reported in a patient from Jeddah, Saudi Arabia. The total number of cases reported by WHO globally including 27 countries were 2279, out of which 806 people lost their lives [5]. The mode of transmission of MERS-CoV as confirmed by RNA sequencing was from dromedary camel to humans. These variants are zoonotic in origin and have genome similarity with bat coronavirus and SARS-CoV-2.

SARS-CoV-2 has been categorized under *Betacoronavirus* along with SARS-CoV and MERS-CoV. It has a genome size of 30 kb. Currently, the ongoing SARS-CoV-2 outbreak is assumed to have originated from the Huanan seafood market which sells seafood, marmots, birds, bats and other wild animals in Wuhan, China [2,6]. The virus has been studied to be of zoonotic origin and preliminary evidence suggest the possible mode of transmission from bats to humans, even though the Malayan pangolins (*Manis javanica*) also in light, are considered to serve as a reservoir of the virus [7].

Human to human transmission occurs from an infected person to a healthy person via droplets that may be released on coughing, sneezing in the open and talking without covering the mouth, and simply by being in close contact [8]. Several modes of transmission of this virus in humans are reported as shown in Fig. 1 [9,10]. Droplet inhalation is one of the most common ways through which the virus enters the body and starts the pathogenic mechanism of infection as it reaches the respiratory tract followed by its attack on the epithelial cells. After entry, the virus replicates and causes infection, leading to various local inflammatory changes that damage the lung tissues as well as essential immune cells such as T cells and macrophages [11]. The clinical symptoms reported are fever, breathing difficulties (dyspnoea), headache, muscular soreness, dry cough and pneumonia. Other atypical symptoms are diarrhea and vomiting.

In addition, COVID-19 infection is also exaggerated by release of pro-inflammatory cytokines [such as Interferons (IFN)- γ , Interleukin (IL)-1B, Monocyte Chemoattractant Protein (MCP1)] in an event called 'cytokine storm', indicating hyperactive response and excessive inflammatory reaction by host immune system. Recently, several studies

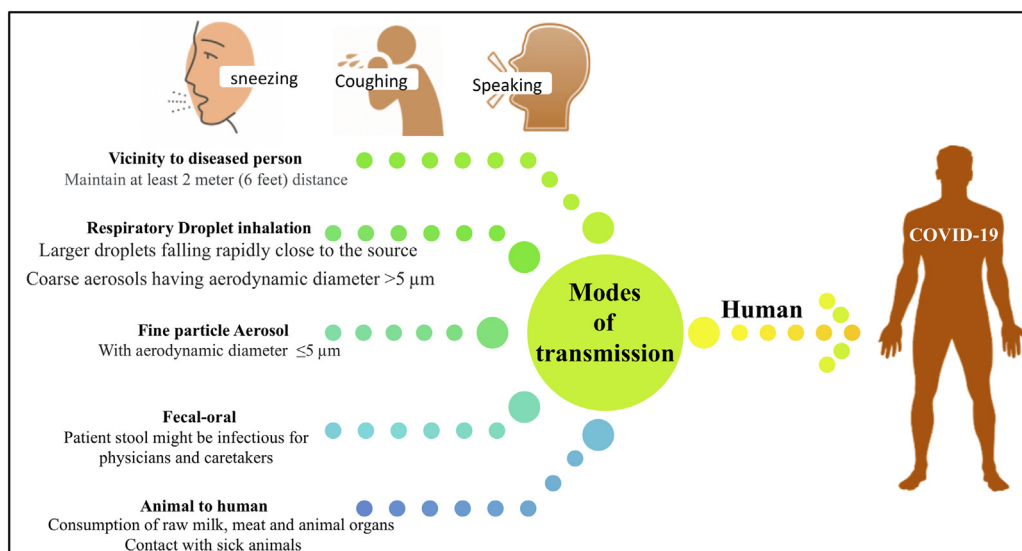


Fig. 1. Modes of transmission of SARS-CoV-2 in humans through various routes.

suggested that cytokine storm is directly responsible for lung damage and multi-organ failure in COVID-19 patients [12].

This pandemic is growing rapidly throughout the world leading to high mortality. According to the WHO reports, the top 5 highly infected countries include the United States of America, Brazil, India, Russian federation and Peru. Along with this, in more than hundred countries, areas and territories on the list, every day new cases of COVID-19 patients are being confirmed. The world map highlighting the top 20 countries which are reported to have the highest number of COVID-19 cases as per WHO is shown in Fig. 2.

A recent research proposes that the full length genomic sequence of human SARS-CoV-2 shares 79.6% identity with SARS-CoV whereas 96% identity with the whole-genome sequence of a bat coronavirus [13]. Another report suggests 89.1% nucleotide identity of SARS-CoV-2 with SARS-like coronaviruses (genus *Betacoronavirus*, subgenus *Sarbecovirus*) on the basis of phylogenetic analysis [14]. These findings of correlation of SARS-CoV-2 with SARS-CoV based on the full-length genome using phylogenetic analysis, the putatively similar cell entry mechanism and human cell receptor usage could be helpful in designing drugs and pre-clinical vaccines to treat this disease [15,16]. The structural and functional relatedness of SARS-CoV-2 with SARS-CoV is well established now [17]. Moreover, it has been found that the novel SARS-CoV-2 uses Angiotensin converting enzyme 2 (ACE2) to target the host, just like SARS-CoV [13].

In this direction, we have performed phylogenetic analysis using NCBI BLAST pairwise alignment to confirm the relatedness of SARS-CoV-2 with other coronaviruses as shown in Fig. 3. BLASTn was performed to compute pairwise alignment between SARS-CoV-2 full genome sequence isolated from a human patient from Wuhan, China (NCBI Reference Sequence: NC_045512.2), and 15 sequences from different CoVs like SARS coronavirus, bat SARS-like coronavirus, pangolin coronavirus, duck

coronavirus and SARS coronavirus civet. Five of these full genome sequences were isolated from human SARS-CoV-2 patients belonging to some of the topmost countries affected by this virus. The statistical method used was fast minimum evolution with a maximum sequence difference of 0.75.

The SARS-CoV-2 genomes from USA (MT893659.1), India (MT434758.1), Russia (MT890462.1), South Africa (MT324062.1) and Brazil (MT126808.1) showed high sequence identity, between 99.5 and 100% to genomic sequence of a patient from Wuhan. Out of these, the sequence isolated from Brazil exhibits the highest genome identity (99.99%), followed by South Africa (99.98%), Russia (99.97%), USA (99.96%) and India (99.95%). The phylogenetic tree clearly reflected that the closest relative of SARS-CoV-2 is a pangolin. The Pangolin coronavirus isolates MP789 (MT121216.1) shares 90.11% whereas isolate PCoV_GX-P1E (MT040334.1) shares around 85.95% identity with SARS-CoV-2. The SARS coronavirus HSR 1 (AY323977.2), SARS coronavirus Tor2 (NC_004718.3), SARS coronavirus TWK (AP006559.1) and SARS coronavirus BJ162 (AY864805.1) showed ~82.30% identity to each other. Other related CoVs according to our analysis are SARS coronavirus civet007 (AY572034.1/82.27%), Bat SARS-like coronavirus (MG772933.1/89.12%, MG772934.1/88.65%). Besides, the Duck coronavirus (NC_048214.1/73.71%) being an outgroup is just 73.71% identical to SARS-CoV-2.

This review aims to facilitate the process of finding a potential vaccine against COVID-19 by providing information about domain organization and the overall structure of SARS-CoV-2 along with its various proteins and their essential roles in causing viral infection. Further, we outline the protein domains which can prove to be good target for designing drugs or vaccines to save thousands of life under risk globally.

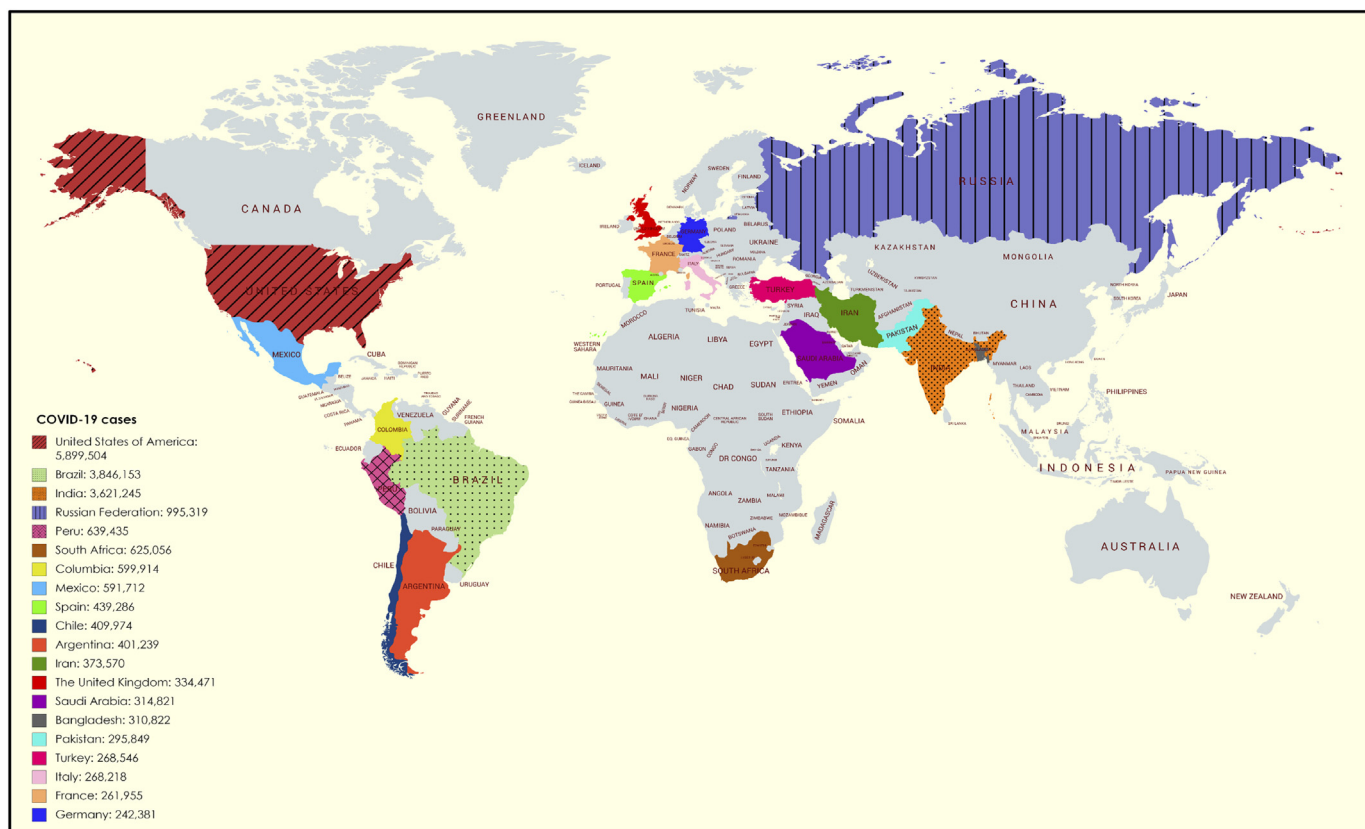


Fig. 2. The map depicts the statistics reported for number of cases among the top 20 countries across the globe severely affected by COVID-19 till date. The race for developing a cure intensifies as cases continue to soar each day. The data as on Sep 1, 2020 has been obtained from WHO website (<https://covid19.who.int/table>) and the map has been created using Mapchart.net. The top 5 highly affected countries include the United States of America, Brazil, India, Russian federation and Peru which are highlighted using shapes like diagonal stripes, patch, texture, vertical stripes and crosshatch respectively.

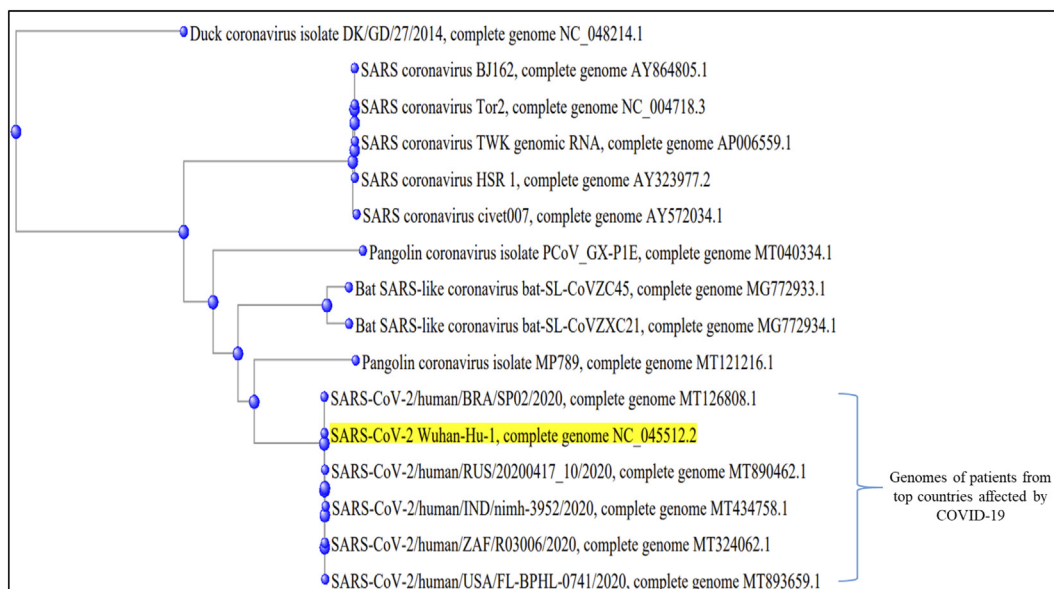


Fig. 3. The tree was constructed using 16 sequences from different coronaviruses to infer the evolutionary history of SARS-CoV-2 by means of fast minimum evolution method having maximum sequence difference of 0.75. NCBI BLAST pairwise sequence alignment was used to generate the tree. Human isolates from different countries have nearly complete identity to the genome isolated from Wuhan in China (in yellow highlight), followed by the Pangolin coronavirus isolate MP789 (MT121216.1) with 90.11% identity. All the 15 sequences were downloaded from NCBI and were used to generate a phylogenetic tree to study the evolutionary relationship.

2. The domain organization of SARS-CoV-2

With a genome size of about 30 kb, the SARS-CoV-2 genes from 5'-end to 3'-end consists of various domains that includes open reading frames (ORFs): ORF1a/b, ORF 2 (referred to as Spike protein), ORF3a, ORF3b, ORF4 (also known as Envelope protein), ORF5 (known as Membrane protein), ORF6, ORF7a, ORF7b, ORF8, ORF9 (referred to as Nucleocapsid protein), and ORF10 [18,19]. A complete genome organization from 5'-end to 3'-end is depicted in Fig. 4 [19]. These ORFs are further categorized into structural proteins, non-structural proteins (nsps) and accessory proteins as shown in Fig. 4. The ORF1a/b is cleaved into 16 nsps by viral proteases-Chymotrypsin like protease (3CL^{pro}), also called main protease (M^{pro}) and papain like protease (PL^{pro}) [17]. After the fusion of virus, the viral genetic material enters the cytoplasm and mRNA then synthesizes two polyproteins- pp1a and pp1ab. These proteins code for these non-structural proteins [20]. The structural proteins consist of Spike (S) protein, Envelope (E) protein, Membrane (M) protein and Nucleocapsid (N) protein whereas accessory proteins consist of ORF 3a, 3b, 6, 7a, 7b, 8a, 8b, 9b and 10. In addition, various functions of these protein factors of SARS-CoV-2 emphasizing their importance for the viral life cycle are described in Fig. 5.

2.1. ORF1a/b

The virus's ORFs code for various non-structural proteins (nsps). The first open reading frame (ORF 1A/B) occupies two-third of the total size of the viral genome and encodes two translational products that are polyproteins 1a and 1ab (pp1a and pp1ab). Here, the main protease (M^{pro}), also called 3CL^{pro} and papain-like protease (PL^{pro}) are cysteine proteases which are reported to be involved in processing pp1a and pp1ab into mature non-structural proteins [13,21]. There are 16 nsps that play a key role in viral RNA synthesis via forming the replication-transcription complexes as described in the following section (Supplementary Fig. 1) [22].

The putative nsp1 of the SARS-CoV-2 is known as leader protein and participates in suppressing host gene expression as well as promotes degradation of host mRNA [23]. Moreover, nsp1 facilitates viral replication and has the ability to evade the host immune response [23]. This nsp is among the first protein to be expressed after viral infection. The cryo-EM structure of SARS-CoV-2 nsp1 and rabbit 40S ribosome complex (PDB code 7JQB) confirmed its function in effective translation of viral protein and provided insights into the mechanism via which it blocks the translation of host mRNA [23]. In SARS-CoV nsp1 was also proven to inhibit host gene expression via targeting 40S ribosome,

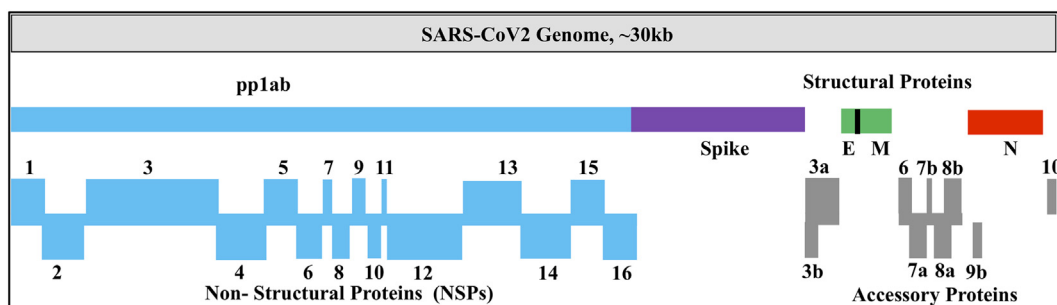


Fig. 4. Complete domain organization of SARS-CoV-2 genome. This represents the genomic arrangement of various proteins of the SARS-CoV-2 from 5'-end to 3'-end.

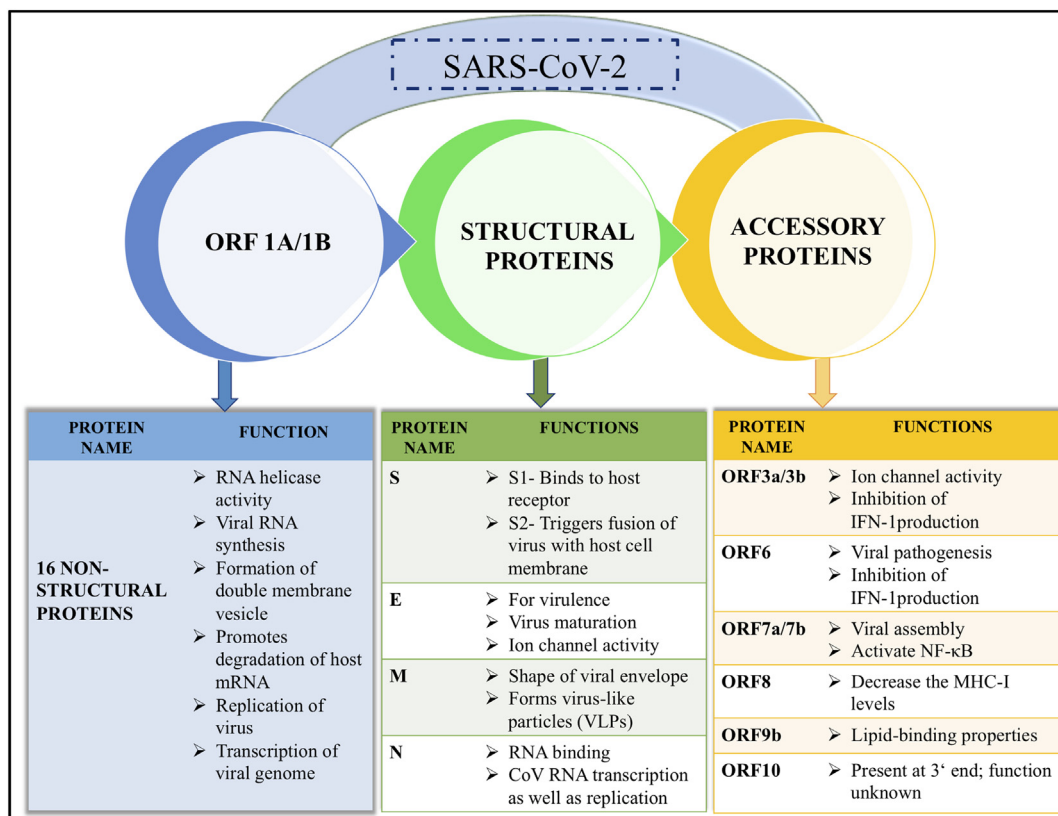


Fig. 5. Categorization of SARS-CoV-2 proteins into three types- Open Reading Frames (ORFs), structural proteins and accessory proteins. Different proteins factors under these three categories are described along with their respective functions in the viral genome.

which is crucial for host protein synthesis and alters the host mRNAs to promote their degradation [24]. The SARS-CoV-2 nsp2 has been reported to bind with two proteins of the host namely prohibitin 1 and prohibitin 2. These two host proteins show involvement in cell migration, progression of cell cycle, apoptosis, cellular differentiation and mitochondrial biogenesis. It is anticipated that the direct contact of nsp2 with proteins prohibitin 1 and prohibitin 2 has some role in disturbing the environment of host cell [25]. The SARS-CoV nsp2 does not involve in the process of viral replication [25]. Further, its examination has accentuated it to be unnecessary for viral replication [26].

Nsp3, a viral cysteine protease also referred as papain-like protease (PL^{pro}), has a molecular mass of approximately 200 kDa. This protein possesses various conserved domains such as protease (papain like protease), transmembrane and ssRNA binding domain etc. The presence of a tetrapeptide motif LXGG has been discovered between the proteins nsp1-2, nsp2-3 and nsp3-4, where X is N, L and L, respectively. This motif is recognized by PL^{pro} which further hydrolyses the G peptide bond present at the carboxyl site, resulting in the release of nsp1, nsp2, nsp3 and nsp4 (Supplementary Fig. 1) [27]. Furthermore, these four nsp3s have been reported to be necessary for the replication of virus [28,29]. Notably, the high-resolution structural analysis of nsp3 macro1 viral domain (mac1) in SARS-CoV-2 (PDB code 6WEY) revealed its binding with ADP-ribose [30]. Here, mac1 domain takes part in removing ADP-ribose from different target proteins; this ability is expected to be associated with cytokine storm syndrome observed in severe cases of COVID-19 [30]. Interestingly, the involvement of SARS-CoV nsp3 has been predicted to hinder the chemokine and cytokine production which results in evading the host immune response by the virus. This makes nsp3 an excellent target for designing antiviral therapeutics against COVID-19 [31]. Recently, two potent inhibitors referred as VIR250 and VIR251 have been developed with high selectivity and are shown to successfully inhibit PL^{pro} of SARS-CoV-2. The structures

of PL^{pro} with VIR250 (PDB code 6WUU) and VIR251 (PDB code 6WX4) underlined the inhibitory mechanism that would be helpful in facilitating future antiviral peptides development against nsp3 [31].

The nsp5, also known as 3C-like proteinase or chymotrypsin-like or main protease (M^{pro}), cleaves the nsp polyprotein at 11 sites leading to the formation of other intermediates or mature non-structural proteins [18,32]. It has been characterized as the best target for drug discovery because of its essential enzymatic ability to process the polypeptides translated from the CoV RNA [33]. In humans, there are no shreds of confirmation regarding presence of proteases with similar cleavage specificity like SARS-CoV-2 main protease. The inhibition of M^{pro} enzyme activity results in blocking virus replication [33]. Additionally, SARS-CoV-2 M^{pro} in apo-form (PDB code 7JPY) has been visualized using X-ray crystallography [34]. Further, crystallographic analysis of this SARS-CoV-2 M^{pro} bound to seven inhibitors accentuated two potent inhibitors i.e., MP15 (PDB code 7JQ2) and MP18 (PDB code 7JQ5) which effectively prevented the cytopathogenic effect of SARS-CoV-2 in Vero E6 cells [34]. The nsp6 is a multiple-spanning transmembrane protein that plays a role in generating autophagosomes and formation of double membrane vesicles. Besides, nsp6 forms a complex with nsp3 and nsp4 [35]. The yeast two hybrid analysis showed the interaction of nsp6 with nsp3 through N-terminal region [36]. The nsp7 has been reported to form a heterodimer with nsp8, which further makes a complex with nsp12 resulting in activating its RNA polymerase activity [37,38].

The non-structural protein 9 (nsp9) of SARS-CoV-2 is involved in viral replication and reproduction of the viral genome. This protein has its counterpart present in the SARS-CoV family. The crystal structure of apo-Nsp9 from SARS-CoV-2 (PDB code 6WXD) has been described in effort to understand its function during viral life cycle [39]. The SARS-CoV-2 Nsp9 shares 97% sequence similarity with nsp9 of SARS-CoV, which has been previously reported to be essential for virulence of

SARS-CoV. Based on high sequence similarity shared by SARS-CoV-2 and SARS-CoV nsp9, they are anticipated to show functional conservation [39].

Nsp10 is exclusively present among the viruses; it contains 140 amino acids in case of SARS-CoV-2. The SARS-CoV nsp10 analysis underlined the necessity of this protein for ensuring stimulation of other proteins such as nsp14 and nsp16, thereby acting as a scaffolding protein [40]. Its role in stimulating the 2'-O-MTase activity of nsp16 and exoribonuclease activity of nsp14 is well known [40]. Moreover, nsp10 plays pivotal role in crucial RNA methylation and in replication of CoV virus. In a recent study, 1.6 Å crystal structure of SARS-CoV-2 nsp10 (PDB code 6ZPE) is reported along with its comparison with SARS-CoV (PDB code 2FYG) and other pathogenic CoV strains [40]. The structural comparison of nsp10 from SARS-CoV-2 and SARS-CoV displayed high similarity with rmsd of ~0.5 Å for all the atoms in the structure. In addition, sequence similarity of nsp10 from SARS-CoV-2 reflected it to be a new variant of SARS-CoV, with the exception of two amino acids i.e., Pro at position 23 mutated to Ala and Arg at 113 mutated to Lys [40].

The nsp12 shows RNA-dependent RNA polymerase (RdRp) activity and can carry out polymerase reaction independently, although the efficacy of this reaction is quite low. However, in the presence of the other two nsps, nsp7 and nsp8, its polymerase activity gets stimulated. Additionally, earlier investigations revealed that a complex of nsp12-nsp7-nsp8 mediates the synthesis of coronavirus RNA [37,38]. Also, nsp12 acts as a key component for directing SARS-CoV-2 replication and transcription. The cryo-EM structure of SARS-CoV-2 nsp12-nsp8-nsp7 complex (PDB code 6M71) affirms its structure to be similar to that of SARS-CoV (PDB code 2AHM) with rmsd value ~0.82 Å, obtained for 1078 C α atoms. These structural details would be helpful for designing antiviral drugs to target RdRp [41].

Intriguingly, SARS-CoV-2 takes maximum advantage of the cell's environment for its replication and to ensure the stability of viral RNA during its life cycle. The genomic RNA stability is crucial for effective translation and survival of SARS-CoV-2 inside the cell [42]. A type 1 cap (Supplementary Fig. 2) with N-methylated guanosine triphosphate and C2'-O-methyl-ribosyladenine is added at the 5'-end of the viral RNA to ensure stability and prevent its degradation. Coronaviruses possess their own enzymes to facilitate this capping at the transcribed viral mRNA [42]. In case of SARS-CoV-2, the cap installation involves the participation of nsp10, nsp13, nsp14, and nsp16 proteins. During the replication process of the virus, nsp13 aids in the unwinding of viral RNA, thus acting as a helicase [42]. In addition, nsp13 cleaves monophosphate at the nascent RNA 5'-end to generate a diphosphate using its triphosphatase activity [42]. Besides, nsp13 has also been discovered to be involved in RNA synthesis as it shows RNA helicase activity. However, how this process takes place is still not clear [43].

The bimodular protein nsp14 exhibits N7-methyltransferase and 3'-to 5'- exoribonuclease activities. Further, in order to stimulate exoribonuclease activity of nsp14, its binding to nsp10 has been examined to be significant [40]. The nsp14 is predicted to directly make contact with the nsp12 to form an elongated complex and direct precise RNA synthesis. Notably, nsp16 has been predicted to form a stable holoenzyme and with the contribution of nsp10 and nsp14, it sets up the capping components in transcription [44,45]. Additionally, nsp15 exhibiting endonuclease activity is important for the biology of coronavirus as it partakes to hide the viral RNA from the defense system of host by degrading it [46]. Two crystal structures of SARS-CoV-2 nsp15 in its apo- and citrate-bound form are available with PDB codes 6VWW and 6W01, respectively [46]. The structure comparison of SARS-CoV-2 nsp15 (citrate-bound form) with nsp15 of SARS-CoV (PDB code 2H85) [47] revealed high similarity with rmsd value 0.52 Å. Additionally, the sequence similarity between nsp15 of SARS-CoV and SARS-CoV-2 was observed to be 88%, suggesting SARS-CoV-2 nsp15 to be close homolog of SARS-CoV and proposing that "inhibitors for SARS-CoV may be effective for SARS-CoV-2 as well" [46]. The viral replication and transcription have been observed to be mediated by a multi-subunit polymerase complex

made by the assembly of these non-structural proteins. Moreover, the complete transcription as well as replication of the viral genome requires other nsps such as nsp10, nsp14, nsp13 and nsp16 etc., whose exact function during viral RNA synthesis is yet to be determined [18,43]. The key functions of these nsps have been mentioned in Fig. 5. Here, we are also delineating the vital roles/functions of individual nsps during SARS-CoV-2 viral life cycle in Table 1.

3. Spike protein

The spike protein encoded by the S gene is a large glycoprotein with a molecular mass of 180 kDa. It is known to bind to the angiotensin receptor and doesn't show much homology to other coronaviruses but shows 93.1% identity with the bat coronavirus RaTG13. The ACE2 used for cell entry by SARS-CoV-2, the causative agent for COVID-19, served as a receptor for SARS-CoV also [48]. The S protein consists of two domains- S1 and S2. The S1 region further possesses an N-terminal and a C-terminal subdomain which bind sialic acid and a proteinaceous receptor, respectively whereas the S2 subunit of this protein comprises of subdomains that are the fusion peptide region (FP), heptad-repeat 1 region (HR1), heptad-repeat-2 region (HR2), the transmembrane (TM) and the C-terminal endodomain (E) as shown in Fig. 6 [6].

The transmembrane spike protein is known to exist as homotrimer on the viral surface. The S1 domain is assumed to facilitate the receptor binding whereas the S2 subunit mediates fusion of the virus to the cell membrane of the host by means of heptad repeats HR1 and HR2 [50]. The cryo-electron microscopy structural analysis of the SARS-CoV-2 Spike ectodomain trimer revealed that it exists in perfusion conformation which is observed to undergo extensive reorganization of its structure for fusion of the virus membrane with the host cell membrane [49]. The S1 subunit binding to the host cell receptor is anticipated to trigger this process of fusion. This receptor binding results in destabilization of the perfusion trimer conformation that leads to shedding of S1 subunit. Further, S2 subunit undergoes a transition to form a stable conformation post fusion. Interestingly, it is noticed for all CoVs that boundary between the two subunits of spike protein- S1 and S2 is the point where cleavage occurs. The distal S1 subunit contains RBD which engages the host cell receptor by undergoing movements adopting hinge-like conformations. There are two states proposed which are referred to as up and down conformation. In the up conformational state the receptor site is accessible, whereas down conformational state represents the state in which the receptor site is inaccessible [49]. In addition, the spike protein is reported to be cleaved via host proteases at site S2' which is positioned upstream of the fusion peptide in most CoVs. This

Table 1
Essential functions of all the non-structural proteins.

Non-structural protein	Function
Nsp1	Suppresses host gene expression and promotes mRNA degradation
Nsp2	Binds with prohibitin 1 and prohibitin 2
Nsp3	Mediates the release of nsp1, 2 and 3 (PL ^{Pro} motif)
Nsp4	Involved in Complex formation with nsp3 and nsp6
Nsp5	3C-like protease (3CL ^{Pro}) cleaves nsp polyprotein at 11 sites
Nsp6	Generates autophagosomes and double membrane vesicles
Nsp7	Complex formation with nsp8
Nsp8	Complex formation with nsp7
Nsp9	Binds DNA/RNA
Nsp10	Stimulates nsp16 methyltransferase activity
Nsp11	Short peptide at ORF1a domain end; function unknown
Nsp12	RNA-dependent RNA polymerase activity
Nsp13	Involvement in RNA synthesis due to RNA helicase activity
Nsp14	Complexes with nsp12 and directs RNA synthesis
Nsp15	Interferes with host defense system by degradation of viral RNA via endonuclease activity
Nsp16	2'-O-ribose methyltransferase

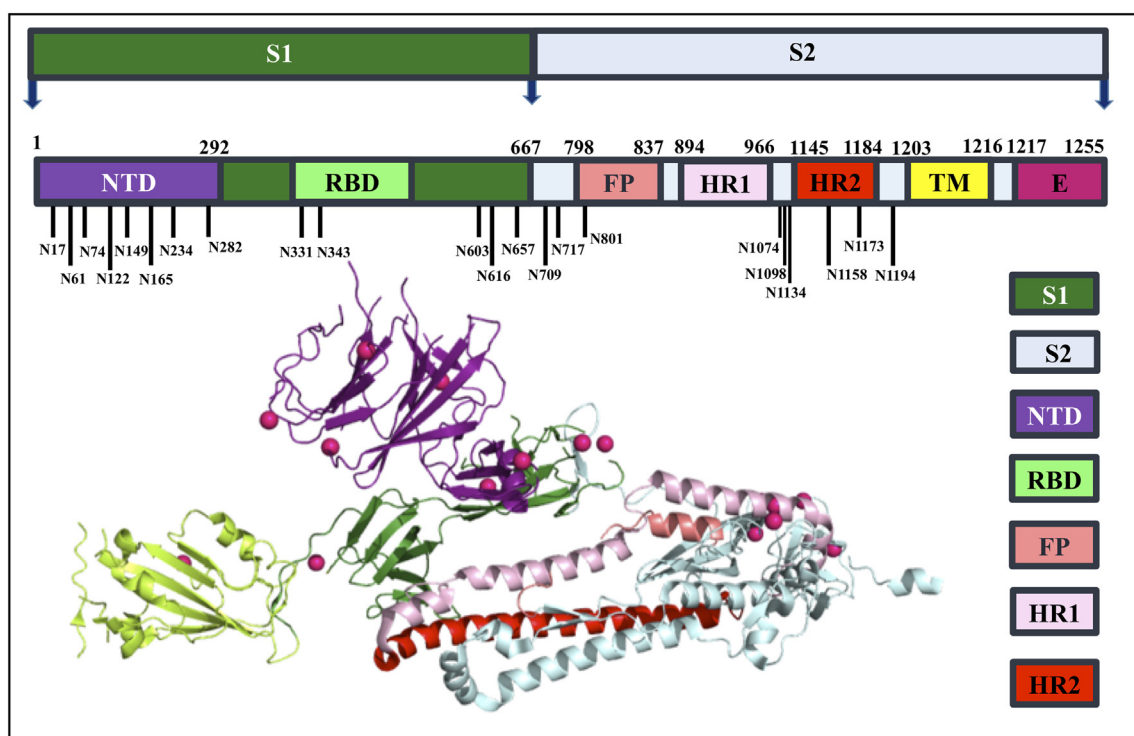


Fig. 6. The domain organization and cryo-EM structure of Spike protein of SARS-CoV-2. Two subunits S1 (dark green) and S2 (light blue) along with the different subdomains are depicted wherein 22 N-glycosites are depicted. The NTD domain, RBD/RBM of S1 are shown in purple and light green/yellow color, respectively whereas FP, HR1, HR2, TM and E region are displayed in punch, light purple, red, orange, and magenta/maroon color, respectively [6,49]. The cryo-EM structure colors are presented according to the color assigned to the domains (PDB code 6VSB); here the observed N-glycosites (16 nos) are displayed with hotpink spheres [49].

cleavage activates the spike protein for membrane fusion by promoting various irreversible changes in the conformation of the structure [51]. Overall the viral entry into the host cell is a complex process that requires concerted actions such as receptor binding and proteolytic processing of the spike protein to happen either concurrently or independently. The aforementioned information about essential functions of spike protein makes it indispensable. All these details of spike protein of SARS-CoV-2 prove that it is a potential target for antibody-mediated neutralization and its atomic level information can be helpful in designing vaccine and/or for drug development [49,51].

The receptor binding domain (RBD) of the S protein of the virus is crucial for interaction with the ACE2 receptor of the host. The structural features of this hot spot including relatively flat surface, glycosylation free site and the presence of salt bridge surrounded by hydrophobic tunnel wall altogether favor the virus easy accessibility and binding [11].

Mass spectroscopy characterization of intact glycopeptides has revealed 22 N-glycosites on the surface of SARS-CoV-2 S protein, which are also found to be preserved in 753 genome sequences of SARS-CoV-2 (Fig. 6). The glycans of spike protein have been highlighted to be extremely significant because of their pivotal role in mechanism by which the human ACE2 receptor attaches with virus. The S protein N-glycosylation plays essential role in directing the proper folding of glycoprotein and priming of protein by host proteases. The identification of site-specific N-glycosylation information might be useful for immunogen design, developing therapeutic antibodies or drug and underlining the precise viral invasion mechanism [52].

Studies have provided the evidence that RBD is capable of folding independently of the S protein and possesses all the information needed for binding to the host receptor [53]. The ACE2 proteins through the epithelial cells of lung alveoli and enterocytes of the small intestine. Further, the presence of ACE2 has also been confirmed in endothelial cells

and smooth muscle cells [54]. Thus, it reiterates the abundance and importance of this protein in humans and its possible implications of binding with SARS-CoV-2 can be far-reaching.

3.1. ORF3

3.1.1. ORF3a

ORF3a encodes an accessory transmembrane protein located between the structural proteins S and E. It is a hydrophobic protein and contains several conserved regions including a cysteine rich domain which provides it ion-channel activity; a tyrosine based sorting domain Yxx ϕ and a diacidic EXD domain having its roles in intracellular trafficking. Besides, these domains also regulate the localization of the SARS-CoV ORF3a in the cell. Even though the Yxx ϕ and cysteine rich domains are conserved in SARS-CoV-2, the EXD domain is not reported, in its place, a SGD domain is present [55,56]. It has been shown to form homodimers, homotetramers and oligomers of 31 kDa in SARS-CoV. Its ion channel activity has been reported in *Xenopus oocytes* [57]. Further, it has been confirmed that ORF3a induces cells to undergo programmed cell death or apoptosis to control the viral spread through several studies carried out in SARS-CoV. Furthermore, research carried out in different cell lines proposes that caspase-3 levels are elevated in the presence of ORF3a. Hence, it is a marker of apoptosis promoted by caspase [56,58]. The cryo-EM structure highlighted that the 3a domain of SARS-CoV-2 forms a dimer which is ~70 Å in length (Supplementary Fig. 3). It possesses a 40 Å long transmembrane region (TR) and a cytosolic region (CR) that extends up to 30 Å; both these are connected through the turn-helix-turn motif. Every protomer in TR consists of three α -helices α 1, α 2 and α 3. Short intercellular linkers connect α 1 to α 2 whereas short extracellular linkers connect α 2 to α 3. The monomer in brown color consists of inner core β 3, β 5, β 8, β 4 and β 7 which in turn interact with another monomer in yellow, possessing several

residues such as V168, V225, F230 and I233 in the inner side forming a hydrophobic core. Each protomer with the help of β sandwiches interacts with the other monomer, thus stabilizing the dimer by forming a stable link as shown in Fig. 7. The structural characterization of ORF 3a from SARS-CoV-2 reveals that it is a class III viroporin [59]. Viroporins are a group of proteins important for viral pathogenesis and formation of viral particles. These are categorized into different classes based on the number of transmembrane domains. The structural studies provide the evidence that ORF3a possesses three transmembrane domains [60]. Further, it provides better understanding of channel activity in proteoliposomes, which may facilitate development of vaccine using this domain [59].

3.1.2. ORF3b

SARS-CoV ORF3b along with ORF6 and nucleocapsid protein is capable of inhibiting the production of type I IFN. This ORF isolated from SARS-CoV-2 is shorter than the one obtained from SARS-CoV. In fact, the length of its C terminal region is linked with the effectiveness of the IFN inhibition. It has been observed that the inhibition of type I IFN response is considerably higher by SARS-CoV-2 than that observed for SARS-CoV [61]. SARS-CoV-2 has been reported to be located in the plasma membrane [56] while SARS-CoV ORF3b containing 154 amino acids is localized in nucleolus and mitochondria [62]. It promotes

necrosis and apoptosis as observed in many cell lines and hinders cell division by inhibiting the transition of cell from G0 or G1 phase to S phase [63].

4. Envelope protein

The Envelope (E) protein is the smallest structural protein with ample expression in the infected cells. This protein is mostly confined around the ER, Golgi complex and ER-Golgi intermediate compartment (ERGIC), where it is involved in assembly of the virus particles and budding [66]. Though it is not a major component of the viral genome, its roles in various stages of replication and infectivity have been reported along with its importance in maturation of the virus [67].

It is a hydrophobic protein and has been shown to undergo various post translational modifications for the functions it performs residing in the ER and ERGIC. This specifically includes palmitoylation and ubiquitination. Addition of palmitoyl group supports in anchoring of the membrane and such proteins are abundant among enveloped viruses [67,68]. The SARS-CoV E-protein has been known to be ubiquitinated and interacts with nsp3 via its ubiquitin-like domain 1 [69]. E-protein may act as an ion channel, specifically like viroporin, which then oligomerizes and enhances membrane permeability [70]. It is crucial for apoptosis and possesses an N-terminal, a transmembrane

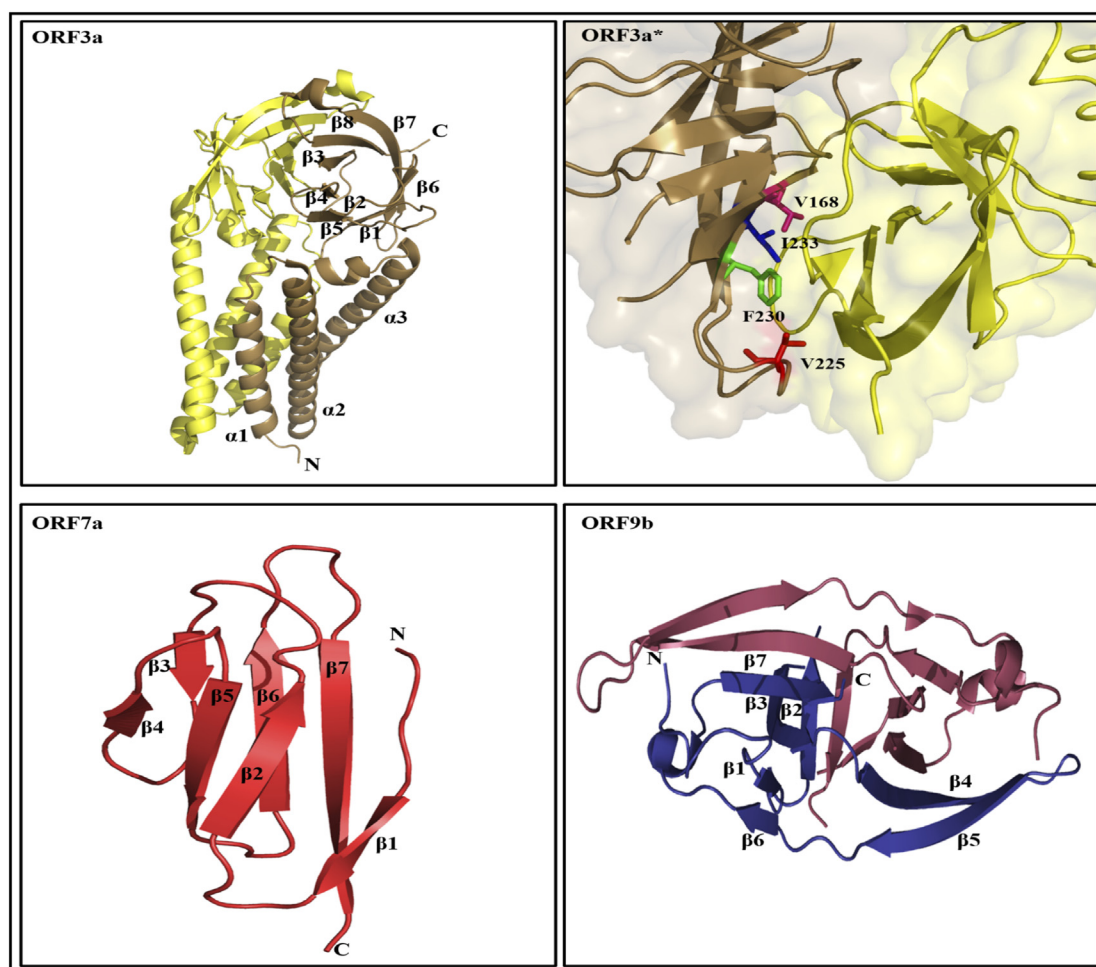


Fig. 7. The top panel represents the cryo-EM structure of ORF3a from SARS-CoV-2 (PDB code 6XDC). The interaction of two monomers (one in yellow and other in brown) results in formation of a stable structure. The three α -helices of the transmembrane region are labelled as $\alpha 1$, $\alpha 2$ and $\alpha 3$. The arrangement of β sandwiches containing 8 β sheets ($\beta 1$ – $\beta 8$) are also shown [59]. The ORF3a* represents the residues labelled as V168, V225, F230 and I233 forming a hydrophobic core in ORF3a. With the help of β sandwiches each protomer interacts with the other monomer, thus forming a dimer. In the bottom panel, crystal structure of ORF 7a of SARS-CoV-2 with seven β -strands core (PDB code 6W37) is shown in bottom left side [64] whereas the bottom right side depicts the X-ray structure of ORF9b from SARS-CoV-2 (PDB code 6Z4U) (manuscript unpublished) [65].

domain and a C-terminal. Further, the transmembrane region of the protein is responsible for homotypic interactions of the protein as analyzed by molecular dynamics (MD) simulations. It was observed that the dimeric, trimeric and pentameric models were conserved through evolution [71]. The experimental structure for SARS-CoV-2 E protein has not been elucidated yet but a homology model of the protein from SARS-CoV-2 which assembles into a viroporin like pentamer has been obtained by Bianchi and co-workers [72].

The viral envelope consists mostly of Spike (S) and Membrane (M) proteins while only a small proportion of the E protein is incorporated in the envelope as observed in coronaviruses like mouse hepatitis virus (MHV), infectious bronchitis virus (IBV) and transmissible gastroenteritis virus (TGEV) [73–75]. The role of E protein in virulence is clear from an observation whereby its deletion led to weakening of the SARS-CoV in both *in vivo* as well as *in vitro* conditions. Such viruses can prove to be good vaccine candidates against SARS-CoV-2 since BALB/C mice immunized with “SARS-CoV sans E protein (rSARS CoV-ΔE)” are protected against MA15 virus-mediated clinical diseases including respiratory and lung damage and show reduced replication of virus [76].

5. Membrane protein

Membrane (M) protein determines the shape of the viral envelope. It possesses three transmembrane domains [77], is the most abundant protein in the viral genome [78] and can bind all other structural proteins [79]. Binding to M protein helps not only in stabilizing the nucleocapsid, but also the N protein-RNA complex, thus completing viral assembly [80]. It is a glycosylated protein and plays a role in activation of Interferon α response [81]. The M protein has been found to be conserved among the β coronaviruses.

Extensive literature suggests that M protein supports assembly of virus particles by interacting with other proteins like ribonucleoprotein and the S protein as well as through M-M protein interactions [82]. The interaction between S and M protein is vital for S protein's retention in ERGIC and ultimately for its integration into virions [83].

The interaction between E and M protein is one of the most studied interactions among different proteins of CoVs. This interaction takes place at the cytoplasmic side of the ERGIC by means of C-termini of E and M proteins. Both these domains are vital for formation of virus-like particles (VLPs), since their deletion reduces the VLP count [84,85]. The actual structure of the protein has not been determined yet. However, a low resolution approximate structure has been predicted using *ab initio* techniques [72].

5.1. ORF6

The accessory protein ORF6 of SARS-CoV is localized within the Golgi apparatus and ER. This protein is among the minor components of SARS-CoV responsible for virulence, but not necessary for replication of the virus. The domain organization of ORF6 contains N-terminal amphipathic motif from 2 to 37 amino acids (aa) which has a role in inducing the rearrangement of intracellular membrane as well as in double membrane vesicles formation [86–88]. The ORF6 protein has been reported to activate the synthesis of DNA and inhibits the co-transfected plasmid expression. Previous examination of the ORF6 has also revealed its crucial function in viral pathogenesis [89].

Moreover, ORF6 localizes along with nsp3 (Papain like proteinase) and acts as a marker for the replication complex [88,89]. Interestingly, ORF6 makes physical interaction with nsp8 as affirmed using co-immunoprecipitation and yeast two hybrid analysis [90]. Additionally, the interaction of ORF6 with 9b accessory protein has also been reported [91], however, how these findings are related to replication of virus and pathogenicity induced by SARS-CoV is yet to be understood.

ORF6 isolated from SARS-CoV-2 inhibits the signaling pathway of type I IFN, (specifically IFN- β), which is a crucial component of host defense system. This ORF is also known to target the Interferon-stimulated

response element (ISRE). The inhibition of interferon activity will ultimately promote replication of the virus [92].

5.2. ORF7

5.2.1. ORF7a

Sequence analysis of ORF7a isolated from SARS-CoV has revealed that it codes for 122 amino acids long, type I transmembrane domain with an N terminal signal peptide and a C terminal domain. Further, it also possesses a distinct luminal domain. This domain folds into a seven-stranded sandwich of β strands, which shows a similar arrangement as seen in the members of Ig superfamily. Previous investigations carried on its structure have revealed that it has also been isolated from other coronaviruses [93,94]. There is little clarity about whether ORF7a localizes with the ERGIC or just the Golgi complex inside the cell. Though its specific function is not reported, it is known to be incorporated inside the mature virion particles and may have some role to play in viral assembly [94]. Its overexpression increases apoptosis and activates nuclear factor kappa B (NF- κ B) as well as c-Jun N terminal kinase (JNK) mediated signaling pathways. Both of these promote the production of pro-inflammatory cytokines like IL-8 [95], thus highlighting the role of ORF7a in virus-host interactions. The structure for SARS-CoV-2 ORF7a has been recently determined as shown in Fig. 7 but detailed information is yet to be published.

In Dali server [96], the SARS-CoV ORF7a encoded X4 protein (PDB code 1YO4) showed maximum similarity of 89% to SARS-CoV-2 ORF7a (PDB code 6W37) with r.m.s.d value of 1.0 Å and z score 10.3, followed by SARS-CoV ORF 7a (PDB code 1XAK) with similarity of 88%, r.m.s.d of 0.4 Å and z score 14.1. The crystal structure of its N terminal ectodomain from SARS-CoV has been determined [93]. The transmembrane protein contains a 15 residue long N terminal signal peptide, a luminal domain, a transmembrane segment and a cytoplasmic tail. It has been studied that the ectodomain contains four residues of cysteine. The luminal domain is composed of seven β -stranded sandwich consisting of two β sheets, organized in an immunoglobulin type domain. The identity to members of Ig superfamily as found by Dali search was in the range of 2–16%. This fold has been observed to be significant for many proteins. Due to the disulphide bond between cysteine residues, both the β sheets lead to the formation of a hydrophobic pocket near the middle region of the β 1 strand. The cytoplasmic tail contains lysine, arginine and lysine at positions 103, 104 and 105, respectively. The lysine at positions 103 and 105 are needed for exit of this ORF from ER. The structural details do not clearly define the functional significance of ORF7a for this virus [93].

5.2.2. ORF7b

The ORF7b codes for a 44 amino acid long protein which is hydrophobic in nature. The ORF7b start codon overlaps with the ORF7a stop codon [94]. The Golgi apparatus has been said to be its site of localization. It codes for a transmembrane protein with an N terminal as well as a cytoplasmic C-terminal. The transmembrane domain of this protein is crucial for localization in the Golgi apparatus. It has been studied that this ORF is not required for viral replication *in vivo* as well as *in vitro* [97].

6. ORF8

Among all the proteins from SARS-CoV-2, the one encoded by ORF8 show only <20% sequence similarity to the SARS-CoV [98]. In a recent study, it has been suggested that overexpression of ORF8 from SARS-CoV-2 contributes to a significant reduction in MHC-I molecules, which serve as hallmarks of the immune system. The ORF8 was shown to localize along with MHC-I and through immunoprecipitation, it was further confirmed that ORF8 binds to exogenous or endogenous MHC-I and promotes degradation via lysosomes. This decline in MHC-I levels seems to be specific to this novel coronavirus [99]. Furthermore, this ORF has not been found to possess any functional domain and based on predictions of the secondary structure it is strikingly suggested

to give rise to a protein with an alpha helix followed by a beta sheet with six strands [17]. Although this ORF doesn't hold any importance for replication of the pathogen's genome [100], the exact reason for frequent mutations in this region is a cause of speculation. Just like protein from ORF 6, its overexpression promotes synthesis of DNA [98,101].

7. Nucleocapsid protein

The Nucleocapsid (N) protein is bound to the viral nucleic acid, the most crucial component of the SARS-CoV located in the Endoplasmic Reticulum and the Golgi complex inside the cell. It thus attaches to the RNA. The N protein is important for replication of the virus and host response to viral infections. It undergoes various post translational modifications that majorly include phosphorylation. This is responsible for structural changes that ultimately increase the protein's affinity for genomic RNA [102]. Further, the SARS-CoV-2 N protein consists of three domains, an N-terminal domain (NTD), a central S/R rich disordered linker region and a C-terminal domain (CTD) (Fig. 8) [103].

These three domains play crucial roles as outlined by previous research [22,102–104]. The N-terminal domain serves a role in RNA-binding whereas the C-terminal domain is responsible for oligomerization. Accumulated knowledge from the previously solved crystal structure of N-terminal domain of N protein of various CoVs such as SARS-CoV, HCoV-OC43 and infectious bronchitis virus suggested some essential functions of this domain. This domain associates with 3'-end of the RNA genome of the virus via electrostatic interactions. Also, R76 and Y94 are involved in binding with RNA by the NTD of SARS-CoV N protein as well as causing viral infection. Substitution of these amino acids by an Alanine residue significantly reduced RNA binding. The middle linker region rich in Ser/Arg (shown in Fig. 8) is necessary for primary phosphorylation [103,105].

The N protein of the virus serves multiple functions during its life cycle that include RNA binding, chaperone activities, direct metabolism of the infected cell and CoV RNA transcription as well as replication. The

initial role of this protein is binding with the RNA genome of the virus and forming the helical ribonucleoprotein complex (RNP complex) during packing of the genome [106]. *In vitro* and *in vivo* analysis suggests that SARS-CoV N protein has affinity to bind with the leader RNA; this ensures maintenance of appropriate RNA conformation which, in turn, facilitates viral RNA genome transcription and replication [103,107]. The expression of N protein is observed to be high during infection. It reacts with sera obtained from patients suffering with SARS-CoV. Distinctively, antibodies to N protein are also observed to form earlier than that for S protein [108]. It also possesses the ability to induce a protective immune response in the human body against SARS-CoV and SARS-CoV-2 [103,109].

Recently, LC-MS studies of the N-protein revealed the presence of N and O-glycans. The N-glycans of high mannose, hybrid and complex type are reported in this study. N-protein NTD has one N-glycosylation site located at N47 and three O-glycosylation sites that are T148, T165 and T166 which regulate the RNA-binding activity. The middle SR-rich linker region of N-protein has N192 and N196 glycosylation sites. The CTD of N protein has two O-glycosylation sites (T245 and T247) and N-glycosylation site at N269 (Fig. 8). The roles of various domains in N-protein emphasize the importance of future investigation on the role played by glycosylation when viral pathogenesis occur as well as its use in vaccine or drug design [110].

Recently, the three-dimensional structure of the SARS-CoV-2 N-terminal domain of the nucleocapsid protein was elucidated using X-ray crystallography [103]. The structural analysis highlighted that one asymmetric unit of this domain consists of four N-NTD monomers packed in orthorhombic crystal form. The arrangement of the orthorhombic crystal form having right handed loops- five antiparallel β -strands- loops of one monomer out of four monomers is depicted in Fig. 8. Here, the β -sheet core of NTD domain had a single 3_{10} helix just before the β_2 strand. In this structure, the β -hairpin extended between β_2 and β_5 has been suggested to be essential for RNA binding. The NTD with its aromatic and basic residues folds in right-hand shape

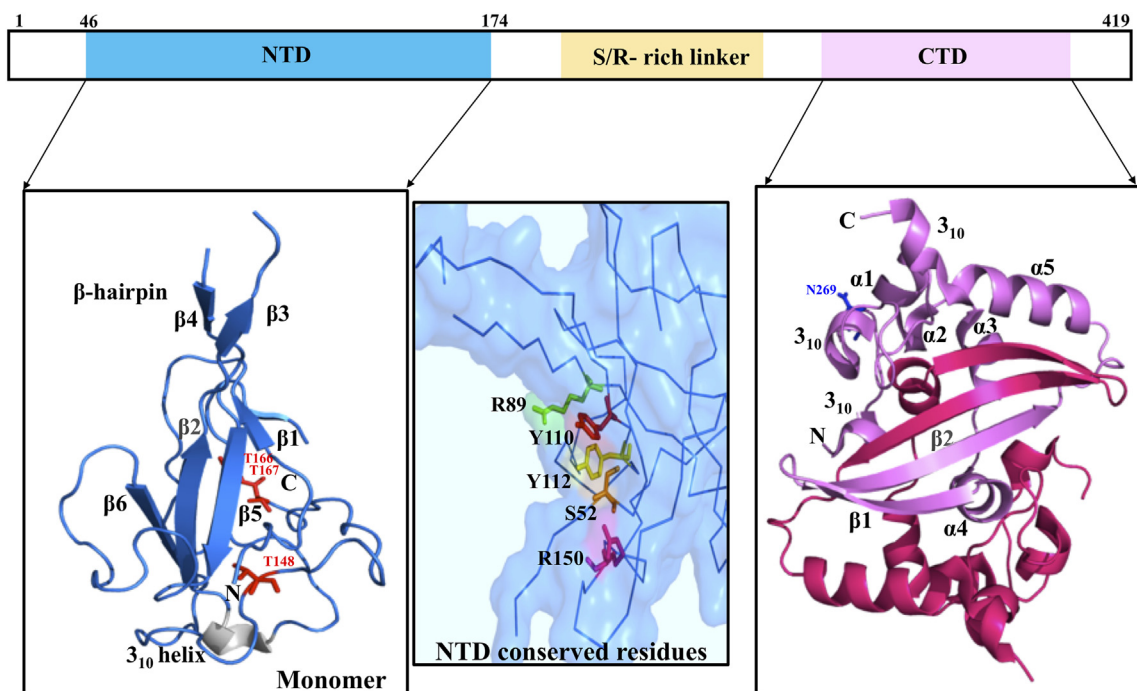


Fig. 8. The domain organization and structure of SARS-CoV-2 N protein. Top panel represents three domains of N protein: NTD (N-terminal RNA binding domain), S/R-rich linker and a CTD (C-terminal domain), shown in magenta, light yellow and purple color, respectively [103]. The crystal structure of the N-terminal RNA binding domain of the N protein of SARS-CoV-2 (PDB code 6M3M) [103], its conserved residues S52, R89, Y110, Y112, and R150 of NTD for ribonucleotide binding site and the cartoon representation of three-dimensional crystal structure of the C-terminal domain of the N protein (PDB code 6WZO) [104] are shown in boxes. In NTD, three O-glycosylation sites, T148, T165 and T166, are represented using red sticks whereas in CTD and N-glycosylation site N269 is shown in blue stick.

resembling a hand having a basic palm and an acidic wrist; here β hair-pins are like fingers. In addition, the interfacial interaction in the NTD of the N protein mostly occurs via interaction of its β -hairpin fingers with the palm region [103,105]. Additionally, this SARS-CoV-2 N-NTD (N-terminal domain of N protein) structure superimposition with N-NTD of SARS-CoV, MERS-CoV, and HCoV-OC43 showed considerable structural similarities. However, several differences in the movement of β -hairpin region backward to the opposite side and forward to the nucleotide binding site in SARS-CoV and MERS-CoV, respectively were seen whereas β -hairpin region is extended less in HCoV-OC43. These differences change the surface charge distribution of protein which brings about flexibility in RNA binding cleft for the RNA based genome. Moreover, multiple sequence alignment of these three CoVs with SARS-CoV-2 N-NTD had revealed conserved residues such as S51, R89, Y110, Y112 and R149 as shown in Fig. 8 [103].

The CTD of SARS-CoV-2 N protein is closely related to homologous domains of various coronaviruses such as SARS-CoV, MERS-CoV, HCoV-NL63 etc. that affirmed the structure of this domain to be a homodimer [104]. The overall structure has ten α helices and four β sheet strands surrounded by 3_{10} helices as shown in Fig. 8. The structure closely resembles the SARS-CoV CTD domain which has compact intertwined dimer structure possessing four β sheets that are centrally located and antiparallel, forming the dimer interface. At the interface, each protomer has two β strands and short α helix extended towards the protomer opposite to it facilitating the packing against the hydrophobic core. The sequence identity for CTD of SARS-CoV-2 with SARS-CoV has been observed to be 96% except for notable changes in five residues; Q268, D291, H335, Q346 and N350 of SARS-CoV to A267, E290, T334, Q345 and N348 in case of SARS-CoV-2, respectively. In order to elucidate the structural similarities, one group overlaid the structure of the SARS-CoV-2 N-protein CTD domain determined by them with CTD structures available with PDB code 7C22 [111] (unpublished manuscript). The outcome revealed high structural similarity with r.m.s.d value of C_{α} in the range 0.15–0.31 Å. Additionally, a positively charged surface area present in the CTD domain has been suggested to facilitate RNA binding. However, the binding affinity of RNA is assumed to vary according to variation in the viral isoform from one another [104].

7.1. ORF9b

This ORF is located in the Nucleocapsid gene and possesses 98 amino acids. Structural studies carried out for SARS-CoV ORF9b provide information that it specifically binds to lipid molecules by means of a hydrophobic cavity located in the center of the structure. Its expression carried out in mammalian cell culture suggests it is membrane bound and associated with vesicular structures inside the cell [112].

The GFP tagged ORF9b fusion protein has been reported to localize in the mitochondria. The same fusion complex expressed in different cell lines has been associated with considerable reduction in Dynamin related protein (DRP) 1 levels which ultimately causes elongation of mitochondria. This can be implicated to the fact that mitochondria protect against SARS-CoV, linked to interferon signaling, which is limited by DRP1 as a defense mechanism [113]. Furthermore, its presence has been reported in VLPs as well as in virions of SARS-CoV. Since both E as well as M proteins are sufficient for formation of VLPs, both these proteins promote the incorporation of ORF9b into these particles [114].

The manuscript elucidating the atomic details of the crystal structure (PDB code 6Z4U) for ORF9b (Fig. 7) and other related aspects is yet to be published. However, on Dali server [96], the SARS-CoV ORF9b (PDB code 2CME) exhibits 81% identity to the SARS-CoV-2 ORF9b with r.m.s.d of 2.4 Å and z score of 9.9. The crystal structure is composed of a symmetrical dimer consisting of two folds made from two adjacent β sheets [112]. Every sheet is made up of β strands which, in turn, are formed by both the monomers resulting in the formation of an intertwined

structure. The dimer thus formed is composed of anti-parallel β -sheets in which the monomers bind each other firmly.

The surface properties analysis of ORF9b showed that one region of the molecule is positively charged while the other is negatively charged. It has a dimeric, tent-like β structure and a hydrophobic cavity in the centre which is surrounded by hydrophobic side chains. This cavity binds lipid molecules and has a unique fold. It is responsible for membrane attachment during viral assembly [112].

7.2. ORF10

SARS-CoV-2 ORF10 is present at 3'-end of the genome. It has been anticipated to possess 38 amino acids and is predicted to code for a transmembrane domain in SARS-CoV-2 [115]. This observation further suggests that SARS-CoV-2 ORF10 might code for a functional protein. However, this hypothesis needs experimental and genomic analysis for clarity. This ORF is particularly peculiar because of its dissimilarity to other proteins in the NCBI database [116]. Its analysis in SARS-CoV lineage suggests that a stop codon disrupts its reading frame [115].

8. Potential targets for vaccine design

Novel coronavirus being highly contagious with a high mortality rate calls for the need to discover vaccines to curb its worldwide spread. To achieve this, researchers across the globe are putting efforts to discover inhibitors against this viral pathogenesis in humans. SARS-CoV-2 has various proteins as mentioned before and any natural product, antiviral compound, nanoparticle derived drugs, peptides etc. which can target these viral proteins can serve as promising drug candidates. This section aims to summarize all the domains of SARS-CoV-2 which are potential targets for vaccine design as suggested by extensive research carried out till date. Presently, the efforts to synthesize antiviral drugs against lethal COVID-19 infection mostly target the spike protein region, the 3C-like protease and papain like protease. Apart from these proteins, RNA polymerase, E protein and helicase proteins of SARS-CoV-2 are also considered to be potential targets for antiviral vaccine or drug development [117,118]. Besides, repurposing of drugs against SARS-CoV-2 is needed but the interaction of such drugs with the host, specifically, drugs used against other human coronaviruses (HCoV) requires in-depth research. Several such immunosuppressive drugs like rapamycin (rapamycin) and agents preventing inflammation like mesalazine etc. are being proposed as treatment options for SARS-CoV-2 [119].

Past experiences with influenza vaccine stress upon the need for development of (a) whole inactivated virus vaccines and (b) live attenuated viral vaccines for SARS-CoV-2 [120]. Agents like formaldehyde, UV irradiation etc. have been successfully used for inactivation of the pathogen in vaccine preparations everywhere. Since all of the structural proteins S, E, M and N stimulate humoral as well as cell mediated immune responses involving production of antibodies as well as CD4⁺ and CD8⁺ T cells, vaccine development may be facilitated [121,122]. Specifically, targeting the S protein as well as the RBD of the virus may lead to generation of neutralizing antibodies, thus providing immunity as studied for SARS-CoV [123]. Moreover, these structural proteins could be used as ideal targets for anti-viral compounds [123].

DNA, RNA and protein based vaccines are being tested against COVID-19 all across the globe. In a recent study, an S protein based mRNA vaccine (mRNA-1273) encoding the spike pre-fusion complex has shown promising results in Phase 1 clinical trials. It is a lipid nanoparticle based vaccine; immune responses against the pathogen were detected in all the patients and no ill effects were noted. As mentioned above, keeping in view the importance of spike protein in leading the viral entry inside the host, many of the vaccines are being developed to target this protein [124]. Most of the nucleic acid based vaccines as well as many recombinant protein vaccines are aiming to target this protein and are in clinical trials [125].

As is widely known, the virus uses its S protein for entry into the host cell which then attaches to ACE2. Some of the vaccines in clinical trials include INO-4800 which is a DNA vaccine and ChAdOx1 nCoV-19, which showed good results in Phase 1 and 2 clinical trials. It is a non-replicating adenovirus based vaccine [125].

Based on these findings, recombinant soluble ACE2 isolated from humans was used to inhibit SARS-CoV-2 in Vero cell lines. This particular ACE2 has earlier been in phase 1 as well as phase 2 testing and is now being thought to be used as a treatment for the current outbreak [126]. The Transmembrane protease serine 2 (TMPRSS2) is known to prepare the S protein and interact with the ACE2. Camostat mesylate is a protease inhibitor and clinically approved to impede the TMPRSS2 activity. As shown recently, camostat mesylate blocked the SARS-CoV-2 entry into cells and hence can be developed as a vaccine candidate or anti-viral compound [127].

The E protein has been implicated in the virulence of this pathogen and its deletion severely cripples the CoV pathogenicity. By appropriately mutating this protein, the feasibility of developing a live attenuated vaccine for SARS-CoV and MERS-CoV has been explored [128]. A similar approach can be used for SARS-CoV-2. In a study by Pang *et al.*, it was found that the antiserum raised against M protein of SARS-CoV had a high neutralization titre. This suggests its possible development as a vaccine for SARS-CoV-2, since it has been found to be evolutionarily conserved [129].

Additionally, finding inhibitors for protease activity of nsp3 can prove to be helpful in vaccine development. For instance, a natural product class called tanshinones has been recently reported to inhibit the protease activity of nsp3 [18,29].

Nsp5 also referred to as M^{pro} is responsible to cleave the viral polyproteins at 11 sites releasing various intermediate as well as mature non-structural proteins, which have crucial roles in virus replication. Therefore, it can also be used as a potent target for antiviral drug development [130]. Previously, inhibitors have been developed to target viral proteases, some of which include peptide aldehyde and its derivatives. The inhibitors like GC373 and GC376 have been used to treat Feline Coronavirus (FCoV) as well as mink and ferret coronaviruses. GC376 covalently binds with cysteine residues of the MERS-CoV M^{pro}. A latest study reports that both bind and reversibly inhibit M^{pro} of SARS-CoV-2 and can effectively be used for antiviral therapy [131]. Moreover, Ebselen (in clinical trial), N3 (pre-clinical but animal testing not done), 13b (pre-clinical; animal testing yet to be done) and 11b (pre-clinical) are examples of already reported inhibitors of SARS-CoV-2 M^{pro} [130,132,133].

RNA-dependent RNA polymerase (Nsp12) [41] of SARS-CoV-2 possesses a groove type structural region which has been suggested to be an active center for RNA synthesis as well as replication of virus. Evidences showed that SARS-CoV-2 RdRp has a sequence similarity of 96% with SARS-CoV, which proposes the inhibitor against the SARS-CoV RdRp could also show similar inhibitory activity against RdRp of SARS-CoV-2 [19,134]. Notably, Nsp12 viral polymerase appears to be a good target for remdesivir [41]. This drug has also been proven as a lead inhibitor according to *in vitro* studies and animal study in rhesus monkeys [135], even though its use against SARS-CoV-2 is still skeptical.

Humoral as well as cellular immune response are heightened during infection in body. The role of T cells is especially coming to the fore because of the generation of a cytokine storm triggered by SARS-CoV-2 in human host. Several studies are already underway to pin-point how these cells can be used for development of an effective vaccine. Both CD4 as well as CD8 cells isolated from convalescent patients have been reported to respond to the structural as well as non-structural region of the SARS-CoV-2. Further, memory T cells have been detected for SARS-CoV even after several years of infection that also exhibited cross-reactivity to SARS-CoV-2 [136], which gives hope for the on-going vaccine

development efforts. Furthermore, trials on artificial Antigen Presenting Cells (aAPCs) with and without cytotoxic (T_C) cells are in progress [137].

Aforementioned many ORFs are involved in the suppression of host defense response, for instance, ORF3b is involved in suppressing the vital type I IFN response. Similarly, ORF8 is associated with reduced MHC-I levels. Such immune molecules are crucial to the host in fight against this pathogen. If one could target these ORFs and inhibit its activity, it would prove to be helpful for disease management.

9. Future perspectives

Designing an effective vaccine to curb the deadly SARS-CoV-2 is the need of the hour. Many vaccine manufacturers are racing against time to deliver a cure for this disease which has claimed many lives across the globe. A plethora of *in vivo* and *in vitro* studies are being carried out simultaneously using various tools and techniques by researchers to unleash the atomic level information, structural as well as biological aspects including epidemiology and prognosis etc. of SARS-CoV-2. The knowledge of correct architecture, identification of biochemical, genetic information as well as structural properties of the SARS-CoV-2 would be helpful in understanding various virus-host interactions. Besides, it will reveal mechanistic details of interactions of different proteins of SARS-CoV-2. Therefore, future studies should focus on determining the structures of those domains which are still not known. This will further pave way for designing potential drugs to effectively treat the disease and reduce mortality.

There is a need to develop proper strategies for treating asymptomatic patients too as otherwise most such strategies are limited to symptomatic patients. Repurposing of drugs used against a variety of other diseases is being proposed as a good strategy to contain the spread of the disease. For instance, the effectiveness of Bacillus Calmette–Guérin (BCG) vaccine used for tuberculosis (TB) treatment is being tested. The study of SARS-CoV-2 at the atomic levels is bound to improve our insights about various virus-host interactions that are involved in disease progression, and to limit harmful consequences and raise awareness about future risks associated with such fatal disease outbreaks. If the immune cells can be boosted in certain ways to respond in an effective manner against the disease caused by this RNA pathogen, it would prove to be a great boon for the suffering masses. Developing a vaccine against this kind of pathogen would be a challenge due to the high mutability of RNA viruses, nevertheless keeping a track of SARS-CoV-2 through extensive research would definitely go a long way in eradicating this virus. Further, diagnostic tests with enhanced sensitivity for detection are the need of the hour and would be key for an effective treatment.

CRedit authorship contribution statement

M.K. and A.S. wrote the initial draft. S.K. and G.S. provided feedback. RPB reviewed, provided feedback and edited.

Declaration of competing interest

No conflict of interest declared.

Acknowledgements

RPB acknowledges the Indian agencies, DST (ECRA/2017/000124) and DBT (BT/PR/27444/BRB/10/1645/2018) for the financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2021.02.212>.

References

- [1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (8) (2020) 727–733.
- [2] B. Chen, E.K. Tian, B. He, L. Tian, R. Han, S. Wang, Q. Xiang, S. Zhang, T. El Arnaout, W. Cheng, Overview of lethal human coronaviruses, *Signal transduction and targeted therapy* 5 (1) (2020), 89.
- [3] G. Lu, Q. Wang, G.F. Gao, Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond, *Trends Microbiol.* 23 (8) (2015) 468–478.
- [4] J.D. Cherry, The chronology of the 2002–2003 SARS mini pandemic, *Paediatr. Respir. Rev.* 5 (4) (2004) 262–269.
- [5] A. Mubarak, W. Alturaiqi, M.G. Hemida, Middle East respiratory syndrome coronavirus (MERS-CoV): infection, immunological response, and vaccine development, *J Immunol Res* (2019) 6491738.
- [6] J.A. Jaimes, N.M. André, J.S. Chappie, J.K. Millet, G.R. Whittaker, Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop, *J. Mol. Biol.* 432 (10) (2020) 3309–3325.
- [7] M.C. Wong, S.J. Javornik Cregeen, N.J. Ajami, J.F. Petrosino, Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019, *bioRxiv* (2020) <https://doi.org/10.1101/2020.02.07.939207>.
- [8] H.A. Rothan, S.N. Byrareddy, The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, *J. Autoimmun.* 109 (2020), 102433.
- [9] E. Housseinzadeh, J. Safoura, M. Farzadkia, F. Mohammadi, H. Hossini, M. Taghavi, An updated min-review on environmental route of the SARS-CoV-2 transmission, *Ecotoxicol. Environ. Saf.* 202 (2020), 111015.
- [10] N.H.L. Leung, D.K.W. Chu, E.Y.C. Shiu, K.H. Chan, J.J. McDevitt, B.J.P. Hau, H.L. Yen, Y. Li, D.K.M. Ip, J.S.M. Peiris, W. Li, M. Seto, G.M. Leung, D.K. Milton, B.J. Cowling, Author correction: respiratory virus shedding in exhaled breath and efficacy of face masks, *Nat. Med.* 26 (6) (2020) 981.
- [11] M.Z. Tay, C.M. Poh, L. Renia, P.A. MacAry, L.F.P. Ng, The trinity of COVID-19: immunity, inflammation and intervention, *Nat Rev Immunol* 20 (6) (2020) 363–374.
- [12] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (10223) (2020) 497–506.
- [13] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (7798) (2020) 270–273.
- [14] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E.C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* 579 (7798) (2020) 265–269.
- [15] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, *Viruses* 12 (3) (2020).
- [16] M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses, *Nat. Microbiol.* 5 (4) (2020) 562–569.
- [17] J.F. Chan, K.H. Kok, Z. Zhu, H. Chu, K.K. To, S. Yuan, K.Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, *Emerg Microbes Infect* 9 (1) (2020) 221–236.
- [18] F.K. Yoshimoto, The proteins of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2 or n-CoV19), the cause of COVID-19, *Protein J.* 39 (3) (2020) 198–216.
- [19] H. Li, L. Yang, F.F. Liu, X.N. Ma, P.L. He, W. Tang, X.K. Tong, J.P. Zuo, Overview of therapeutic drug research for COVID-19 in China, *Acta Pharmacol. Sin.* (2020) 1–8.
- [20] A.H. de Wilde, E.J. Snijder, M. Kikkert, M.J. van Hemert, Host factors in coronavirus replication, *Curr. Top. Microbiol. Immunol.* 419 (2018) 1–42.
- [21] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, *The Journal of general virology* 83 (Pt 3) (2002) 595–599.
- [22] Y. Cong, M. Ulasli, H. Schepers, M. Mauthe, P. V'kovski, F. Kriegenburg, V. Thiel, C.A.M. de Haan, F. Reggiori, Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronavirus life cycle, *J. Virol.* 94 (4) (2020).
- [23] S. Yuan, L. Peng, J.J. Park, Y. Hu, S.C. Devarkar, M.B. Dong, Q. Shen, S. Wu, S. Chen, I.B. Lomakin, Y. Xiong, Nonstructural protein 1 of SARS-CoV-2 is a potent pathogenicity factor redirecting host protein synthesis machinery toward viral RNA, *Mol. Cell* 80 (6) (2020) 1055–1066 e6.
- [24] C. Huang, K.G. Lokugamage, J.M. Rozovics, K. Narayanan, B.L. Semler, S. Makino, SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage, *PLoS Pathog.* 7 (12) (2011), e1002433.
- [25] C.T. Cornillez-Ty, L. Liao, J.R. Yates 3rd, P. Kuhn, M.J. Buchmeier, Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling, *J. Virol.* 83 (19) (2009) 10314–10318.
- [26] R.L. Graham, A.C. Sims, R.S. Baric, M.R. Denison, The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication, *Adv. Exp. Med. Biol.* 581 (2006) 67–72.
- [27] Y. Sakai, K. Kawachi, Y. Terada, H. Omori, Y. Matsuura, W. Kamitani, Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication, *Virology* 510 (2017) 165–174.
- [28] J. Lei, Y. Kusov, R. Hilgenfeld, Nsp3 of coronaviruses: structures and functions of a large multi-domain protein, *Antivir. Res.* 149 (2018) 58–74.
- [29] Y.M. Báez-Santos, S.E. St John, A.D. Mesecar, The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds, *Antivir. Res.* 115 (2015) 21–38.
- [30] D.N. Frick, R.S. Virdi, N. Vuksanovic, N. Dahal, N.R. Silvaggi, Molecular basis for ADP-ribose binding to the Mac1 domain of SARS-CoV-2 nsp3, *Biochemistry* 59 (28) (2020) 2608–2615.
- [31] W. Rut, Z. Lv, M. Zmudzinski, S. Patchett, D. Nayak, S.J. Snipas, F. El Oualid, T.T. Huang, M. Bekes, M. Drag, S.K. Olsen, Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: a framework for anti-COVID-19 drug design, *Sci. Adv.* 6 (42) (2020).
- [32] S. Tomar, M.L. Johnston, S.E. St John, H.L. Osswald, P.R. Nyalapatla, L.N. Paul, A.K. Ghosh, M.R. Denison, A.D. Mesecar, Ligand-induced dimerization of Middle East respiratory syndrome (MERS) coronavirus nsp5 protease (3CLpro): IMPLICATIONS FOR nsp5 REGULATION AND THE DEVELOPMENT OF ANTIVIRALS, *J. Biol. Chem.* 290 (32) (2015) 19403–19422.
- [33] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors, *Science (New York, N.Y.)* 368 (6489) (2020) 409–412.
- [34] K.S. Yang, X.R. Ma, Y. Ma, Y.R. Alugubelli, D.A. Scott, E.C. Vatansever, A.K. Drelich, B. Sankaran, Z.Z. Geng, L.R. Blankenship, A quick route to multiple highly potent SARS-CoV-2 main protease inhibitors, *ChemMedChem* (2020).
- [35] M.M. Angelini, M. Akhlaghpour, B.W. Neuman, M.J. Buchmeier, Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles, *mBio* 4 (4) (2013).
- [36] I. Imbert, E.J. Snijder, M. Dimitrova, J.C. Guillemot, P. Lécine, B. Canard, The SARS-coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein, *Virus Res.* 133 (2) (2008) 136–148.
- [37] A.J. te Velthuis, S.H. van den Worm, E.J. Snijder, The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension, *Nucleic Acids Res.* 40 (4) (2012) 1737–1747.
- [38] L. Subissi, C.C. Posthuma, A. Collet, J.C. Zevenhoven-Dobbe, A.E. Gorbalenya, E. Decroly, E.J. Snijder, B. Canard, I. Imbert, One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities, *Proc. Natl. Acad. Sci. U. S. A.* 111 (37) (2014) E3900–E3909.
- [39] D.R. Littler, B.S. Gully, R.N. Colson, J. Rossjohn, Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9, *iScience* 23 (7) (2020) 101258.
- [40] A. Rogstam, M. Nyblom, S. Christensen, C. Sele, V.O. Talibov, T. Lindvall, A.A. Rasmussen, I. André, Z. Fisher, W. Knecht, F. Kozielski, Crystal structure of non-structural protein 10 from severe acute respiratory syndrome Coronavirus-2, *Int. J. Mol. Sci.* 21 (19) (2020).
- [41] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus, *Science (New York, N.Y.)* 368 (6492) (2020) 779–782.
- [42] P. Krafčikova, J. Silhan, R. Nencka, E. Bura, Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin, *Nat. Commun.* 11 (1) (2020) 3717.
- [43] Q. Peng, R. Peng, B. Yuan, J. Zhao, M. Wang, X. Wang, Q. Wang, Y. Sun, Z. Fan, J. Qi, G.F. Gao, Y. Shi, Structural and biochemical characterization of the nsp12-nsp7-nsp8 Core polymerase complex from SARS-CoV-2, *Cell Rep.* 31 (11) (2020) 107774.
- [44] Y. Wang, Y. Sun, A. Wu, S. Xu, R. Pan, C. Zeng, X. Jin, X. Ge, Z. Shi, T. Ahola, Y. Chen, D. Guo, Coronavirus nsp10/nsp16 methyltransferase can be targeted by nsp10-derived peptide in vitro and in vivo To reduce replication and pathogenesis, *J. Virol.* 89 (16) (2015) 8416–8427.
- [45] Y. Ma, L. Wu, N. Shaw, Y. Gao, J. Wang, Y. Sun, Z. Lou, L. Yan, R. Zhang, Z. Rao, Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex, *Proc. Natl. Acad. Sci. U. S. A.* 112 (30) (2015) 9436–9441.
- [46] Y. Kim, R. Jedrzejczak, N.I. Maltseva, M. Wilamowski, M. Endres, A. Godzik, K. Michalska, A. Joachimiak, Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2, *Protein science: a publication of the Protein Society* 29 (7) (2020) 1596–1605.
- [47] S. Ricagno, M.P. Egloff, R. Ulferts, B. Coutard, D. Nurizzo, V. Campanacci, C. Cambillau, J. Ziebuhr, B. Canard, Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family, *Proc. Natl. Acad. Sci. U. S. A.* 103 (32) (2006) 11892–11897.
- [48] H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E.C. Holmes, A.C. Hughes, Y. Bi, W. Shi, A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein, *Curr. Biol.* 30 (11) (2020) 2196–2203 (e3).
- [49] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science (New York, N.Y.)* 367 (6483) (2020) 1260–1263.
- [50] S. Xia, M. Liu, C. Wang, W. Xu, Q. Lan, S. Feng, F. Qi, L. Bao, L. Du, S. Liu, C. Qin, F. Sun, Z. Shi, Y. Zhu, S. Jiang, L. Lu, Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike

- protein that harbors a high capacity to mediate membrane fusion, *Cell Res.* 30 (4) (2020) 343–355.
- [51] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Velesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, *Cell* 181 (2) (2020) 281–292 (e6).
- [52] Y. Zhang, W. Zhao, Y. Mao, S. Wang, Y. Zhong, T. Su, M. Gong, X. Lu, J. Cheng, H. Yang, Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins using high-resolution mass spectrometry, *bioRxiv: the preprint server for biology* (2020) (03.28.013276).
- [53] F. Li, W. Li, M. Farzan, S.C. Harrison, Structure of SARS coronavirus spike receptor-binding domain complexed with receptor, *Science (New York, N.Y.)* 309 (5742) (2005) 1864–1868.
- [54] I. Hamming, W. Timens, M.L. Bulthuis, A.T. Lely, G. Navis, H. van Goor, Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis, *J. Pathol.* 203 (2) (2004) 631–637.
- [55] C.M. Chan, H. Tsoi, W.M. Chan, S. Zhai, C.O. Wong, X. Yao, W.Y. Chan, S.K. Tsui, H.Y. Chan, The ion channel activity of the SARS-coronavirus 3a protein is linked to its pro-apoptotic function, *Int. J. Biochem. Cell Biol.* 41 (11) (2009) 2232–2239.
- [56] Y. Ren, T. Shu, D. Wu, J. Mu, C. Wang, M. Huang, Y. Han, X.Y. Zhang, W. Zhou, Y. Qiu, X. Zhou, The ORF3a protein of SARS-CoV-2 induces apoptosis in cells, *Cell Mol Immunol* 17 (2020) 881–883.
- [57] W. Lu, B.J. Zheng, K. Xu, W. Schwarz, L. Du, C.K. Wong, J. Chen, S. Duan, V. Deubel, B. Sun, Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release, *Proc. Natl. Acad. Sci. U. S. A.* 103 (33) (2006) 12540–12545.
- [58] P.T.W. Law, C.H. Wong, T.C.C. Au, C.P. Chuck, S.K. Kong, P.K.S. Chan, K.F. To, A.W.I. Lo, J.Y.W. Chan, Y.K. Suen, H.Y.E. Chan, K.P. Fung, M.M.Y. Waye, J.J.Y. Sung, Y.M.D. Lo, S.K.W. Tsui, The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells, *The Journal of general virology* 86 (Pt 7) (2005) 1921–1930.
- [59] D.M. Kern, B. Sorum, C.M. Hoel, S. Sridharan, J.P. Remis, D.B. Toso, S.G. Brohawn, Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs, *bioRxiv: the preprint server for biology* (2020) <https://doi.org/10.1101/2020.06.17.156554>.
- [60] J.M. Hyser, M.K. Estes, Pathophysiological consequences of calcium-conducting viroporins, *Annu Rev Virol* 2 (1) (2015) 473–496.
- [61] Y. Konno, I. Kimura, K. Uriu, M. Fukushi, T. Irie, Y. Koyanagi, S. Nakagawa, K. Sato, SARS-CoV-2 ORF3b is a Potent Interferon Antagonist whose Activity Is further Increased by a Naturally Occurring Elongation Variant, *bioRxiv: The Preprint Server for Biology*, 2020 (2020.05.11.088179).
- [62] S.R. Schaecher, A. Pekosz, SARS coronavirus accessory gene expression and function, *Molecular Biology of the SARS-Coronavirus* (2009) 153–166.
- [63] X. Yuan, Y. Shan, Z. Zhao, J. Chen, Y. Cong, G0/G1 arrest and apoptosis induced by SARS-CoV 3b protein in transfected cells, *Virol. J.* 2 (2005) 66.
- [64] C.A. Nelson, et al., Structure of the SARS-CoV-2 ORF7a encoded accessory protein, <http://www.rcsb.org/structure/6W372020>.
- [65] S.D. Weeks, et al., X-ray crystallographic structure of Orf9b from SARS-CoV-2, <http://www.rcsb.org/structure/6Z4U02020>.
- [66] J.L. Nieto-Torres, M.L. Dediego, E. Alvarez, J.M. Jiménez-Guardaño, J.A. Regla-Nava, M. Llorente, L. Kremer, S. Shuo, L. Enjuanes, Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein, *Virology* 415 (2) (2011) 69–82.
- [67] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, *Virol. J.* 16 (1) (2019), 69.
- [68] J. Sobocińska, P. Roszczenko-Jasińska, A. Ciesielska, K. Kwiatkowska, Protein palmitoylation and its role in bacterial and viral infections, *Front. Immunol.* 8 (2017) 2003.
- [69] E. Alvarez, M.L. DeDiego, J.L. Nieto-Torres, J.M. Jiménez-Guardaño, L. Marcos-Villar, L. Enjuanes, The envelope protein of severe acute respiratory syndrome coronavirus interacts with the non-structural protein 3 and is ubiquitinated, *Virology* 402 (2) (2010) 281–291.
- [70] Y. Liao, J. Lescar, J.P. Tam, D.X. Liu, Expression of SARS-coronavirus envelope protein in *Escherichia coli* cells alters membrane permeability, *Biochem. Biophys. Res. Commun.* 325 (1) (2004) 374–380.
- [71] J. Torres, J. Wang, K. Parthasarathy, D.X. Liu, The transmembrane oligomers of coronavirus protein E, *Biophys. J.* 88 (2) (2005) 1283–1290.
- [72] M. Bianchi, D. Benvenuto, M. Giovanetti, S. Angeletti, M. Ciccozzi, S. Pascarella, SARS-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? *Biomed. Res. Int.* 2020 (2020), 4389089.
- [73] X. Yu, W. Bi, S.R. Weiss, J.L. Leibowitz, Mouse hepatitis virus gene 5b protein is a new virion envelope protein, *Virology* 202 (2) (1994) 1018–1023.
- [74] D.X. Liu, S.C. Inglis, Association of the infectious bronchitis virus 3c protein with the virion envelope, *Virology* 185 (2) (1991) 911–917.
- [75] M. Godet, R. L'Haridon, J.F. Vautherot, H. Laude, TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions, *Virology* 188 (2) (1992) 666–675.
- [76] J. Netland, M.L. DeDiego, J. Zhao, C. Fett, E. Álvarez, J.L. Nieto-Torres, L. Enjuanes, S. Perlman, Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease, *Virology* 399 (1) (2010) 120–128.
- [77] J.K. Locker, J.K. Rose, M.C. Horzinek, P.J. Rottier, Membrane assembly of the triple-spanning coronavirus M protein. Individual transmembrane domains show preferred orientation, *J. Biol. Chem.* 267 (30) (1992) 21911–21918.
- [78] C.A. de Haan, H. Vennema, P.J. Rottier, Assembly of the coronavirus envelope: homotypic interactions between the M proteins, *J. Virol.* 74 (11) (2000) 4967–4978.
- [79] Y.-C. Hsieh, H.-C. Li, S.-C. Chen, S.-Y. Lo, Interactions between M protein and other structural proteins of severe, acute respiratory syndrome-associated coronavirus, *J. Biomed. Sci.* 15 (6) (2008) 707–717.
- [80] K. Narayanan, A. Maeda, J. Maeda, S. Makino, Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells, *J. Virol.* 74 (17) (2000) 8127–8134.
- [81] C.A. de Haan, M. de Wit, L. Kuo, C. Montalto-Morrison, B.L. Haagmans, S.R. Weiss, P.S. Masters, P.J. Rottier, The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain, *Virology* 312 (2) (2003) 395–406.
- [82] B.W. Neuman, G. Kiss, A.H. Kunding, D. Bhella, M.F. Baksh, S. Connelly, B. Droese, J.P. Klaus, S. Makino, S.G. Sawicki, S.G. Siddell, D.G. Stamou, I.A. Wilson, P. Kuhn, M.J. Buchmeier, A structural analysis of M protein in coronavirus assembly and morphology, *J. Struct. Biol.* 174 (1) (2011) 11–22.
- [83] D.J. Opstelten, M.J. Raamsman, K. Wolfs, M.C. Horzinek, P.J. Rottier, Envelope glycoprotein interactions in coronavirus assembly, *J. Cell Biol.* 131 (2) (1995) 339–349.
- [84] E. Corse, C.E. Machamer, The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction, *Virology* 312 (1) (2003) 25–34.
- [85] K.P. Lim, D.X. Liu, The missing link in coronavirus assembly. Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins, *J. Biol. Chem.* 276 (20) (2001) 17515–17523.
- [86] J. Netland, D. Ferraro, L. Pewe, H. Olivares, T. Gallagher, S. Perlman, Enhancement of murine coronavirus replication by severe acute respiratory syndrome coronavirus protein 6 requires the N-terminal hydrophobic region but not C-terminal sorting motifs, *J. Virol.* 81 (20) (2007) 11520–11525.
- [87] L. Pewe, H. Zhou, J. Netland, C. Tangudu, H. Olivares, L. Shi, D. Look, T. Gallagher, S. Perlman, A severe acute respiratory syndrome-associated coronavirus-specific protein enhances virulence of an attenuated murine coronavirus, *J. Virol.* 79 (17) (2005) 11335–11342.
- [88] H. Zhou, D. Ferraro, J. Zhao, S. Hussain, J. Shao, J. Trujillo, J. Netland, T. Gallagher, S. Perlman, The N-terminal region of severe acute respiratory syndrome coronavirus protein 6 induces membrane rearrangement and enhances virus replication, *J. Virol.* 84 (7) (2010) 3542–3551.
- [89] D.X. Liu, T.S. Fung, K.K. Chong, A. Shukla, R. Hilgenfeld, Accessory proteins of SARS-CoV and other coronaviruses, *Antivir. Res.* 109 (2014) 97–109.
- [90] P. Kumar, V. Gunalan, B. Liu, V.T. Chow, J. Druce, C. Birch, M. Catton, B.C. Fielding, Y.J. Tan, S.K. Lal, The nonstructural protein 8 (nsp8) of the SARS coronavirus interacts with its ORF6 accessory protein, *Virology* 366 (2) (2007) 293–303.
- [91] E. Calvo, M.L. DeDiego, P. García, J.A. López, P. Pérez-Breña, A. Falcón, Severe acute respiratory syndrome coronavirus accessory proteins 6 and 9b interact in vivo, *Virus Res.* 169 (1) (2012) 282–288.
- [92] J.Y. Li, C.H. Liao, Q. Wang, Y.J. Tan, R. Luo, Y. Qiu, X.Y. Ge, The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway, *Virus Res.* 286 (2020), 198074.
- [93] C.A. Nelson, A. Pekosz, C.A. Lee, M.S. Diamond, D.H. Fremont, Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein, *Structure* 13 (1) (2005) 75–85.
- [94] A. Pekosz, S.R. Schaecher, M.S. Diamond, D.H. Fremont, A.C. Sims, R.S. Baric, Structure, expression, and intracellular localization of the SARS-CoV accessory proteins 7a and 7b, *Adv. Exp. Med. Biol.* 581 (2006) 115–120.
- [95] N. Kanzawa, K. Nishigaki, T. Hayashi, Y. Ishii, S. Furukawa, A. Niino, F. Yasui, M. Kohara, K. Morita, K. Matsushima, M.Q. Le, T. Masuda, M. Kannagi, Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF- κ B activation, *FEBS Lett.* 580 (30) (2006) 6807–6812.
- [96] L. Holm, DALI and the persistence of protein shape, *Protein science: a publication of the Protein Society* 29 (1) (2020) 128–140.
- [97] S.R. Schaecher, M.S. Diamond, A. Pekosz, The transmembrane domain of the severe acute respiratory syndrome coronavirus ORF7b protein is necessary and sufficient for its retention in the Golgi complex, *J. Virol.* 82 (19) (2008) 9477–9491.
- [98] T.G. Flower, C.Z. Buffalo, R.M. Hooy, M. Allaire, X. Ren, J.H. Hurley, Structure of SARS-CoV-2 ORF8, a Rapidly Evolving Coronavirus Protein Implicated in Immune Evasion, *bioRxiv: The Preprint Server for Biology*, 2020.
- [99] Y. Zhang, J. Zhang, Y. Chen, B. Luo, Y. Yuan, F. Huang, T. Yang, F. Yu, J. Liu, B. Liu, Z. Song, J. Chen, T. Pan, X. Zhang, Y. Li, R. Li, W. Huang, F. Xiao, H. Zhang, The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I, *bioRxiv: The Preprint Server for Biology*, 2020 (2020.05.24.111823).
- [100] B. Yount, R.S. Roberts, A.C. Sims, D. Deming, M.B. Frieman, J. Sparks, M.R. Denison, N. Davis, R.S. Baric, Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice, *J. Virol.* 79 (23) (2005) 14909–14922.
- [101] P.Y. Law, Y.M. Liu, H. Geng, K.H. Kwan, M.M. Waye, Y.Y. Ho, Expression and functional characterization of the putative protein 8b of the severe acute respiratory syndrome-associated coronavirus, *FEBS Lett.* 580 (15) (2006) 3643–3648.
- [102] M. Surjit, S.K. Lal, The SARS-CoV nucleocapsid protein: a protein with multifarious activities, infection, genetics and evolution: (journal of molecular epidemiology and evolutionary genetics in infectious diseases 8 (4) (2008) 397–405.
- [103] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, Y. Yan, C. Zhang, H. Shan, S. Chen, Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, *Acta Pharm. Sin. B* 10 (2020) 1228–1238.
- [104] Q. Ye, A.M.V. West, S. Silletti, K.D. Corbett, Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein, *Protein science: a publication of the Protein Society* 29 (2020) 1890–1901.

- [105] Y.W. Tan, S. Fang, H. Fan, J. Lescar, D.X. Liu, Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells, *Nucleic Acids Res.* 34 (17) (2006) 4816–4825.
- [106] R. McBride, M. van Zyl, B.C. Fielding, The coronavirus nucleocapsid is a multifunctional protein, *Viruses* 6 (8) (2014) 2991–3018.
- [107] T.K. Tang, M.P. Wu, S.T. Chen, M.H. Hou, M.H. Hong, F.M. Pan, H.M. Yu, J.H. Chen, C.W. Yao, A.H. Wang, Biochemical and immunological studies of nucleocapsid proteins of severe acute respiratory syndrome and 229E human coronaviruses, *Proteomics* 5 (4) (2005) 925–937.
- [108] B. Shang, X.-Y. Wang, J.-W. Yuan, A. Vabret, X.-D. Wu, R.-F. Yang, L. Tian, Y.-Y. Ji, V. Deubel, B. Sun, Characterization and application of monoclonal antibodies against N protein of SARS-coronavirus, *Biochem. Biophys. Res. Commun.* 336 (1) (2005) 110–117.
- [109] S.J. Liu, C.H. Leng, S.P. Lien, H.Y. Chi, C.Y. Huang, C.L. Lin, W.C. Lian, C.J. Chen, S.L. Hsieh, P. Chong, Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates, *Vaccine* 24 (16) (2006) 3100–3108.
- [110] N.T. Supekar, A. Shahajan, A.S. Gleinich, D. Rouhani, C. Heiss, P. Azadi, SARS-CoV-2 Nucleocapsid protein is decorated with multiple N- and O-glycans, *bioRxiv: the preprint server for biology* (2020) <https://doi.org/10.1101/2020.08.26.269043>.
- [111] R. Zhou, R. Zeng, A. von Brunn, J. Lei, Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein, *Molecular Biomedicine* 1 (1) (2020), 2.
- [112] C. Meier, A.R. Aricescu, R. Assenberg, R.T. Aplin, R.J. Gilbert, J.M. Grimes, D.I. Stuart, The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus, *Structure* 14 (7) (2006) 1157–1165.
- [113] C.S. Shi, H.Y. Qi, C. Boullaran, N.N. Huang, M. Abu-Asab, J.H. Shelhamer, J.H. Kehrl, SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome, *J. Immunol.* 193 (6) (2014) 3080–3089.
- [114] K. Xu, B.J. Zheng, R. Zeng, W. Lu, Y.P. Lin, L. Xue, L. Li, L.L. Yang, C. Xu, J. Dai, F. Wang, Q. Li, Q.X. Dong, R.F. Yang, J.R. Wu, B. Sun, Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein, *Virology* 388 (2) (2009) 279–285.
- [115] R. Cagliani, D. Forni, M. Clerici, M. Sironi, Coding potential and sequence conservation of SARS-CoV-2 and related animal viruses, *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* 83 (2020) 104353.
- [116] T. Koyama, D. Platt, L. Parida, Variant analysis of SARS-CoV-2 genomes, *Bull. World Health Organ.* 98 (7) (2020) 495–504.
- [117] P. Calligari, S. Bobone, G. Ricci, A. Bocedi, Molecular investigation of SARS-CoV-2 proteins and their interactions with antiviral drugs, *Viruses* 12 (4) (2020).
- [118] M. Prajapat, P. Sarma, N. Shekhar, P. Avti, S. Sinha, H. Kaur, S. Kumar, A. Bhattacharyya, H. Kumar, S. Bansal, B. Medhi, Drug targets for corona virus: a systematic review, *Indian journal of pharmacology* 52 (1) (2020) 56–65.
- [119] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, *Cell Discovery* 6 (1) (2020), 14.
- [120] L.A. Grohskopf, L.Z. Sokolow, K.R. Broder, E.B. Walter, A.M. Fry, D.B. Jernigan, Prevention and control of seasonal influenza with vaccines: recommendations of the advisory committee on immunization practices—United States, 2018–19 influenza season, *MMWR. Recommendations and reports: morbidity and mortality weekly report, Recommendations and reports* 67 (3) (2018) 1–20.
- [121] S. Jiang, Y. He, S. Liu, SARS vaccine development, *Emerg. Infect. Dis.* 11 (7) (2005) 1016–1020.
- [122] J.A. Regla-Nava, J.L. Nieto-Torres, J.M. Jimenez-Guardeno, R. Fernandez-Delgado, C. Fett, C. Castano-Rodríguez, S. Perlman, L. Enjuanes, M.L. DeDiego, Severe acute respiratory syndrome coronaviruses with mutations in the E protein are attenuated and promising vaccine candidates, *J. Virol.* 89 (7) (2015) 3870–3887.
- [123] L. Du, Y. He, Y. Zhou, S. Liu, B.J. Zheng, S. Jiang, The spike protein of SARS-CoV—a target for vaccine and therapeutic development, *Nat Rev Microbiol* 7 (3) (2009) 226–236.
- [124] L.A. Jackson, E.J. Anderson, N.G. Roupael, P.C. Roberts, M. Makhene, R.N. Coler, M.P. McCullough, J.D. Chappell, M.R. Denison, L.J. Stevens, A.J. Pruijssers, A. McDermott, B. Flach, N.A. Doria-Rose, K.S. Corbett, K.M. Morabito, S. O'Dell, S.D. Schmidt, P.A. Swanson, M. Padilla 2nd, J.R. Mascola, K.M. Neuzil, H. Bennett, W. Sun, E. Peters, M. Makowski, J. Albert, K. Cross, W. Buchanan, R. Pikaart-Tautges, J.E. Ledgerwood, B.S. Graham, J.H. Beigel, R.N.A.S.G., m, An mRNA vaccine against SARS-CoV-2 - preliminary report, *N. Engl. J. Med.* 383 (2020) 1920–1931.
- [125] P.M. Folegatti, K.J. Ewer, P.K. Aley, B. Angus, S. Becker, S. Belij-Rammerstorfer, D. Bellamy, S. Bibi, M. Bittaye, E.A. Clutterbuck, C. Dold, S.N. Faust, A. Finn, A.L. Flaxman, B. Hallis, P. Heath, D. Jenkin, R. Lazarus, R. Makinson, A.M. Minassian, K.M. Pollock, M. Ramasamy, H. Robinson, M. Snape, R. Tarrant, M. Voysey, C. Green, A.D. Douglas, A.V.S. Hill, T. Lambe, S.C. Gilbert, A.J. Pollard, J. Aboagye, K. Adams, A. Ali, E. Allen, J.L. Allison, R. Anslow, E.H. Arbe-Barnes, G. Babbage, K. Baillie, M. Baker, P. Baker, I. Baleanu, J. Ballaminut, E. Barnes, J. Barrett, L. Bates, A. Batten, K. Beadon, R. Beckley, E. Berrie, L. Berry, A. Beveridge, K.R. Bewley, E.M. Bijker, T. Bingham, L. Blackwell, C.L. Blundell, E. Bolam, E. Boland, N. Borthwick, T. Bower, A. Boyd, T. Brenner, P.D. Bright, C. Brown-O'Sullivan, E. Brunt, J. Burbage, S. Burge, K.R. Buttigieg, N. Byard, I. Cabera Puig, A. Calvert, S. Camara, M. Cao, F. Cappuccini, M. Carr, M.W. Carroll, V. Carter, K. Cathie, R.J. Challis, I. Chelysheva, J.-S. Cho, P. Cicconi, L. Cifuentes, H. Clark, E. Clark, T. Cole, R. Colin-Jones, C.P. Conlon, A. Cook, N.S. Coombes, R. Cooper, C.A. Cosgrove, K. Coy, W.E.M. Crocker, C.J. Cunningham, B.E. Damratoski, L. Dando, M.S. Dattoo, H. Davies, H. De Graaf, T. Demissie, C. Di Maso, I. Dietrich, T. Dong, F.R. Donnellan, N. Douglas, C. Downing, J. Drake, R. Drake-Brockman, R.E. Drury, S.J. Dunachie, N.J. Edwards, F.D.L. Edwards, C.J. Edwards, S.C. Elias, M.J. Elmore, K.R.W. Emery, M.R. English, S. Fagerbrink, S. Felle, S. Feng, S. Field, C. Fixmer, C. Fletcher, K.J. Ford, J. Fowler, P. Fox, E. Francis, J. Frater, J. Furze, M. Fuskova, E. Galiza, D. Gbesemetse, C. Gilbride, G. Gorini, L. Goulston, C. Grabau, L. Gracie, Z. Gray, L.B. Guthrie, M. Hackett, S. Halwe, E. Hamilton, J. Hamlyn, B. Hanumunthadu, I. Harding, S.A. Harris, A. Harris, D. Harrison, C. Harrison, T.C. Hart, L. Haskell, S. Hawkins, I. Head, J.A. Henry, J. Hill, S.H.C. Hodgson, M.M. Hou, E. Howe, N. Howell, C. Hutlin, S. Ikram, C. Isitt, P. Iveson, S. Jackson, F. Jackson, S.W. James, M. Jenkins, E. Jones, K. Jones, C.E. Jones, B. Jones, R. Kailath, K. Karampatsas, J. Keen, S. Kelly, D. Kelly, D. Kerr, S. Kerridge, L. Khan, U. Khan, A. Killen, J. Kinch, T.B. King, L. King, J. King, L. Kingham-Page, P. Klenerman, F. Knapper, J.C. Knight, S. Koleva, A. Kupke, C.W. Larkworthy, J.P.J. Larwood, A. Laskey, A.M. Lawrie, A. Lee, K.Y. Ngan Lee, E.A. Lee, H. Legge, A. Lelliott, N.-M. Lemm, A.M. Lias, A. Linder, S. Lipworth, X. Liu, S. Liu, R. Lopez Ramon, M. Lwin, F. Mabesa, M. Madhavan, G. Mallett, K. Mansatta, I. Marcal, S. Marinou, E. Marlow, J.L. Marshall, J. Martin, J. McEwan, G. Meddaugh, A.J. Mentzer, N. Mirtorabi, M. Moore, E. Moran, E. Morey, V. Morgan, S.J. Morris, H. Morrison, G. Morshead, R. Morter, Y.F. Mujajidi, J. Muller, T. Munera-Huertas, C. Munro, A. Munro, S. Murphy, V.J. Muster, P. Mweu, A. Noé, F.L. Nugent, E. Nugent, K. O'Brien, D. O'Connor, B. Oguti, J.L. Oliver, C. Oliveira, P.J. O'Reilly, M. Osborn, P. Osborne, C. Owen, D. Owens, N. Owino, M. Pacurar, K. Parker, H. Parracho, M. Patrick-Smith, V. Payne, J. Pearce, Y. Peng, M.P. Peralta Alvarez, J. Perring, K. Pfaffert, D. Pipini, E. Plested, H. Pluess-Hall, K. Pollock, I. Poulton, L. Presland, S. Probstgaard-Morys, D. Pulido, K. Radia, F. Ramos Lopez, J. Rand, H. Ratcliffe, T. Rawlinson, S. Rhead, A. Riddell, A.J. Ritchie, H. Roberts, J. Robson, S. Roche, C. Rohde, C.S. Rollier, R. Romani, I. Rudiansyah, S. Saich, S. Sajjad, S. Salvador, L. Sanchez Riera, H. Sanders, K. Sanders, S. Sapaun, C. Sayce, E. Schofield, G. Scream, B. Selby, C. Semple, H.R. Sharpe, A. Shea, H. Shelton, S. Silk, L. Silva-Reyes, D.T. Skelly, H. Smees, C.C. Smith, D.J. Smith, R. Song, A.J. Spencer, E. Stafford, A. Steele, E. Stefanova, L. Stockdale, A. Szigei, A. Tahiri-Alaoui, M. Tait, H. Talbot, R. Tanner, I.J. Taylor, V. Taylor, R. Te Water Naude, N. Thakur, Y. Themistocleous, A. Themistocleous, M. Thomas, T.M. Thomas, A. Thompson, S. Thomson-Hill, J. Tomlins, S. Tonks, J. Towner, N. Tran, J.A. Tree, A. Truby, K. Turkentine, C. Turner, N. Turner, S. Turner, T. Tuthill, M. Ulaszewska, R. Varughese, N. Van Doremalen, K. Veighey, M.K. Verheul, I. Vichos, E. Vitale, L. Walker, M.E.E. Watson, B. Welham, J. Wheat, C. White, R. White, A.T. Worth, D. Wright, S. Wright, X.L. Yao, Y. Yau, Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial, *The Lancet*.
- [126] V. Monteil, H. Kwon, P. Prado, A. Hagelkruys, R.A. Wimmer, M. Stahl, A. Leopoldi, E. Garreta, C. Hurtado Del Pozo, F. Prosper, J.P. Romero, G. Wirnsberger, H. Zhang, A.S. Slutsky, R. Conder, N. Montserrat, A. Mirazimi, J.M. Penninger, Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2, *Cell* 181 (4) (2020) 905–913 e7.
- [127] M. Hoffmann, G. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Muller, C. Drosten, S. Pohlmann, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, *Cell* 181 (2) (2020) 271–280 e8.
- [128] R.L. Graham, E.F. Donaldson, R.S. Baric, A decade after SARS: strategies for controlling emerging coronaviruses, *Nat Rev Microbiol* 11 (12) (2013) 836–848.
- [129] H. Pang, Y. Liu, X. Han, Y. Xu, F. Jiang, D. Wu, X. Kong, M. Bartlam, Z. Rao, Protective humoral responses to severe acute respiratory syndrome-associated coronavirus: implications for the design of an effective protein-based vaccine, *The Journal of general virology* 85 (Pt 10) (2004) 3109–3113.
- [130] K. Anand, J. Ziebuhr, P. Wadhvani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs, *Science (New York, N.Y.)* 300 (5626) (2003) 1763–1767.
- [131] W. Vuong, M.B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H.A. Saffran, R.T. McKay, M.J. van Belkum, M.A. Joyce, H.S. Young, D.L. Tyrrell, J.C. Vederas, M.J. Lemieux, Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication, *Nat. Commun.* 11 (1) (2020) 4282.
- [132] A. Gahlawat, N. Kumar, R. Kumar, H. Sandhu, I.P. Singh, A. Sjöstedt, P. Garg, Structure-based virtual screening to discover potential lead molecules for the SARS-CoV-2 main protease, *J. Chem. Inf. Model.* 60 (2020) 5781–5793.
- [133] C. Ma, M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, J. Wang, Boceprevir, GC-376, and Calpain Inhibitors II, XII Inhibit SARS-CoV-2 Viral Replication by Targeting the Viral Main Protease, *bioRxiv*, 2020.
- [134] J.S. Morse, T. Lalonde, S. Xu, W.R. Liu, Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV, *Chembiochem: a European journal of chemical biology* 21 (5) (2020) 730–738.
- [135] B.N. Williamson, F. Feldmann, B. Schwarz, K. Meade-White, D.P. Porter, J. Schulz, N. van Doremalen, I. Leighton, C.K. Yinda, L. Pérez-Pérez, A. Okumura, J. Lovaglio, P.W. Hanley, G. Saturday, C.M. Bosio, S. Anzick, K. Barbian, T. Cihlar, C. Martens, D.P. Scott, V.J. Munster, E. de Wit, Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2, *Nature* 585 (2020) 273–276.
- [136] N. Le Bert, A.T. Tan, K. Kunasegaran, C.Y.L. Tham, M. Hafezi, A. Chia, M.H.Y. Chng, M. Lin, N. Tan, M. Linster, W.N. Chia, M.I.C. Chen, L.-F. Wang, E.E. Ooi, S. Kalimuddin, P.A. Tambyah, J.G.-H. Low, Y.-J. Tan, A. Bertoletti, SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls, *Nature* 584 (2020) 457–462.
- [137] Y.F. Tu, C.S. Chien, A.A. Yarmishyn, Y.Y. Lin, Y.H. Luo, Y.T. Lin, W.Y. Lai, D.M. Yang, S.J. Chou, Y.P. Yang, M.L. Wang, S.H. Chiou, A review of SARS-CoV-2 and the ongoing clinical trials, *Int. J. Mol. Sci.* 21 (7) (2020).

Glossary

Angiotensin converting enzyme 2 (ACE2): Enzyme which attaches to the cell membranes of cells present in lungs, heart, kidney, intestines and arteries. The human ACE2 referred as hACE2 is responsible for binding of various coronaviruses into cells.

3CL^{pro}: Chymotrypsin-like proteases which cleaves the nsp polyprotein at 11 sites leading to the formation of other intermediates as well as mature non-structural proteins.

Domain: A part of protein sequence that can exist and function independently.

ER-Golgi intermediate compartment (ERGIC): an organelle located between the Golgi and rough endoplasmic reticulum on the secretory pathway in eukaryotic cells.

Human Coronaviruses (HCoVs): RNA viruses that cause respiratory tract infections and diseases in humans such as COVID-19.

Major Histocompatibility Complex (MHC): A set of polymorphic genes present on vertebrate DNA coding for cell proteins involved in adaptive immune response

Non-structural proteins (nsps): Proteins encoded by the virus but do not form part of the viral particle.

Nuclear factor kappa B (NF- κ B): A protein complex which regulates the cytokine production, survival of cells and DNA transcription.

Nucleocapsid protein: Abundant protein of coronaviruses having variable molecular mass ranging from 45 to 60 kDa.

Open Reading Frame (ORF): It refers to the part of DNA that can be translated and contains no stop codons.

Receptor binding domain (RBD): Spike protein of the virus has this key region which docks with ACE2 receptor of host for entering into the cell, further causing infection.

Ribonucleoprotein complex (RNP complex): A complex formed by RNA binding protein and RNA.

Tumor Necrosis Factor (TNF): A multifunctional cytokine responsible for key roles in different cellular events like cell proliferation, survival, death and differentiation.

Virus like particles (VLPs): Entities that resemble viruses, but are non-infectious due to lack of viral genetic material.

Zoonotic: of animal origin.