



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, infection, transmission, transcription, translation, proteins, and treatment: A review



Jahangir Emrani<sup>a,\*</sup>, Maryam Ahmed<sup>b</sup>, Liesl Jeffers-Francis<sup>c</sup>, John C. Teleha<sup>d</sup>, Nathan Mowa<sup>b</sup>, Robert H. Newman<sup>c</sup>, Misty D. Thomas<sup>c</sup>

<sup>a</sup> Department of Chemistry, North Carolina A&T State University, Greensboro, NC 27411, United States of America

<sup>b</sup> Department of Biology, Appalachian State University, Boone, NC 28608, United States of America

<sup>c</sup> Department of Biology, North Carolina A&T State University, Greensboro, NC 27411, United States of America

<sup>d</sup> Department of Reference and Instruction, North Carolina A&T State University, Greensboro, NC 27411, United States of America

## ARTICLE INFO

## Keywords:

Enzymes

Proteins

Peptides: Nucleic acids

DNA

RNA

Drugs

Vaccines

SARS-COV-2

COVID-19

CORONA virus

Epidemiology

Antibody

Transmission

## ABSTRACT

In this review, we describe the key molecular entities involved in the process of infection by SARS-CoV-2, while also detailing how those key entities influence the spread of the disease. We further introduce the molecular mechanisms of preventive and treatment strategies including drugs, antibodies, and vaccines.

## 1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also known as “the novel coronavirus” due to genome variation relative to previously identified coronaviruses, is a positive sense RNA virus and the etiological agent of COVID-19. SARS-CoV-2 is a member of the viral family, *Coronaviridae*, and subfamily, *Coronavirinae*, which are large, enveloped, single-stranded RNA viruses between 65 and 125 nm in diameter. Under electron microscopy, virions take on a roughly spherical shape and possess finger like extensions (glycoproteins), commonly referred to as spikes, on their surface (Fig. 1) [1–4]. These spikes resemble a crown (or a “corona” in Latin) that is the basis for the name of this family of viruses [5]. Members of the subfamily *Coronavirinae* infect a variety of mammals, causing diverse clinical syndromes. They are grouped into four sub-groups based on sequence homology: alpha, beta, gamma and delta [6]. To date, only seven coronaviruses infect humans. Four (two  $\alpha$ -CoVs and two  $\beta$ -CoVs) cause mild cold-like illnesses but

three, including severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and the recently discovered SARS-CoV-2 lead to serious or fatal disease. SARS-CoV and MERS-CoV emerged in the human population in 2002 and 2012, respectively, while SARS-CoV-2 emerged in 2019, leading to the current global pandemic [7,8]. The 2002 SARS-CoV infected approximately 8096 people, with a mortality rate of ~10%, whereas MERS-CoV infected over 2500 people with a mortality rate of ~36% [9–12]. As of August 20, 2021, the COVID-19 pandemic caused by the SARS-CoV-2 coronavirus has infected over 210 million people and has killed close to 4.5 million people worldwide [13–15]. Genomic analysis shows patterns of molecular divergence between SARS-CoV-2 and other coronaviruses, leading to a disease that is distinct from that of both SARS-CoV and MERS-CoV [5,16]. In this article, we will focus on describing the molecular basis of SARS-CoV-2 infection and antiviral drugs that target the virus. More detailed medical, epidemiological, and technologies used to identify and combat SARS-CoV-2 have been

\* Corresponding author.

E-mail address: [emrani@ncat.edu](mailto:emrani@ncat.edu) (J. Emrani).

<https://doi.org/10.1016/j.ijbiomac.2021.10.172>

Received 30 September 2021; Accepted 21 October 2021

Available online 28 October 2021

0141-8130/© 2021 Elsevier B.V. All rights reserved.

reviewed elsewhere [17–26].

## 2. Pathogenicity and transmission

Like other coronaviruses, transmission of SARS-CoV-2 can occur primarily via infected respiratory droplets through direct, indirect, or close contact with infected individuals [27]. The virus targets and infects cells at the nasal, conjunctival, or oral mucosa once respiratory particles are inhaled or deposited on these locations. Receptors for SARS-CoV-2, which are found mainly in the human respiratory tract epithelium, facilitate virus entry and allow the virus to efficiently replicate and spread to various organs. Approximately five days after infection, the host is contagious and can transmit the virus easily through exhalation to potentially infect others. Symptoms associated with the virus infection are due to transient damage to susceptible cells in various organs, including the lung, intestines, kidneys, and blood vessels [28].

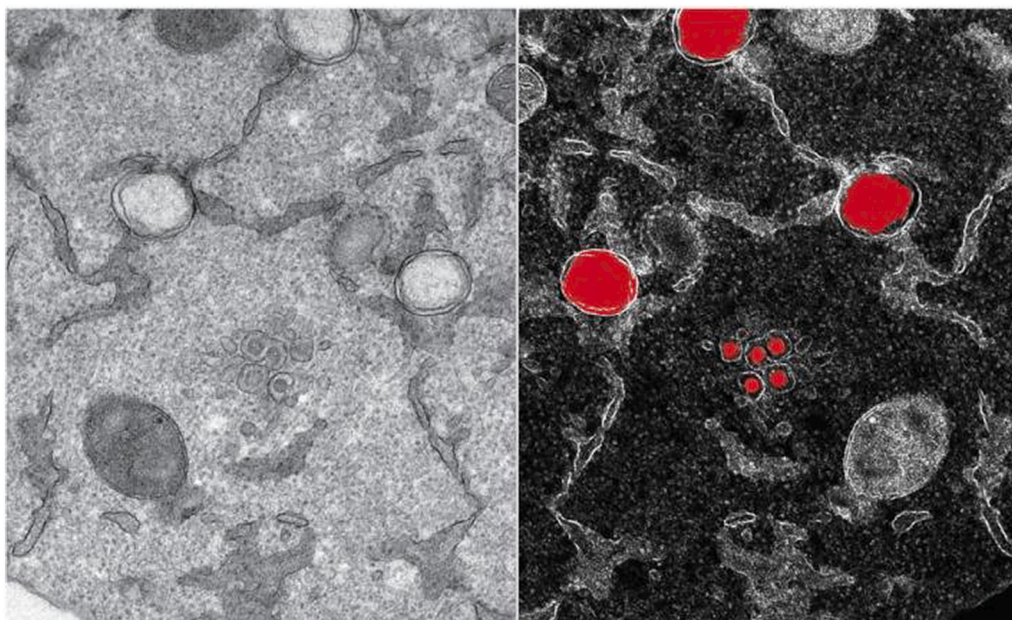
Infection by the virus triggers an initial inflammatory immune response during which white blood cells originating from the lymph nodes, such as helper T-cells and cytotoxic T-cells, infiltrate to the site of infection to eliminate virus-infected cells [29]. In addition, body temperature is raised, causing fever in an attempt to kill (i.e., denature) the virus while facilitating the immune response [30,31]. In patients who develop more severe disease, SARS-CoV-2 induces an aberrant host immune response. Overproduction of proinflammatory cytokines leads to a so-called “cytokine storm” which, if not remedied with the help of immunomodulators that regulate cytokine production or drugs like dexamethasone or blood thinners, can lead to death [32–35]. This “cytokine storm”, can adversely affect several organs, including the liver, kidneys, and lungs [32,36]. The latter leads to breakdown of lung epithelial cells that line the air sacs in lungs, causing the air sacs to fill with fluid. This induces pneumonia and, in severe cases, acute respiratory distress syndrome (ARDS) that can cause lung failure and death. Studies have shown that levels of certain cytokines can be correlated with the severity of COVID-19 symptoms [32].

SARS-CoV-2 has an enhanced rate of transmission compared to the 2002 SARS-CoV. This is due to several characteristics, both of the virus

(i.e., SARS-CoV-2) and its associated disease (i.e., COVID-19). For instance, active virus can remain viable on some surfaces for up to 9 days and in the indoor air for several hours, thus increasing the chances of exposure [37]. Likewise, differences in the severity and onset of symptoms of COVID-19 compared to SARS may lead to increased contact with infected individuals. For example, during the SARS-CoV epidemic of 2002, mortality rates were nearly twice that currently observed for COVID-19 pandemic (9.6% vs. 5.4%, respectively) [10,12], and infected individuals exhibited clear symptoms of respiratory distress almost immediately [38–40]. In contrast, people infected with SARS-CoV-2 may not show signs of infection or illness for a week or longer, despite being contagious and shedding virus [41]. Furthermore, asymptomatic individuals and people with mild symptoms may carry large amounts of virus in the upper respiratory tract, thus contributing to the rapid spread of SARS-CoV-2. In fact, in some situations, even after symptoms disappear, SARS-CoV-2-infected individuals may be shedding virus. In addition, re-infection may occur in specific cases, especially given that human coronaviruses are well equipped to subvert host immunity. For instance, the emergence of new variants of the virus that are more contagious and may carry an increased risk of death, such as the South African variant (B.1.351), the UK variant (B.1.1.7), the Indian variant (B.1.617.2, also commonly known as the delta variant), or the Brazil/Japan variant (P.1), suggest that SARS-CoV-2 is rapidly evolving mechanisms to increase contagiousness and subvert existing immune defenses [42,43]. Clearly, additional information is needed to understand the global risks these new variants pose.

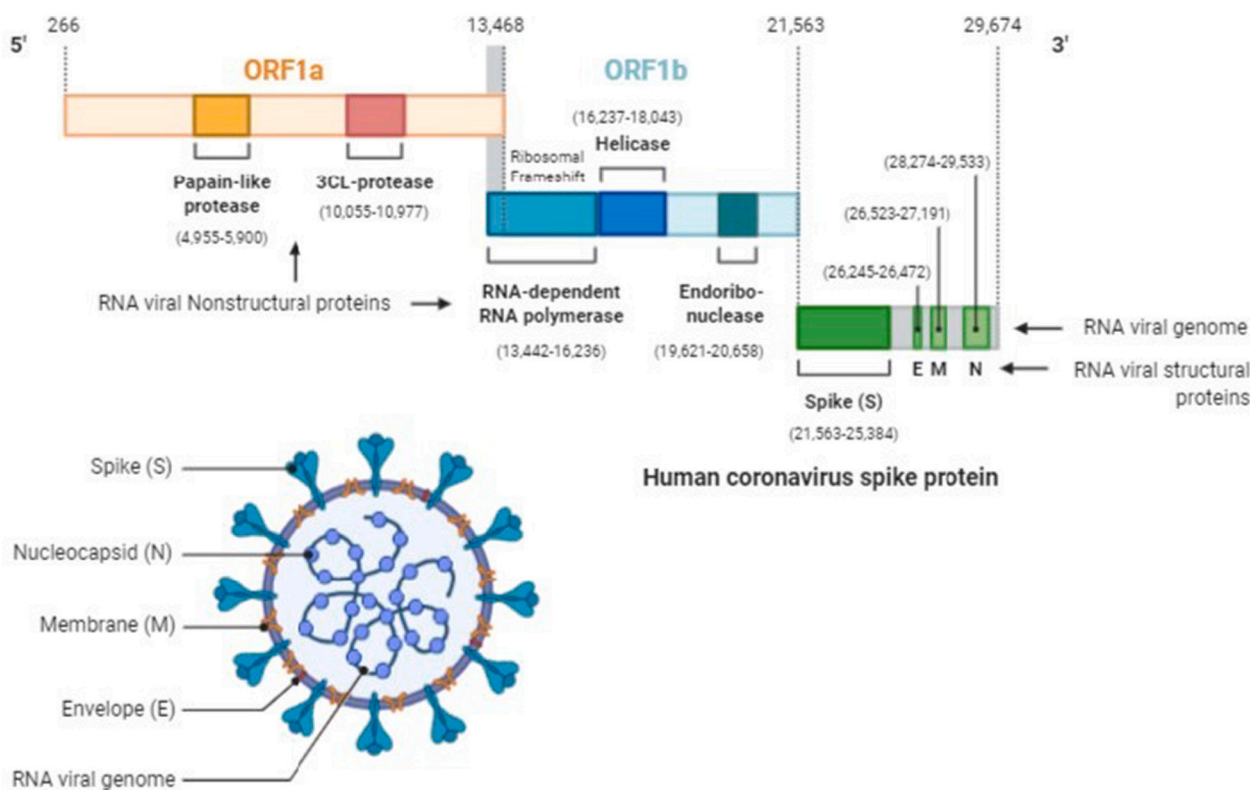
### 2.1. Genome structure

After the initial sequencing of the SARS-CoV-2 genome in December 2019 [16], the viral genome has been sequenced numerous times in order to detect mutations and identify variants [44–46]. The genome of SARS-CoV-2 is comprised of a single-stranded RNA molecule housed inside a fatty acid membrane (Fig. 2), known as the envelope [47]. The viral RNA, which is capped and polyadenylated, contains a 5' leader sequence of 70 bases, with 7–10 of those bases being transcription-regulatory sequences (TRS-L), and encodes 13–15 open reading frames



**Fig. 1.** Transmission electron micrograph of SARS-CoV-2 viral particle, isolated from a patient. Large red circles represent newly formed genome copies and the small red balloons are new virions formed by budding at the interface of Golgi apparatus and endoplasmic reticulum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Credit: Yannick Schwab/EMBL in Ref. [4].



## Coronavirus genome and Proteins

Created in BioRender.com

**Fig. 2.** The genome of SARS-CoV-2 contains over 29,000 bases and codes for 29 proteins. SARS-CoV-2 has four structural proteins: The proteins E and M that form the viral envelope, protein N binds to the RNA genome of SARS-CoV-2 and S protein, which binds to host ACE2 receptor to initiate infection. The nonstructural proteins (NsPs) initially form as two long polypeptides, which by the action of the virus' main protease (MP), Nsp5, aside from autocatalytically releasing itself, also releases the RNA-dependent RNA polymerase (RdRp), Nsp12. Because of its important role in the replication and transcription of the virus, MP can be targeted by antiviral drugs against SARS-CoV-2. Ref: [47].

(ORFs). ORF1a is the longest region of the genome, comprising ~13,200 of the 29,674 bases in the sequence and encoding two genes, the papain-like protease and the 3CL protease. ORF1b, which overlaps with ORF1a [47], encompasses bases 14,442–21,563 and encodes enzymes important for viral replication, such as a RNA-dependent RNA polymerase (RdRp), a helicase, and an endonuclease. Subsequent regions spanning from nucleotides 21,562–29,674 encode the structural proteins: the spike (S) protein, membrane (M) protein, envelope (E) protein, nucleocapsid (N) protein, and accessory proteins. Each ORF also has its own transcriptional-regulatory sequences (TRSs), which are located immediately adjacent to the ORFs. The RdRp pauses when it crosses a TRS in the body (TRS-B) and switches the template to the TRS in the leader (TRS-L) during replication. This leads to discontinuous transcription and the fusion of negative (or complementary) strand RNAs that serve as templates for generation of genomic strand RNAs.

It is important to note that while most living organisms possess DNA as their fundamental genetic material, using RNA as the basis of the genome is unique to viruses [16,48]. There are several important differences between DNA and RNA that can affect their cellular function. For instance, RNA uses the nitrogenous base uracil in place of the thymine found in DNA. Likewise, RNA nucleotides are composed of ribose sugar molecules that contain a hydroxyl moiety in the 2' position. In contrast, DNA contains a hydrogen atom at the 2' position, making it a deoxyribose. The absence of a hydroxyl at the 2' position causes the ribose sugar to adopt a C3'-endo sugar pucker vs. the C2'-endo pucker typically found in B-form DNA. As a consequence, rather than aligning neatly with the helical axis like in B-DNA, the nucleobases of RNA are

tilted relative to the helical axis. This reduces the stability of complementary base pairs in RNA molecules, preventing RNA from forming long double helices. Instead, RNA tends to exhibit more single-stranded character than DNA, which is predominantly double-stranded. In DNA-based organisms, gene expression involves the transcription of DNA to RNA using RNA polymerase. RNA viruses, on the other hand, use RNA as templates for the transcription of additional copies of RNA. Due to this unique transcriptional requirement, virally encoded RdRps facilitate this process.

The genomic structure and content of SARS-CoV-2 has been shown to promote its survival in host cells and may contribute to the progression of disease. Coronaviruses, including SARS-CoV-2, contain cytosine-phosphate-guanine (CpG) nucleotides as part of their genome. CpG sequences allow cellular detection of viruses through association with toll-like receptor 9 (TLR9) in endosomes to trigger a type I interferon (IFN) antiviral immune response in the host. This antiviral mechanism may be driven by the zinc-finger antiviral protein (ZAP), which restricts numerous viral pathogens by targeting CpG-rich RNA sequences. Interestingly, studies have shown that during replication of SARS-CoV-2, there is a selective pressure to lower CpG content as a means of evading the host immune response [49]. Interestingly, as another survival mechanism, the SARS-CoV-2 viral genome contains large-scale internal RNA base pairing, or genome-scale ordered RNA structures (GORS), that are associated with shielding viral RNA recognition in the host cell and may contribute to persistent mechanisms [45].

Many secondary structures (SS) are also built into the genome. These structures are generated once the viral RNA has been synthesized and

released into the cytoplasm of the host cell. The various forms of SS include hairpin-forming inverted repeats (IR), quadruplex sites, and a slippery sequence with a downstream pseudoknot, all of which are needed to equip the virus for replication. For instance, a pseudoknot structure in ORF1ab controls the frameshift during the overlapped transition of ORF1a and ORF1ab. This is known as a programmed -1 ribosomal frameshift (-1 PRF) signal. The -1 PRF signal also contains a slippery sequence and linker region that play a role in protein translation [50,51].

## 2.2. Viral replication

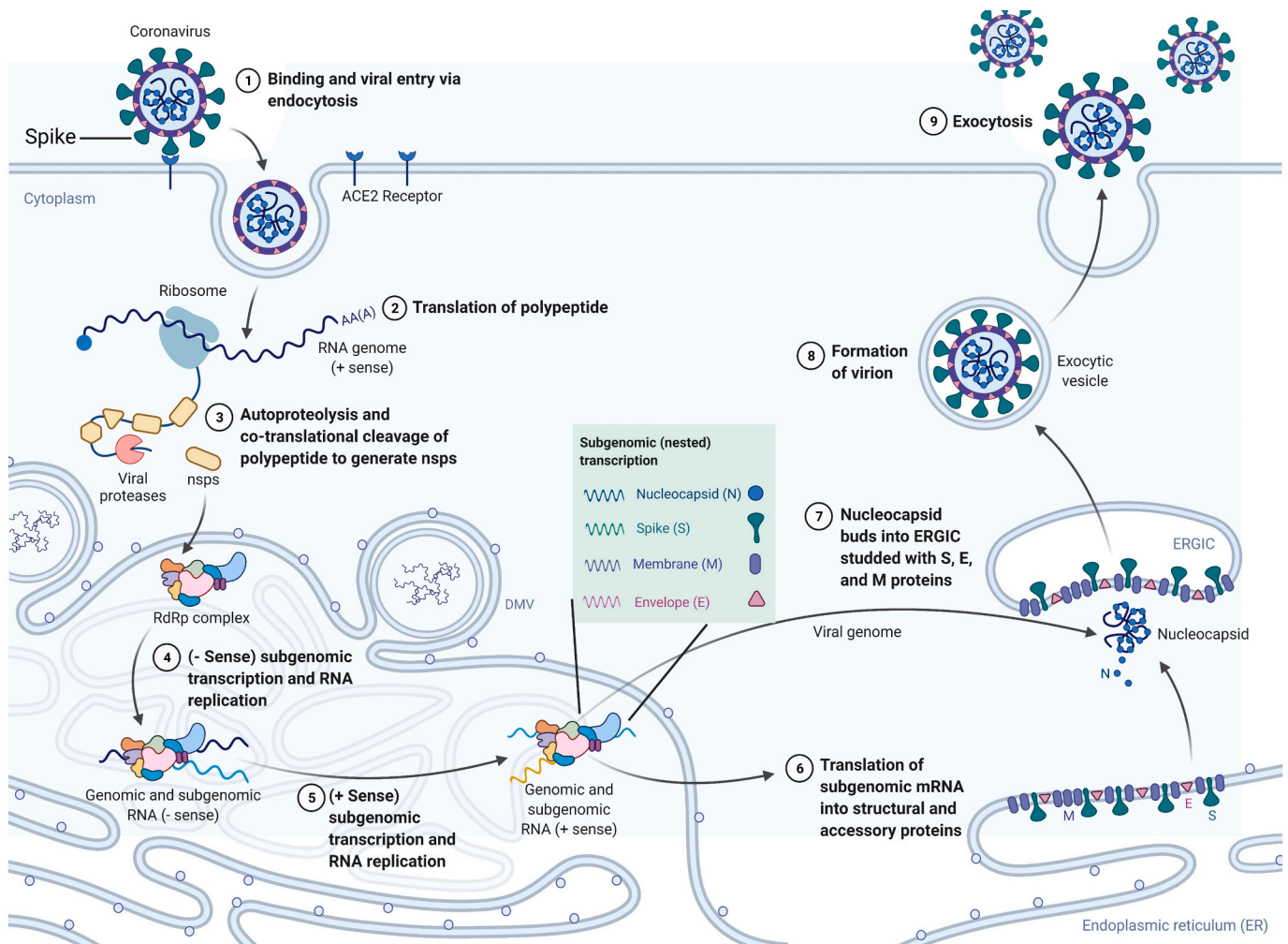
### 2.2.1. Entry into host cells

All viruses, including SARS-CoV-2, enter host cells and efficiently modulate cellular factors to facilitate replication (Fig. 3) [52]. SARS-CoV-2 has been shown to infect cells in two ways: entry through endocytic mechanisms or through fusion of the viral envelope with the host cell plasma membrane [46]. In both cases, the spike protein subunits, S1 and S2, mediate attachment and entry by binding with the cell surface protein, angiotensin-converting enzyme 2 (ACE2). Recent structural

data obtained via high resolution cryo-electron microscopy has revealed the simultaneous binding of two S glycoprotein trimers to an ACE2 dimer [53,54]. Binding of the viral glycoproteins to ACE2 is further facilitated by heparan sulfate (HS). Finally, cleavage by the serine protease, TMPRSS2, promotes fusion with the host membrane. Other cellular proteases (e.g., furin) facilitate pH-dependent entry through the endocytic pathway. Studies have shown that the main entry routes are dependent on the presence of select proteases in specific cell types [55]. Once the virus has entered the host cell, the RNA genome is released into the cytoplasm in a process known as uncoating [56].

### 2.2.2. Translation of viral proteins

The release of the coronavirus genome into the host cytoplasm triggers a complex and highly regulated program of viral gene expression [46]. The two polyproteins (pp1a and pp1ab) are produced through translation of ORF1a and ORF1b from the genomic RNA. As mentioned above, pp1ab is expressed from a programmed-1 ribosome frameshift between ORF1a and ORF1b. The efficiency of the frameshift can be measured through ribosomal profiling and, in the case of SARS-CoV-2, pp1a is expressed at levels approximately 1.4–2.2 times greater than



**Fig. 3.** SARS-CoV-2 life cycle. SARS-CoV-2 spike (S) glycoprotein binds to the ACE2 receptor on the host cell surface and the virus enters the cell via endocytosis. Viral genomic RNA is released into the cytoplasm of the host cell, and the single-stranded positive-sense genome is transcribed and translated to produce nonstructural proteins (nsps) including replicase polyproteins (RNA-dependent RNA polymerase and helicase) to create an RdRp complex. Subgenomic transcription and RNA replication occur within the RdRp complex to synthesize negative-strand guide RNA (gRNA) and a set of subgenomic RNAs for viral replication and transcription. The newly produced subgenomic RNAs are translated into viral structural proteins such as the spike (S), nucleocapsid (N), membrane (M), and envelope (E) proteins. These proteins are inserted into the membrane of the rough endoplasmic reticulum (ER) and then transported to the ER-Golgi intermediate compartment (ERGIC) to assemble with the N protein-encapsidated RNA to form viral particles. Virions are then released from the cell through exocytosis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pp1ab [57]. Through co-translational and post-translation mechanisms, including processing by viral proteases nsp3 and nsp5, sixteen non-structural proteins are released from pp1a (nsp1–11) and pp1ab (nsp1–10, nsp12–16), of which fifteen compose the viral replication and transcription complex (RTC). The RTC includes RNA-processing and RNA-modifying enzymes as well as an RNA proofreading function. Of these enzymes, RNA synthesis is performed by the viral RdRp (nsp12) together with its two cofactors, nsp7 and nsp8, which has a 3'-terminal adenyltransferase activity. The bifunctional nsp14 contains both a 3' to 5' exonuclease (ExoN) domain and a guanine-N7-methyltransferase (N7-MTase) domain. Previous studies have shown that the ExoN function of nsp14 is critical for boosting longer-term replication fidelity for SARS-CoV. However, recently, it was revealed through biochemical evaluation of MERS-CoV and SARS-CoV-2 mutants that, for these viruses, ExoN plays a more direct role in RNA synthesis in addition to safeguarding the long-term fidelity of the viral genome [58]. In addition to these enzymes, non-structural proteins also provide important support functions to accommodate the RTC, such as modulating intracellular membranes, host immune evasion, and providing cofactors to facilitate replication. The 3' one third of the SARS-CoV-2 genome expresses genes for viral structural proteins and accessory proteins, including the spike (S), nucleocapsid (N), membrane (M) and envelope (E) proteins. The S1 subunit of S protein mediates attachment to ACE2 while S2 promotes membrane fusion. Likewise, the N, M and E proteins are critical for different aspects of viral functioning, maturation and assembly and will be described in more detail in the sections below [59].

### 2.2.3. Transcription of the new genetic material

The early expression of viral genes leads to the biogenesis of viral replication organelles in the cytoplasm of infected cells that create a protected microenvironment for viral genomic RNA expression and the transcription of subgenomic mRNAs [46]. Viral genome replication is initiated by the transcription of full-length negative-sense genomic copies that serve as templates for the production of positive-sense genomic RNAs. These newly synthesized genomes also generate more nsps and RTCs or are packaged into progeny virions during the assembly process. In addition, all coronaviruses, including SARS-CoV-2, have a conserved leader sequence in their 5'-ends that function as *cis*-acting elements for the transcription of subgenomic mRNAs directed by viral RdRp. The TRS-L sequence described previously constitutes a signal for the transcription of subgenomic mRNA while each transcriptional unit on the genomic RNA are preceded by TRS-B, additional transcription regulatory elements [60]. Studies have also demonstrated that *trans*-acting factors encoded by the virus as well as cellular proteins play a role in subgenomic mRNA synthesis [61].

Once subgenomic mRNAs are expressed, nsp16 in conjunction with nsp10, methylates the 5'-end of viral mRNAs to create a 5'-methyl-cap [62]. This mechanism serves to not only promote translation of virally encoded messages by the host translational machinery, but it is also an essential strategy to protect the virus from the host immune response. RNA capping in coronaviruses involves several nonstructural proteins, including nsp13, nsp14, and nsp16. Nsp16, a cap ribose 2'O methyltransferase, is known to form a complex with nsp10 to convert mRNA species from the Cap-O (<sup>me7</sup>G<sub>0</sub>pppA<sub>1</sub>) to the Cap-1 form (<sup>me7</sup>G<sub>0</sub>pppA<sub>1m</sub>) by methylation of the 2'O ribose of the first nucleotide. Mass spectrometry studies have revealed that an essential antiviral factor, IFT1, has high affinity for unmethylated Cap-O forms of RNA and impairs binding of eukaryotic translation initiation factors to 2'O-unmethylated RNA templates, resulting in the inhibition of translation [63]. Therefore, the specificity of IFT1 for the Cap-O form of mRNA serves as a potent antiviral mechanism against viruses that lack 2'O methyltransferase activity. Since coronaviruses actively generate Cap-1 forms of viral mRNAs, this not only promotes translation, but also serves as a viral strategy to avoid detection by the innate immune response.

## 3. Viral proteins

The SARS-CoV-2 genomes contains 29,811 nucleotides and codes for a total of 29 different viral proteins (see Fig. 2) [64]. For comparison, the human genome contains 3.2 billion DNA nucleotides, which contains ~20,000 protein-coding genes [65]. Each of the SARS-CoV-2 viral proteins is classified into one of three groups: 1) structural proteins, 2) non-structural proteins, or 3) accessory proteins (Table 1). Many structures of SARS-CoV-2 proteins have been described in the literature and have been compared with those of other coronaviruses [66]. The variation between these proteins accounts for the observed differences in each virus's contagiousness and infectivity in human cells.

### 3.1. Structure and function of the SARS-CoV-2 proteins

Structural analysis of proteins and protein complexes can provide

**Table 1**  
List of names, functions, and length of proteins.<sup>a</sup>

Name	Name-function	Length (aa)	Name	Name-function	Length (aa)
ORF1ab	ORF1ab polyprotein	7096	nsp14	3'-to-5' exonuclease, proof reading during RNA synthesis	527
ORF1a	ORF1a polyprotein	4405	nsp15	endoRNase, cleaves RNA at polyuridylylate sites	346
nsp1	leader protein, inhibits host protein translation	180	nsp16	2'-o-ribose methyltransferase, in RNA capping	298
nsp2	nsp2, disrupts host cell cycle	638	nsp11	nsp11	13
nsp3	Protease, lowers host immune response, and host translation	1945	S	Surface glycoprotein	1273
nsp4	nsp4	500	ORF3a	ORF3a	275
nsp5	3C-like proteinase	306	E	Envelope protein	75
nsp6	nsp6	290	M	Membrane glycoprotein	222
nsp7	nsp7, primer synthesis and RNA replication	83	ORF6	ORF6	61
nsp8	nsp8, primer synthesis and RNA replication	198	ORF7a	ORF7a	121
nsp9	nsp9, interacts with DDX5 protein to facilitate replication of virus	113	ORF7b	ORF7b	43
nsp10	nsp10, mRNA cap methylation	139	ORF8	ORF8	121
RdRp	RNA-dependent RNA polymerase, RNA replication	932	N	Nucleocapsid phosphoprotein	419
nsp13	Helicase, activity during RNA replication	601	ORF10	ORF10	38

<sup>a</sup> Adapted from Kadam et al. [64].

key insights into the molecular basis for their observed function. Therefore, researchers have examined the structure of several SARS-CoV-2 proteins and related complexes. Traditionally, X-ray crystallography has been used to determine high resolution three-dimensional protein structures important for determining biological function and drug design [67–69]. Though X-ray crystallography is a very detailed and powerful method, it is often limited by difficulties associated with protein crystallization caused by size and flexibility constraints. In addition, due to their flexibility, it is difficult to observe glycosylation patterns on the surface of crystallized proteins. Finally, since crystals require static positioning within a crystal lattice, it is very difficult to observe and predict conformational changes and protein dynamics without assuming movement between crystals in different static conformational states. Therefore, as an alternative technique, many structural biologists have begun to use cryogenic electron microscopy (cryo-EM) to determine protein structure [70,71]. During a cryo-EM experiment, the protein is deposited frozen onto a metal grid in a single molecule layer. The deposited layer is then irradiated with low energy electron beams and a 2-D image of the protein is obtained. By adding additional layers and using a computer to collect and sort the data, a 3D image of the protein is obtained. With cryo-EM, conformational changes and protein dynamics can also be recorded. In recent years, the quality of the images from cryo-EM experiments has improved to resolutions that are comparable to those obtained using crystallography [72–74]. Because of these technologies, several structures of SARS-CoV-2 S-protein in a complex with the ACE2 receptor have been solved [8,53,75,76].

### 3.1.1. The S-protein

The S-protein of SARS-CoV-2 is a 150 kDa transmembrane protein embedded into the surface of the viral envelope. As mentioned previously, S-protein binds to ACE2 receptors on the surface of host cells to facilitate viral entry [8,77]. Several groups have solved the three-dimensional structure of the S-protein, thus providing valuable insights into its overall shape, plasticity, and specific target interactions [8,53,54,72,75]. The S-protein consists of an S1 subunit that binds the host cell receptor and an S2 subunit that is responsible for fusion of the viral envelope with host-cell membranes. In addition, S-protein has a receptor binding domain (RBD), composed of approximately 230 amino acids, that binds to specific sequences of the ACE2 receptor. The protein is fully glycosylated and forms a homotrimer made of A, B and C chains embedded within the viral envelope (in Fig. 9A and B, glycosylation sites are shown in dark grey). Interestingly, the S1 subunit has several domains that have been shown to vary in function depending on the specific viral strain. This variation leads to its ability to use alternative host cell recognition targets.

Recently a cryo-EM structure of the SARS-CoV-2 S-protein ectodomain trimer was solved [8,54]. The result suggests that the SARS-CoV-2 S-protein adopts conformations similar to those reported for both the SARS-CoV and MERS-CoV S-proteins. As observed in other  $\beta$ -coronavirus S-glycoproteins, the SARS-CoV-2 S1 subunit takes on a V-shape and harbors three human ACE2-recognition motifs in its closed state (known as the receptor binding domain or RBD shown in Fig. 9C). Structural information shows that these motifs are buried, and therefore an opening of the structure is likely required for interactions with ACE2. This interaction would initiate a conformational change that would subsequently allow S2 protease cleavage, leading to membrane fusions and viral entry. In comparison to SARS-CoV, which uses S-protein residues Y422, L472, N479, D480, T487, and Y4911 for binding to ACE2, the SARS-CoV-2 S-protein may use amino acids L455, F486, Q493, S494, N501, Y505 for binding to ACE2 [78–81]. The differences in the amino acids promote stronger binding interactions between the SARS-CoV-2 S1 RBD and ACE2 receptors on the host cell. While cryo-EM is a beneficial strategy to elucidate the overall protein structure of S1, it is not an optimal method to identify the structure of the sugars that are covalently linked to the surface of the protein [82–84]. The sugars play important

roles in stabilizing the proteins to aid in folding. In addition, they have been shown to contribute to viral host immune evasion tactics. Because of their importance, the molecular composition and the structure of these sugars have been analyzed through mass spectrometry [85].

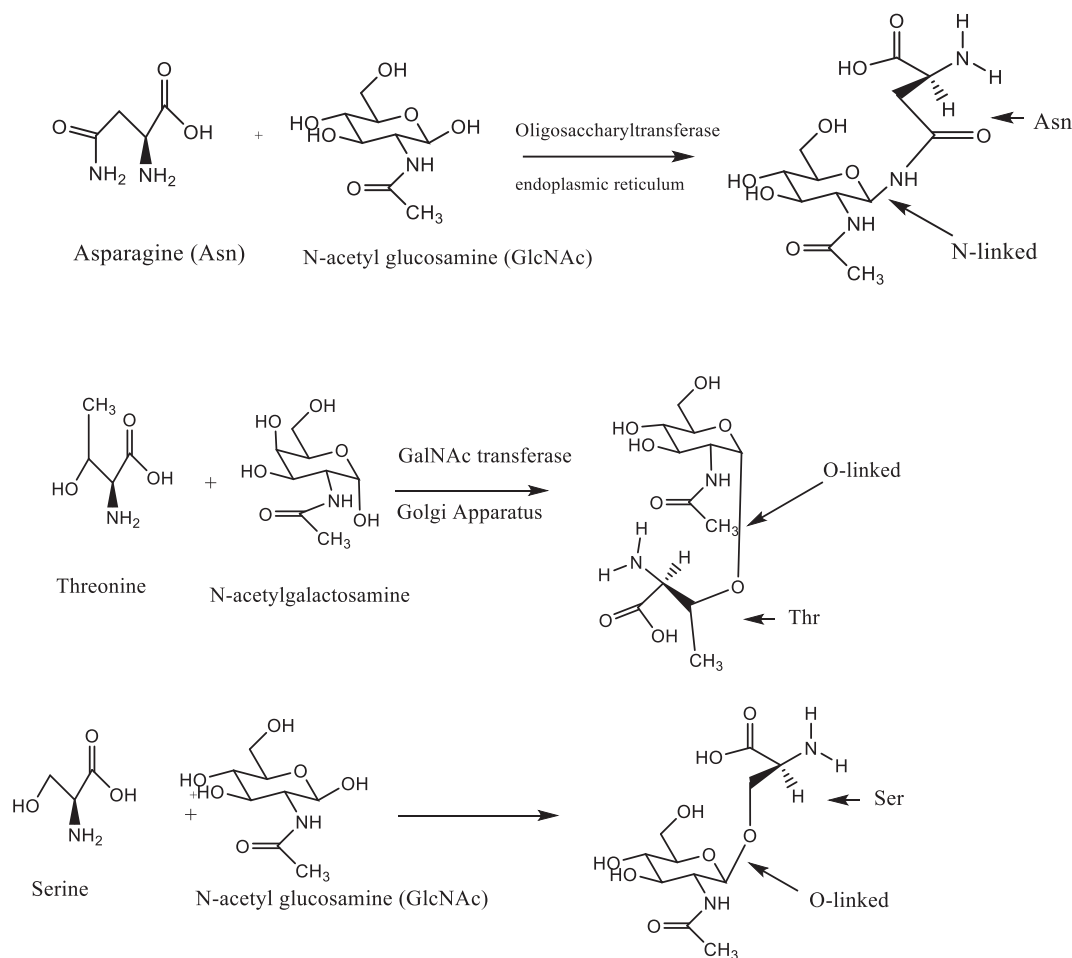
S-protein is heavily glycosylated, undergoing glycosylation on multiple asparagine (Asn) or serine (Ser) residues via N- and O-glycosylation, respectively. The formation of the N-glycan, which involves conjugation of N-acetylglucosamine to asparagine residues, is catalyzed by oligosaccharyltransferases present in the endoplasmic reticulum (ER) while the formation of O-glycans is mediated by O-GlcNAc transferase (OGT), likely in the ER or the Golgi apparatus [86] (Fig. 4).

After being conjugated to Asn or Ser, N-acetyl glucosamine (GlcNAc) is further glycosylated by the enzyme GTase to produce D-galactose-GlcNAc [87] (Figs. 5–6). After this initial reaction, more sugars are added by other glycosyltransferases, eventually leading to the formation of long chains of various polysaccharides.

The SARS-CoV-2 S-protein has 22 putative N-linked glycosylation sites and 4 possible O-linked glycosylation sites (Fig. 7A and B). Interestingly, mass spectral analysis shows differential glycosylation patterns in different cell types [88–91]. It appears that, depending on the tissue type, the overall level of glycosylation, as well as the relative ratios of N- to O-glycosylation, is different. In O-glycans, the sugar molecule is attached to the oxygen atom of serine (Ser) or threonine (Thr) of the protein (Figs. 5 & 6) after the protein is synthesized. O-glycosylation occurs in the ER, Golgi apparatus, and even the cytoplasm. In humans, the S1 protein has been shown to be O-glycosylated on Thr323, Ser325, Ser673, Thr678 and Ser686 [92]. The presence of O-glycosylation at Thr323 and Ser325 of the S1 subunit of SARS-COV-2 has been correlated with the active structure for the protein, which leads to the biological function [93]. These differences can have a bearing on vaccine production, as they can influence how the S-protein interacts with ACE2 receptor [94,95].

### 3.1.2. Changes in S-Protein due to mutations

Mutations are commonplace in viruses. As viruses replicate, mutations cause changes in the genomes of the newly synthesized viral particles, which carry the mutation(s) to hosts during subsequent infections. Only some mutations produce more aggressive, more deadly diseases. Changes in the genome causes changes in the viral proteins and since S-protein is the instrument of infection for SARS-COV-2, changes in the S-protein determines how easily virus can infect its host, or how rapidly infection spreads among populations. Similarly, mutations can impact how effective an antibody is to the S-protein. Antibodies bind to the N-terminal domain of the S-protein while the receptor binding domain (RBD) associates with the ACE-2 receptor at the surface of host cells. In the B.1.1.7 Variant that was first detected in the UK, nine mutations in the S-protein, including N501Y in the RBD and deletion of 3 amino acids in positions 69, 70, and 144 in the N-terminal domain. Tighter binding of the S-protein with ACE-2 due to these mutations makes the virus 50–70% more infectious. In the B.1.351 Variant, which first appeared in South Africa, of the nine amino acid changes, three were involved in N501Y as in the B.1.1.7 Variant, the other two were K417N (lysine to asparagine) and E484K (switching glutamic acid to lysine). These changes made the virus 50% more infectious than the original variant. In P.1 Variant initially detected in Brazil, 11 mutations occur in the S-protein, which include N501Y and E484K seen in other variants and K417T, which involves changing lysine to threonine. These mutations also increase the binding between the RBD with the ACE2 receptor on the host cells. In the B.1.427 and B.1.429 Variants, which were first detected in California, there are four mutations in the spike protein, including L452R in the RBD. This variant, which was about 20% more aggressive, is disappearing [96]. Delta and Kappa Variants first identified in India both have two mutations, E484Q and L452R. In addition, the Delta Variant contains a T478K mutation. These mutations in S1 increase the binding affinity to ACE2, causing higher transmissibility and virulence [44].



**Fig. 4.** The amine functional group of asparagine (Asn) functions as a nucleophile to promote N-linked glycosylation, in which a carbohydrate chain (e.g., N-acetylglucosamine (GlcNAc)) is added to the protein chain. In general, a carbohydrate chain is added to an Asn residue when it is flanked C-terminally by X-serine or X-threonine, where X is any amino acid other than proline. In the case of O-linked glycosylation, the hydroxyl group of threonine or serine acts as the nucleophile to promote conjugation of GlcNAc or N-acetyl galactosamine (GalNAc). See ref. [277].

### 3.1.3. E-protein

The SARS-CoV-2 envelope protein (E-protein) is a 10 kDa protein with an N-terminal transmembrane domain and a C-terminal cytoplasmic tail [97]. In SARS-CoV, E-protein mediates the budding and release of progeny viruses and also activates the host inflammasome. Studies have shown that the E-protein of SARS-CoV-2 is structurally and functionally similar to that of SARS-CoV and is highly conserved in their N-terminal regions [98,99]. E-protein oligomerizes to form viroporins, which are small membrane-embedded proteins present on different viruses with ion-conducting properties. They form cation selective channels across the ER-Golgi intermediate compartment to aid in the viral production, maturation and release processes, which are regulated by ion homeostasis of cellular organelles [99–101]. This protein has also been found to influence folding of other proteins, which has an effect on the survivability of the host cells [101–105].

### 3.1.4. M-protein

The most abundant structural protein of coronaviruses is the membrane (M)-glycoprotein. M-protein is a membrane bound glycoprotein with protease activity and three transmembrane domains. The M-protein can interact with viral RNA and bind to the other structural proteins, including S-, E- and N-proteins [106]. Binding with M-protein helps to stabilize N-proteins to promote viral assembly. In addition, M-protein also aids the S-protein during the attachment phase and may be involved in immune evasion and pathogenicity functions of the virus [107–109]. The C-terminus of the M-protein resides inside the cell with

the N-terminus on the outside. The third transmembrane domain, common to all coronaviruses, contains both polar and non-polar ends. The presence of an N-terminal Ser at position 70 distinguishes the M-protein of SARS-CoV-2 from M-proteins of other CoVs and may have allowed SARS-CoV-2 to jump from bats to humans [110]. An additional change is substitution of Ala for Thr at position 285 in SARS-CoV-2 M-protein, which may have increased the protease activity of the SARS-CoV-2 enzyme [84].

### 3.1.5. N-protein (nucleocapsid protein)

The N-protein has a molecular weight of 47 kDa and binds to gRNA. It has three domains, termed the N-terminal domain (NTD), central linker (CL), and C-tail. This protein is involved in the replication and transcription of viral RNA and formation of the ribonucleoprotein (RNP) complex [108]. It also influences the host cell cycle to enhance viral multiplication [111]. The CTD domain of SARS-CoV-2 N-protein, which is required for dimerization and oligomerization, interacts with the M-protein through the CL region [112].

The NTD binds to RNA and contains many positively-charged amino acids residues, which aid in binding the negatively-charged RNA. The amino acid sequence of SARS-CoV-2 N-protein is 90% similar to that of SARS-CoV [113]. Most of the differences are due to the presence of these positively-charged amino acids in the SARS-CoV-2, which are absent in other CoVs.



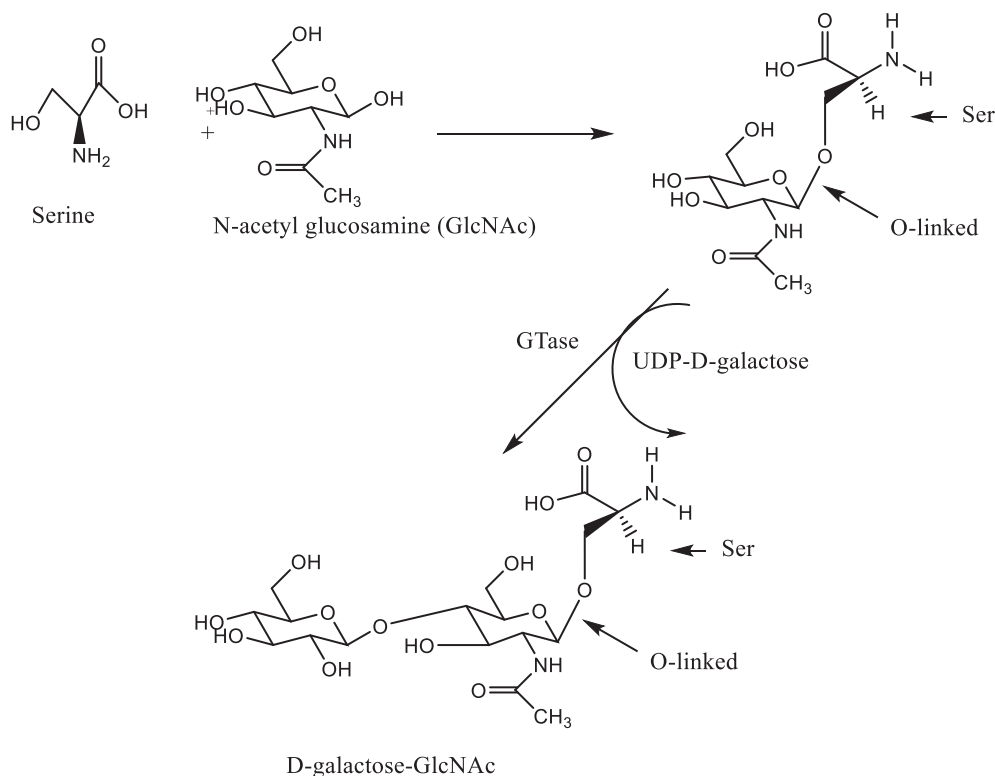


Fig. 5. Addition of D-galactose to GlcNAc-serine. A similar reaction may be observed with threonine-*N*-acetylgalactosamine.

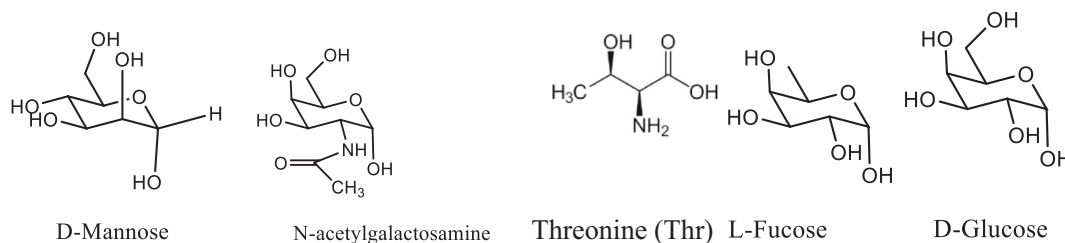


Fig. 6. Components of O-linked glycans.

### 3.2. Viral non-structural proteins

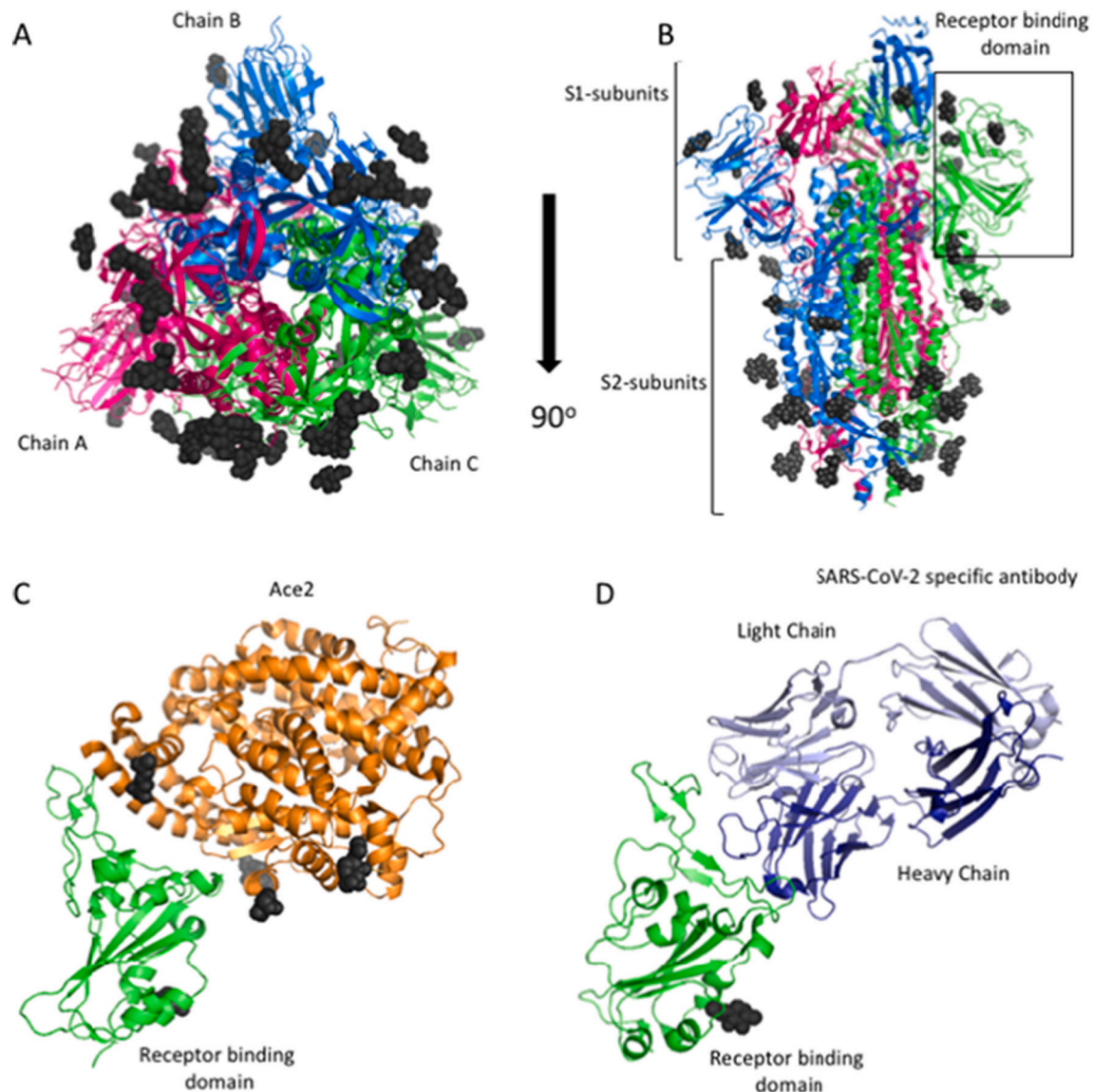
In addition to the structural proteins, there are also 16 non-structural proteins (Nsps) in SARS-CoV-2, including polyproteins, nucleoproteins, and membrane proteins, each with varying functions (Fig. 8). For instance, Nsp1 and Nsp2 are involved in modulating the immune response [114]. Nsp1 has 180 amino acids and binds to the small subunit of the ribosome, stopping synthesis of antiviral proteins. In addition, Nsp1 halts all host protein synthesis by cleaving host mRNA [47,115]. As a result, production of type-1 interferons (which provide innate immunity) is halted, disabling the host defense mechanisms against the virus. Nsp2, on the other hand, influences cell division by altering the host cell cycle (Fig. 3) [116].

Nsp3, or papain-like protease (PLpro), is the largest of all Nsps and plays many roles. Along with Nsp5, it is responsible for cleavage of viral polyproteins into functional units [117–118]. Nsp3 also plays a role in the onset of cytokine storm, assembly of viral particles, and, via interactions with Nsp4, in formation of double membrane vesicles [119]. Finally, Nsp3 both inhibits the host enzyme poly-[ADP-ribose] polymerase (PARP), whose role is to block viral replication, and suppresses the expression of interferon genes (required for inhibition of viral infections).

Nsp9 is a single-stranded RNA-binding protein that plays a role in viral replication. Nsp12 is an RNA-dependent RNA polymerase (RdRp) and nsp13 is a helicase [80,118]. Both are essential for RNA replication. Nsp12, along with Nsp7 and Nsp8, recognize the RNA template and facilitate processivity [120] (Fig. 8A [121]). In SARS-CoV-2, it has been shown that the hetero-dimer Nsp7–Nsp8 and individual Nsp8 can bind with the RdRp [82].

The nsp13 helicase contributes to RNA synthesis and 5'-RNA capping by unwinding double stranded RNA in a 5'-3' direction to enable elongation by the RdRp. The RdRp (Nsp12) from SARS-COV-2 has the structure of a typical polymerase, which resembles a right hand with finger, thumb, and palm domains [122]. When overlaid with the RdRp of SARS-CoV, there is only a root mean squared distance (rmsd) of 0.82 for 1078 of the C-alpha for SARS-COV-2, indicating that there are very few structural differences between the two proteins [82]. Despite these small differences, the amino acid sequence and the active sites of the RdRps from the two viruses are nearly identical [123]. This has enabled the rapid development of inhibitors against SARS-CoV-2 using the lessons learned from the previous inhibition studies on SARS-CoV.

Nsp14 is an exonuclease protein that acts as a methyl transferase for N7 guanine in mRNA cap synthesis. For activation, Nsp14 requires Nsp10 to form a heterodimer [124]. Nsp15 is a nidoviral RNA uridylylate



**Fig. 7.** Structures of the SARS-CoV-2 S-glycoprotein. A) Cartoon diagram of the 3D structure of the S-glycoprotein solved using cryo-EM to 3.46 Å (PDB ID: 6VSB). The S-glycoprotein forms a trimer composed of an S1-subunit (pink), an S2-subunit B (green) and an S3-subunit (blue). This structure represents the prefusion confirmation and was found to have 22 glycosylation sites (dark grey). B) 90° rotation of A. Here, the receptor binding domain (RBD) is located at the top of the structure. The RBD of the subunit C is boxed in black. C) Cartoon diagram of the RBD of the S-glycoprotein (green) bound to the host receptor ACE2 (orange) (PDB ID: 6M0J). This structure was solved to 2.45 Å using X-ray protein crystallography. D) Cartoon diagram of the SARS-CoV-2 RBD (green) bound to a SARS-CoV-2 specific antibody (the antibody heavy and light chains are colored dark blue and light blue, respectively). This structure was solved using X-ray crystallography to 2.85 Å (PDB ID: 7BWJ). It is important to note that the antibody binding site is at the same position as the ACE2 binding site, thereby blocking interactions between the virus and the host receptor. All figures were made using PyMol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

specific endoribonuclease (NendoU) that protects from attacks by the host immune system [125]. Meanwhile, Nsp16 is a 2'-O-methyltransferase, which uses SAM for its methyl source. It complexes with nsp10 but has two S-adenosyl-L-methionine (SAM) binding sites of its own. Binding between SAM and nsp16 is mediated by both hydrophobic and ionic interactions [126,127,123,128,129].

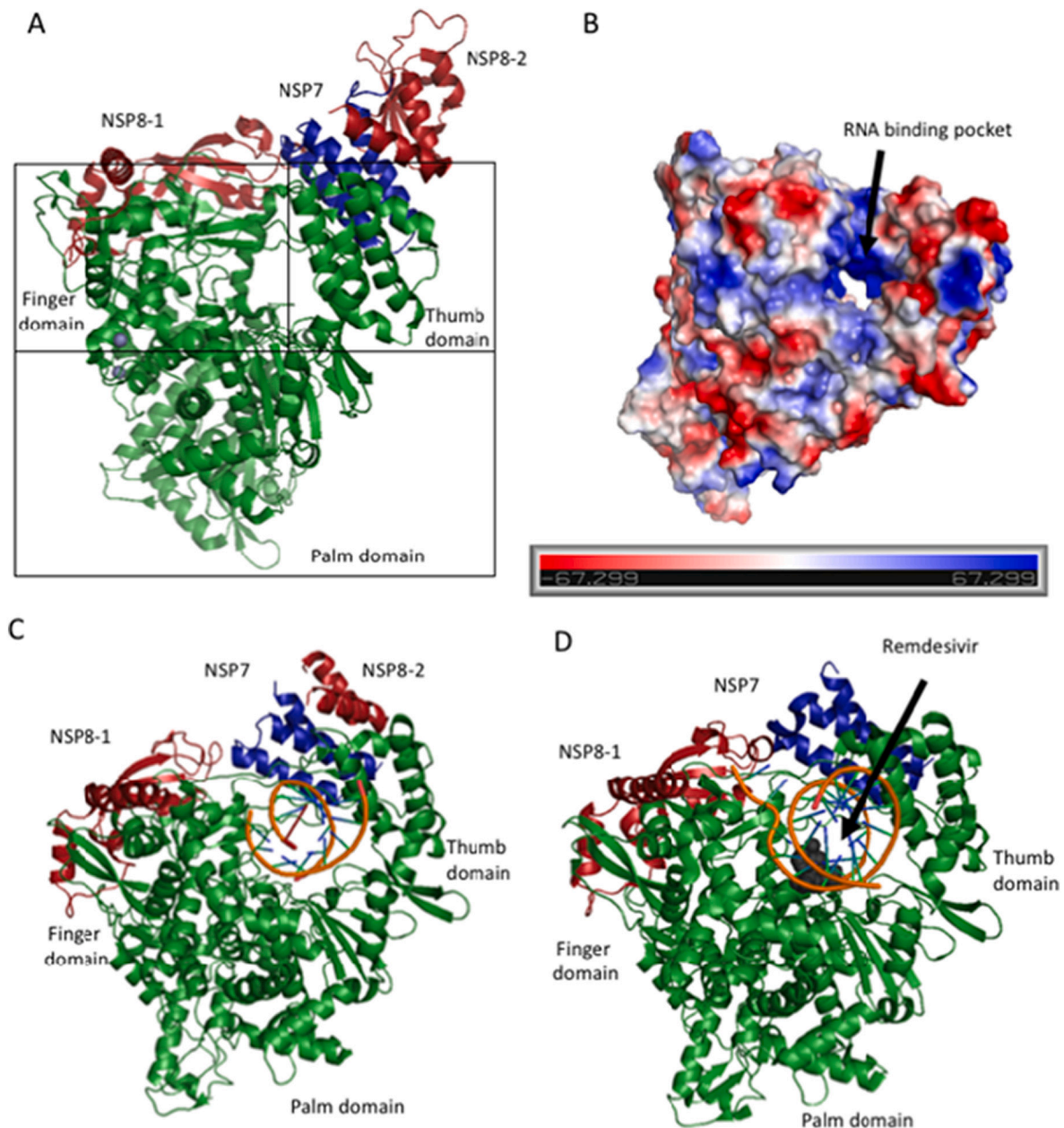
### 3.3. Accessory proteins

Accessory proteins are encoded primarily on the 3'-end of the viral RNA. They, like many of the other SARS-CoV-2 proteins, help to protect the virus against the host immune system. For instance, accessory proteins 3a and 7a are thought to be involved in induction of inflammation. The six accessory proteins, 3a, 6, 7a, 8a and 9b, are similar to those in

SARS-COV. The only exception is 8b, which has an additional 37 amino acids not observed in SARS-COV proteins [5,130–132]. Because the structure of most SARS-CoV-2 proteins is very similar to the structure of the proteins of coronaviruses studied previously [133], similar techniques can be applied to the new virus during the development of the best inhibitors for target enzymes [71,72,134].

### 3.4. Target receptors

As alluded to above, respiratory droplets and aerosols carrying the SARS-CoV-2 virus produced by an infected person target the host nasal goblet epithelial cells, type II pneumocytes, and enterocytes cells [64,135,136] via interactions between ACE2 and the viral S-protein, which is primed by transmembrane serine protease 2 (TMPRSS2)



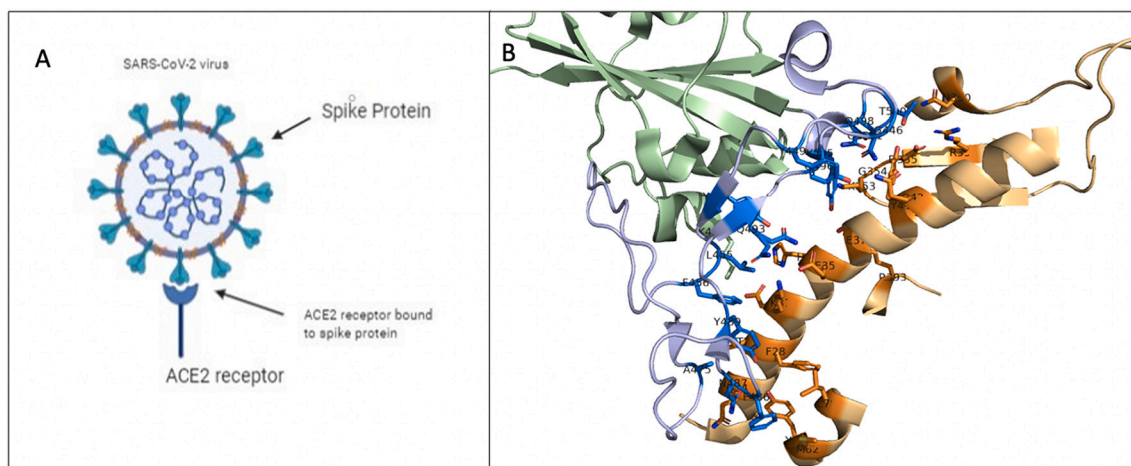
**Fig. 8.** The SARS-CoV-2 RNA dependent RNA polymerase (RdRp). The structure of the RdRp was solved to 2.95 Å using cryo-EM. A) Cartoon diagram of the RdRp (green) in complex with its required co-factors, NSP7 (blue), Nsp8-1 (light red) and Nsp8-2 (dark red). In addition, there are two molecules of Zn bound to the RdRp (PDB ID: 7BTF). The structure takes on a typical polymerase fold, with a thumb, palm and finger domain that cradles the RNA substrate during synthesis. B) Electrostatic surface diagram of Nsp12 alone. Here, the hole traversing between the domains is highly positively charged in order to provide a complementary binding site for the negatively charged RNA substrate. C) SARS-CoV-2 Nsp12/RNA complex solved using Cryo-EM to 3.53 Å (PDB ID: 6X2G). D) Cryo-EM structure solved to 2.5 Å of Nsp12, its cofactors, template RNA, and the drug Remdesivir. When Remdesivir is covalently incorporated into the RNA primer, it terminates chain elongation and therefore inhibits the polymerase activity of the enzyme (PDB ID: 7BV2). Figures made using PyMol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

located on the surface of the cells (Fig. 2) [36,137].

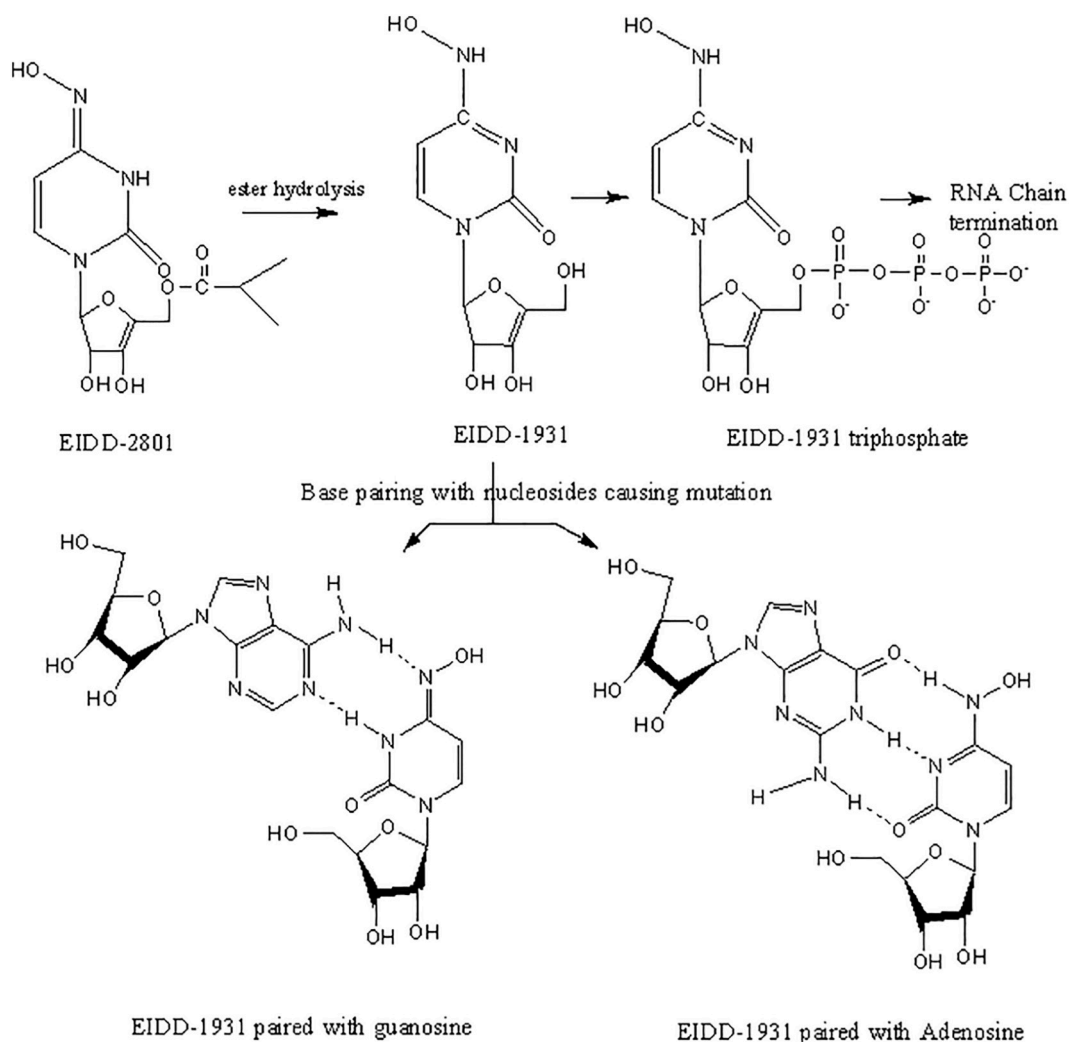
The receptor binding motif (RBM) is the specific region within the RBD of the S-protein that binds to the ACE2 receptor. The RBM is made up of residues 438–506 and is located in an extended insertion between  $\beta$ 4 and  $\beta$ 7. The concave surface of the RBD contacts the N-terminal helix of ACE2. There is significant sequence variation in this region of the S-protein compared to SARS-CoV, with 17 total residues of SARS-CoV-2 S-protein making contact with 20 residues of ACE2 (vs. 16 of SARS-CoV S-protein to 20 of ACE2). Of the 20 ACE2 residues making contact, 17 are shared between the proteins. These changes could account for the observed differences in binding affinity for SARS-CoV-2 (4.7 nM) and

SARS-CoV (31 nM), despite their overall structural similarities (rmsd of 1.3 Å) [108,138].

The TMPRSS2 protein, which is also expressed in the prostate, salivary glands, colon, and stomach, is a chymotrypsin family serine protease (492 aa) composed of three domains [64,139–140]. It mediates the cleavage of S-protein within the S1–S2 linker (R685) and at S2' (R815). The S1-S2 cleavage site is facilitated by the presence of several repeated basic arginine residues [64,137,141]. Additional help is provided to the virus by Furin, a host cell protease that cleaves at the polybasic furin cleavage site (e.g., with the PQRESRRKK/GLF sequence of amino acids) in S-protein, promoting cell fusion and entry of the virus into human



**Fig. 9.** Depiction of SARS-CoV-2 using its spike protein (S-protein) to bind to angiotensin converting enzyme 2 (ACE2) receptor to initiate infection. A) Schematic diagram depicting interactions between SARS-CoV-2 S-protein and ACE2. B) Co-crystal structure highlighting the binding interface between S-protein (purple) and ACE2 (orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 10.** The oxime form of EIDD-1931 mimics uridine, base pairing with adenosine (left), while the other tautomer mimics cytidine and base pairs with guanosine (right).

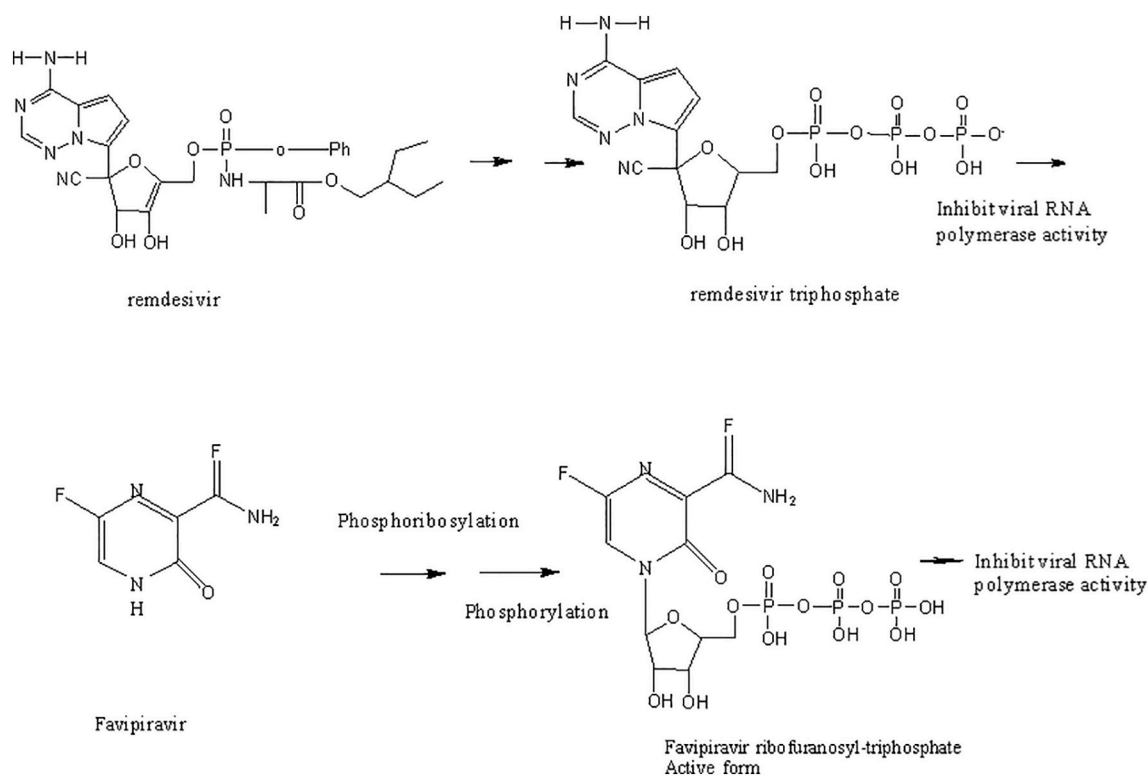


Fig. 11. Comparison of the mechanism of actions of antivirus Remdesivir, favipiravir, EIDD-2801, and EIDD-1931 through incorporation into DNA [227].

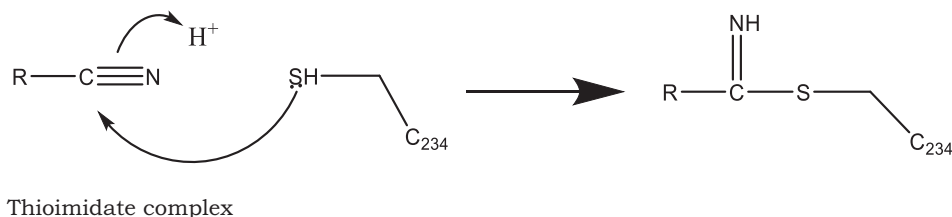


Fig. 12. Interaction of Cysteine 234 and nitrile leads to the formation of a thioimidate complex, which inhibits cathepsin C. See Ref. [230].

cells [64,79,135,136,141]. After cleavage at the S2' site, the fusion peptide is then inserted into the host membrane. The two HR regions, HR1 and HR2, in the S2 domain form antiparallel six-helix bundles (6-HB). Once inside the host cell, the virus can effectively replicate and initiate the process of infection and disease development [136,142]. The RBD from SARS-CoV-2 S-protein has higher binding affinity ( $\sim 5$  kcal/mol) towards human ACE2 than the corresponding RBD from SARS-CoV S-protein, which is higher than its affinity for ACE2 from other animals [64,143]. This higher affinity, according to contact map analysis, is due to greater electrostatic interactions between the virus's S-protein and ACE2 receptor [144]. Computational alanine scanning analysis has identified key residues responsible for binding of the three RBDs of SARS-CoV-2 with ACE2 [145]. Accordingly, SARS-CoV-2 uses almost all (90%) of the tyrosine (Tyr) and glycine (Gly) residues that are present at its ACE2 interface for binding. Molecular dynamics simulation on membrane-bound ACE2 supports the presence of about seven additional hydrogen-bonded contacts between the SARS-CoV-2 RBD and ACE2 [146]. Compared to human ATR1, preferential electrostatic interactions between ACE2 and the RBD of SARS-CoV-2 are due to 7 additional hydrogen bonds, which is equivalent to  $\sim 3.4$  kcal/mol of binding energy. According to mutagenesis analysis, residues 484 and 498 of the S-protein in SARS-CoV-2 are involved in recognizing human ACE2 [147].

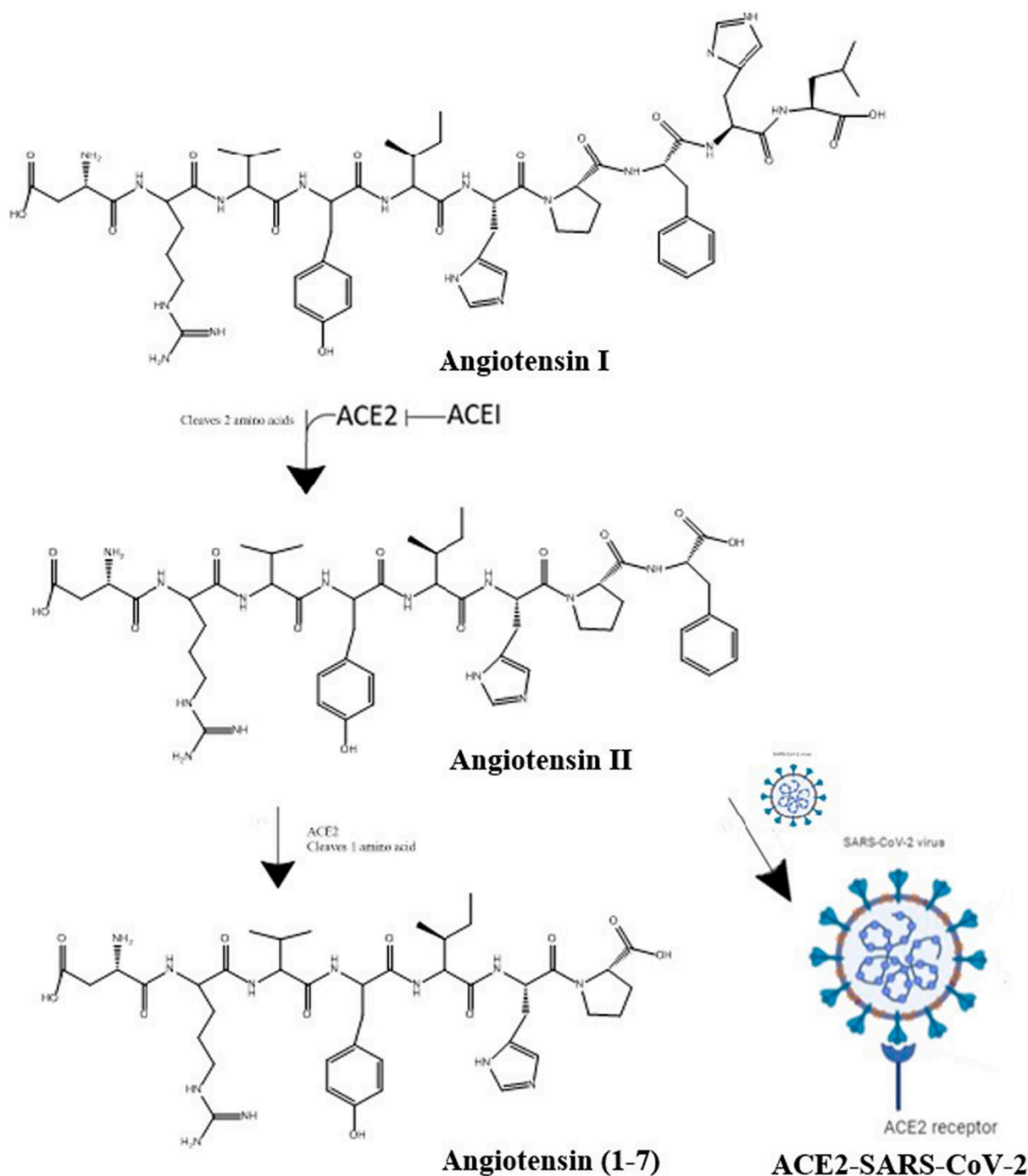
As stated above, in addition to the Type II alveolar cells of lungs, ACE2 and TMPRSS2 are also expressed in the outer layer of the nasal and

esophagus epithelium, the absorptive enterocytes from ileum and colon, and cornea [36,148,149]. In fact, single-cell RNA sequencing has shown high ACE2 mRNA expression in Type II alveolar cells of lungs, myocardial cells, esophagus upper and stratified epithelial cells, and digestive tract and kidneys [53,150–152]. The initial contact may also occur on the skin, throat, or in the gastrointestinal (GI) tract [36,153–157].

### 3.5. Comparing SARS-CoV-2, SARS-CoV, and MERS-CoV receptor binding

The SARS-CoV-2 genome is different from all previously identified coronaviruses, including SARS-CoV and MERS-CoV. These differences likely account for differences in cellular targets, as SARS-CoV-2 targets host nasal epithelium in addition to many other organs, in contrast SARS-CoV, which does not target the nasal epithelium but targets lung, trachea/bronchus, stomach, small intestine, distal convoluted renal tubule, sweat gland, parathyroid, pituitary, pancreas, adrenal gland, liver and cerebrum [158]. The SARS-CoV-2 S-protein exhibits even less homology with the MERS-CoV S-protein, which recognizes dipeptidyl peptidase-4 (DPP-4) as a specific receptor for host cell entry (as opposed to ACE2) [79,141,159,160].

Aside from molecular factors, other factors such as age, smoking, and gender may also influence the expression of ACE2 and TMPRSS2 in the

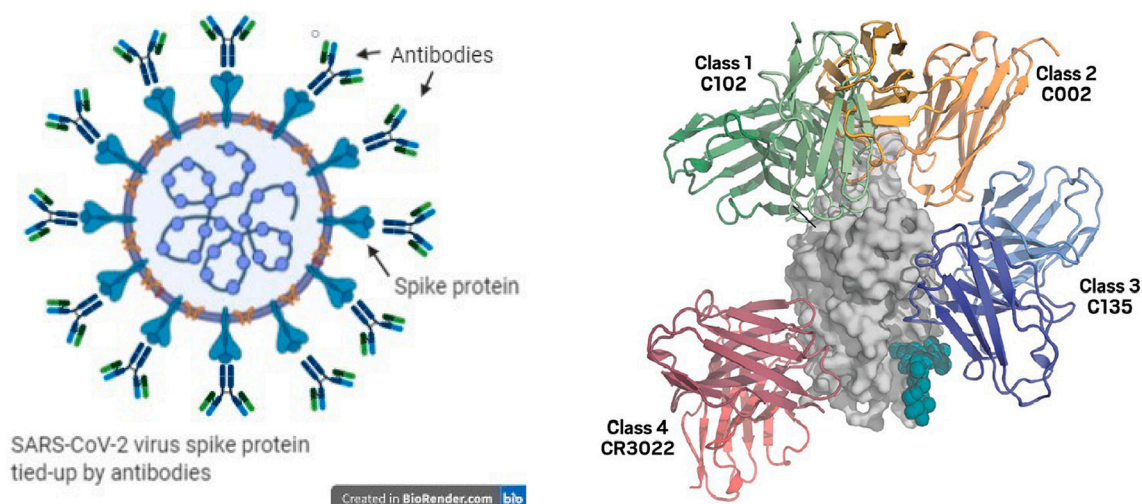


**Fig. 13.** Blood pressure is regulated by the renin-angiotensin system. To enter target cells, SARS-CoV-2 binds to ACE2, preventing it from breaking down angiotensin II into angiotensin 1–7.

organs and tissues and thus influence the severity and rate of infection of SARS-CoV-2 [137,161,162]. There is evidence that by directly infecting macrophages and T cells, SARS-CoV represses the body's immunity. But it is not yet clear if this is also the case for SARS-CoV-2, as macrophages and monocytes in the lung do not express a large number of the ACE2 receptors [163].

The SARS-CoV-2 genome sequence was first published in January 2020 and is 79.5% similar to the genome of SARS-CoV [66,80,164,165]. Both SARS-CoV and SARS-CoV-2 have S-proteins that are nearly 80% identical in composition [166–169]. Interestingly, the relatively small number of differences between the two S-proteins has a dramatic influence on parameters such as affinity between the S-protein and its

receptor (ACE2) [170,171]. The increased affinity of its S-protein for ACE2 allows SARS-CoV-2 to attach to human cells much more strongly than SARS-CoV. This characteristic of SARS-CoV-2 enhances both the contagiousness and infectivity of the virus [54,168]. Importantly, for infection, viral S-protein binds to host ACE2, therefore people with increased numbers of ACE2 sites in their lungs are more susceptible to the infection by this virus [172]. One additional characteristic that sets SARS-CoV-2 apart from other coronaviruses is the presence of a furin cleavage site within the S-protein. Furin is a host cell protease and has been found to be important for S-protein-mediated cell-cell fusion and entry into human lung cells [79,141].



**Fig. 14.** Depiction of how antibodies block the spike protein of SARS-CoV-2, preventing the virus from attaching its spike protein to ACE2 receptors to initiate infection (left) [252,254]. Four different ways that antibodies bind to receptor binding domain of spike protein from SARS-COV-2 virus (right). (From Ref [74]).

#### 4. Strategies to fight SARS-COV-2

Strategies to fight SARS-CoV-2 include vaccines, drug repurposing, novel antivirals, and therapeutic antibodies [171,173–176]. For effective development, all these strategies require knowledge of the viral genome and/or proteins, along with an understanding of the basic biology of its replication cycle. To this end, interdisciplinary teams of scientists use computer modeling, structure determination, metabolism/pharmacokinetics, biochemical assays, and synthetic chemistry protocols to design and optimize effective therapeutics. As a result, some drugs have been designed to inhibit capsid formation through formation of van der Waals forces, hydrogen bonds, and electrostatic interactions with the capsid directly. Meanwhile, others change the forces that hold the capsid together, disrupting their ability to attach and enter host cells. Many of the non-structural proteins, including the S-protein, M-protein, and RdRp, are also targeted by small molecule inhibitors, which bind to the enzyme active site. In some cases, these drugs are already available. For instance, several *in silico* screening efforts have shown similarities between SARS-CoV-2 and other non-coronaviruses, enabling the use of a number of available antivirals and drugs that are currently on the market [177–180].

##### 4.1. Spike protein as target for inhibitors (potential drugs)

Much of the current efforts in combating SARS-CoV-2 is focused on finding ways to block the interactions between ACE2 and the viral S-protein, sometimes through specialized maneuvering that stabilizes the spike [181]. The difficulty in doing this is that ACE2 is also present on the cells in other parts of the body that are not targets of the virus. Therefore, non-tissue specific inhibition of ACE2 could lead to unwanted side effects [182–183]. ACE2's physiological role in the cell is in the maturation of angiotensin, which is a hormone that controls blood pressure by regulating blood vessel constriction. Consequently, the inhibition of ACE2 can lead to several complications, including cardiovascular disease and hypertension [53]. As a result, alternative strategies are being explored to block the interaction between the SARS-CoV-2 S-protein and ACE2 directly using therapeutic antibodies. The ideal therapeutic antibody would specifically bind the S-protein and prevent it from interacting with the ACE2 receptor, achieving a similar result as blocking ACE2 directly without the negative side effects [77]. Alternatively, designing small peptides or small molecules that bind to different subunits of the SARS-CoV-2 S-protein will prevent the subunits from interacting with ACE2 and viral entry into the cell. As previously

stated, the S-protein is initially cleaved into S1 and S2 subunits by the host cell protease, furin. S1 facilitates receptor binding by engaging with the host cell membrane while S2 provides structural support and regulates the fusion with the host's epithelial cell membrane. For this reason, blocking the cleavage of the S-protein could also be a potential strategy for inhibiting viral entry [8,54,77,184–186].

Using knowledge gained from HKU1, a common cold virus, and MERS virus, scientists found two small loops of amino acids that held two  $\alpha$ -helices in the spike together. When the S-protein binds to human cells, the coil-like structure is released and the two helices and one loop are elongated to bring the human cell and the virus close together for fusion. By adding up to six proline residues to the SARS-CoV-2 S-protein, scientists believe that they can prevent the “spring” from being released, which would prevent the fusion of SARS-CoV-2 with human cells [181].

##### 4.2. Chemistry of selected antiviral compounds

Due to the relatively small size of the SARS-CoV-2 genome and recent rapid advances in whole genome sequencing technologies, scientists were able to sequence the genome of the SARS-CoV-2 rather quickly once it was isolated [80,165]. As a result, the sequence has been used to determine the primary structure (i.e., linear amino acid sequence) of the viral proteins. Once a protein's primary sequence has been determined, computational modeling programs can be used to deduce its secondary and tertiary structures. These programs are based largely on homology modeling (which uses the structures of closely related proteins to model the structure of the protein-of-interest). This is either in lieu of 3D structure determination, as discussed previously, or as a complement to confirm what was observed *in vitro*. Once a putative structure is determined, different therapies can be designed to target the protein's active site and/or other protein surfaces important for enzyme function [118]. Of course, similar strategies can be used once a bona fide structure of the enzyme-of-interest is solved experimentally using structural biology methods, such as X-ray crystallography, NMR spectroscopy or, more recently, cryo-EM. The structural information can, in turn, be used to develop drugs that target those proteins involved in infection, viral replication, or transmission of the disease. In the process, computer models of the viral proteins are constructed and their interactions with human proteins are studied at the molecular level [74,187]. At the same time, high-throughput screening may be performed for the compounds that look promising based on *in silico* docking studies [83,179].

At the onset of the COVID-19 pandemic, a list of 50,000 compounds with potential antiviral properties was provided by the American

Chemical Society (ACS) Central Science. The compounds were extracted from the CAS registry, which contains 160 million compounds [188]. The identified compounds were anti-infective, enzyme inhibitors, or have shown an effect on the respiratory system [189].

Recently, the cryo-EM structure of the RdRp was determined (Fig. 8) [74,82,190]. The structural information led to computational simulations (in silico) of the interactions between RdRp and potential drug candidates. Similarly, knowledge of the structure of the S-protein and main protease (M-protein or M<sup>pro</sup>) has enabled similar studies [82,170,191,192]. To make these studies possible, computational resources around the world have been made readily available to COVID-19 researchers by major international computer centers inside and outside the U.S. during the COVID-19 pandemic [193].

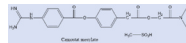
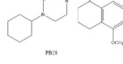
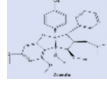
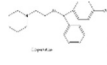
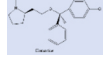
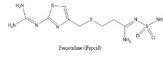

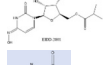

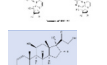

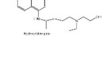


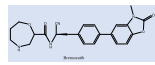
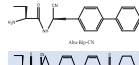
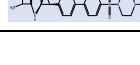
In addition to in silico analysis, animal models have been used to model infection, including mice that express the necessary receptors and succumb to similar symptoms as humans [189,194]. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. This affords the opportunity to devise appropriate treatments through docking studies [83,135].

Researchers at Novartis had quickly discovered that the M<sup>pro</sup> from SARS-CoV-2 (cysteine-main protease) was very similar to the protease from the previous corona viruses. This includes the active site cysteine, which serves as a target for identifying inhibitors. As a result, sixteen potential inhibitors were identified using computational methods [195–197]. A select number of identified inhibitors were synthesized and tested in models. Chemical structures were then further refined based on the results of the inhibition studies. A new smaller set of inhibitors were then obtained and some of the final compounds have entered randomized clinical trials [198,199]. Computational studies have suggested that lopinavir, oseltamivir, and ritonavir may bind to SARS-CoV-2 M<sup>pro</sup> [198,199]. A similar approach could be used to develop inhibitors for all other viral proteins, including RdRp and the S-protein. More recently, it was discovered that drug repurposing studies for COVID-19 provided compounds that were harmful to the body as their mechanism of action was misunderstood in the initial studies [200]. As a result, some of such drugs like hydroxychloroquine, Amiodarone, and Sertraline, cause phospholipidosis in the infected host cells which mistakenly in the studies was correlated with antiviral efficacy.

As previously discussed, RdRp is responsible for replication of the viral RNA (Fig. 8) [201]. Remdesivir (See Table 2) (an inhibitor of this enzyme), was originally developed as anti-Ebola drug [202–205], and soon entered the Phase I clinical trial against COVID-19 [205,206]. In addition, remdesivir was tested in combination with chloroquine and the combination compared with chloroquine alone [194,207,208]. Remdesivir in combination with interferon beta was also shown to be effective against SARS-CoV-2 [208,209]. Based on clinical trial results, and the compassionate use of this drug [18,203,210], remdesivir received Emergency Use Authorization by the FDA [205,211–214]. However, similar studies conducted in China suggested that remdesivir did not offer any benefit compared to placebo [205,215–217]. Considering all the results, clinicians think of remdesivir as a helpful treatment for shortening recovery time but not a cure [215,218]. Phase III Clinical trials of RdRp inhibitors, are currently underway. Remdesivir was recently approved by the FDA [205,206,213].

Two other inhibitors of viral RdRp's, favipiravir and the ribonucleoside analog, EIDD-2801 (Molnupiravir, See Table 2), were originally developed against HIV and hepatitis B [219–221]. In animals infected with SARS-CoV and MERS-CoV, EIDD significantly reduced viral titer and loss of body weight. Also, in vitro studies using primary human airway epithelial cells (HAECs) suggested that EIDD reduced SARS-CoV-2 reproduction in a dose-dependent manner [189,222]. Remdesivir (administered through IV) and EIDD-2801 (administered orally) both hamper viral replication by targeting RdRp [223]. Interestingly, EIDD, unlike other antiviral compounds, cannot be overcome by resistance through mutations in the virus. EIDD-2801 has a pyrimidine as its base structure altered with a N-hydroxycytidine (Fig. 10). As a result, the

**Table 2**List of compounds tested against SARS-CoV-2 and their chemical structures.<sup>a</sup>

Name of the compound	Chemical structure	Function	Reference
Camostat mesylate		Protease inhibitor	[137]
PB28		Replication of SARS-CoV-2 by inhibiting the translation of viral RNA, or by modulation cell signaling	[232,233]
Zotatifin		Replication of SARS-CoV-2 by inhibiting the translation of viral RNA, or by modulation cell signaling	[232,233]
Cloperastine		Antihistamines showed good antiviral properties	[233]
Clemastine		Antihistamines showed good antiviral properties	[233]
Famotidine (Pepcid)		Targets histamine H2 receptors which are abundant in all parts of the body release of histamine by the mast cells is blocked and the overreaction of immune system dampened	[234] [235]
Remdesivir		Inhibitors of viral RdRp's	
EIDD-2801 (Molnupiravir)		Inhibitors of viral RdRp's	[219–221]
Favipiravir		Inhibitors of viral RdRp's	[219–221]
Tautomers of EIDD-1931		Antiviral against RNA viruses	[189]
Dexamethasone		Suppresses the overreaction of the immune system	[236]
Hydroxychloroquine		Anti-malarial drug. Inhibits viral replication in vitro Now not recommended for COVID-19 treatment	[237–240]
Baricitinib		Anti-inflammation rheumatoid arthritis drug Could prevent the cytokine storm	[241–243]
Ruxolitinib		A cancer drug Could prevent the cytokine storm	[241–243]
Brensocatib			[231]
Abu-Bip-CN			[231]
Icat <sub>C<sub>xp</sub>z41</sub>			[231]



<sup>a</sup> Structures created with ChemDraw.

RdRp incorporates EIDD-2801 into its genome. However, since EIDD-2801 can tautomerize, the RdRp erroneously reads EIDD-2801 as uridine instead of cytidine when it replicates. Incorporation of adenosine instead of guanosine is a mutation that becomes part of the viral genome and affects all future generations of the virus by dictating how viral genomes are synthesized and copied. Since all RNA-based viruses have RdRp enzyme, this drug should work on SARS-CoV-2 and other similar viruses [218].

The drug EIDD-1931 has two tautomers (Fig. 10). One tautomer is an oxime that mimics uridine and, upon incorporation into the genome (Fig. 11), base pairs with adenosine, while the other tautomer mimics cytidine and base pairs with guanosine [219]. As a result, the RdRp incorporates EIDD-2801 into the SARS-CoV-2 genome. However, since EIDD-2801 can tautomerize, the RdRp erroneously reads EIDD-2801 as uridine instead of cytidine when it replicates. Incorporation of adenosine instead of guanosine is a mutation that becomes part of the viral genome and affects all future generations of the virus by dictating how viral genomes are synthesized and copied. Since all RNA-based viruses have RdRp enzyme, this drug should work on SARS-CoV-2 and other similar viruses [224]. However, in practice, initially it did not work, when monkeys were dosed with this drug in their GI, the CH<sub>2</sub>OH group in the molecule of EIDD-2801 was phosphorylated and the phosphorylated drug was readily excreted without being absorbed to work against the virus. To prevent this, the CH<sub>2</sub>OH group on the sugar moiety of the molecule was first converted to 2-methyl propionate (prodrug formation) before dosing. This protected the CH<sub>2</sub>OH group against pre-mature phosphorylation. The 2-methyl propionate moiety of the prodrug is broken back into the CH<sub>2</sub>OH upon absorption and then along with two other hydroxyl groups is phosphorylated to make the active drug. EIDD-2801 is developed in the United Kingdom (UK) by Merck. Phase I clinical trials of EIDD-2801 began on April 10, 2020 in the UK [225]. EIDD-2801 is recommended to be given to patients as soon as infection is detected. Because of its mechanism of action, this drug is likely to work against many RNA viruses, including MERS-CoV, HIV, Ebola, H1N1, and SARS-CoV. However, since other compounds that are similar to EIDD-2801 have been linked to birth defects in animals, expecting mothers or those who intend to become pregnant should avoid this drug [226]. The mechanism of incorporation of these drugs into RNA is shown in Fig. 11 [227].

Other potential drugs under study against COVID-19 include interferon, ribavirin, tocilizumab, sarilumab, lopinavir, and chloroquine. To date, no benefit has been observed with lopinavir-ritonavir treatment of patients [118,228,229].

As additional potentially useful drug target, protein kinases also have a crucial role in entry, spread, and replication of the virus. AK-1 inhibitors, caffeic acid and its ester, propolis, ketorolac, and triptolide may help patients by inhibition of these crucial stages of virus life cycle. However, these compounds are not easily available to the body due to their solubility problems [176,230]. Cathepsin C inhibition has also been pursued as an option to save overburdened lungs in serious cases [231]. Structures of three such inhibitors, Brensocatib, Abu-Bip-CN, IcatC<sub>xpz41</sub> are included in Table 2. In terms of mechanism, in the inhibitor compound containing nitrile group, the nitrile group is said to interact with the cysteine 234 of Cathepsin C, resulting in a thioimidate complex and inhibition (Fig. 12) [231].

#### 4.3. Blood pressure drugs

Blood pressure drugs are thought to increase the expression of ACE2 receptors (though this idea has been disputed by more recent studies) [244–246]. Therefore, initially, it was thought that people taking these drugs would have an increased number of ACE2 receptors on their cell surfaces, thus providing more sites for the virus to attach and making the

patient more susceptible to infection. However, more recent evidence suggests that blood pressure drugs actually protect patients against the induced lung injury, as described in more detail below [18,244,247,248].

Since COVID-19 is more serious disease in older patients, they are also more likely to take ACE inhibitors (ACEIs) and angiotensin II receptor blockers (ARBs) for hypertension, heart disease, and cardiovascular disease. These drugs regulate the renin-angiotensin equilibrium. ACEIs prevent ACE1 from converting angiotensin I to angiotensin II. Angiotensin II controls lung injury and severe inflammation, preventing angiotensin II from binding to its receptors. To lower blood pressure, these drugs keep the level of angiotensin II at a desired level by effecting the renin-angiotensin equilibrium (Fig. 13) [244,247]. In the lungs, heart, kidneys, nose, and GI tract, ACE2 helps regulate the angiotensin II levels by shifting this equilibrium to the right or the left, as needed. For instance, when angiotensin II levels are too high, ACE2 converts angiotensin II to angiotensin 1–7, which is inactive. When angiotensin II levels are too low, more angiotensin II is produced [53]. If the level of angiotensin II gets too high, fibrous tissue will form in the heart, kidney, and lungs, which can lead to severe fibrosis in these organs. Antihypertensive drugs like captopril, Lisinopril, and losartan either stop the production of angiotensin II or make it inactive. Towards the beginning of COVID-19 pandemic, many scientists warned against using ACEIs and ARBs in COVID-19 patients [249], [even though a prolonged absence of these drugs could lead to serious health problems or even death in the patients by increasing their hypertension]. Subsequently, it was discovered that using these drugs does not, in fact, harm COVID-19 patients [244]. In fact, they may actually help. This is because, during COVID-19 infection, the virus binds to host ACE2 and occupies host ACE2 sites. As a result, the ACE2 needed to convert angiotensin II to angiotensin 1–7 is no longer available, causing angiotensin II levels to go up. The increase in angiotensin II levels causes cell death and injury in lung and heart tissue. The injury in lung causes the release of cytokines (small proteins), which enhance the inflammation reaction and leads to more cell death [33,35]. Fibroblast's form scar tissues and pneumonia ensues, filling the lung with fluid. The blood pressure drugs stop all these negative consequences of angiotensin build-up by increasing the number of ACE2 sites, which help the patient by converting excess angiotensin II to angiotensin 1–7 through the removal of one amino acid.

In a study on mice infected with SARS-CoV-2 virus, levels of angiotensin II increased while levels of ACE2 decreased. Giving the blood pressure drug, losartan, to the animals reduced injury. The effect of blood pressure drugs is under study about in 30 clinical trials. In related studies, to study the effect of ACE2 on viral infection, by making changes in the translation, ACE2 levels are lowered [250].

#### 4.4. Antibody-based therapies

Antibodies are proteins of the immune system that bind to a specific antigen (such as virus before it infects a host cell). The process involves both the innate and adaptive immune responses [171]. Antibodies bind to the antigens on the surface of infected cells (see Fig. 14) and mark them for destruction by killer T cells. Among the SARS-CoV-2 and SARS-CoV proteins, the S-protein elicits an antibody response, as it resides on the external surface of the virus, while the structural proteins, M-, E-, and N-protein (in SARS-CoV) do not. As a result, over 90% of antibodies are directed against S-protein [188,251]. Antibodies are made of a constant region (which interacts with other host immune proteins) and a variable region (that targets the specific antigen). The variable region is constructed to match the shape and chemical properties of their target antigen with high specificity. Construction of the variable region is generated through a process called VDJ recombination. One type of specific antibodies, “neutralizing antibodies” can be used as therapies where they capture antigens like bacteria and viruses to prevent them from infecting our cells. Neutralizing antibodies for therapy have been developed against COVID-19 [252–254].

Due to similarity of viruses, the antibody raised against the SARS-CoV S-protein can also prevent infection by binding S-proteins from SARS-CoV-2 and MERS-CoV [254–257,251]. However, because the S-protein in SARS-CoV has fewer sugar moieties on its surface, antibody binding is weaker to the SARS-CoV-2 S-protein. As a result, for treatment purposes, antibodies are isolated from convalescent plasma collected from patients who recently recovered from COVID-19. Currently, scientists are working to identify the gene in the B cells that code for the SARS-CoV-2 antibody to then recombinantly express those genes as therapeutics. This approach has already worked in the case of SARS [26,188].

In another approach used by Regeneron, genetically modified mice were injected with the S-protein. B cells were then isolated from the mouse and among the B cells, those that produced the best monoclonal antibody were isolated, compared with best antibody from patients, and then mass produced through cell culture [251]. Palivizumab, a monoclonal antibody produced in this manner, is effective against respiratory syncytial virus. With financial support from the U.S. government, using VelociMouse technology, a mouse genetically modified to have human immune system, Regeneron developed monoclonal antibody against the S-protein of SARS-CoV-2, which received EUP from FDA approval [74,258]. Tychan, a company from Singapore, started its clinical trial in China. Working with AbCellera, Eli Lilly also developed its monoclonal antibody (Ly-CoV555), a neutralizing antibody, designed to prevent SARS-CoV-2 from entering the cells. The antibody targets the S-protein of the virus and thus prevents infection by clearing the virus from the body. One out of three doses were given to people with mild symptoms of COVID-19 or placebo. In these studies, the rate of hospitalization in those patients who received the real dose was reduced to 1.7% compared to 6% for the placebo [259]. Another antibody against S-protein (JS016) is a human monoclonal neutralizing antibody developed by Junshi Biosciences. When combined with Ly-CoV555, this product makes a strong antibody [260]. Some of the difficulty for these techniques lies in scaling production. As a result, with the immediate threat of these viruses, it is important to note that these companies have been able to cut the time of the development and evaluation of these therapies from 10 to 12 months to possibly 5–6 months [261].

#### 4.5. SARS-CoV-2 vaccines

Vaccines trick our bodies to perceive the vaccine as an antigen, causing our immune system to mount an attack to destroy the antigen [262]. The body's defense is presented in the form of antibodies which are proteins created to match the structure of the antigen in addition to secondary defensive mechanisms involving T-cells. Close to 200 vaccine development programs have been working to create vaccines against SARS-CoV-2 virus. Some, like Moderna and Pfizer, use mRNA as vaccine to generate antibodies against the S-protein inside the human cells [263]. Other strategies use the S-protein or the inactivated virus itself as the vaccine. Vaccination of the populations against SARS-CoV-2 has been successful in preventing the infection and death in many countries. However, some people refuse the vaccines while many do not have access to the vaccines. Initially, the major challenge for the vaccination, especially in the U.S., was the creation and production of the vaccines. However, after vaccination of about 50% of the population, the challenge has shifted to convincing the unvaccinated people to accept the vaccine. In the meantime, the virus has sufficient hosts to mutate to new variants including alpha, beta, gamma, and delta. The emergence of newer variants resistant to current vaccines may require the development and production of booster vaccines [264].

#### 5. Summary and conclusions

The fight against SARS-CoV-2 has been a major challenge for our best scientists, massive investments of pharmaceutical companies and governments, and our available technology. In a very short period of time,

significant advances were made in finding ways to lessen the spread of the virus and to prevent and treat the viral infection through therapeutics and vaccines. Understanding the mechanism of interaction between the viral S-protein and ACE2, along with the structure and function of the RdRp, M<sup>Pro</sup>, viral RNA, and various proteases, have been instrumental in discovering novel treatment methods [171,265]. We have seen that biochemically, SARS-CoV-2 shows higher affinity towards the ACE2 receptor compared to SAR-CoV. Despite these differences, fortunately there are also many similarities between these two viruses, which has facilitated the discovery and development of antibody therapies and vaccines to help control the COVID-19 pandemic [256]. Vaccine development used many different techniques, some very new, but all focused on inducing the production of neutralizing antibodies against the spike protein [266–269]. In addition to developing vaccines, scientists also explored other potential strategies that have not been discussed here. They include targeting the viral proteases, therapies based on cytokines, and nucleic acids. Knowledge of these and other relevant technologies is essential, not only for fighting SARS-CoV-2 today, but also to combat future viruses and pathogens.

Despite development of so many prevention and treatment tools in a short time period, mutation in SARS-CoV-2 RNA may lead to drastic changes in the viral genome and render all or some of these strategies useless [5,16]. Fighting a mutating virus is possible, but as experience has shown, may be challenging. Thus, the most effective way of combating the virus is to vaccinate the whole population and follow safety guidelines, such as social distancing, wearing masks, sanitization, and contact tracing. However, experience has shown that for political or other reasons, these are not fully achievable [270–276].

Just as in the case of the Spanish flu of 1918, SARS-CoV-2 will continue to dominate the news for years to come. In fact, it is expected that this virus, like other viruses such as flu virus, will remain within the human population for the foreseeable future. As younger people acquire immunity, older adults will remain susceptible until herd immunity is established within the population. The potential for childhood vaccines for long-term protection, as in the case of other childhood vaccines, is still uncertain as we continue to understand the immune response of all segments of the population including children. In the meantime, new experimental treatments, along with social distancing and proper hygiene will be necessary in conjunction with vaccines to protect susceptible populations. In the process, cooperation of national and international leaders is important for fighting the current pandemic.

#### Abbreviations

3CLpro	3C-like protease or main protease
ACE2	angiotensin-converting enzyme II
Ang	angiotensin
ARDS	acute respiratory distress syndrome
CatC	cathepsin C
CCR2	CC chemokine receptor 2
CCL2	CC chemokine ligand 2
CCR5	CC chemokine receptor 5
COVID-19	coronavirus disease 2019
CPG	CpG is shorthand for 5'—C—phosphate—G—3' [cytosine and guanine separated by one phosphate group]
E protein	envelope protein
ELISA	enzyme-linked immunosorbent assay
FP	fusion peptide
HA	hemagglutinin
HCoV	human coronavirus
HCV	hepatic C virus
HIV	human immune deficiency virus
HTS	high-throughput screen
IFA	immunofluorescence assay
IFN-β	interferon β
M protein	membrane protein

















- plant restarts, and more, ACS Chem. Health Saf. 27 (4) (2020) 204–208, <https://doi.org/10.1021/acs.chas.0c00069>.
- [276] C.A. Bettenhausen, A chemist's guide to disinfectants, C&EN Glob. Enterp. 98 (20) (2020) 24–25, <https://doi.org/10.1021/cen-09820-feature3>.
- [277] Protein glycosylation in: Lodish H, Berk A, Zipursky SL, et al. New York: W. H. Freeman; 2000. found at : <https://www.ncbi.nlm.nih.gov/books/NBK21744/>.