

Structural, molecular biology and the immunopathology of SARS-CoV-2: An updated review

Rahul Mallick¹ and Asim K Duttaroy²

¹University of Eastern Finland School of Medicine

²University of Oslo Faculty of Medicine

July 6, 2020

Abstract

Coronaviruses are a group of enveloped viruses with non-segmented, single-stranded, and positive-sense RNA genomes. Human coronavirus infection causes respiratory diseases with mild to severe outcomes. In December 2019, a new outbreak of the novel coronavirus disease 2019 (COVID-19) emerged in Wuhan, China and spread around the world. Genomic analysis revealed that severe acute respiratory syndrome coronavirus (SARS-CoV-2) is phylogenetically related to SARS-like bat viruses. The intermediate source of origin of SARS-CoV-2 and its transfer to humans is not known; however, it acquired efficient human-to-human transmissibility while retaining human pathogenicity. Spike protein of SARS-CoV-2 has the potential furin-like cleavage site may play a significant role in virus entry. Receptor binding domain (RBD) of SARS-CoV-2 attaches with angiotensin-converting enzyme -2 (ACE2) of epithelial cells. The SARS-CoV-2 genome encodes four major structural proteins: the spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and the envelope (E) protein are involved in assembly, budding, envelope formation, and pathogenesis. Notably, E protein act as viroporin and there is no mutation found on E protein among SARS-CoV-2 strains. At present, the case fatality rate is estimated to range from 6 to 7%. COVID-19 is now a public health emergency of international concern. There is no clinically approved antiviral drug or vaccine available against COVID-19. This review summarized the latest information on the structural and molecular biology infectivity, host immune response and molecular immunopathology of the SARS-CoV-2.

Abstract

Coronaviruses are a group of enveloped viruses with non-segmented, single-stranded, and positive-sense RNA genomes. Human coronavirus infection causes respiratory diseases with mild to severe outcomes. In December 2019, a new outbreak of the novel coronavirus disease 2019 (COVID-19) emerged in Wuhan, China and spread around the world. Genomic analysis revealed that severe acute respiratory syndrome coronavirus (SARS-CoV-2) is phylogenetically related to SARS-like bat viruses. The intermediate source of origin of SARS-CoV-2 and its transfer to humans is not known; however, it acquired efficient human-to-human transmissibility while retaining human pathogenicity. Spike protein of SARS-CoV-2 has the potential furin-like cleavage site may play a significant role in virus entry. Receptor binding domain (RBD) of SARS-CoV-2 attaches with angiotensin-converting enzyme -2 (ACE2) of epithelial cells. The SARS-CoV-2 genome encodes four major structural proteins: the spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and the envelope (E) protein are involved in assembly, budding, envelope formation, and pathogenesis. Notably, E protein act as viroporin and there is no mutation found on E protein among SARS-CoV-2 strains. At present, the case fatality rate is estimated to range from 6 to 7%. COVID-19 is now a public health emergency of international concern. There is no clinically approved antiviral drug or vaccine available against COVID-19. This review summarized the latest information on the structural and molecular biology infectivity, host immune response and molecular immunopathology of the SARS-CoV-2.

Introduction

During the 21st century, five respiratory viruses (3 coronavirus strains and 2 influenza virus strains) outbreaks have occurred worldwide. Human respiratory tract infection by coronaviruses first characterized in 1960 (1). A recent rapid outbreak of infection around the globe caused by a novel coronavirus, which was first emerged in December 2019 in Wuhan city of Hubei province, People's Republic of China. It was declared a pandemic on 11th March 2020 by World Health Organization (WHO) (2–4). The COVID-19 disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (5). Within two decades COVID-19 is the 3rd outbreak of highly contagious coronavirus in humans, followed by a severe acute respiratory syndrome-related coronavirus (SARS-CoV) in 2002-03 and the Middle East respiratory syndrome-related coronavirus (MERS-CoV) in 2012 (6). More than 2,600,000 people from 210 countries and territories along with 2 international conveyance are affected by SARS-CoV-2 as of April 23, 2020 (7). The case fatality rate for COVID-19 has been estimated at around 4-5% so far (7). Although some countries including China have seen a decrease in new cases of COVID-19; however, the global situation remains serious. As of now, the origin of COVID-19 has not been determined and no specific antiviral treatment or vaccine is currently available. During the last few months, substantial data poured in on identifying the SARS-CoV-2, basic structure and mechanism of infection, and immunological challenges. Based on the published data, this review systematically discusses a general overview of COVID-19 with latest pre-clinical issues, structural and molecular biology, host immune response. It is believed that this review will provide the latest information to the concerned people involved in the prevention and control of the COVID-19 epidemic.

1. SARS-CoV-2 virus
2. Phylogenetics and taxonomy of SARS-CoV-2

Enveloped positive-sense single-stranded RNA virus, SARS-CoV-2 belongs to a group of viruses known as coronaviruses that cause respiratory, enteric, hepatic and neurological diseases in mammals and birds (8). So far, 7 coronavirus species are identified those can infect humans(9). Four of them (HKU1, HCoV-OC43, HCoV-NL63, and HCoV-229E) are known to infect immunocompromised individuals (8). Other two zoonotic origin species (SARS-CoV and MERS-CoV) from *Betacoronavirus* genus are lethal to humans (10,11). SARS-CoV-2 is the new strain of SARS-CoV species (12). In December 2019, SARS-CoV-2 caused an outbreak of coronavirus disease 2019 (COVID-19). Phylogenetic analysis has shown that like other SARS-related coronavirus strains (including the strain of 2003 SARS outbreak), SARS-CoV-2 is also the member of subgenus *Sarbecovirus* and species SARS-CoV (13,14). SARS-CoV-2 genome has the following similarities: 96% identity to the sequence of Yunnan-RaTG13 (nearest bat precursor), 90% identity to the sequences of Guangdong-1/PS2 (nearest pangolin precursors), 88% identity to the sequences of SL-CoVZC45/ZXC21 (bat precursor), 80% SARS-CoV and only 50% identity to the sequence of MERS-CoV (14,15). Even despite mutations at the genome or amino acid (AA) level, SARS-CoV-2 strains are highly homologous (~99.99%) (16).

Structural biology of SARS-CoV-2

SARS-CoV-2 virions are around 50-200 nm in diameter, contains four main structural proteins: spike protein (S protein), envelope protein (E protein), nucleocapsid protein (N protein) and membrane protein (M protein) (17–19). The single-stranded RNA genomes of SARS-CoV and MERS-CoV encode two large genes, the ORF1a and ORF1b. **Figure-1** shows the structural proteins of SARS-CoV and SARS-CoV-2. S, E and M proteins form viral envelope/surface together, while N protein holds RNA genome (18). SARS-CoV-2 virus nucleotide is around 30k long (14). Like SARS-CoV, SARS-CoV-2 genome contains around ten open reading frames (ORFs), and 2/3 of the viral genome is first ORFs (ORF1a/b), those are translated into polypeptides required for viral replication and transcription (20). One-third of SARS-CoV-2 genome contains other ORFs, which encode both structural, and accessory proteins. These accessory proteins interrupt the innate immune response of the host(20). 13 variation sites in ORFs (1a, 1b, S, 3a, M, 8 and N regions) have so far been identified in SARS-CoV-2 (16). The functions of all accessory proteins encoded by ORFs of SARS-CoV-2 are still not well elucidated.

Non-structural proteins (NSPs) of SARS-CoV-2

Table-1 summarises the non-structural proteins of SARS-CoV-2 virus. ORF1 encodes pp1a and pp1ab polyproteins those are responsible for the formation of replicase-transcriptase complex (20,21). Following the virus entry, 16 non-structural proteins (NSPs) are produced from these polyproteins by essential viral proteases, e.g. main protease and papain-like protease (20,22). Main protease and papain-like protease act like scissors, snipping the links between the different NSPs and freeing them to do their jobs.

Unique class I spike glycoprotein of SARS-CoV-2

Primary coronavirus tropism, the crown (Latin "corona") shaped S protein (~1200 AA) long) resides on the surface of SARS-CoV-2, which allows the virus to attach and fuse with the host cell membrane (18,23,24). Structural proteins also help assemble and release of new copies of the virus. Like other human infecting coronaviruses, following cleavage class I, S protein of SARS-CoV-2 generates 2 subunits: S1 (attachment domain) subunit and S2 (fusion & transmembrane domain) subunit (24–26). **Figure-2** describes the S protein of SARS-CoV-2 with a focus on putative favourable sites. SARS-CoV-2 virus is highly pathogenic and transmissible comparing with other *betacoronaviruses* of lineage b, due to its polybasic furin-like cleavage site in S protein (24,26,27). Also, the receptor-binding domain (RBD) of SARS-CoV-2 S protein has a high affinity for human angiotensin converting enzyme 2 (ACE2) receptor, which increases the virulence (26). Also, S protein of SARS-CoV-2 can interact with CD147 for cellular invasion (28). So, targeting spike protein of SARS-CoV-2 could be potential for drug or vaccine designing.

Attachment capability of S1 subunit of SARS-CoV-2

The S protein mediates viral entry into host cells by first binding to a host receptor through the RBD in the S1 subunit. N terminal S1 subunit comprises 2 distinct domains: N-terminal domain(NTD) and RBD (24,29). Details of various domains of S1 subunit are described in **Figure-2**. Both domains are critical as receptor recognition (29). Although the RBD is known to bind the cell surface receptor ACE2, the function of NTD of SARS-CoV-2 is not well investigated. NTD has been suggested to facilitate the initial attachment by recognizing sialic acids (as receptor determinants) in the host cell membrane (25,30,31). Further studies however are required for definitive conclusions.

Fusion capability and viral infectivity of S2 subunit of SARS-CoV-2

Several viral proteins such as fusion peptide (FP), second proteolytic site (S2'), internal fusion peptide (IFP), heptad repeat 1& 2 (HR1 & 2) and transmembrane domain (TM) form S2 subunit complex (24). FP and IFP are thought to be responsible for virus entry (32). Like another type I fusion proteins such as Ebola GP and HIV gp41, SARS-CoV-2 HR1 and HR2 may form canonical six-helix bundle (32). The detailed functions of FP, IFP, HR1 & 2 and TM are still not known to conclude whether they work individually or synergistically in the fusion with the host. Further investigation is required for better understanding of functionality of S2 subunit.

1. Multiple cleavage sites of SARS-CoV-2 glycoprotein
2. Identical polybasic furin-like cleavage sites of SARS-CoV-2

Proteolytic cleavage regulates processes in viral-host interaction. One key player is the ubiquitously expressed serine protease furin, which cleaves a plethora of proteins at polybasic recognition motifs. The enzymatic activity of furin is exploited by numerous viral pathogens to enhance their virulence and spreading capacity. Seven mammalian secretory preprotein convertases (PC1, PC2, furin, PC4, PC5, PACE4 and PC7) cleave precursor proteins at specific single or paired basic amino acids (AA) within the canonical motif (**(R/K)-(2X)n-(R/K)**—, where **n** = 0, 1, 2, or 3 and **X** = AA.(33). The preprotein convertases, especially furin (highly expressed in the lung) is known to cleave viral envelope glycoproteins to enhance cellular tropism and pathogenesis (24). And polybasic motif in viral envelope glycoproteins is furin-like cleavage specific and associated with the hyper-virulence of the SARS-CoV-2 (24). Different pathogenic RNA viruses, e.g. HIV, Ebola virus with polybasic envelop protein cleavage site(s) are on focus due to their higher pathogenicity, as shown in **Table-2**. Similar to MERS-CoV, 2 polybasic furin-like cleavage sites (S1/S2 and S2') are discovered in SARS-CoV-2 (24). But the role of furin-like cleavage sites in S protein of SARS-CoV-2 on viral

pathogenesis or replication is still not well investigated.

Other protease-mediated S1/S2 cleavage site of SARS-CoV-2

Like SARS-CoV, SARS-CoV-2 contains the conserved site sequence **AYT—M** between RBD and FP, which can be cleaved by other proteases (e.g. cathepsin L, TMPRSS2) rather than proprotein convertases (24,25). The conserved sequence **AYT—M** may be cleaved following furin cleavage at S1/S2 site. This possibly demonstrates if S protein of SARS-CoV-2 would not cleave at furin-like cleavage sites during viral endocytosis, S protein would certainly cleave at conserved sequence **AYT—M**.

Glycobiology of SARS-CoV-2

Insertion of proline before furin cleavage site (-RRAR-) results in the addition of O-linked glycans which flank at the cleavage site (S1/S2) (25,26). **Figure-3** shows the cleavage site of closely related *betacoronaviruses*. The function of O-linked glycan of SARS-CoV-2 has not been studied much. It is predicted that O-linked glycan could create mucin-like domain to protect epitopes or key residues on S protein. Different viruses use the mucin-like domain to conserve immune evasion (34). Therefore, O-linked glycan could be a crucial determinant for vaccine discovery of SARS-CoV-2.

Mutations in the receptor-binding domain

Surface proteins of SARS-CoV-2 have around 98% and 76% identical AA sequence of Yunnan-RaTG13 (bat precursor) and SARS-CoV, respectively (14). The most variable part of the coronavirus genome seems to be RBD in the S protein (26). **Figure-4** describes the RBD of closely related *betacoronaviruses*. With 30 mutations (13 interface mutations) in RBD indicates how changes could allow switching the host from bat to human (14). As positive-sense RNA virus, COVID-19 virus can mutate in human for its adaptation. Although second closest relative, Guangdong-1/PS2 (pangolin precursors) has approximately 91% similarity with S protein of SARS-CoV-2, RBD of Guangdong-1/PS2 is much closer to SARS-CoV-2 (99% identical) than Yunnan-RaTG13 (14,35,36). So, the probable pangolin origin of COVID-19 outbreak may not be ruled out. Seemingly natural selection in animal host before zoonotic transfer or natural selection in humans following zoonotic transfer can be the reason of origin of human SARS-CoV-2 infection (26). Additionally, these genomic comparisons suggest that the SARS-CoV-2 could be the result of a recombination between Yunnan-RaTG13 and Guangdong-1/PS2 (pangolin precursors) viruses. For recombination, two divergent positive-sense RNA viruses must have infected the same organism simultaneously (37,38). The recombination mechanism was described earlier section to explain the origin of SARS-CoV (39). Enough data are still not available to conclude whether the recombination occurred in bat, pangolin, or any another species. Further research is needed for definitive conclusions. However, the two features of the SARS-CoV-2 virus such as its distinct backbone and the mutations in the RBD of the S protein, probably rule out laboratory manipulation as a potential origin for SARS-CoV-2.

Promising envelop protein of SARS-CoV-2

Like in SARS-CoV, E protein may be involved in SARS-CoV-2 assembly, budding and envelope formation (40). Interestingly no mutation has been observed so far among E protein of SARS-CoV-2 strains (16). Similar to SARS-CoV, E protein of SARS-CoV-2 may form channels in ERGIC/Golgi membranes to permeabilize calcium ions to trigger NOD-like receptor 3 (NLRP3) inflammasome, leading to IL-1 β overproduction (41). E protein latches onto proteins that help turn host genes on and off. Ultimately, E protein has a significant role in the viral life cycle which could be a promising target for vaccine development (42).

Nucleocapsid protein of SARS-CoV-2

Nucleocapsid protein (N protein) protects the viral RNA by stabilising it inside the virus (43). N protein induces apoptosis and inhibits type I interferon production, cell cytokinesis and proliferation (44). N protein also regulates several pathways, e.g. AP-1 signal transduction pathway, nuclear factor kappa B (NF κ B) pathway and transforming growth factor-beta signalling (44). These cytokines associated with inflammation in COVID-19 may be useful in developing the therapeutic strategy.

Accessory proteins of SARS-CoV-2

Accessory proteins modulate the environment in infected cells to make easier for viral replication. Therefore, their roles in pathogenesis are important. The roles of the accessory proteins such as ORF3a, ORF6, ORF7a, ORF8, ORF9b and ORF10 of SARS-CoV-2 in viral replication are yet to be elucidated fully. KAUST metagenomic 154 analysis platform (KMAP, www.cbrc.kaust.edu.sa/kmap) has shown that accessory proteins of SARS-CoV-2 encoded by 5 gene clusters (ORF3, ORF6, ORF7, ORF8, ORF9) are similar to those of other *betacoronaviruses* (40). The endoplasmic reticulum-induced ORF3a protein is thought to be responsible for apoptosis and suppression of type I interferon (IFN) signalling through the PERK pathway (45). Whereas ORF3a upregulates fibrinogen expression and activates NF κ B and NLRP3 inflammasome to trigger inflammation (45,46). ORF3a makes holes in the infected cell membrane for new viruses to escape out. Vesicle-bound ORF6 may play a vital role in the viral replication (47). It also blocks some of host cells's virus-fighting proteins (48). ORF7a binds lymphocyte function-associated antigen 1 (LFA-1), induces apoptosis and activates NF κ B and IL-8 promoters (47). ORF7a cuts down infected cell's supply of tetherin to allow more of the viruses to escape. ORF8 of SARS-CoV-2 was distinct from other *betacoronaviruses* (49). Intracellular stress pathways and NLRP3 inflammasomes triggering aggregation motif **VLLVLL** is absent in the SARS-CoV-2 ORF8 (49). More research is needed to define the function of ORF8. Cellular protein Crml interacts with ORF9b encoded protein to induce apoptosis (47). ORF9b can block interferon. Unlike SARS-CoV, host immune response modulating ORF3b seems to be absent in SARS-CoV-2 due to overlap the same sequence of ORF3a. Multi-localized accessory protein encoded by ORF3b upregulates cytokines (except type I IFNs) and chemokines by modulating transcription factor RUNX1b, induces AP1 transcriptional activity through activation of JNK and ERK pathways, arrests cell growth in G0/G1 phase and induces apoptosis and necrosis (47). Surprisingly lacking of ORF3b doesn't affect much in the infective ability of the virus (47). ORF7b overlaps the same sequence of ORF7a in RNA genome, but it is not clear what, if anything, the gene does. Although ORF7b encoded non-structural (type III integral transmembrane) protein is responsible for apoptosis, IFN dependent reporter gene expression and type I IFN response, but deletion of ORF7b enhances viral growth may explain the reason of more tissue damage by SARS-CoV-2 (51). The non-conserved ORF7b in SARS-CoV-2 may affect its infection property. However, unique ORF10 may be an artefact annotation for SARS-CoV-2 (40).

Polarity of the RNA of SARS-CoV-2

SARS-CoV-2 belongs to the group IV according to the Baltimore classification (52). Positive-sense (5'-3') viral RNA genome directly serves as messenger RNA, which is capable of translating into viral proteins (skipping transcription step) in the host cell during viral replication. Vulnerable SARS-CoV-2 RNA genome protects it's genome integrity from environmental factors and host factors (attack by nucleases and other RNA-modifying enzymes that comprise the cellular intrinsic or innate immune response) through different mechanisms (37).

1. Transmission immunopathology of SARS-CoV-2 virus
2. Reservoir of SARS-CoV-2

Bat is thought to be the primary reservoir for all coronavirus related human outbreaks (10), but the original source of SARS-CoV-2 is yet to be established (53,54). SARS-CoV-2 shares 96% of its genome with RaTG13 from *Rhinolophus affinis* in Yunnan province, China (55). Coronavirus recovered from malayan pangolin also have higher similarity with SARS-CoV-2, which indicates significant recombination event and more complex origin of SARS-CoV-2 (56).

Routes of transmission of SARS-CoV-2

Transmission of COVID-19 occurs via respiratory droplets (within the range of approximately 27 feet) and fomites during unprotected contact between the infector and infectee (54,57,58). Although the possibility of airborne transmission cannot be ruled out (59,60). Still, it is not known how much concentration of viable SARS-CoV-2 is needed to infect a human being. van Doremalen *et al.* showed that viability of SARS-CoV-2 in the air (up to 3 hours), copper (4 hours), cupboard (24 hours), plastic and stainless steel (up to 3 days)

(61). Chin *et al.* could not detect viable infectious SARS-CoV-2 from printing and tissue paper after 3 hours, wood or cloths on day 2 and glass or banknote on day 4, but infectious SARS-CoV-2 was detectable on the outer surface of the surgical mask on day 7 (62). COVID-19 transmission was suggested to be reduced at high temperature and high humidity (63). SARS-CoV-2 is extremely stable at 4°C, but at 70°C SARS-CoV-2 virus inactivates within 5 min (62). As viable SARS-CoV-2 is present on stool sample and ultimately sewage water, so the faecal-oral route may be considered for COVID-19 transmission (54,64). COVID-19 transmission may occur at subclinical infection or asymptomatic conditions (65–67).

1. Molecular biology of infection
2. Entry of SARS-CoV-2 virus

In order to reproduce and establish infection the virus must enter the host organism. S protein of SARS-CoV-2 is the vital determinant of viral entry, which binds to ACE2 expressing host epithelial cell through RBD (**Figure- 5**) (20). S protein of SARS-CoV-2 has 10-20 times stronger affinity for human ACE2 receptor than that of SARS-CoV (68). ACE2 regulates renin-angiotensin system homeostasis (69). Along with angiotensin II receptor type 2, ACE2 protects from severe acute lung injury (69). The potential target ACE2 is expressed in upper and lower respiratory tract (polarized epithelial cells and type II pneumocytes), oral cavity (epithelial cells), oesophagus (upper and stratified epithelial cells), intestine (enterocytes), gall bladder (cholangiocytes), heart (myocardial cells), breast, adrenal gland, kidney (proximal tubular cells), urinary bladder (urothelial cells), eye (conjunctival fibroblast & epithelial cells and corneal epithelial cells), brain (neuronal cells) and male organ (testis and seminal vesicle) (70–76). Direct or indirect contact with mucus membrane in the nose, mouth and eyes can transmit COVID-19 virus (77). Droplets and fomites containing SARS-CoV-2 find their prospective target ACE2 receptors at respiratory tract primarily as a portal of entry. But due to the highest expression of ACE2 receptor into enterocytes (and cholangiocytes), the intestinal wall cannot be ignored as the portal of entry of COVID-19 virus (70). Studies found that stools were positive for SARS-CoV-2 for 1 to 12 days in patients following negative results in respiratory samples (78). Gastric pH may reduce the infective capacity of SARS-CoV-2, as previous studies showed how pH influenced the viability of SARS-CoV in the stool (79). Although Chin *et al.* claimed of extreme stability of SARS-CoV-2 in a wide range of pH values (3–10) at room temperature (22°C) (62).

1. Replication of SARS-CoV-2
2. Two-step cleavage of S Protein of SARS-CoV-2

Following binding to ACE2 receptor, 2 steps sequential protease cleavage occurs in S protein of SARS-CoV-2 by host proteases: priming cleavage on S1/S2 and activating cleavage on S2' sites (20,31) Priming cleavage on S1/S2 site is regulated by host proteases, e.g. cathepsin L and TMPRSS2 (20,25,80). Also, ACE2 receptor is proteolytically cleaved by TMPRSS2 and/or ADAM17 to propagate SARS-CoV-2 coronavirus entry (81).

Syncytium formation and virus entry

Figure-5 shows the schematic diagram of SARS-CoV-2 entry to the host cells. Phosphatidylinositol-3,5-bisphosphate (PI(3,5)P₂) is important for early to late lysosomal maturation. As the critical SARS-CoV-2 viral entry regulators, phosphatidylinositol 3-phosphate 5-kinase primarily synthesizes PI(3,5)P₂ in the early endosome, while two-pore channel subtype 2 (TPC2) in endosome is one of the fundamental downstream effector of PI(3,5)P₂ (31). Cathepsin L (highly expressed in lung and bronchus) is essential for endosomal cell entry and might be responsible for activation of membrane fusion through proteolysis on S1/S2 site (31,71,82). The FP can expose and insert into the host cell membrane due to the acidic environment of the endosome (80). HR1 and HR2 domains of S2 subunit form canonical six-helical fusion core between the viral and host cell endosomal membrane for syncytium formation (68). Trypsin priming or activation of S protein may trigger directly by endocytosis for rapid disease progression (31). Membrane-bound TMPRSS2 may play the role for viral entry in the gastrointestinal tract (25,71). At neutral pH, transmembrane TMPRSS2 activates S protein that results in syncytium formation and the release of the viral RNA genome into the host cell cytoplasm (25,68,80). The involvement of furin/furin-like protease in SARS-CoV-2 entry is obvious. Similar to MERS-CoV, the SARS-CoV-2 has polybasic furin-like cleavage sites and expected to have a similar

mechanism of action (24). Endogenous furin expression in various tissues (e.g. upper respiratory tract, salivary gland) may influence viral entry through cleavage of S2' site during fusion mediated viral entry (71,83,84). However, one cannot rule out the possibility of cross-talk between heparan sulphate proteoglycans in the host cell surface, and furin-like cleavage sites in SARS-CoV-2 (84,85), as high level of heparan sulphate proteoglycans may cause coagulopathy during COVID-19 (86). More studies on furin-like cleavage sites are urgently needed.

Virus replication

Virus replication process starts after hijacking the host's translational machinery (22). The released SARS-CoV-2 RNA genome in host cell cytoplasm is translated into 2 polyproteins, structural and accessory proteins (20,22). This translational microenvironment protects viral RNA from host nucleases and innate immune response (22). The cleavage of the polyproteins forms 16 NSPs (20,49). The NSPs assembled replicase-transcriptase complex in double-membraned vesicles is responsible for synthesizing a full-length negative RNA strand template (20,21). The RNA strand replicates to genomic RNA and individual sub-genomic RNA continuously (21,87). Sub-genomic RNA serves as mRNA, which directly translated to viral structural proteins and trafficked into the endoplasmic reticulum-Golgi intermediate compartment (21). It is noteworthy to mention that furin is thought to cleave S1/S2 site during S protein biosynthesis (83).

Packaging and release of SARS-CoV-2

S protein, E protein, N protein and relatively abundant membrane protein (M protein) along with newly generated complete RNA genome form virion in budding Golgi vesicles (21). Virion containing vesicles fuse with the host cell membrane, and mature SARS-CoV-2 is released by exocytosis (20,87).

Local spread and dissemination at the portal of entry

The newly formed SARS-CoV-2 virus attacks another neighbouring cell, and this goes on. Immune and lung systems, pulmonary epithelium and immune cells are the primary targets for SARS-CoV-2 (20,88). Similar to SARS-CoV and MERS-CoV, the studies showed widespread immunopathology and/or extrapulmonary dissemination and replication of SARS-CoV-2 among infected patients (88,89). Post mortem study analysis of deceased SARS-CoV-2 and SARS-CoV infected patients found other potential target sites such as neurons, epithelium of renal tubules, intestinal mucosa, macrophages in various organs (88–91). As part of the upper respiratory tract, the olfactory epithelium has also come into focus due to COVID-19 associated anosmia (92).

1. Host immune response
2. Antigen presentation in SARS-CoV-2 infection

Currently, there is very limited knowledge of the host immune response to SARS-CoV-2. However, based on the accumulated clinical and experimental data on these previous viruses some suggestions can be made. Following virus entry into the host cell, the antigen presentation is the key determinant of T-cell immune response. T cells identify major histocompatibility complex (MHC) / human leukocyte antigen (HLA) bound processed cell surface antigens only. Mainly with the help of MHC class I (e.g. HLA-A, HLA-B, HLA-C) molecules virus-infected host cells are presented to CD8+ / cytotoxic T cells, but the MHC class II (HLA-DM, HLA-DR, HLA-DP, HLA-DQ, HLA-DOA and HLA-DOB) also contributes in antigen presentation to CD4+ / helper T cells (93,94). Not only macrophages, dendritic cells or B cells, but also aerodigestive epithelial cells (enterocytes, columnar ciliated epithelial cells and type II pneumocytes) present MHC class II (93). Variability of HLA genes makes the differences in immune response and disease severity among individuals (95). Same as SARS-CoV, studies across 145 HLA alleles showed that individuals with HLA-B*46:01 allele is vulnerable to COVID-19 disease (20,95). HLA-A*25:01 and HLA-C*01:02 alleles are also shown to be related to the susceptibility of SARS-CoV-2, while HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03 might be related to the protection from SARS-CoV-2 infection (95). Due to limited comprehensive COVID-19 pathogenesis study. HLA-DR B1*1202 and HLA-DR0301 are least and top, binding MHC class II molecules for SAR-CoV (20). Even HLA-DRB1*11:01 and HLA-DQB1*02:0 are associated with MERS-CoV infection (20).

Lymphocytopenia with hyperactivity

Following antigen presentation, B and T cell-mediated humoral and cellular immune response occurs respectively to eliminate the virus and disease progression. The humoral immune response plays a vital role to limit infection at later phase and prevent reinfection in the future. Like other acute viral infections, a typical pattern of IgM and IgG antibodies against SARS-CoV-2 are reported. SARS-CoV-2 induced seroconversion starts gradually from approximately day 5 and switches to IgG by around day 14 (96). At the end of week 12, the SARS-specific IgM antibodies disappears, while the IgG antibody against SARS-specific S, M, E and N proteins can last long (20,97). *In vitro* cross-reactivity of SARS-CoV-2 sera with SARS-CoV suggests possible mounting of the humoral immune response (97). As COVID-19 disease is associated with lymphocytopenia, the number of blood CD4+ and CD8+ T cells of COVID-19 patients are significantly reduced (20,98). But CD4+ and CD8+ T cells are overactivated (88). Similar results were observed in acute phase response in SARS-CoV virus infection where CD4+ and CD8+ memory T cells sustained for 4 years in recovered patients to perform T cell proliferation, DTH response and production of IFN- γ (20). CD8+ T cell responses are higher in SARS-CoV infection (which is crucial for lung pathology) than CD4+ T cell responses (97). In severe COVID-19 cases, higher levels of multifunctional CD4+ T cells (IFN γ , TNF α , and IL-2), CD8+ T cells (IFN γ , TNF α and degranulated state) and Th2 cytokines (IL-4, IL-5, IL-10) were detected (97).

The first line of defence against SARS-CoV-2

Only a few studies are available on host innate immune response to SARS-CoV-2 infection. Gradual increment of total neutrophil counts, C-reactive protein and many innate cytokines suggesting highly pro-inflammatory condition in COVID-19 disease progression and severity (97). First-line defence against RNA viruses is comprised of toll-like receptors (TLRs), retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs) and NLRPs to detect viral genome and its replication intermediates (99). Innate immune cells recognise viral invasion by either the endosomal TLRs (TLR3, TLR7 and TLR8) and the cytosolic double-stranded RNA sensor RIG-I/MDA-5 (97,100,101). Adaptor proteins (MyD88 and MAVS) induced downstream signalling are recruited upon recognition by TLRs and RLRs (101). As the first line of defence, the recognition events lead to interferon regulatory transcription factor 3 or 7 (IRF3/7), and NF κ B induced type I IFN (IFN- α and IFN- β) and other pro-inflammatory cytokines (e.g. TNF- α , IL-6) expression (97,101). Type I induction is essential in the early phase of the disease to halt virus propagation within the host and modulate innate and adaptive immune responses. STAT1/2-IRF9 complex-mediated IFN-stimulated genes (e.g. anti-viral enzyme RNase L, pro-inflammatory chemokine CXCL10) transcription follows activation of the JAK-STAT pathway via type I IFN (97). During the early phase of infection, SARS-CoV-2 viral (structural and non-structural) proteins could use multiple strategies to dampen type I IFN mediated immune response (97). But the influx of hyperinflammatory neutrophils, monocytes, and macrophages due to delayed-type I IFN immune response compensates early viral control could result in deteriorating consequences to the infected host. Additionally, SARS-CoV-2 E and ORF3a proteins may trigger inflammasome sensor NLRP3 to secrete IL-1 β and induce pyroptosis (an inflammatory form of cell death) (41,102). Simply saying, the innate immune response could be protective or destructive for SARS-CoV-2 infected patients and may open the window for immune intervention. Due to highly effective innate immunity, very few severe cases were reported in young patients. So, the innate immune response may be the key determinant for disease outcome

Immune evasion by SARS-CoV-2

Clinical and experimental data showed that SARS-CoV-2 deals with host immune response (97). This virus can evade immune detection and dampen human immune responses, which explain why the incubation period of SARS-CoV-2 is higher than the influenza virus. Innate immune responses are inhibited at the level of type I interferon recognition and signalling, while adaptive immune evasion is processed by MHC class I and class II downregulation markedly (97). Like other betacoronaviruses, intermediate products of SARS-CoV-2 may avoid host recognition within double-membraned vesicles during replication process (103,104). SARS-CoV-2 M protein may suppress RIG-I-induced IRF3 triggering, while deubiquitinase (DUB) activity in the infected cell and IRF3 inhibitory activity of SARS-CoV-2 NSP3 may play a vital role in immune evasion (105–108).

NSP1 and NSP6 of SARS-CoV-2 may block the phosphorylation of STAT1 and the translocation of the STAT1/2/IRF9 complex to block antiviral activity and IFN response (109,110). Therefore, downregulating IFN may cause an imbalanced production of pro-inflammatory cytokines and infiltration of inflammatory cells that observed in severe COVID-19.

Cytokine release syndrome on COVID-19

Acute respiratory distress syndrome (ARDS) and multiple organ dysfunction are the fundamental causes of COVID-19 deaths (20). As the initiator of ARDS, deadly uncontrolled cytokine storm results from the release of large amounts of pro-inflammatory cytokines (IFN- α , IFN- γ , IL-1 β , IL-6, IL-12, IL-18, IL-33, TNF- α , TGF β etc.) and chemokines (CCL2, CCL3, CCL5, CXCL8, CXCL9, CXCL10 etc.) by the host immune effector dendritic cells, macrophages, lymphocytes, and natural killer cells (20).

Conclusions

A better understanding of structural, molecular biology of SARS-CoV-2 is critically important to develop effective prevention and therapeutic strategies such as vaccine and antiviral drugs against COVID-19 disease. Rapid response to the current COVID-19 outbreaks with diagnostic and therapeutic measures are urgently required to save thousands of lives. However, there are still several unanswered questions remain regarding the COVID-19 despite massive pouring of research and clinical data on a daily basis. The presence of most significant percentages of COVID-19 are asymptomatic or mildly symptomatic infections underscore the critical importance of elucidating the structure, function of the SARS-CoV-2 and the host-immune response. This sort of information can be exploited in the development of vaccine against COVID-19. For example, E protein can be exploited for vaccine preparation as it is being involved in multiple aspects of the viral replication cycle: from assembly and induction of membrane curvature to scission or budding and release to apoptosis, inflammation and even autophagy.

In this review, we summarized the latest information in these aspects, although information regarding the virology, epidemiology, and transmission of SARS-CoV-2 continues to evolve.

Acknowledgements

This study was supported in part by the Faculty of Medicine, University of Oslo

Conflicts of interest

Both the authors declare no conflict of interest.

References

1. Kahn JS, McIntosh K. History and Recent Advances in Coronavirus Discovery. *Pediatr Infect Dis J*. 2005;
2. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*. 2020;
3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - J Am Med Assoc*. 2020;
4. (WHO) (Press release). WHO Director-General's opening remarks at the media briefing on COVID-19 -11 March 2020 [Internet]. World Health Organization. 2020. Available from: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>
5. WHO. Naming the coronavirus disease (COVID-19) and the virus that causes it [Internet]. World Health Organization. 2020 [cited 2020 Apr 1]. Available from: [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)

6. De Wit E, Van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews Microbiology*. 2016.
7. <https://www.worldometers.info/coronavirus/>. COVID-19 CORONAVIRUS OUTBREAK [Internet]. <https://www.worldometers.info/coronavirus/>. 2020 [cited 2020 Apr 23]. Available from: <https://www.worldometers.info/coronavirus/>
8. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;
9. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends in Microbiology*. 2016.
10. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*. 2019.
11. Perlman S, Netland J. Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology*. 2009.
12. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;
13. Nextstrain. Phylogeny of SARS-like betacoronaviruses including novel coronavirus SARS-CoV-2 [Internet]. nextstrain. 2020. Available from: <https://nextstrain.org/groups/blab/sars-like-cov?p=full>
14. GISAID EpifluDB. Newly emerging coronavirus, hCoV-19 [Internet]. GISAID EpifluDB. 2020. Available from: <https://platform.gisaid.org/epi3/frontend#1ea49f>
15. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;
16. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 2020;
17. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China Life Sciences*. 2020.
18. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B*. 2020;
19. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;
20. Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *J Pharm Anal*. 2020;
21. Fehr AR, Perlman S. Coronaviruses: An overview of their replication and pathogenesis. In: *Coronaviruses: Methods and Protocols*. 2015.
22. Báez-Santos YM, St. John SE, Mesecar AD. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antiviral Research*. 2015.
23. Woo PCY, Huang Y, Lau SKP, Yuen KY. Coronavirus genomics and bioinformatics analysis. *Viruses*. 2010.
24. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res*. 2020;

25. Hoffmann M, Kleine-Weber H, Schroeder S, Mü MA, Drosten C, Pö S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor Article SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* [Internet]. 2020;181:1–10. Available from: <https://doi.org/10.1016/j.cell.2020.02.052>
26. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The Proximal Origin of SARS-CoV-2. *Virological*. 2020;
27. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020;
28. Wang K, Chen W, Zhou Y-S, Lian J-Q, Zhang Z, Du P, et al. SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. *bioRxiv* (preprint). 2020;
29. Li F. Evidence for a Common Evolutionary Origin of Coronavirus Spike Protein Receptor-Binding Subunits. *J Virol*. 2012;
30. Schwegmann-Weßels C, Herrler G. Sialic acids as receptor determinants for coronaviruses. *Glycoconjugate Journal*. 2006.
31. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of 2019- nCoV on virus entry and its immune cross- reactivity with spike glycoprotein of SARS-CoV. *Nat Commun* [Internet]. 2020;(2020):1–38. Available from: <http://dx.doi.org/10.1038/s41467-020-15562-9>
32. Lu G, Wang Q, Gao GF. Bat-to-human: Spike features determining “host jump” of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in Microbiology*. 2015.
33. Seidah NG, Chretien M. Proprotein and prohormone convertases: A family of subtilases generating diverse bioactive polypeptides. *Brain Res*. 1999;
34. Bagdonaite I, Wandall HH. Global aspects of viral glycosylation. *Glycobiology*. 2018.
35. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of 2019-nCoV associated with outbreak of COVID-19. *SSRN eLibrary*. 2020.
36. Cyranoski D. Did pangolins spread the China coronavirus to people? *Nature*. 2020;
37. Barr JN, Fearn R. How RNA viruses maintain their genome integrity. *Journal of General Virology*. 2010.
38. Cheng VCC, Lau SKP, Woo PCY, Kwok YY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*. 2007.
39. Graham RL, Baric RS. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *J Virol*. 2010;
40. Alam I, Kamau A, Kulmanov M, Arold ST, Arnab P, Gojobori T, et al. Functional pangenome analysis provides insights into the origin, function and pathways to therapy of SARS-CoV-2 coronavirus. *bioRxiv* (preprint) [Internet]. 2020; Available from: <https://www.biorxiv.org/content/10.1101/2020.02.17.952895v1.full.pdf>
41. Nieto-Torres JL, Verdía-Báguena C, Jimenez-Guardeño JM, Regla-Nava JA, Castaño-Rodríguez C, Fernandez-Delgado R, et al. Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology*. 2015;
42. Schoeman D, Fielding BC. Coronavirus envelope protein: Current knowledge. *Virology Journal*. 2019.
43. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein - Forms and functions. *Antiviral Research*. 2014.
44. Zebin L, Qian F, Jinlian M, Lishi Z, Yu Q, Tian C, et al. The Nucleocapsid Protein of SARS-CoV-2 Abolished Pluripotency in Human Induced Pluripotent Stem Cells. *bioRxiv* (preprint). 2020;

45. Minakshi R, Padhan K, Rani M, Khan N, Ahmad F, Jameel S. The SARS coronavirus 3a protein causes endoplasmic reticulum stress and induces ligand-independent downregulation of the Type 1 interferon receptor. *PLoS One*. 2009;
46. Siu KL, Yuen KS, Castano-Rodriguez C, Ye ZW, Yeung ML, Fung SY, et al. Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB J*. 2019;
47. McBride R, Fielding BC. The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses*. 2012;4(11):2902–23.
48. Lokugamage KG, Hage A, Schindewolf C, Rajsbaum R, Menachery VD. SARS-CoV-2 is sensitive to type I interferon pretreatment. *bioRxiv*. 2020;
49. Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*. 2020;
50. Schaefer SR, Mackenzie JM, Pekosz A. The ORF7b Protein of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Is Expressed in Virus-Infected Cells and Incorporated into SARS-CoV Particles. *J Virol*. 2007;
51. Pfefferle S, Kräling V, Ditt V, Grywna K, Mühlberger E, Drosten C. Reverse genetic characterization of the natural genomic deletion in SARS-Coronavirus strain Frankfurt-1 open reading frame 7b reveals an attenuating function of the 7b protein in-vitro and in-vivo. *Virol J*. 2009;
52. Baltimore D. Expression of animal virus genomes. *Bacteriol Rev*. 1971;
53. Cohen J. Wuhan seafood market may not be source of novel virus spreading globally. *Science* (80-). 2020;
54. Aylward, Bruce (WHO); Liang W (PRC). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). WHO-China Jt Mission Coronavirus Dis 2019. 2020;
55. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;
56. Wong MC, Cregeen SJJ, Ajami NJ, Petrosino JF. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *bioRxiv*. 2020;2013:2020.02.07.939207.
57. Centers for Disease Control and Prevention. How COVID-19 Spreads [Internet]. U.S. Centers for Disease Control and Prevention (CDC). 2020. Available from: https://www.cdc.gov/coronavirus/2019-ncov/prepare/transmission.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fabout%2Ftransmission.html
58. Bourouiba L. Turbulent Gas Clouds and Respiratory Pathogen Emissions: Potential Implications for Reducing Transmission of COVID-19. *JAMA - Journal of the American Medical Association*. 2020.
59. Morawska L, Cao J. Airborne transmission of SARS-CoV-2 : The world should face the reality. *Environ Int* [Internet]. 2020;139. Available from: <https://doi.org/10.1016/j.envint.2020.105730>
60. Santarpia JL, Rivera DN, Herrera V, Morwitzer MJ, Creager H, Santarpia GW, et al. Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center. *medRxiv*. 2020;
61. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *N Engl J Med*. 2020;
62. Chin AWH, Chu JTS, Perera MRA, Hui KPY, Yen H-L, Chan MCW, et al. Stability of SARS-CoV-2 in different environmental conditions. *The Lancet Microbe*. 2020;

63. Wang J, Tang K, Feng K, Lv W. High Temperature and High Humidity Reduce the Transmission of COVID-19. SSRN Electron J. 2020;
64. DutchWaterSector. Sewage water as indicator for spreading of COVID-19 [Internet]. Dutch Water Sector. 2020. Available from: https://www.dutchwatersector.com/news/sewage-water-as-indicator-for-spreading-of-covid-19?fbclid=IwAR3G2QNSnpRkdd7mGfs_WwWu4Ka4jM7KVNffu4NyZ-tCT_jdB622natAuWU
65. Kupferschmidt K. Study claiming new coronavirus can be transmitted by people without symptoms was flawed. *Science* (80-). 2020;
66. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020;
67. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020;
68. Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, et al. Inhibition of SARS-CoV-2 infection (previously 2019-nCoV) by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res* [Internet]. 2020; Available from: <https://www.nature.com/articles/s41422-020-0305-x>
69. Cheng H, Wang Y, Wang G-Q. Organ-protective Effect of Angiotensin-converting Enzyme 2 and its Effect on the Prognosis of COVID-19. *J Med Virol*. 2020;
70. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci*. 2020;
71. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science* (80-). 2015;
72. Ren X, Glende J, Al-Falah M, de Vries V, Schwegmann-Wessels C, Qu X, et al. Analysis of ACE2 in polarized epithelial cells: Surface expression and function as receptor for severe acute respiratory syndrome-associated coronavirus. *J Gen Virol*. 2006;
73. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV. *bioRxiv* [Internet]. 2020;2:2020.01.26.919985. Available from: <https://www.biorxiv.org/content/10.1101/2020.01.26.919985v1>
74. Xin P. Binding action between SARS-CoV S666 protein and ACE2 receptor in eyes. *Recent Adv Ophthalmol*. 2007;
75. Doobay MF, Talman LS, Obr TD, Tian X, Davisson RL, Lazartigues E. Differential expression of neuronal ACE2 in transgenic mice with overexpression of the brain renin-angiotensin system. *Am J Physiol - Regul Integr Comp Physiol*. 2007;
76. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med*. 2020;
77. Zheng J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *Int J Biol Sci*. 2020;
78. Gu J, Han B, Wang J. COVID-19: Gastrointestinal manifestations and potential fecal-oral transmission. *Gastroenterology*. 2020;
79. WHO. First data on stability and resistance of SARS coronavirus compiled by members of WHO laboratory network [Internet]. WHO. 2003. Available from: https://www.who.int/csr/sars/survival_2003.-05_04/en/?fbclid=IwAR2tORhSZ2_YDCDI-w8NjcAWc5z_BTzgVha-8kIc8dzbt_BbXDTJfKTDC8Q

80. Shen LW, Mao HJ, Wu YL, Tanaka Y, Zhang W. TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. *Biochimie*. 2017.
81. Simmons G, Zmora P, Gierer S, Heurich A, Pohlmann S. Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. *Antiviral Research*. 2013.
82. El Najjar F, Lampe L, Baker ML, Wang LF, Dutch RE. Analysis of cathepsin and furin proteolytic enzymes involved in viral fusion protein activation in cells of the bat reservoir host. *PLoS One*. 2015;
83. Mille JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A*. 2014;
84. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A*. 2009;
85. Pasquato A, Dettin M, Basak A, Gambaretto R, Tonin L, Seidah NG, et al. Heparin enhances the furin cleavage of HIV-1 gp160 peptides. *FEBS Lett*. 2007;
86. Wang S, Qi C, Liu Z, Xu T, Yao C. Endogenous Heparin-Like Substances May Cause Coagulopathy in a Patient with Severe Postpartum Hemorrhage. *Transfus Med Hemotherapy*. 2019;
87. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Mil Med Res*. 2020;
88. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med*. 2020;
89. Gu J, Gong E, Zhang B, Zheng J, Gao Z, Zhong Y, et al. Multiple organ infection and the pathogenesis of SARS. *J Exp Med*. 2005;
90. Liu Q, Wang R, Qu G, Wang Y, Liu P, Zhu Y, et al. Anatomy of a New Coronavirus Pneumonia Death Corpse System (Chinese). *J forensic Med China*. 2020;36(1):21–3.
91. Poyiadji N, Shahin G, Noujaim D, Stone M, Patel S, Griffith B. COVID-19–associated Acute Hemorrhagic Necrotizing Encephalopathy: CT and MRI Features. *Radiology*. 2020;
92. Brann DH, Tsukahara T, Weinreb C, Logan DW, Datta SR. Non-neural expression of SARS-CoV-2 entry genes in the olfactory epithelium suggests mechanisms underlying anosmia in COVID-19 patients. *bioRxiv* (preprint) [Internet]. 2020; Available from: [file:///C:/Users/rahulma/Downloads/Non-neural expression of SARS-CoV-2 entry genes in the olfactory epithelium suggests mechanisms underlying anosmia in COVID-19 patients .pdf](file:///C:/Users/rahulma/Downloads/Non-neural%20expression%20of%20SARS-CoV-2%20entry%20genes%20in%20the%20olfactory%20epithelium%20suggests%20mechanisms%20underlying%20anosmia%20in%20COVID-19%20patients.pdf)
93. Wosen JE, Mukhopadhyay D, MacAubas C, Mellins ED. Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Frontiers in Immunology*. 2018.
94. Hansen TH, Bouvier M. MHC class i antigen presentation: Learning from viral evasion strategies. *Nature Reviews Immunology*. 2009.
95. Austin Nguyen, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, et al. Human leukocyte antigen susceptibility map for SARS-CoV-2. *medRxiv* (preprint) [Internet]. 2020; Available from: <https://www.medrxiv.org/content/10.1101/2020.03.22.20040600v1>
96. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin Infect Dis*. 2020;
97. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pacific J allergy Immunol*. 2020;
98. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, et al. COVID-19 infection: the perspectives on immune responses. *Cell Death Differ*. 2020;

99. Jensen S, Thomsen AR. Sensing of RNA Viruses: a Review of Innate Immune Receptors Involved in Recognizing RNA Virus Invasion. *J Virol.* 2012;
100. Artis D, Spits H. The biology of innate lymphoid cells. *Nature.* 2015.
101. Lim Y, Ng Y, Tam J, Liu D. Human Coronaviruses: A Review of Virus–Host Interactions. *Diseases.* 2016;
102. Chen IY, Moriyama M, Chang MF, Ichinohe T. Severe acute respiratory syndrome coronavirus viroporin 3a activates the NLRP3 inflammasome. *Front Microbiol.* 2019;
103. Knoops K, Kikkert M, Van Den Worm SHE, Zevenhoven-Dobbe JC, Van Der Meer Y, Koster AJ, et al. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* 2008;
104. Oudshoorn D, Rijs K, Limpens RWAL, Groen K, Koster AJ, Snijder EJ, et al. Expression and cleavage of middle east respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. *MBio.* 2017;
105. Frieman M, Ratia K, Johnston RE, Mesecar AD, Baric RS. Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease Ubiquitin-Like Domain and Catalytic Domain Regulate Antagonism of IRF3 and NF- B Signaling. *J Virol.* 2009;
106. Yang X, Chen X, Bian G, Tu J, Xing Y, Wang Y, et al. Proteolytic processing, deubiquitinase and interferon antagonist activities of Middle East respiratory syndrome coronavirus papain-like protease. *J Gen Virol.* 2014;
107. Li SW, Wang CY, Jou YJ, Huang SH, Hsiao LH, Wan L, et al. SARS coronavirus papain-like protease inhibits the TLR7 signaling pathway through removing Lys63-linked polyubiquitination of TRAF3 and TRAF6. *Int J Mol Sci.* 2016;
108. Lui PY, Wong LYR, Fung CL, Siu KL, Yeung ML, Yuen KS, et al. Middle East respiratory syndrome coronavirus M protein suppresses type I interferon expression through the inhibition of TBK1-dependent phosphorylation of IRF3. *Emerg Microbes Infect.* 2016;
109. Wathelet MG, Orr M, Frieman MB, Baric RS. Severe Acute Respiratory Syndrome Coronavirus Evades Antiviral Signaling: Role of nsp1 and Rational Design of an Attenuated Strain. *J Virol.* 2007;
110. Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P. Severe Acute Respiratory Syndrome Coronavirus Open Reading Frame (ORF) 3b, ORF 6, and Nucleocapsid Proteins Function as Interferon Antagonists. *J Virol.* 2007;
111. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol.* 2020;
112. Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio.* 2013;

Table 1: Prospective functions of SARS-CoV-2 virus non-structural proteins (NSPs)

| Proteins | Functions |
|-----------------------------|---|
| NSP1 (cellular saboteur) | * Degrades host mRNA, blocks host translation and innate immune responses |
| NSP2 (mystery protein) | * Binds with prohibitin proteins (21). |

| Proteins | Functions |
|---|---|
| NSP3 (large multi-domain transmembrane protein) (untagging and cutting) | * Interact with N protein (through ubiquitin-like 1 and glutamic acid-rich ac * Promotes cytokine expression (through ADP-ribose-1'-phosphatase domain * Cleaves viral polyprotein and blocks host innate immune response (throug * Binds oligoneucleotides known to form G-quadruplexes (by SARS-unique * Nucleic acid chaperon function (by nucleic acid binding domain) (22) * Putative metal binding by transmembrane domains (22) * Unknown function (Ubiquitin-like 2, G2M, Y domains) (21) |
| NSP4 (transmembrane scaffold protein) (bubble marker) | Responsible for double-membrane vesicles assembly (21) |
| NSP5 (main protease) (protein scissor) | Cleaves viral polyprotein (21) |
| NSP6 (transmembrane scaffold protein) (bubble factory) | Able to proliferate membrane, induces perinuclear vesicles localization (112) |
| NSP7 (copy assistant) | Acts as clamp for RNA polymerase by forming complex with NSP8 (21) |
| NSP8 (copy assistant) | Acts as clamp for RNA polymerase by forming complex with NSP7 (21) |
| NSP9 (interest on nucleus) | RNA binding protein (21) |
| NSP10 (genetic camouflage) | Forms heterodimer with NSP14 & NSP16 to stimulate viral exoribonuclease |
| NSP11 | Short peptide at the end of ORF1a (49) |
| NSP12 (copy machine) | RNA-dependent RNA polymerase (21) |
| NSP13 (unwinding) | RNA helicase domain and RNA 5' triphosphatase activity (21) |
| NSP14 (proof-reader) | * N7 Methyltransferase activity to add 5' cap to viral RNAs (21) * Viral exoribonuclease activity as proofreading of viral genome (21) |
| NSP15 (cleaning up) | Viral endoribonuclease (21) |
| NSP16 (camouflage assistant) | 2'-O-methyltransferase activity to shield viral RNA from melanoma different |

Table 2: Comparison of canonical envelop protein polybasic cleavage sites of different RNA viruses SARS-CoV-2 and other RNA viruses display the canonical (R/K)-(2X)n-(R/K)— motif (red bold alphabets) at cleavage site(s). The hydrophobic residues (except alanine) (black bold alphabets) at P2' position of cleavage site favours the cleavage by furin/furin-like proteases.

| Virus | Envelop protein | Cleavage site |
|--------------------|-----------------|---------------------|
| SARS-CoV-2 | S protein | SPRRAR—SVAS (S1/S2) |
| SARS-CoV-2 | S protein | SKPSKR—SF (S2') |
| SARS-CoV | S protein | LKPTKR—SF (S2') |
| MERS-CoV | S protein | TPRSCR—SVPG (S1/S2) |
| HKU1 | S protein | SRRKRR—SISA (S1/S2) |
| HCoV-OC43 | S protein | KNRRSR—GAITT |
| HIV | Gp160 | VQREKR—AV |
| Human CMV | gB | HKRTKR—ST |
| Influenza virus H5 | HA | RKRKKR—GL |
| Avian H5N1 A/HK/98 | HA | RERKRKKR—GL |

| Virus | Envelop protein | Cleavage site |
|--------------------|-----------------|---------------|
| Yellow Fever Virus | PrM | SRRSRR—AI |
| Zika Virus | PrM | ARRSRR—AV |
| Ebola Virus | GP | GRRTRR—EA |

Figure legends

Figure 1: Structure of SARS-CoV-2

Schematic representation of the genome organization and functional domains of SARS-CoV-2. The single-stranded RNA genomes of SARS-CoV and SARS-CoV-2 encode two large genes, the ORF1a and ORF1b genes. ORF1a/b (2/3 of the SARS-CoV-2 genome) in the 5'-terminal encodes polypeptides, which form the viral replicase-transcriptase complex. The other ORFs encode four main structural proteins: spike protein (S), envelope protein (E), nucleocapsid protein (N) and membrane protein (M), as well as accessory proteins.

Figure 2: S protein of SARS-CoV-2 with a focus on favourable putative site

S protein consists of S1 subunit (N-terminal domain (NTD) and receptor binding domain (RBD)) and S2 subunit (fusion peptide (FP), internal fusion peptide (IFP), heptad repeat 1/2 (HR1/2), and transmembrane domain (TM)) along with signal peptide (SP). Cleavage sites of SP, S1/S2 and S2' are indicated by arrows. The figure is modified and adapted from Coutard *et al*.

Figure 3: Polybasic cleavage site of closely related betacoronaviruses

Alignment of SARS-CoV-2 against the most closely related viruses. Polybasic furin cleavage site (light green box) and predicted related O-linked glycans (dark green box) at S1/S2 junction of SARS-CoV-2 genome sequence. The figure is modified and adapted from Andersen *et al*.

Figure 4: RBD (ACE2 contact residue) of closely related betacoronaviruses

Alignment of SARS-CoV-2 against the most closely related viruses. Key residues in the RBD of S1 subunit that make contact to the ACE2 receptor are marked with blue boxes in both SARS-CoV-2 and related viruses. 6 key AA residues involved in interaction with ACE2 receptor are 99% similar between Guangdong-1/PS2 (Pangolin) and SARS-CoV-2. But only 1 key AA residue is similar between Yunnan-RaTG13 (Bat) or SARS-CoV and SARS-CoV-2. The figure is modified and adapted from Andersen *et al*.

Figure 5: SARS-CoV-2 virus life cycle

(A) SARS-CoV-2 binds with host cell ACE2 receptor resulting in fusion, either (B1) endocytosis or (B2) priming cleavage by TMPRSS2; shedded ACE2 also augments viral entry by either activated ADAM17 or TMPRSS2. (B1a) Endocytosed virus S protein is also cleaved by cathepsin L; (C) Virus RNA genome is liberated into cytoplasm and (D) translation into polypeptides takes place into endoplasmic reticulum derived double membrane vesicles; (E) Polypeptides are cleaved into NSPs; (F) Translation of structural proteins and RNA replication; (G) Trafficking of newly synthesized proteins and RNA from endoplasmic reticulum to Golgi body; (H) Packaging of virion in the budding vesicle; (I) Mature virus releases via exocytosis; (J) SARS-CoV-2 antigens presentation to antigen presenting cells via MHCs / HLAs; (K) Stimulated antigen presenting cells releases cytokines to enhance proinflammatory response. [In both pathways, furin / heparan sulphate could cleave S2' site for successful viral entry and replication]. (Adapted and modified from https://www.genetex.com/Research/Overview/infectious_diseases/SARS-CoV-2?fbclid=IwAR0T8T8J75gdjt1z6Uuvh6KEdsJNDO6Ja8xuhso5Q0SDlkdkwATY077cMxo)

Figure 1: Structure of SARS-CoV-2

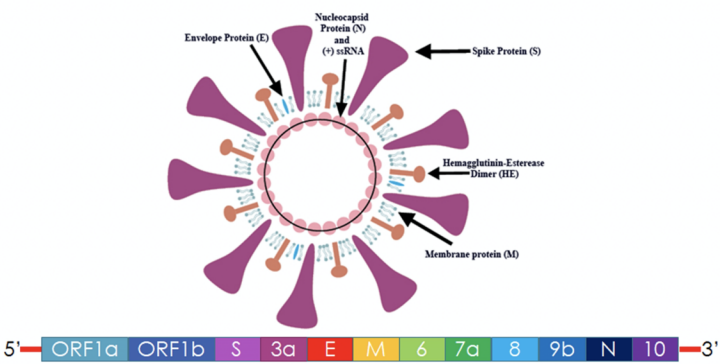


Figure 2: S protein of SARS-CoV-2 with a focus on favourable putative site

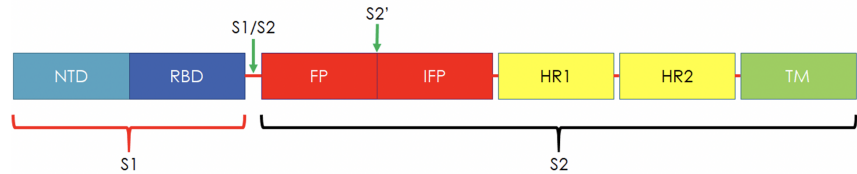


Figure 3: Polybasic cleavage site of closely related betacoronaviruses

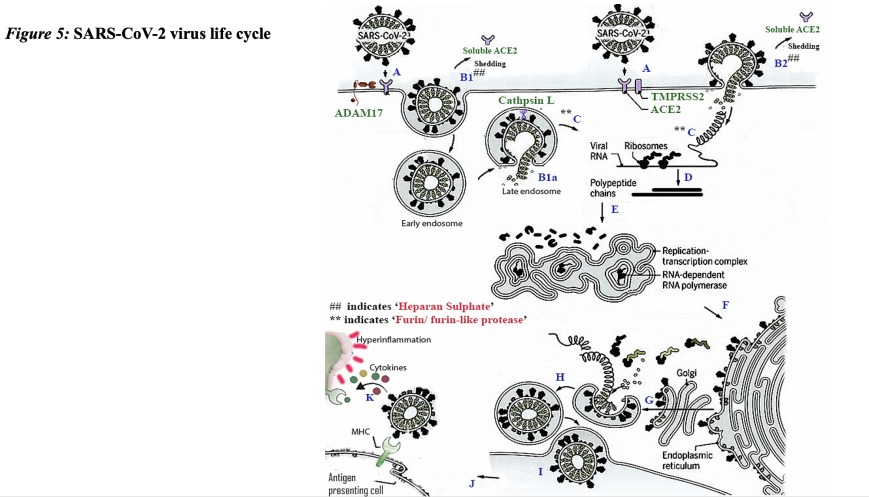
Polybasic cleavage site

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SARS-CoV-2 (Human) | G | A | G | I | C | A | S | Y | Q | T | Q | T | N | S | P | R | R | A | R | S | V | A | S | Q | S | I | I |
| Yunnan-RaTG13 (BAT) | G | A | G | I | C | A | S | Y | Q | T | Q | T | N | S | - | - | - | - | R | S | V | A | S | Q | S | I | I |
| Gúangdong-1/PS2 (Pangolin) | G | A | G | I | C | A | S | Y | Q | T | Q | T | N | S | - | - | - | - | R | S | V | S | S | X | A | I | I |
| SARS-CoV (Human) | G | A | G | I | C | A | S | Y | H | T | V | S | L | L | - | - | - | - | R | S | T | S | Q | K | S | I | V |

Figure 4: RBD (ACE2 contact residue) of closely related **betacoronaviruses**

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SARS-CoV-2 (Human) | Y | L | Y | R | L | F | R | K | S | N | L | K | P | F | E | R | D | I | S | T | E | I | Y | Q | A | G | S | T | P | C | N | G | V | E | G | F | N | C | Y | F | P | L | Q | S | Y | G | F | Q | P | T | N | G | V | G | Y | Q | P | Y | R |
| Yunnan-RaTG13 (BAT) | Y | L | Y | R | L | F | R | K | A | N | L | K | P | F | E | R | D | I | S | T | E | I | Y | Q | A | G | S | K | P | C | N | G | Q | T | G | L | N | C | Y | F | P | L | Y | R | Y | G | F | P | T | D | G | V | H | Q | P | Y | R | | |
| Guangdong-1/PS2 (Pangolin) | Y | L | Y | R | L | F | R | K | S | N | L | K | P | F | E | R | D | I | S | T | E | I | Y | Q | A | G | S | T | P | C | N | G | V | E | G | F | N | C | Y | F | P | L | Q | S | Y | G | F | H | P | T | N | G | V | G | Y | Q | P | Y | R |
| SARS-CoV (Human) | Y | L | Y | R | L | R | H | G | K | L | R | P | F | E | R | D | I | S | N | V | P | F | S | P | D | G | K | P | C | T | P | P | - | A | L | N | C | Y | W | P | L | N | D | Y | G | F | Y | T | T | G | I | G | Y | Q | P | Y | R | | |

Figure 5: SARS-CoV-2 virus life cycle



Hosted file

Table 1.docx available at <https://authorea.com/users/337825/articles/463390-structural-molecular-biology-and-the-immunopathology-of-sars-cov-2-an-updated-review>

Hosted file

Table 2.docx available at <https://authorea.com/users/337825/articles/463390-structural-molecular-biology-and-the-immunopathology-of-sars-cov-2-an-updated-review>