

Available online at [GSC Online Press Directory](https://www.gsconlinepress.com/)

GSC Biological and Pharmaceutical Sciences

e-ISSN: 2581-3250, CODEN (USA): GBPSC2

Journal homepage: <https://www.gsconlinepress.com/journals/gscbps>

(REVIEW ARTICLE)



## Role of structural and functional proteins of SARS -COV-2

Endriyas Kelta Wabalo \*, Abebe Dukessa Dubiwak, Tariku Sime Gizaw and Urge Gerema Kotu

*Department of Biomedical sciences, Faculty of Medical Sciences, Institute of Health, Jimma University, Jimma, Ethiopia.*

Publication history: Received on 23 August 2020; revised on 19 September 2020; accepted on 23 September 2020

Article DOI: <https://doi.org/10.30574/gscbps.2020.12.3.0275>

### Abstract

Severe acute respiratory corona virus-2 (SARS-CoV-2) is a ribonucleic acid (RNA) virus with enveloped no-segmented positive sense belonging to a beta ( $\beta$ ) - corona virus family. It has 29,903 nucleotides sized genome with 10 open reading frames (ORF). ORF1 (ab) encodes two polypeptides pp1a and pp1b cleaved into 16 functional proteins, which are mainly intended to form replication transcription complex (RTC). The cleavage process of pp1a and pp1b polypeptides to 16 functional proteins of SARs-CoV-2 is mainly facilitated by main protease and papain-like protease. The replication transcription complex (RTC) formed by the action of 16-functional proteins of SARs-CoV-2 is mainly involved as viral RNA synthesis machinery in the transcriptional and replication process of viral RNAs. ORF (2-10) encodes for structural (for example: spike (S), membrane (M), nucleocapsid (N), and envelop (E)) and accessory proteins of SARs-CoV-2. The main functions of structural proteins are viral assembly, viral coating, viral entry into host cells and assembly of the RNA genome. Accessory proteins are proteins that are not involved in the viral synthesis machinery, as 16 functional proteins, and in the viral assembly, coating, entry into host cells and packaging of Viral RNAs, as structural proteins. Rather, these are proteins that may play central role by enhancing viral assembly process, virulence and pathogenesis of SARs-CoV-2. Our aim in the current review was to elaborate the specific role of these structural and functional proteins on viral genomic replication and transcription, viral assembly, host cell attachment and pathogenesis. Multiple literatures have been reviewed to achieve the objective of this review.

**Keywords:** SARS COV-2; Structural Proteins; Functional Proteins; Open Reading Frames; Therapeutic Target

### 1. Introduction

Corona viruses are viruses that found in the order, Nidovirales, and family, Corona viridae. They were first discovered in the 1960s; the earliest one's infected chickens: infectious bronchitis virus. The name "Corona" is a Latin name meaning "crown-like" protein spikes found on the surfaces of Corona viruses surrounding the central core of viral genome [1-3].

#### 1.1. Classifications and sub-classifications of corona viruses

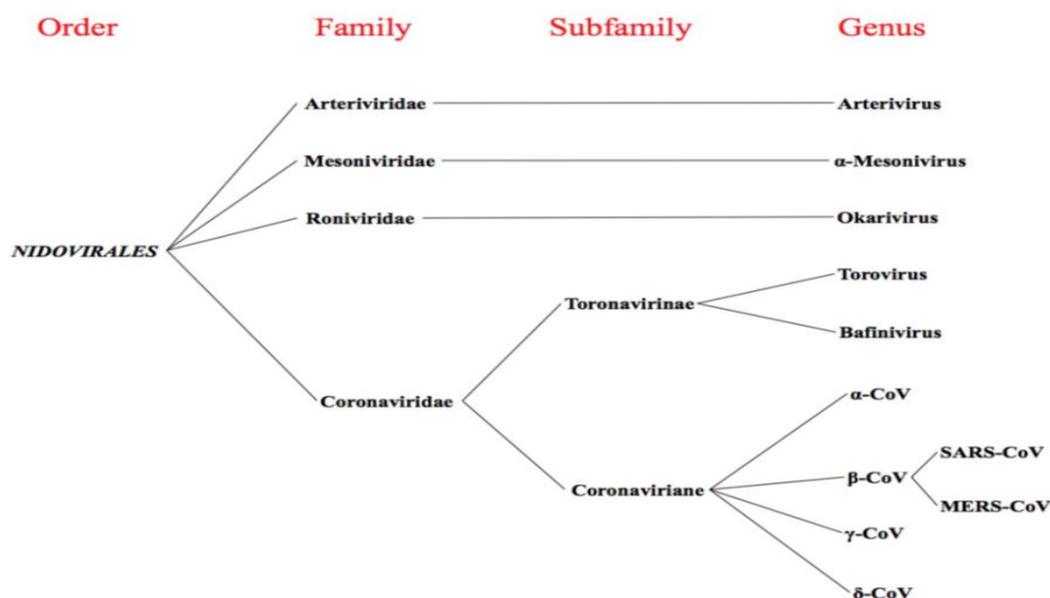
Currently, corona viruses are classified into four genera in the subfamily Coronavirinae. These are Alpha ( $\alpha$  -Cov), Beta ( $\beta$ - CoV), Gamma ( $\gamma$  -CoV) and Delta ( $\delta$  -CoV) corona viruses. Most of mammalian corona viruses, including human corona viruses, are usually from alpha (for example: Human corona virus (HCoV)-229E and HCoV-NL63) and beta (for example: Human corona virus identified by Hong Kong University (HCoV- HKU1), HCoV-OC43, SARS-CoV and Middle East Respiratory Syndrome corona virus (MERS -CoV), and utilize bat as natural reservoir for their evolution and dissemination. However, the bird corona viruses are usually from  $\gamma$ -CoVs and  $\delta$ -CoVs lineages [3]. There are four  $\beta$ -corona virus lineages (A, B, C and D). The SARS-CoV, SARS-CoV-2 and over 200 published virus sequences belong to

\*Corresponding author: Endriyas Kelta Wabalo

Department of Biomedical sciences, Faculty of Medical Sciences, Institute of Health, Jimma University, Jimma, Ethiopia.

Lineage B of beta corona viruses. The MERS-CoV and over 500 viral sequences belong to lineage C of beta corona viruses [4].

In the mid-1960's, Centers for Disease Control and Prevention (CDC) reported six corona viruses from alpha and beta corona virus genera that can infect people with different levels of pathogeniety. HCoV-229E and HCoV-NL63 are among alpha corona viruses that can infect people with level of pathogeniety, and HCoV-HKU1 and HCoV-OC43 are among beta corona viruses that can infect with mild respiratory symptoms just like that of common cold. SARS-CoV and MERS-CoV are the most serious human corona viruses among beta corona virus that may cause more severe and potentially fatal respiratory symptoms [5, 6]. These serious beta corona virus species are originally, the animal corona viruses, transferred from animal to animal and rarely, pass from animal to people and named as novel corona viruses. For example; SARs-CoV was originally the animal corona virus, until it has been identified in China on November 2002 and named as novel human corona Virus. Similarly, MERs-CoV is another animal corona Virus which has been named as novel human corona virus during its first discovery in Saudi Arabia in 2012. The current corona Virus (2019-nCoV) of Wuhan, China also falls under this group [5, 6].



**Figure 1** Classifications and sub-classifications of corona viruses [7].

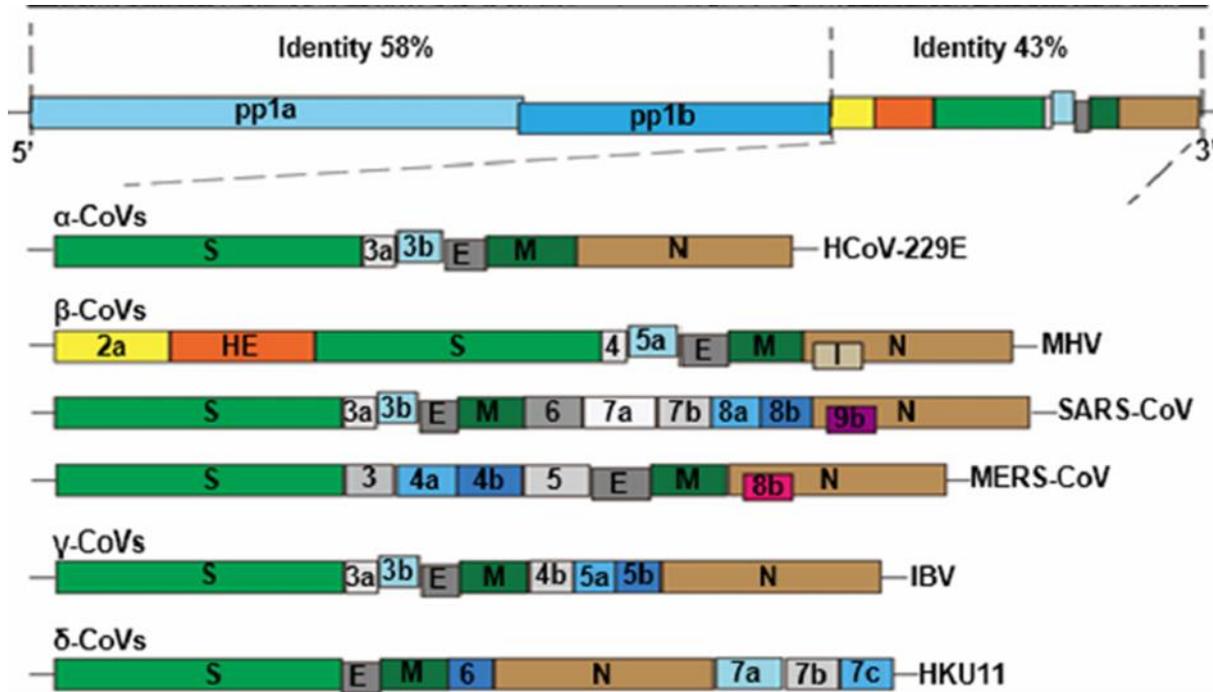
## 1.2. Genomic structure and phylogenetic tree of corona viruses

Corona viruses (CoVs) are RNA viruses with positive-sensed enveloped RNA as genetic material. The genomic sizes of Corona viridae, particularly in SARS COVs, range from 26 to 32 kb (approximately 29,700 nucleotides) with 5'-cap and 3'-poly-A tail un-translated regions (UTRs). In the whole genomic structure of COVs there are 14 open reading frames (ORFs) encoding for 28 proteins that fall in three distinct classes [8-10].

ORF1ab is about 2/3<sup>rd</sup> of whole genome of the virus and encodes for two large polypeptides (pp1a and pp1ab), during viral RNA synthesis. Sixteen functional (nsp1-nsp16) proteins are obtained by enzymatic cleavage of these two large polypeptides in all genera of corona virus. The cleavage process into (nsp1-nsp16) is facilitated by virally encoded chymotrypsin-like protease (3CLpro) or main protease (Mpro) and one or two papain-like protease. The γ-CoV genera are the exceptional corona virus that lacks nsp1 [7, 11].

The other ORFs (ORF 2-14) are about 1/3<sup>rd</sup> of genome size of the virus and encodes for four structural proteins and around eight (8) auxiliary proteins, Auxiliary proteins are proteins that are not involved in the viral synthesis machinery, as 16 functional proteins, and in the viral assembly, coating, entry into host cells and packaging of Viral RNAs, as structural proteins. Rather, these are proteins that may play central role by enhancing viral assembly process, virulence and pathogenesis of SARs-CoV-2The hemagglutinin esterase (HE) protein is among the accessory proteins and it presents only in some beta corona viruses [7-11].

The ORF1ab region of genomic RNA is directly used as template during translation process into polypeptides 1a/1ab (pp1a/pp1ab) encoding for 16 functional proteins (nsps) that form the replicase transcriptase complex (RTC) in a double-membrane vesicle (DMV). Then, the newly formed RTC synthesizes a nested set of sub-genomic RNAs (sgRNAs) in a manner of non-continuous RNA transcription. Thus, main functions of RTC are summarized as sub-genomic RNA synthesis, processing and solving the challenge of host cell innate immune system. As usual mRNAs, the sub-genomic mRNA has 5'-leader and 3'-terminal sequences, and transcription regulatory sequences in between ORFs for termination of transcription process and acquisition of a leader RNA. The sgRNAs of CoVs are responsible for the translation process of all structural and auxiliary proteins [7, 9, 11, 12].



**Figure 2** The genomic structure of four genera of corona viruses [11].

## 2. Severe acute respiratory corona virus-2 (SARS COV-2)

SARS-CoV-2 is the seventh corona virus that is known to cause human disease and it is a new strain of corona virus that has not been previously detected in human. In December 2019 in Wuhan, China outbreak of the source of unexplained pneumonia reported, out of the first 41 people with pneumonia who were identified as having this new strain corona virus infection by 2 January 2020, two-thirds had been associated with Huanan Seafood Market, the largest wholesale market of live animal and seafood in Jiangshan District, Wuhan, Hubei province, China [2, 13]. On February 11th, 2020, the new corona virus was officially renamed “SARS-CoV-2” from “2019-nCoV”. The disease caused by SARS-CoV-2 was called “corona virus disease 2019” (COVID-19) [13].

The structure of SARS-CoV-2 is found to be similar to SARS-CoV with virion size ranging from 70 to 90 nm. It is enveloped viruses with a positive-sense single-strand RNA of around 29,903 nucleotides, belonging to genera, beta corona virus. It was found that the genome sequence of SARS-CoV-2 is 96.2% and 85.5% to 92.4% identical to the virus isolate from bat, a bat CoV RaTG13 and pangolin, novel pangolin CoV respectively, that is why the definitive host of SARS COV-2 is most probable of bat CoV RaTG13 from *Rhinolophus affinis* transmitted to human via unknown intermediate host [6, 13, 14, 15].

SARS-CoV-2 is also closely related and shares a 79.5% sequence identity to SARS-CoV. For some encoded proteins like corona virus main proteinase (3CLpro), papain-like protease (PLpro), and RNA-dependent RNA polymerase (RdRp), the sequence identity is even higher and can be as high as 96% between SARS-CoV-2 and SARS-CoV. Therefore, it was thought that SARS-CoV-2 would function in a similar way to SARS-CoV in the human-infection and pathogenic mechanism [6, 16].

According to virus genome sequencing results and evolutionary analysis, bat has been suspected as natural host of virus origin, it is clear now that SARS-CoV-2 could use angiotensin-converting enzyme2 (ACE2), the same receptor as SARS-CoV, to infect humans despite intermediate host that transmitted from bat to human to infect unclearly yet [6].

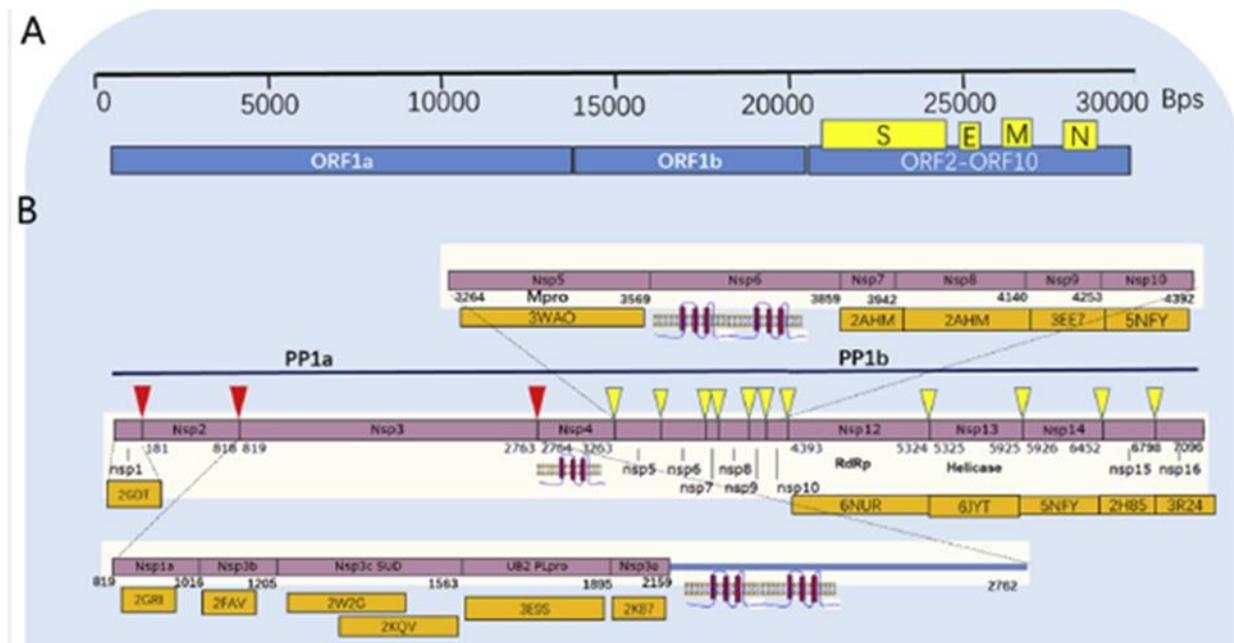
Full-genome sequencing and phylogenic analysis indicated that the corona virus that causes COVID-19 is a beta corona virus in the same subgenus as the severe acute respiratory syndrome (SARS) virus (as well as several bat corona viruses), but in a different clade. The structure of the receptor-binding gene region is very similar to that of the SARS corona virus, and the virus has been shown to use the same receptor, the angiotensin-converting enzyme 2 (ACE2), for cell entry. Thus, Corona virus Study Group of the International Committee on Taxonomy of Viruses has proposed that this virus be designated severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) [5,6].

In a phylogenetic analysis of 103 strains of SARS-CoV-2 from China, two different types of SARS-CoV-2 were identified, designated type L (accounting for 70 percent of the strains) and type S (accounting for 30 percent). The L type predominated during the early days of the epidemic in China, but accounted for a lower proportion of strains outside of Wuhan than in Wuhan. The clinical implications of these findings are uncertain [6, 17]. Due to outbreak of this virus currently population of the world faced to cumbersome health problem. The appropriate vaccine and drug haven't been prepared so far, hence to invent the drugs and vaccines it needs more study on the role of structural and functional proteins of SARS COV-2. The aim of this review is to emphasize the role of structural and functional proteins of SARS COV-2.

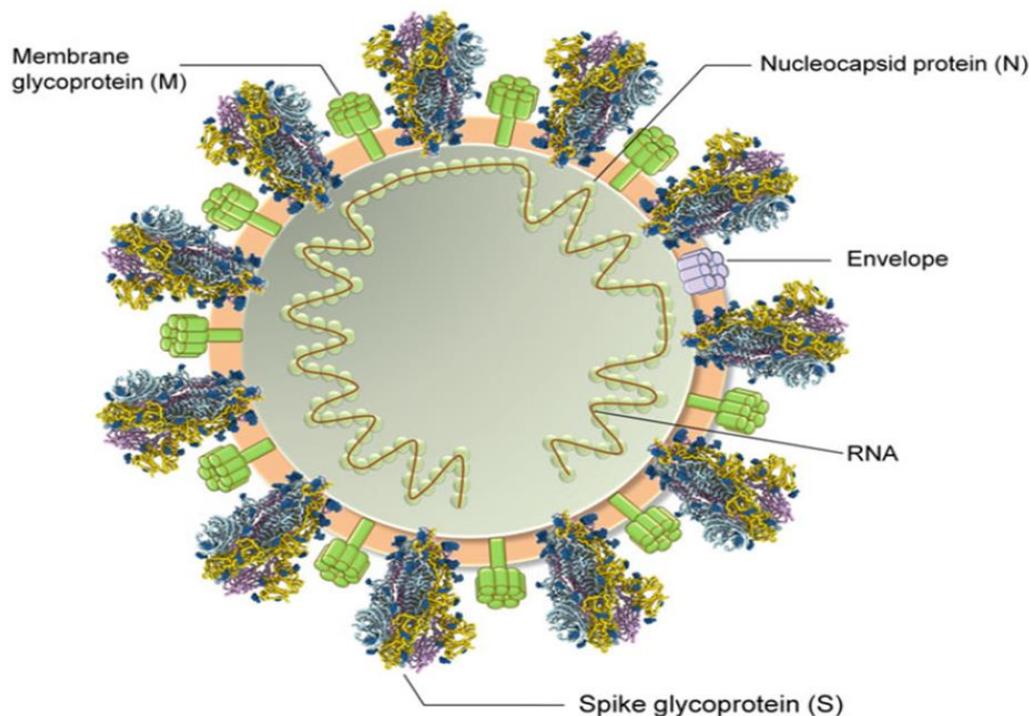
### 2.1. Structural and functional proteins of SARS COV-2

SARS-CoV-2 genome has 10 open reading frames (ORF). ORF1ab encodes replicase polyprotein 1 ab. After cleaved by two proteases, replicase proteins showed multifunction involved in transcription and replication of viral RNAs. ORF2-10 encodes viral structural proteins such as S, M, N, and E proteins, and eight auxiliary proteins (Figure 3 A and B) [6].

The S, M, and E proteins are involved in the formation of the viral coat, and the N protein is involved in the packaging of the RNA genome (Figure 5). Generally, SARS-CoV-2, viral particles comprise four main structural proteins: the spike, membrane, envelope protein, and nucleocapsid, and eight accessory proteins, and 16 non-structural proteins (nsps) that are responsible for virus replication [18-22].



**Figure 3** Overall genome and protein analysis of SARS-CoV-2. (A) The whole Genome of SARS-CoV-2. (B) Replicase polyprotein (PP1ab). The purple rectangle represents the replicase polyprotein PP1ab which contains 7096 amino acids. The position of red triangle means cleavage site of PLpro, and position of yellow triangle means the site of 3CLpro, the numbers below the purple rectangle represent the residues that bound the individual proteins or domains. Orange rectangle below the individual proteins or domains means exactly homology 3D structure was detected [21].



**Figure 4** Schematic representation of the viral surface proteins, spike, envelope and membrane, are embedded in a lipid bilayer. The single-stranded positive-sense viral RNA is associated with the nucleocapsid protein [23].

## 2.2. Role of structural proteins

### 2.2.1. Role of spike (s) protein

Spike is the main structural protein of corona virus, on the surface of the virus as a homo-trimer and assembles into a special corolla structure. S protein of corona virus is made of three identical chains with 1273 amino acid each, and organized into two regions. The first region is outer subunit (S1) favoring host cell recognition and the receptor binding. The S1 region also includes an N-terminal domain (NTD) and three C-terminal domains (CTD1, CTD2, and CTD3). The second region is transmembrane subunit (S2) favoring membrane fusion [16, 22, 24].

In SARS CoV-2 there are notable structural feature, one is polybasic furin cleavage site at the junction of S1 and S2 regions of spike protein. It allows effective cleavage by furin and other proteases and determines viral infectivity. The functional consequence of the polybasic cleavage site in SARS-CoV-2 is unknown, and it might be important to determine its impact on transmissibility and pathogenesis in animal models [15]. Hence, corona viruses use the surface spike (S) glycoprotein on the corona virus envelope to attach host cells and mediate host cell membrane and viral membrane fusion during infection [16, 22].

Different studies revealed that for SARS-CoV and SARS CoV-2, the receptor binding domain (RBD) is located in the CTD1 of the S1 region and bind to the same site of the human ACE2 receptor. This does not mean that RBD of each virus is merely the same; RBD in the spike protein is the most variable part of the corona virus genome. Six RBD amino acids have been shown to be critical for binding to ACE2 receptors. Five of six residues differ between SARS- CoV-2 and SARS-CoV [15,16]. On the basis of structural studies and biochemical experiments, SARS-CoV-2 seems to have an RBD that binds with high affinity to ACE2 from humans [15].

SARS-CoV attaches the human host cells through the binding of the RBD protein to the angiotensin-converting enzyme II (ACE2). Therefore, the interaction between RBD and ACE2 is a prerequisite for the human infection with SARS-CoV. Given the high homology between SARS-CoV and SARS-CoV-2, it was confirmed that SARS-CoV-2 would also use the ACE2 molecule as the receptor to enter into human cells. Thus, S-protein is integrated over the surface of the virus, promotes attachment of the virus to the host cell surface receptors and fusion between the viral and host cell membranes to facilitate viral entry into the host cell [15, 16, 21, 22, 24, 25].

### 2.2.2. Role of membrane (M) protein

The M protein is the most abundant structural protein and defines the shape of the viral envelope. It has three transmembrane domains, which shapes the virions, promotes membrane curvature, transport of nutrients across cell membranes and binds to the nucleo-capsid. It is also regarded as the central organizer of CoV assembly, interacting with all other major corona viral structural proteins [11, 25, 26].

Homotypic interactions between the M proteins are the major driving force behind virion envelope formation but, alone, it is not sufficient for virion formation. Interaction of S with M is necessary for retention of S in the ER-Golgi intermediate compartment (ERGIC)/Golgi complex and its incorporation into new virions, but dispensable for the assembly process. Binding of M to N stabilizes the nucleo-capsid (N protein-RNA complex), as well as the internal core of virions, and, ultimately, promotes completion of viral assembly. Together, M and E make up the viral envelope, and their interaction is sufficient for the production and release of VLPs [25, 27].

### 2.2.3. Role of nucleocapsid (N) protein

The N proteins are phosphoproteins capable of binding to helix and possess flexible structure of viral genomic RNA. N protein is the structural component of CoV localizing in the endoplasmic reticulum-Golgi region that bound to the nucleic acid material of the virus. Once, the protein bounds to RNA, it is involved in processes related to the viral genome, the viral replication cycle, and the cellular response of host cells to viral infections. N protein is also heavily phosphorylated and suggested to lead to structural changes enhancing the affinity for viral RNA [8, 28].

The capsid is the protein shell, and inside the capsid, there is nuclear capsid or N-protein which is bound to the virus single positive strand RNA that allows the virus to hijack human cells and turn them into virus factories. The N protein is, therefore, required for RNA genome assembly [25, 29]. The N-protein coats the viral RNA genome which plays a vital role in its replication and transcription as the N-terminus of the N-protein binds to genomic and sub-genomic RNAs in MHV and IBV virions [7].

Although N protein is known to be necessary for corona virus replication, the specific role it plays in this process remains unknown. However, many studies suggest that N protein interaction with nsp3 plays a critical role in the virus replication early in infection. This is one of the important open research area in the development of an effective drug targeting to prevent the contacts between N-terminal of N-protein and single positive RNA strand which can stop viral replication and transcription [8, 25].

### 2.2.4. Role of envelope (E) protein

E-protein is a short and integral membrane protein that contains 76-109 amino acid residues, with the size ranging between 8.4 and 12 kDa, and consists of 35  $\alpha$ -helices and 40 loops. The protein has short hydrophilic N-terminus consisting of 7-12 amino acid residues, followed by a large hydrophobic transmembrane domain of 25 amino acid residues and long hydrophilic C-terminal domain [25].

Different studies revealed three roles for the CoV E protein. First one is, its role in viral assembly; via making interaction with the cytoplasmic tails of the M protein, as the result of this drives vesicle like particle (VLP) production. The Second one is, it's crucial role in the release of virions; through its hydrophobic trans-membrane domain (TMD). The third one is its role implicated in the pathogenesis of the virus; through interaction with host cell proteins [25, 27].

Its hydrophobic region can generate oligomerization and form anion-conductive pore in membranes, and plays a significant role in the assembly of the viral genome. Thus, E protein is involved in several aspects of the virus life cycle, such as assembly, budding, envelope formation, membrane permeability of the host cell and virus host cell interaction and pathogenesis. The E protein's ion channel activity is found in the transmembrane region of the protein. The membrane potential has been regulated by E-protein controlling the ion flow between the intracellular and extracellular environment. The ion conductivity triggered by E-protein via the manipulation of COVID-19 genome seems to be a novel route involved in virus pathogenesis [25].

E-protein's ion channel activity and the alteration of corona virus cell ion balance by E-protein is a necessary process for virus production but the effect of E-protein ion channel activity in virus pathogenesis remains elusive. Few efforts have been found that mutation of the E-protein in the extracellular membrane could disrupt the ion-conductivity and the normal viral assembly; hence control of the E protein dynamics is a promising target for preventing pathogenesis associated with the COVID-19 [25].

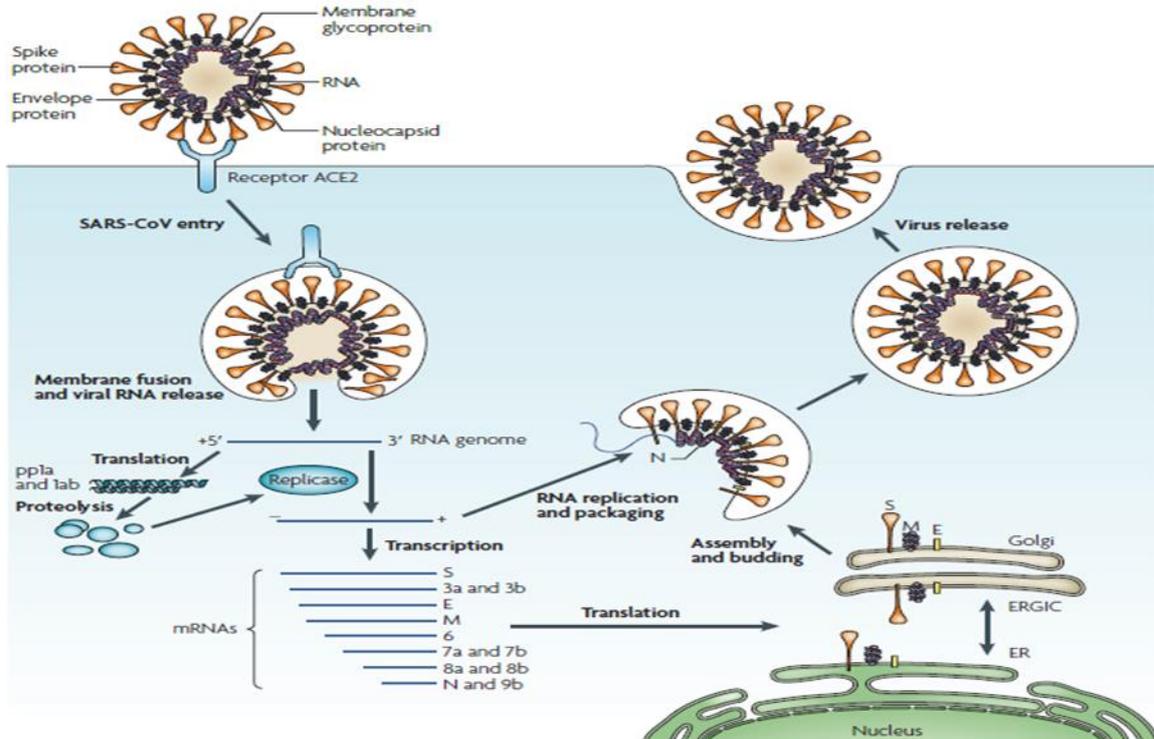
### 2.2.5. Role of hemagglutinin-esterase (HE) protein

Hemagglutinin-esterase (HE) is a dimeric protein that has been located on the surface of the virus. The genome of SARS-CoV-2 lacks the hemagglutinin esterase gene [23], and it only presents in some beta corona viruses. The HE protein may be involved in virus entry, is not required for replication, but appears to be important for infection of the natural host-cell [7, 25].

### 2.3. Role of functional proteins of SARS COV-2

The replication of corona virus occurs in host cell cytoplasm. This virus can enter the human body through its receptors, ACE2, found in various host organs such as heart, lungs, kidneys, and gastrointestinal tract, thus facilitating viral entry into target cells. The viruses primarily bind to the receptor on the cell surface via the spike (S) protein. This attachment occurs in the binding domain of S protein of SARS-CoV-2 receptors which are present at 331 to 524 amino acid residues, and can bind strongly to human ACE2 [8,28]. When S protein is bound to the receptor, a conformational change occurs in the structure and the process of virus entry into the host cell begins [8].

After successful entry into host cell cytoplasm, the viral particle releases its single stranded RNA genome (viral genetic material) into the host cell cytoplasm. In the host cell cytoplasm, subsequently, the ORF, gene segments that encodes to two large polyproteins, pp1a and pp1ab, directly translated by using host cell proteins translation machinery. These polyproteins also called replicase polyprotein1ab, of about 7096 amino acids residues. After cleavage by papain-like proteases (PLpro) and Mpro (chymotrypsin-like protease (3CLpro)) protease, replicase proteins like RdR polymerase and other non-structural proteins released. These cleavage products are involved in the transcription and replication of viral RNAs [21, 25, 28, 30], mediated by the so-called replication/transcription complex (RTC) [8, 25, 28], formed by so many of the functional proteins (nsps) in the double-membrane vesicles (DMVs). RTC are mainly an assembly by RNA-dependent RNA polymerase (RdRp) and helicase containing subunits, the canonical RdRp domain residing of CoV nsp12 and Avian Virus nsp9. Furthermore, the complex transcribes an endogenous genome template of viral entry to negative-sense genes of both the progeny genome and sub-genomic RNA as intermediate products and followed by transcription to positive-sense mRNAs that are mainly mediated by RdRp [28]. RdRp also synthesizes a full-length negative-strand RNA template to be used by RdRp to make more viral genomic RNA [30].



**Figure 5** The schematic diagram of the mechanism of COVID-19 entry and viral replication and viral RNA packing in the human cell [12].

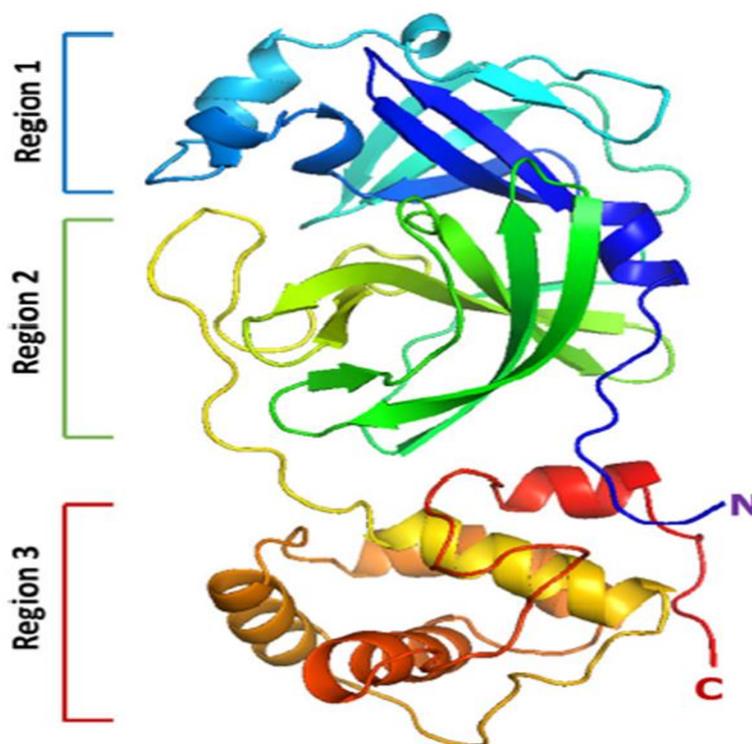
### 2.3.1. Role of SARS COV-2 main protease (Mpro)

Corona virus 3C-like protease (3CLP) or coronavirus main protease (Mpro) is a homodimeric cysteine protease and a member of a family of enzymes found in the corona virus polyprotein (nsps5). SARS-CoV-2 Mpro was predicted to contain 306 amino acids residues (located in the polyprotein, from 3264-3569 amino acid residues). It cleaves the replicase polyproteins into individual polypeptides that are required for replication and transcription. The amino acid sequence of SARS PP1ab of restriction cleavage (14 specific proteolytic) sites recognized by Mpro and PLpro, and 11/14 sites are cleaved by Mpro, and the remaining 3/14 cleavage sites are cleaved by PLpro [21, 29, 31].

Following the translation of the messenger RNA to yield the polyproteins, the Mpro is first auto-cleaved from the polyproteins to become a mature enzyme. The mature Mpro then cleaves all the 11 cleavage sites at the C-terminus of replicase polyprotein 1ab, and releases the key replicative functions, such as the RNA-dependent RNA polymerase and the helicase, from the polyprotein precursors. Mpro is the best characterized corona virus proteolytic enzyme, plays a central role in the viral replication cycle and is an attractive target against the human SARS virus [29, 32].

The COVID-19 replicase gene encoded two polyproteins, pp1a and pp1ab with molecular weight 450 and 750 KD respectively. In the proteolytic process, the functional polypeptides of spike, membrane, envelop, nucleocapsid protein, replicase and polymerase are released from polyproteins. This process was carried out by a chymotrypsin-fold proteinase namely main protease (Mpro). The main protease (Mpro) is essential in processing the polyproteins that are translated from the viral RNA and virus maturation. Hence, it is considered to be an attractive target for antiviral drug design as an approach toward COVID-19 treatment. The Mpro operates at 11 cleavage sites on the large polyprotein 1ab; the recognition sequence at most sites is Leu-Gln↓ (Ser, Ala, Gly) (↓ marks the cleavage site). Inhibiting the activity of this enzyme would block viral replication. Because no human proteases with similar cleavage specificity are known, such inhibitors are unlikely to be toxic. One of the best-characterized drug targets among corona viruses is the main protease [25, 33].

The crystal structure of COVID-19 Mpro possess three regions: (Blue) region of 8-101 residues, (Green) of 102-184 residues and (Orange) of 201-303 residues. The blue and green regions have an anti-parallel b-sheet structure, while orange region has five alpha helices arranged into a cluster. The orange region is connected to gree region through a long-loop region of 185-200 residues, (Figure 6). Mass spectroscopy determined weight of COVID-19 Mpro is 33797.0 Da [25].



**Figure 6** The 3-D structure of SARS-CoV-2 Mpro [25]

### 2.3.2. Role of papain-like protease (PLpro)

All corona viruses encode two (paralogous) papain-like cysteine proteinases (PL1pro and PL2pro). The PLpro is essential for processing the replicase polyproteins that are translated from the viral RNA. It is responsible to cleave three sites of polyprotein regions at 181-182, 818-819, and 2763-2764 amino acid residues at the N-terminus and releases three essential functional proteins for correcting virus replication (Nsp1, Nsp2 and Nsp3). PLpro is also confirmed to be significant to antagonize the host's innate immunity [21, 32, 33].

There is a -1 frame shift between ORF1a and ORF1b, leading to production of two polypeptides: pp1a and pp1ab by the action of papain-like proteinases (PL1pro and PL2pro). These polypeptides are processed not only by papain-like proteinases (PL1pro and PL2pro), but also by virally encoded chymotrypsin-like protease (3CLpro) or main protease (Mpro). Mpro cleaves the central and C-proximal regions at 11 conserved sites into 16 functional (nsps) proteins. As an indispensable enzyme in the process of corona virus replication and infection of the host, PLpro has been a popular target for corona virus inhibitors. It is very valuable for targeting PLpro to treat corona virus infections, but no inhibitor has been approved by the FDA for marketing [11, 21, 32].

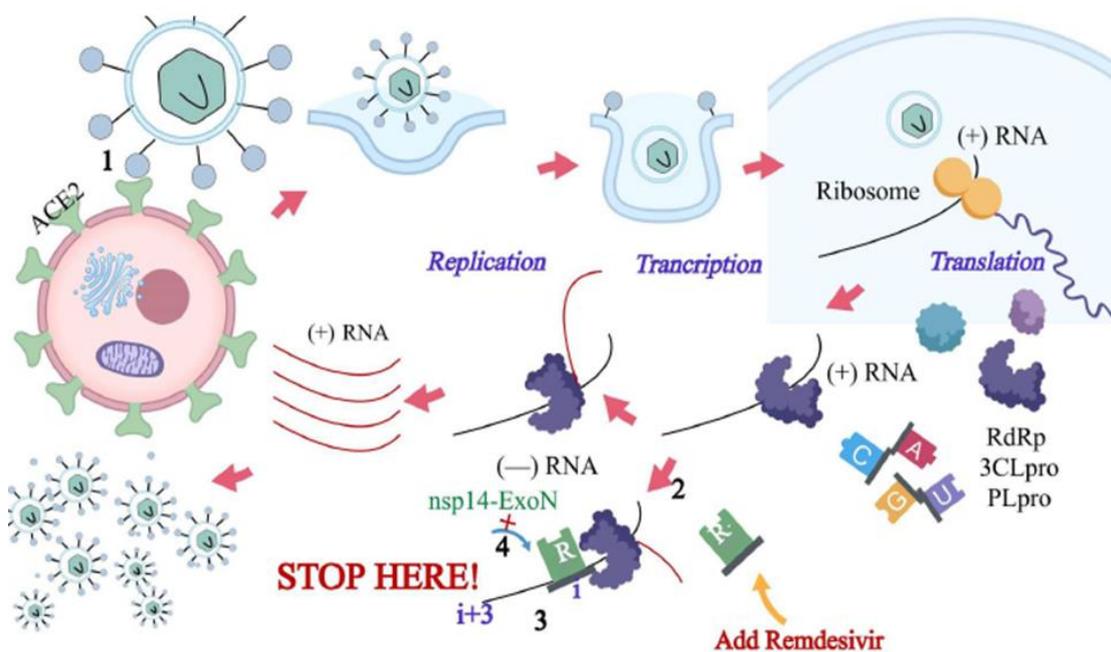
### 2.3.3. Role of RNA dependent RNA polymerase (RdRp)

The SARS-CoV-2 Nsp12 polymerase was predicted to contain 932 amino acids residues (located in the polyprotein, from 4393-5324 amino acids residues). RNA-dependent RNA polymerase (RdRp) (Nsp12) is a conserved protein in corona virus, and the vital enzyme of corona virus replication/transcription complex [21, 33].

The RNA-dependent RNA polymerase (RdRp) (Nsp12) comprised of N-terminal (1-397 amino acids residues) and 398-919 amino acids residues polymerase domain at the C-terminus. The RNA-dependent RNA polymerase (RdRp) (Nsp12) is comprised of a finger (398-581 and 628-687 amino acids residues), a palm (582-627 and 688-815 amino acids residues), and a thumb sub-domain(816–919 amino acids residues).The finger and the thumb sub-domains of nCoV-RdRp contact each other, and configure the RdRp active site in the center for the substrate access through template entry, template-primer exit, and NTP tunnels. Within the SARS-CoV-2, the RNA-dependent RNA polymerase (RdRp) (Nsp12) also revealed seven conserved motifs (A-G) arranged in the polymerase active site chamber, involved in a template and nucleotide binding and catalysis [21, 31].

The Nsp8 can de novo synthesize up to six nucleotides in length, that can be used as a primer for RNA-dependent RNA polymerase (RdRp) (Nsp12) RNA synthesis. Further, the Nsp7- Nsp8 complex increases the binding of RNA-dependent RNA polymerase (RdRp) (Nsp12) to RNA and enhances the RdRps enzyme activity of Nsp12. In the research of SARS-CoV and MERS-CoV inhibitors, Nsp12-RdRp has been used as a very important drug target. In principle, targeted inhibition of Nsp12-RdRp could not cause significant toxicity and side effects on host cells [21].

There are a number of studies revealed that Remdesivir (nucleoside analog) as an RNA-dependent RNA polymerase (RdRp) inhibitor [26, 34-38]. A bioinformatics model of the catalytic core of the SARS-CoV RdRp has been proposed allowing insight into the structure and function of this key viral enzyme during RNA synthesis. The RdRp is the protein complex CoVs use to replicate their RNA-based genomes [35, 39].



**Figure 7** SARS-CoV-2 invasion process and mechanisms of action of remdesivir [26].

#### 2.3.4. Role of helicase

The SARS-CoV-2 Nsp13 contains 596 amino acids residues, and is predicted to be released from pp1ab by 3CLpro-mediated cleavages of the 5301Gln|Ala5302 and 5902Gln|Ala5903 dipeptide bonds. Helicase (Nsp13) is a multi-functional protein comprising five domains. These are two “RecA-like” domains (1A (261-441 amino acids residues) and 2A(442-596 amino acid residues)), 1B (150-260 amino acid residues) domain, N-terminal Zinc binding(ZBD) (1-99 amino acid residues) domain and stalk (100-149 amino acid residues) domain. N-terminal structure contains 26 cysteine residues to form a Zn<sup>2+</sup> binding domain and helicase domain with a conserved motif at the C-terminus. Nsp13 can unravel double-stranded (ds) DNA and RNA along the 5'---3'direction in an NTP-dependent manner [21, 31, 32].

Even though, helicases were initially thought as molecular engines that unwind nucleic acids during replication, recombination, and DNA repair, various reports have shown that they are also taking part in other biological processes, including displacement of proteins from nucleic acid, movement of Holliday junctions, chromatin remodeling, catalysis of nucleic acid conformational changes, several aspects of RNA metabolism, including transcription, mRNA splicing, mRNA export, translation, RNA stability and mitochondrial gene expression [40].

Importantly, it has been reported that the SARS-Nsp13 sequence is conserved and indispensable, and is a necessary component for the replication of corona virus. Therefore, it has been identified as a target for anti-viral drug discovery, but there are few reports about Nsp13 inhibitors [21, 32].

#### 2.3.5. Role of other SARS COV-2 functional proteins (Nsp)

Some functional proteins, including Nsp3b, Nsp3e, Nsp7-Nsp8 complex, Nsp9, Nsp10, Nsp14, Nsp15, and Nsp16, also play an important role in the virus RNA synthesis and replication, and may suggested to be useful targets for the anti-viral drug discovery [21]. However, the functions of some of the nsp are unknown or not well understood. The known functions of the 16 nsp were summarized in (Table below).

**Table 1** Sixteen (16) functional proteins of corona viruses and their functions [3, 10, 11, 28].

Non-structural (nsps)	Functions
Nsps-1	Cellular mRNA degradation, inhibiting IFN signaling & with nsp-3 blocking of host innate immune response by promotion of cellular degradation and blocks translation of host's
Nsps-2	Binding to pro-inhibition protein.
Nsps-3	PLP, viral polypeptides cleaving, blocking host innate immune response, promoting cytokine expression
Nsps-4	DMV formation (with nsp-6 Contribute to structure of DMVs as transmembrane scaffold protein)
Nsps-6	Restricting autophagosome expansion, DMV formation
Nsps-7	Cofactor with nsp8 and nsp12 Nsps-7/8 complex-Process clamp for RNA polymerase by orms hexadecameric complex
Nsps-8	Cofactor with nsp7 and nsp12, primase
Nsps-9	Dimerization and RNA binding protein phosphatase
Nsps-10	Scaffold protein for nsp14 and nsp16 and Stimulation of ExoN and 2'-O-MT activity
Nsps-11	unknown
Nsps-14	Exoribonuclease, N7-MTase, Proofreading of viral genome
Nsps-15	Viral endoribonuclease, evasion of dsRNA sensors and chymotrypsin-like protease
Nsps 16	Ribose 2'-O-MT activity

---

### 3. Conclusion

Currently, population of the world faced too cumbersome health problem that is outbreak of corona virus infectious disease 2019 (COVID-19) caused by severe acute respiratory syndrome corona virus-2 (SARS COV-2). SARS COV-2 belongs to family of Nidovirus, positive single stranded RNA genome. Genomic size approximately 29903 nucleotides and contain at least 10 open reading frames (ORF). ORF1ab is two third of their genome translated to two polypeptide pp1a and pp1b, subsequently these polypeptides processed and cleaved into 16 non-structural protein. Those proteins have tremendous role in viral RNA transcription and replication and pathogenesis. One third of viral genome encodes four structural proteins spike, membrane, envelope and nucleocapsid and other accessories proteins. The role of structural protein to cell recognition and attachment for entry of cell, define the shape of the viral cell, uptake nutrient to the cell, viral assembly, hijack host cell and pathogenesis.

---

### 4. Abbreviations

AV: avian virus, COV: corona virus, COVID-19: corona virus disease-19, 3CLpro: chymotrypsin- like protease, DMV: double-membrane vesicle, dsRNA: double-stranded RNA viruses, E: envelope protein, HCoV; Human coronavirus, HE: Hemagglutinin-esterase, MHV: Murine hepatitis virus, HKU: Corona viruses identified by Hong Kong University, IFN: interferon, IBV: infectious bronchitis virus, M: Membrane protein, MERS COV: Middle East Respiratory Syndrome corona virus, Mpro: main protease, mRNA: messenger RNA, N: Nucleo-capsid, Nsp: non-structural protein, 2'-O-MTase: 2' -O- methyl transferase, PLP: papain like protease, Pp1a: Polyprotein 1a, Pp1b: Polyprotein 1b, RdRp: RNA dependent RNA polymerase, S: Spike protein, SARS COV-2: severe acute respiratory syndrome coronavirus-2, SARS COV: severe acute respiratory syndrome corona virus

---

## Compliance with ethical standards

### *Acknowledgments*

First of all, I pay my wholehearted gratitude to the Almighty God who being my strength and make me healthy, without His grace I will not here where I'm now. I would like thanks to my co-authors Abebe Dukesa Dubiwak, Tariku Sime Gizaw and Urge Gerema Kotu for their valuable engagement and commitments on preparing this review.

### *Disclosure of conflict of interest*

The authors of this study declare that they have no competing interests.

### *Authors' contributions*

All Authors had made equal and substantial contributions to searching and reviewing different literatures. All authors read and approved the final version of the manuscript.

---

## References

- [1] Wick BG. Coronavirus 2020 what is really happening and how to prevent it. 2020.
- [2] Family H. Novel Coronavirus Manual:A Compilation of Things to Know and Do to Protect Yourself from the Wuhan Coronavirus Outbreak. United States by HaLaDi Fam. 2020; 1–52.
- [3] Makino S. Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus Res.* 2014;1–12.
- [4] Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol.* 2020; 5(March).
- [5] McIntosh K, Hirsch MS, Bloom A. Coronavirus disease 2019 (COVID-19). *UpToDate Terms Use.* 2020.
- [6] Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Mil Med Res.* 2020;7:1–10.
- [7] Tok TT, Tatar G. Structures and Functions of Coronavirus Proteins : Molecular Modeling of Viral Nucleoprotein - . *Int J Virol Infect Dis.* 2 June 2017; 1–7.
- [8] Tatar G, Tok TT. Structures and Functions of Coronavirus Proteins : Molecular Modeling of Viral Nucleoprotein - . *Int J Virol Infect Dis.* 1 June 2017 1; 001–001.
- [9] Staup AJ, Silva IU De, Catt JT, Tan X, Hammond RG, Johnson MA. Structure of the SARS-Unique Domain C From the Bat Coronavirus HKU4. *sagepub.* 2019; 1–11.
- [10] Sutton G, Fry E, Carter L, Sainsbury S, Walter T, Nettleship J, et al. The nsp9 Replicase Protein of SARS-Coronavirus , Structure and Functional Insights. 2004; 12: 341–353.
- [11] Chen Y, Chen Y, Liu Q, Guo D. Emerging coronaviruses : Genome structure, replication and pathogenesis. *Med Virol WILEY.* January 2020.
- [12] Du L, He Y, Zhou Y, Liu S, Zheng BJ. The spike protein of SARS-CoV — a target for vaccine and therapeutic development. *Nature.* 7 March 2009; 226–36.
- [13] Xu J, Zhao S, Teng T, Elgaili , Abualgasim Abdalla Zhu W, Xie L, Wang Y, et al. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. 2020; 12:1–17.
- [14] Liao Y, Wei W, Cheung WY, Li W, Li L, Leung GM, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature.* 2020; 1–5.
- [15] Mers-Cov. The proximal origin of SARS-CoV-2. *Nat Med.* 26 April 2020; 450–52.
- [16] He J, Tao H, Yan Y, Huang S, Xiao Y. Molecular Mechanism of Evolution and Human infection with SARS-Cov-2. 2020
- [17] Terms U, Province H. Corona virus disease 2019 (COVID-19). *Up-to-date Terms Use.* 2020; 1–32.

- [18] Law AHY, Lee DCW, Cheung BKW, Yim HCH, Lau ASY. Role for Nonstructural Protein 1 of Severe Acute Respiratory Syndrome Corona virus in Chemokine Dysregulation. 2007; 81(1):416–22.
- [19] Tang X, Li G, Vasilakis N, Zhang Y. Differential Stepwise Evolution of SARS Coronavirus Functional Proteins in Different Host Species. BMC Evol Biol. April 2009; 1–15.
- [20] Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriou N, et al. The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. Am Soc Microbiol. 2008; 82(22):11318–30.
- [21] Wu C, Liu Y, Yang Y, Zhang P, Li X, Zheng M, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B. 2020.
- [22] Hofmann H, Po S. Cellular entry of the SARS coronavirus. TRENDS Microbiol. 2004; 12(10):466–472.
- [23] Kumar S, Nyodu R, Maurya VK, Saxena SK. Morphology, Genome Organization, Replication, and Pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2. Med Virol from Pathog to Dis Control. 2020; 2:23–31.
- [24] Li F. Structure, Function and Evolution of Coronavirus Spike Proteins. Annu Rev of Virology. 2016; 237–264.
- [25] Boopathi S, Poma AB, Kolandaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. J Biomol Struct Dyn. 2020; 1–10.
- [26] Cao Y, Deng Q, Dai S. Remdesivir for Severe Acute Respiratory Syndrome Coronavirus 2. Travel Med Infect Dis. 2020; 10:16–47.
- [27] Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Schoeman Feilding Virol J. 2019; 1–22.
- [28] Astuti DM, ndwiani, Ysrafil, M. Biomed. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An Overview of Viral Structure and Host Response. Diabetes Metab Syndr Clin Res Rev. 2020; 2.
- [29] Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Dmitry S, et al. Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. BMC. 2020; (2).
- [30] Liu C, Zhou Q, Li Y, Garner L V, Watkins SP, Carter LJ, et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. Am Chem Soc. 2020; 6:315–31.
- [31] Mirza MU, Froeyen M. Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. J Pharm Anal. 2020.
- [32] Hertzog T, Thiel V, Ivanov KA, Schelle B, Bayer S, Weißbrich B, et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol. 19 June 2003; 8:2305–15.
- [33] Zhang L, Lin D, Sun X, Curth U, Drosten C. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. April 2020; 4(12): 409–412.
- [34] Al-tawfiq JA, Al-homoud AH, Memish ZA. Remdesivir as a possible therapeutic option for the COVID-19. Travel Med Infect Dis. 2020; 10:15.
- [35] Amirian ES, Levy JK. Current knowledge about the antivirals remdesivir (GS-5734) and GS-441524 as therapeutic options for coronaviruses. One Heal. 9 March 2020; 100128.
- [36] Choy K, Wong AY, Kaewpreedee P, Sia S. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication *in vitro*. Antiviral Res. 2020; 104786.
- [37] To L, Editor THE. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. January 2020; 19–21.
- [38] Agostini ML, Andres EL, Sims AC, Graham RL, Sheahan TP, Lu X, et al. Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. 1–15.
- [39] Brockway SM, Denison MR. Molecular targets for the rational design of drugs to inhibit SARS coronavirus. 2004; 1(2):205–9.
- [40] Adedeji AO, Marchand B, Velthuis AJW, Snijder EJ, Weiss S, Eoff RL, et al. Mechanism of Nucleic Acid Unwinding by SARS-CoV Helicase. PLoS One. 2012; 7(5).