Peer-reviewed version available at *Plos Pathogens* **2020**, *16*; doi:10.1371/journal.ppat.1008762

COVID-19 Pandemic: Insights into Structure, Function, and hACE2 Receptor Recognition by SARS-CoV-2

Anshumali Mittal[1]*, Kavyashree Manjunath[2], Rajesh Kumar Ranjan[3], Sandeep Kaushik[4], Sujeet Kumar[5], Vikash Verma[6]*

1. Formerly at the Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India
2. Formerly at Institute for Stem Cell Science and Regenerative Medicine, Bangalore, India
3. Department of Environmental Science, Central University of South Bihar, Gaya, India
4. Formerly at the 3B's Research Group**,** Research Institute on Biomaterials, Biodegradables and Biomimetics of University of Minho, Guimarães, Portugal
5. Centre for Proteomics & Drug Discovery, Amity Institute of Biotechnology, Amity University, Maharashtra, India
6. Biology Department, University of Massachusetts, Amherst, USA

*Correspondence: Anshumali Mittal, mittalans@gmail.com
Vikash Verma, vverma@umass.edu

Abstract

SARS-CoV-2 is a newly emerging, highly transmissible, and pathogenic coronavirus in humans, which has caused global public health emergency and economic crisis. To date, millions of infections and thousands of deaths have been reported worldwide, and the numbers continue to rise. Currently, there is no specific drug or vaccine against this deadly virus; therefore, there is a pressing need to understand the mechanism through which this virus enters the host cell. Viral entry into the host cell is a multistep process in which SARS-CoV-2 utilizes the receptor binding domain of the spike glycoprotein (S) to recognize ACE2 receptors on the human cells; this initiates host cell entry by promoting viral-host cell membrane fusion through large scale conformational changes in the S protein. Receptor recognition and fusion are critical and essential steps of viral infections and are key determinants of the viral host range and cross-species transmission. In this review, we summarize the current knowledge on the origin and evolution of SARS-CoV-2 and the roles of key viral factors. We discuss the RNA dependent RNA polymerase structure of SARS-CoV-2, its significance in drug discovery, and explain the receptor recognition mechanisms of coronaviruses. We provide a comparative analysis of the SARS-CoV and SARS-CoV-2 S proteins, receptor-binding specificity, and discuss the differences in their antigenicity based on biophysical and structural characteristics.

Key words: COVID-19, SARS-CoV, SARS-like coronavirus, 2019-nCoV, SARS-CoV-2, angiotensin-converting enzyme 2 (ACE2), RdRp, Remdesivir, and neutralizing antibody.

1. Introduction

Before 2003, only two human coronaviruses, HCoV-229E and HCoV-oC43, causing mild illness were known [1,2,3]. However, the emergence of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) changed the view worldwide, because coronaviruses can cause life- threatening infections[4,5,6]. The ongoing pandemic of a novel strain of coronaviruses, Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), is posing an unforeseen public health and economic threats worldwide. As of June 27, 2020, SARS-CoV-2 has infected more than 9.65 million people with 491,115 deaths reported from 215 countries and territories[7], of which there are 2,407,590 confirmed cases of COVID-19 and 124,161 deaths in the USA alone[8]. Recombination, mutator alleles and mutational robustness are some of the evolutionary mechanisms [9], which make coronaviruses capable of expanding their host ranges, including humans . Therefore, understanding the virology of the coronaviruses at a structural level is of utmost importance, because the health threats from these zoonotic viruses are constant and long term.

Coronaviruses are large, enveloped, positive-stranded RNA viruses, responsible for infecting a wide variety of mammalian and avian species[10]. These viruses contain spike-like projections of glycoproteins on their surface, which appear as a crown under the electron microscope; hence, they are referred to as coronaviruses... The coronavirus genome encodes several structural and nonstructural proteins. The structural proteins are responsible for host infection[11], membrane fusion[12], viral assembly[13], morphogenesis, and release of virus particles[14] among other functions, and the non-structural proteins facilitate viral replication and transcription[15,16]. The membrane (M), the envelope (E), and the spike protein (S) makeup

the structural proteins and are associated with the envelope. Among these structural proteins, the trimeric spike proteins protrude from the virus envelope and are the key machinery that facilitate virus entry into the host cell[10,17].

The spike proteins are clove-shaped, type-I transmembrane proteins and have three segments: a large ectodomain, a single-pass transmembrane, and an intracellular tail. The ectodomain of spike proteins consist of the S1 subunit containing a receptor binding domain (RBD) and the membrane fusion subunit (S2). The host cell receptor recognition by the RBDs on spike proteins is the initial step of viral infection, and binding interactions between coronavirus spike and its receptor is one of the most critical factors for host range and cross-species transmission. Human coronaviruses recognize a variety of host receptors, specifically HCoV-NL229 recognizes aminopeptidase N (APN)[18], MERS-CoV binds dipeptidyl peptidase-4 (DPP4)[19], HCoV-OC43 and HCoV-HKU1 bind certain types of O-acetylated sialic acid[20], and HCoV-NL63 and SARS-CoV recognize angiotensin-converting enzyme 2 (ACE2)[21,22]. Recent structures along with functional studies, have suggested that the SARS-CoV-2 spike proteins utilize ACE2 and Transmembrane Serine Protease 2 (TMPRSS2) for host cell entry, which are very similar to the mechanisms exploited by SARS-CoV[23]. See section 5 of this review for detailed information on the mechanism of coronavirus cell entry mediated by the viral spike-glycoproteins. The spike proteins, common among all coronaviruses, are a major target for eliciting antibodies; therefore, structural and molecular details of spike protein and its interactions with cognate receptor would be vital in developing vaccines and anti-viral drugs against SARS-CoV-2.

In this review, we discuss the coronavirus classification, details of SARS-CoV-2 emergence, morphology, and key virulence factors. We specifically explain the RNA dependent

RNA polymerase structure of SARS-CoV-2 and its significance in drug discovery. Further, structure, function and antigenicity of spike glycoproteins and its interactions with ACE2 receptor are discussed.

2. Emergence of SARS-CoV and SARS-CoV-2

In November 2002, SARS began spreading from the Guangdong province of Southern China, but its reservoir was unknown. In the past, Nipah and Hendra, both zoonotic viruses, originated from bats and this motivated researchers to find if bats are the natural reservoirs of SARS-CoV[24,25]. In 2005, two research groups independently reported that bats (horseshoe bats in particular) are the natural host of genetically diverse coronaviruses, and closely related to those responsible for the SARS outbreak[26,27]. These viruses were termed SARS-like coronaviruses, and they displayed considerable genetic similarities to SARS-CoV isolated from human or civets. This suggested that the virus responsible for SARS outbreak was a member of SARS-like coronaviruses group[26]. In Saudi Arabia MERS-CoV emerged in 2012, when humans were infected through direct or indirect contacts with infected dromedary camels. However, genome analysis suggested that MERS-CoV might have also originated in bats and was transmitted to camels in distant past[28] (Figure 1).

In December 2019, severe pneumonia patients of unknown cause were reported in Wuhan, China and a novel coronavirus strain was detected from the lower respiratory tract of four patients [29]. Viruses were isolated from these clinical samples, and their genomes were analyzed by deep sequencing [30,31,32]. Phylogenetic analysis of 2019-nCoVs genomes and other coronaviruses were used to establish the evolutionary history and infection sources. Interestingly, this indicated that 2019-nCoV (GenBank: MN908947.3) shares about 96% nucleotide sequence identity to Bat coronavirus RaTG13 (GenBank: MN996532.1), with 79.5%

and 55% identity to SARS-CoV BJ01 (GenBank: AY278488.2) and MERS-CoV HCoV-EMC (GenBank: MH454272.1), respectively and belongs to the same family of viruses that caused SARS and MERS (Figure 2). This suggests that bat is possibly the host of 2019-nCoV origin, and it might have transmitted either directly from bat or through an unknown intermediate host to infect humans[29,33,34,35]. Despite high sequence similarities, a few notable and conserved variations arose in 2019-nCoVs genomes that were not previously seen in betacoronaviruses. These notable features, which establish this virus different from SARS-CoV and SARS-like coronaviruses are: (i) multiple mutations in the RBDs of spike protein that may interact with ACE2 receptor, (ii) a polybasic furin-like protease site (RRAR/S) at the boundary of S1/S2 subunits rather than a single arginine observed in SARS-CoV, and (iii) addition of three predicted O-linked glycans flanking the protease site[36,37]. Of note, a furin-like protease site is a signature of several highly pathogenic avian influenza viruses and pathogenic Newcastle disease virus[38,39].

Originally this virus was called “2019-novel coronavirus” (2019-nCoV), but later the International Committee on Taxonomy of Viruses on February 11, 2020, officially named it “Severe Acute Respiratory Syndrome Coronavirus 2” (SARS-CoV-2) due to its genetic similarity to SARS-CoV[33]. SARS-CoV-2 causes the respiratory illness and WHO named it coronavirus disease-2019 (COVID-19). It is a contagious, primarily transmitted among people through respiratory droplets and contact routes [40,41],and more than nine million COVID-19 cases are confirmed worldwide. WHO declared the COVID-19 outbreak a public health emergency of international concern first, and later characterized as a pandemic on Mach 11, 2020[42].

3. Classification of coronaviruses

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

The coronavirus study group of the International Committee on Taxonomy of Viruses has classified coronaviruses under the family *Coronaviridae*, subfamily *Coronavirinae*. Based on genotypic and serological characterization, *Coronavirinae* is divided into four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* [43,44,45,46] (Figure 3A). Only six Human Coronavirus species (HCoV) were known until December 2019 that cause human disease. Four of them cause common cold symptoms in immunocompromised individuals, which are HCoV-229E and HCoV-OC43 first identified in mid-1960s[1,2,3], HCoV-NL63 in 2004[47,48], and HCoV-HKU1 in 2005[49]. The other two strains, which cause fatal illness, are namely severe acute respiratory syndrome coronavirus (SARS-CoV) first identified in 2003[4,5] and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012[6]. SARS-CoV-2 has 96% nucleotide sequence similarity to Bat coronavirus RaTG13, a SARS-like coronavirus; therefore, belongs to *Betacoronavirus* genera (Figure 2).

Forsters et al. performed phylogenetic network analysis of 160 complete SARS-CoV-2 genomes sampled from across the world to understand the evolution of this virus in humans and infection sources. They named these closely related genomes in three lineages, namely A, B, and C based on amino acid changes. The lineage A was named for the original bat coronavirus that caused COVID-19, but surprisingly it was not the dominant virus type in Wuhan. The A and C types were found largely in American and Europeans, respectively, while the B types was mostly prevalent in East Asia and had acquired mutations before spreading outside East Asia. The lineage C differs from its parent lineage B by a mutation at amino acid position 26144 and was prevalent in France, Italy, Sweden, England, California, Brazil, Singapore, Hong Kong, Taiwan and South Korea but absent from mainland Chinese samples. This kind of phylogenetic

classification has a potential to accurately trace the infection routes and will prove helpful in designing treatments and vaccines development[50].

4. Morphology, genomic structure, and key viral factors of SARS-CoV-2

Coronaviruses are non-segmented, enveloped viruses with ssRNA ranging between 26 to 32 kb in length. At this length, the coronaviruses genome is the largest among RNA viruses. Electron microscopy of negative-stained SARS-CoV-2 particles revealed their spherical shape with diameter ranging from 60-140 nm and an outer surface studded with distinctive 9-12 nm long spikes that gave virions the appearance of a solar corona[29] (Figure 3B). The observed morphology of SARS-CoV-2 is consistent with other members of the Coronaviridae family.

SARS-CoV-2 Wuhan-Hu-1 isolate (GenBank: MN908947.3) was among the first complete genome sequenced and comprises a 29903 bp long RNA. It is 5'-capped, 3'-polyadenylated, consists of two flanking untranslated regions (UTRs), and contains several open reading frames (ORFs) that encode multiple proteins. The genome is arranged in the order of a non-coding 5'-UTR – replicase genes (orf1ab) – structural proteins (S, E, M, and N) – non-coding 3'-UTR[51] (Figure 3D). Notably, it lacks the hemagglutinin-estrase gene, which is a common feature of lineage A *Betacoronaviruse*[31]. The orf1a/b, located at the 5'-end of the genome, is the largest open reading frame and it encodes 15 nonstructural proteins (nsp1-10 and nsp12-nsp16)[52]. Briefly, the orf1a/b has overlapping orfs and produces two polypeptides, pp1a and pp1ab, due to ribosomal frameshifting. The virus genome encodes two cysteine proteases, a papain-like protease (PL2pro) or nsp3 and a 3C-like protease (3CLpro) or nsp5. These proteases cleave pp1a and pp1ab polypeptides into 15 nsps. Specifically, PL2pro is responsible for cleaving between nsp1|2, nsp2|3 and nsp3|4 sites and the 3CLpro cleaves at the LQ↓SAG sites to produce nsp4 through nsp16[31,53]. RNA-dependent RNA polymerase (nsp12)

in complex with nsp7, nsp8, helicase (nsps13), and exonuclease (nsp14) are critical enzymes among these nsps responsible for the transcription and viral RNA replication.

The 3'-terminus of the SARS-CoV-2 genome contains four structural proteins, which are responsible for virus-host cell receptor binding, virion assembly, morphogenesis, and release of virus particles from the host cell. The envelope (E) protein of SARS-CoV-2 is the smallest of all structural proteins found in the viral membrane and localizes to the ER and Golgi complex in the host cells[54]. The E protein along with M and N are known to facilitate virus-like particle formation[14]. The membrane (M) glycoprotein is a transmembrane protein located in the viral membrane and is the most abundant structural protein in a virion, almost ~100 times higher than E protein. The M protein plays a major role in the viral assembly along with E and N proteins[13,14,55]. The N-protein is responsible for packaging the viral genome RNA (gRNA) into a helical ribonucleocapsid (RNP). SARS-CoV-2 also has eight accessory proteins derived from sub-genomic RNA: 3a, 3b, 6, 7a, 7b, 8b, 9b and orf14 (based on the NCBI annotation NC_045512.2, and reference 57), and they are distributed among the structural genes [51,52,56].

Phylogenetic tree-based analysis of the whole genomes and individual genes suggest that SARS-CoV-2 is closer to SARS-like bat coronaviruses than to SARS-CoVs. Specifically, the spike gene of SARS-CoV-2 is closer to SARS-like bat coronaviruses, although the 3a and 8b accessory genes are closer to SARS-CoVs[52,57]. In a recent study based on available genomic sequences, it was observed that SARS-CoV-2 (106 sequences) genome has a much lower mutation rate and genetic diversity than SARS-CoV (39 sequences), and in particular the spike protein-coding gene is relatively more conserved than other protein-encoding genes[58].

4.1. Structure of the SARS-CoV-2 RNA dependent RNA polymerase complex

Coronaviruses use an RNA-dependent RNA polymerase (RdRp) complex for replication of their genome and transcription of their genes[16]. The SARS-CoV-2 RdRp complex is composed of a catalytic subunit nsp12, and two accessory subunits: nsp7 and nsp8, which increases RdRp template binding and processivity[59]. The mechanism of replication and inhibition of SARS-CoV-2 RdRp has been elucidated by several groups using cryo-EM structures of the RdRp-nps7-nsp8 complex[15], its complex with RNA[60], and Remdesivir[61]. The overall structure of the SARS-CoV-2 nsp12-nsp7-nsp8 complex highly resembles that of SARS-CoV, with a global root mean square deviations (RMSD) of ~1Å for the α-carbon atoms[15,62]. The SARS-CoV-2 RdRp complex structure reveals that the nsp12 core catalytic subunit is bound to a heterodimer of nsp7-nsp8 and an additional nsp8 subunit at a different binding site (Figure 4A, B, C)[62]. The N-terminus of nsp12 contains nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain followed by an interface domain, and a C-terminal RdRp domain (Figure 4B)[15]. The RdRp domain includes seven conserved motifs (A-G), which are distributed in the finger, palm, and thumb subdomains (Figure 4A, C). The palm subdomain is formed by five conserved motifs A-E; the motif C contains a critical SDD sequence (“Ser-Asp-Asp” residues 759-761), which forms the catalytic active center. Both D760 and D761 coordinate with two magnesium ions at the catalytic center. The F and G motifs are located within the finger subdomain and direct the template strand RNA into the active site and the thumb subdomain intersects the extensions from the finger subdomain to hold the first turn of RNA [59,60,61,62]. The residues involved in RNA binding as well as forming the catalytic active site are highly conserved among different RNA viruses, which highlight the conserved mechanism of genome replication used by RdRp[61].

The RNA polymerase of the viruses is an established target for inhibiting the viral replication and has pre-established values for clinical engagements by the broad-spectrum

nucleotides, such as prodrug Remdesivir. These drugs have shown therapeutic efficacies against several viruses from different families, including Ebola, Nipah, MERS and SARS-CoV[63,64]. The cell-based studies in Vero E6 cells (ATCC-1586) have shown that Remdesivir is able to potently block the SARS-CoV-2 viral infections at very low concentrations ($EC_{50} = 0.77$ μM) *in vitro*[65]. The cryo-EM structure of the RdRp-Remdesivir complex suggests that Remdesivir inhibits the viral RdRp activity through non-obligate RNA chain termination, a mechanism that converts the prodrug to the active drug in the triphosphate form[61]. Besides Remdesivir, Flavipiravir, Ribavirin, Galidesivir and EIDD-2801 have been shown to inhibit SARS-CoV-2 replication in cell-based assays. Specifically, EIDD-2801 is 3-10 times more potent than Remdesivir in blocking SARS-CoV-2 replication[66]. The cryo-EM structure of the RdRp-Remdesivir complex (PDB ID 7BV2) provides the mechanism of Remdesivir binding as well as a blueprint for designing more potent antiviral therapeutics to combat the vicious infection of SARS-CoV-2[61].

5. Structure, function, antigenicity and ACE2 recognition by the SARS- CoV-2 Spike Glycoprotein

The S protein is a multifunctional molecular machine that plays key roles in the early steps of viral infection by interacting with host susceptibility factors, including receptors and proteases. These interactions subsequently infect human cells, which contain angiotensin-converting enzyme 2 (hACE2) transmembrane proteins[67]. The SARS-CoV-2-S is a transmembrane glycoprotein composed of S1 regions containing the NTD and CTD, S2, a transmembrane region, and a short cytoplasmic domain (Figure 3C, D). Both cryo-EM and crystallographic methods have been used to determine multiple structures of the SARS-CoV-2 spike protein alone, including the ectodomain of S protein (SARS-CoV-2-S), receptor binding

domain (SARS-CoV-2- RBD), or in complex with full length ACE2 or soluble ACE2/ B°AT1, in a very short time. These structural studies have enabled us to understand the molecular basis of SARS-CoV-2 entry into human cells displaying ACE2 receptors [17,68,69,70]. Several structures of SARS-CoV-2-S were observed in multiple states (the prefusion, closed and partially open conformations and in complex with hACE2 receptor) with the receptor-binding domains (RBDs) either in an "up" or "down" conformation (Figure 5A, B). Of note: to engage the ACE2 receptor, the RBDs of S1 undergo hinge-like movements that either hide or expose the receptor binding regions and these conformations are referred to as "up" (receptor accessible) or "down" (receptor inaccessible) conformations, respectively. SARS-CoV-2-S structures show that the protein adopts a clover shaped homotrimeric structure, with three S1 heads that recognize a cognate cell surface receptor and a membrane-anchored trimeric S2 stalk, which contains the fusion machinery and is primarily α-helical[17] (Figure 5C, D). In the prefusion conformation of SARS-CoV-2-S protein, the RBDs rest above the trimeric S2 stalk, exhibiting two protomers in the "down" conformation and one protomer in the "up" conformation, which is a receptor-accessible state required for binding to a ACE2 receptor[17]. Overall the SARS-CoV-2-S ectodomain resembles the closely related SARS-CoV-S structure with an RMSD of 3.8Å over 959 Cα atoms, with a high degree of structural homology when individual domains of SARS-CoV-S and SARS-CoV-2-S were aligned[17].

5.1. SARS-CoV-2-RBD interactions with human ACE2 receptor

Multiple structures of SARS-CoV-2-RBD in complex with either full-length hACE2 or soluble hACE2 have shown that the extracellular peptidase domain (PD) of ACE2 recognizes the RBDs of S protein mainly through polar interactions[68,69]. Similar to other betacoronaviruses, the SARS-CoV-2-RBD structure suggested that it contains two subdomains: a

core subdomain containing a twisted five-stranded antiparallel β sheet (β1, β2, β3, β4 and β7) with a conserved disulfide bond between β2-β4. The other subdomain is receptor binding motif (RBM), located between β4 and β7 strand as an extended insertion (Figure 5E).

The RBM forms a gently concave surface that accommodates the N-terminal α-helix of the hACE2, and a series of hydrophilic residues that form a solid network of H-bond and salt bridges interactions (Figure 5F). In brief, strong polar contacts include CTD residues A475, N487, E484, Y453 that interact S19, Q24, K31, H34 of α1 helix of hACE2, respectively[11]. In addition, residues Q498, T500, N501 on the bulged loop forms a network of H-bonds with Y41, Q42, K353, R357 from ACE2[68]. Thus, overall virus-receptor interactions are dominated by polar contacts mediated by hydrophilic residues[11,68,69] (Figure 5G).

5.2. Comparison of the SARS-CoV-2-RBD and SARS-CoV-RBD interactions with human ACE2 receptor

The majority of the secondary structure elements shared between SARS-CoV- RBD (PDB ID: 2AJF) and SARS-CoV-2-RBD (PDB ID: 6LZG, 6M17) are well superimposed, with an RMSD of 0.475Å over 128 Cα atoms, with the exception of the receptor binding loop. Interestingly, these structures revealed that the majority of binding sites of SARS-CoV RBD in hACE2 also overlap with the SARS-CoV-2-RBD binding sites suggesting that the SARS-CoV-2-RBD: hACE2 complex is strikingly similar to the SARS-CoV-RBD: ACE2 structure with an RMSD of 0.431Å over 669 Cα atoms (Figure 5G, H). However, despite the overall similarity, a number of sequence variations were observed at the binding interface that may account for the difference in the affinities for hACE2 receptors. The detailed comparison of the receptor binding interfaces suggested that the SARS-CoV-2-RBD: ACE2 complex (PDB ID: 6VW1, 6M17) has larger buried surface areas (1773 $Å^2$ versus 1686 $Å^2$), has additional contacts (21 versus 17),

more Van der Waals interactions (288 versus 213) as wells as H-bonds (16 versus 1) than the SARS-CoV-RBD: hACE2 (PDB ID: 2AJF) complex[69]. Additionally, residue F486 in SARS-CoV-2-RBD forms stronger aromatic-aromatic interactions with Y83 of hACE2 than I472 of SARS-Co-V-RBD. Residue E484 in the SARS-CoV-2-RBD also forms stronger ionic interactions with K31 compared to P470 of SARS-CoV-RBD[69]. A sample collected from the state of Kerala in India on January 27, 2020, revealed a Arg408→ Ile408 mutation in the SARS-CoV-2-S protein (GenBank ID: MT012098.1), which otherwise is a strictly conserved residue in SARS-CoV, SARS-CoV-2, and bat SARS-like CoVs. Residue R408 is located near to the binding interface of both, the SARS-CoV-2-RBD: hACE2 (PDB: 6VW1) and SARS-CoV-RBD: hACE2 (PDB: 2AJF), complexes but appears not to be interacting directly with hACE2 in either case. However, R408 does forms a H-bond (3.3Å) with the glycan attached to N90 from hACE2; thus, potentially contributes to higher affinities observed for SARS-CoV2-RBD: hACE2 interactions than the SARS-CoV-RBD: ACE2 complex, where the corresponding R395 is located relatively further away (6.1Å) from N90 of hACE2. The Arg408→ Ile408 mutation that emerged in SARS-CoV-2 strain (GenBank ID: MT012098.1) suggested that the loss of H-binding capacity could potentially reduce ACE2 binding affinity. The equilibrium dissociation constants (K_D) for hACE2 interacting with the S proteins have indicated that the binding affinity of SARS-CoV-2-S is several fold higher than that of SARS-CoV[11,17,69].

Using cryo-EM the structure of full-length hACE2 in complex with SARS-CoV-2-RBD and B°AT1 (neutral amino acid transporter) was determined.This structure revealed that the ACE2: B°AT1complex is assembled as a dimer of heterodimers, where the collectrin-like domain of hACE2 drives homodimerization (PDB ID: 6M17)[68]. The SARS-CoV-2-RBD is recognized by the extracellular PD of ACE2 as described previously. Further it demonstrates that

a homodimeric ACE2 can accommodate two S protein trimers, each through a monomer of ACE2[68]. Interestingly, a superimposition of the ternary complex on RBD in the "down" conformation has indicated that PD clashes with the S protein, whereas in the "up" conformation (PDB 6VSB) no clashes are observed. This suggests that the "up" confirmation of RBD is a receptor-accessible state and is essential for the ACE2-receptor binding. Taken together, the overall interface between SARS-CoV2-RBD: ACE2 is very similar to the previously known SARS-CoV-RBD: ACE2 interface, and is dominated by the polar interactions as reported by different investigations[11,68,69]. These observations further suggest that SARS-CoV-2-RBD has increased atomic interactions with hACE2, which results in higher affinities compared to the SARS-CoV-RBD: hACE2 complex, which might be one of the reasons for enhanced human-to-human transmission of SARS-CoV-2.

5.3. SARS-CoV-2 Exhibits Distinct Epitope Features on the RBD from SARS-CoV

Numerous binding and neutralization epitopes have been identified on the spike protein of coronaviruses making the S protein an essential target for vaccine design[71,72,73]. Soon after the emergence of COVID-19 pandemic, some of the initial efforts were focused on screening SARS-CoV-S specific antibodies to find neutralizing antibody/antibodies for vaccine and drug development against SARS-CoV-2. The hypothesis behind these studies was based on significant sequence as well as structural similarities and, moreover, both viruses bind to the same receptor with overlapping epitopes. Therefore, it was expected that SARS-CoV specific antibody/antibodies alone or in combination can interfere or even inhibit SARS-CoV-2 and hACE2 receptor interactions.

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

It has been shown *in vitro* as well in animal models that monoclonal antibodies, such as 80R[74], CR3014[75], S230.15[76] and m396[76] can block binding of the S1 domain and hACE2 receptors by potently neutralizing SARS-CoV. However, CR3022[77] alone did not show neutralization but the mixture of CR3022 and CR3014 both showed neutralization of SARS-CoV in a synergistic fashion by recognizing different epitopes on the RBDs[75]. Of note, a report suggests that CR3022 can also neutralize SARS-CoV alone[78]. Interestingly, researchers from China tested several published SARS-CoV specific monoclonal antibodies and found that CR3022 can bind to the RBDs of SARS-CoV-2 with a K_D of 6.3 nM, whereas other antibodies, such as m396, CR3014 and S230.15 failed to bind to the SARS-CoV-2-S protein[17,79]. However, a low level of binding to SARS-CoV-2-S was observed with a SARS-CoV-S1 specific polyclonal antibody T62 (#40150-T62, Sino Biological Inc., Beijing, China) and it could poorly neutralize SARS-CoV-2-S protein mediated virus entry. Further analysis revealed that the epitope for T62 was likely located on the RBDs of SARS-CoV-2-S, but detailed information is lacking [34]. In an exciting study, the Wilson laboratory determined the crystal structure of CR3022 antibody in complex with SARS-CoV-2-RBD (PDB ID: 6W41) and revealed that CR3022 binds a highly conserved epitope that is distantly located from receptor-binding site, which enables cross-reactive binding, but could not neutralize SARS-CoV-2 *in vitro*[78] (Figure 5I). However, whether CR3022 can synergize with other SARS-CoV-2-RBD binding antibodies for neutralization requires further evaluation and study.

The SARS-CoV (GenBank: AY278488.2) and SARS-CoV-2 (GenBank: MN908947.3) spike proteins share about 76% amino acids sequence identity suggesting that the remaining 24% amino acids sequences, which are non-conserved might be responsible for antigenic differences between these two proteins. In the quest to find novel antibody binding epitopes on spike

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

proteins, Zheng et al. performed antibody epitope analysis, and surface epitope accessibility using bioinformatic tools to identify both weak and strong epitopes, which might be otherwise experimentally ignored[80]. Their analysis identified five shared epitopes along with 40 and 29 unique epitopes on the spike proteins of SARS-CoV and SARS-CoV-2, respectively. Among these unique epitopes, 92.7% originated from non-conserved regions, which might explain why most of the SARS-CoV specific antibodies discussed in this review did not bind to the spike protein of SARS-CoV-2[80]. Taken together, these results suggest the necessity to develop SARS-CoV-2 specific antibodies and vaccine candidates.

5.4. ACE2-independent receptors in viral pathogenesis

It is now established that both SARS-CoV and SARS-CoV-2 exploit angiotensin-converting enzyme 2 (ACE2) receptor to gain a host cell entry [10,23]; however, some studies indicate that in addition to ACE2, SARS-CoV might exploit other host factors such as vimentin (a cytoskeleton protein) and lectins (a glycoprotein) to mediate viral entry [81,82]. It is important to note that the precise role of lectins in SARS-CoV infection has not been explored extensively and the topic remains controversial. Jeffers et al., reported that SARS-CoV might use both ACE-2 and CD209L, a C-type lectin to invade host cell [82]. In contrast, Zhou et al., and others indicated that mannose-binding lectins interfere with viral entry, potentially by blocking other interactions [83,84].

The spike protein seems to be heavily glycosylated; however, the role of glycosylation in SARS-CoV-2 infection also remains an unexplored area. A recent *in vitro* study reported interactions between SARS-CoV-2 spike protein and C-type lectins as well as sialic-acid binding lectins; however, a major limitation of this study is that it doesn't provide any cell-based *in vivo* data and proper controls are missing [85]. SARS-CoV-2 seems to infect a diverse range of cell

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

types; therefore, it is reasonable to speculate that ACE2-independent interactions might provide an alternate route for viral invasion. A recent study reported that two T lymphocyte cell lines (MT-2 and A3.01), expressing very low levels of ACE2 were susceptible to SARS-CoV-2 infection, but not SARS-CoV [86]. This means that SARS-CoV-2 is capable of infecting cells, even when ACE2 expression is very low or it exploits another interaction partner(s). Notably, CD147 was recently reported to be a novel interaction partner of the SARS-CoV-2 spike protein, which facilitates host cell entry [87]. This observation could be interpreted to mean that in cell types where ACE2 expression is non-existent, SARS-CoV-2 utilizes another receptor to invade cells, although efficacy these studies should be further tested in primary human T cells. Given the importance of this topic and its massive impact on human lives, future studies will have to carefully evaluate, whether non-ACE2 interactions compete with ACE2 to inhibit viral entry or ACE2-independent interactions produce a synergistic effect with ACE2-mediated entry to exacerbate the symptoms of COVID-19.

8. Conclusions

The recent global outbreak of COVID-19 has killed almost 425 thousand[7] people and threatened the global economy, causing economic hardships to millions of people. Extensive progress has been made in terms of structure and function of the spike glycoproteins. Specifically, decade-long structural studies on the spike proteins of SARS-coronaviruses have designated six key residues (Y442, L472, N479, D480, T487 and Y491 for SARS-CoV)[67] in the RBDs that are critical for the host cell ACE2 receptor binding as well as for playing important roles in the cross-species transmission. Notably, five out of these six residues differ between the RBDs of SARS-CoV and SARS-CoV-2 S proteins, which have exhibited enhanced

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

binding between the RBDs of SARS-CoV-2 and ACE2 receptors. This might be one of the reasons behind widespread human-to-human transmission of SARS-CoV-2. In addition there are likely to be others factors that contribute to infectivity and pathogenicity of SARS-CoV-2, which are required to be investigated.

The trimeric prefusion structure of the SARS-CoV-2 spike protein was obtained in an asymmetric conformation where one protomer was observed in the “up” and other two in the “down” conformations. This phenomenon known as protein “breathing” was observed in the S1 domain while determining the trimeric prefusion structure, which suggested the mechanism used by the CR3022 antibody to access a cryptic epitope on the trimeric S protein that is otherwise not possible. Interestingly, a similar breathing phenomenon identified unique and conserved epitopes in the trimeric interface of influenza hemagglutinin protein recently. The antibodies binding to these cryptic epitopes did not inhibit viral infection *in vitro* but conferred *in vivo* protection[88,89]. A similar phenomenon was observed in case of the CR3022 monoclonal antibody; therefore, further *in vivo* studies are required as soon as a suitable animal model is established for SARS-CoV-2 studies. In the course of writing this review, two exciting reports became available: (i) an antibody 47D11 that is reported to neutralize SARS-CoV-2 as well as SARS-CoV in cell culture through an unknown mechanism, which is different from the virus neutralization process[90], (ii) an inactivated novel coronavirus vaccine (PiCoVacc) that is able to induce SARS-CoV-2 specific neutralizing antibodies in mice, rats and non-human primates. Additionally, data demonstrate that PiCoVacc vaccine provides partial to complete protection in macaques against SARS-CoV-2 challenge[91]. Future investigations are required to understand the mechanism of neutralization by these antibodies.

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

Last but not the least, glycosylation has been an important measure of virus antigenic properties and plays a critical role for the manufacturing of effective vaccines against HIV and influenza. Notably, the SARS-CoV-2 spike protein is densely decorated by host-derived heterogenous N-linked glycans as indicated by a site-specific glycosylation analysis undertaken by mass spectrometry. Specifically, each SARS-CoV-2 spike trimer displays 66 N-linked glycosylation sites with an elevation in oligomannose- and hybrid-type glycans compared to typical host-derived glycoproteins[92]. Finally, glycan profiling will be an important addition to measure antigen quality, and should be examined while producing glycoprotein-based vaccine candidates for COVID-19.

Though it is observed that SARS-CoV-2 binds to its receptor on the host cells with higher affinities than SARS-CoV, the fatality rate caused by SARS-CoV-2 (3.4%,) is significantly less than the reported rate of SARS-CoV (9-11%), as reported by the WHO. The reasons behind these differences remain elusive and future research will shed light on these variations. Recent sequencing data indicate that SARS-CoV-2 mutation rate is ~25 mutations per year. If these mutations enable more efficient virus spread and increased pathogenicity, then vaccine development will be a challenging task. Hopefully, future studies will be able to resolve these questions and come up with medications as well as vaccines against this deadly virus. Even with the vaccine and medications against this virus, future outbreaks of similar viruses and pathogens are likely to continue. Therefore, apart from curbing this outbreak, government policies and efforts should be made to formulate thorough measures to prevent future outbreaks of viruses and bacteria (there is already a significant threat from antibiotic-resistant bacteria).

References

1. Hamre D, Procknow JJ (1966) A new virus isolated from the human respiratory tract. Proc Soc Exp Biol Med 121: 190-193.
2. Kahn JS, McIntosh K (2005) History and recent advances in coronavirus discovery. Pediatr Infect Dis J 24: S223-227, discussion S226.
3. Tyrrell DA, Bynoe ML (1966) Cultivation of viruses from a high proportion of patients with colds. Lancet 1: 76-77.
4. Centre WM (2020) Coronavirus never before seen in humans is the cause of SARS.
5. Holmes KV (2003) SARS-associated coronavirus. N Engl J Med 348: 1948-1951.
6. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367: 1814-1820.
7. WHO (2020) Coronavirus Disease (COVID-19) Dashboard.
8. WHO (2020) Coronavirus Disease (COVID-19) Dashboard, United States of America.
9. Peck KM, Burch CL, Heise MT, Baric RS (2015) Coronavirus Host Range Expansion and Middle East Respiratory Syndrome Coronavirus Emergence: Biochemical Mechanisms and Evolutionary Perspectives. Annu Rev Virol 2: 95-117.
10. Li F (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol 3: 237-261.
11. Lan J, Ge J, Yu J, Shan S, Zhou H, et al. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581: 215-220.
12. Hofmann H, Pohlmann S (2004) Cellular entry of the SARS coronavirus. Trends Microbiol 12: 466-472.
13. Vennema H, Godeke GJ, Rossen JW, Voorhout WF, Horzinek MC, et al. (1996) Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. EMBO J 15: 2020-2028.
14. Siu YL, Teoh KT, Lo J, Chan CM, Kien F, et al. (2008) 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. J Virol 82: 11318-11330.
15. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, et al. (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science 368: 779-782.
16. Snijder EJ, Decroly E, Ziebuhr J (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. Adv Virus Res 96: 59-126.
17. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, et al. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367: 1260-1263.
18. Wentworth DE, Holmes KV (2001) Molecular determinants of species specificity in the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation. J Virol 75: 9741-9752.
19. Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, et al. (2013) Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495: 251-254.
20. Hulswit RJG, Lang Y, Bakkers MJG, Li W, Li Z, et al. (2019) Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proc Natl Acad Sci U S A 116: 2681-2690.
21. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, et al. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426: 450-454.

22. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, et al. (2005) Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci U S A 102: 7988-7993.
23. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, et al. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181: 271-280 e278.
24. Yob JM, Field H, Rashdi AM, Morrissy C, van der Heide B, et al. (2001) Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. Emerg Infect Dis 7: 439-441.
25. Halpin K, Young PL, Field HE, Mackenzie JS (2000) Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. J Gen Virol 81: 1927-1932.
26. Li W, Shi Z, Yu M, Ren W, Smith C, et al. (2005) Bats are natural reservoirs of SARS-like coronaviruses. Science 310: 676-679.
27. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, et al. (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A 102: 14040-14045.
28. WHO (2019) Middle East respiratory syndrome coronavirus (MERS-CoV).
29. Zhu N, Zhang D, Wang W, Li X, Yang B, et al. (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 382: 727-733.
30. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270-273.
31. Chan JF, Kok KH, Zhu Z, Chu H, To KK, et al. (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9: 221-236.
32. Lu R, Zhao X, Li J, Niu P, Yang B, et al. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395: 565-574.
33. WHO (2020) Naming the coronavirus disease (COVID-19) and the virus that causes it
34. Ou X, Liu Y, Lei X, Li P, Mi D, et al. (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 11: 1620.
35. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, et al. (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 7: 11.
36. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, et al. (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res 176: 104742.
37. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. Nat Med 26: 450-452.
38. Klenk HD, Garten W (1994) Host cell proteases controlling virus pathogenicity. Trends Microbiol 2: 39-43.
39. Steinhauer DA (1999) Role of hemagglutinin cleavage for the pathogenicity of influenza virus. Virology 258: 1-20.
40. Chan JF, Yuan S, Kok KH, To KK, Chu H, et al. (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395: 514-523.

41. Liu J, Liao X, Qian S, Yuan J, Wang F, et al. (2020) Community Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, Shenzhen, China, 2020. Emerg Infect Dis 26: 1320-1323.
42. BBC (2020) Coronavirus: Confirmed global cases pass one million.
43. Coronaviridae Study Group of the International Committee on Taxonomy of V (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5: 536-544.
44. Fung TS, Liu DX (2019) Human Coronavirus: Host-Pathogen Interaction. Annu Rev Microbiol 73: 529-557.
45. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, et al. (2012) Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol 86: 3995-4008.
46. Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17: 181-192.
47. Fouchier RA, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, et al. (2004) A previously undescribed coronavirus associated with respiratory disease in humans. Proc Natl Acad Sci U S A 101: 6212-6216.
48. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, et al. (2004) Identification of a new human coronavirus. Nat Med 10: 368-373.
49. Lau SK, Woo PC, Yip CC, Tse H, Tsoi HW, et al. (2006) Coronavirus HKU1 and other coronavirus infections in Hong Kong. J Clin Microbiol 44: 2063-2071.
50. Forster P, Forster L, Renfrew C, Forster M (2020) Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci U S A 117: 9241-9243.
51. Wang C, Liu Z, Chen Z, Huang X, Xu M, et al. (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol.
52. Wu A, Peng Y, Huang B, Ding X, Wang X, et al. (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 27: 325-328.
53. Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, et al. (2004) Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. J Virol 78: 13600-13612.
54. Nieto-Torres JL, Dediego ML, Alvarez E, Jimenez-Guardeno JM, Regla-Nava JA, et al. (2011) Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology 415: 69-82.
55. Voss D, Kern A, Traggiai E, Eickmann M, Stadler K, et al. (2006) Characterization of severe acute respiratory syndrome coronavirus membrane protein. FEBS Lett 580: 968-973.
56. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY (2016) Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov 15: 327-347.
57. Menachery VD, Yount BL, Jr., Debbink K, Agnihothram S, Gralinski LE, et al. (2015) A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med 21: 1508-1513.
58. Yong Jia GS, Yujuan Zhang, Keng-Shiang Huang, Hsing-Ying Ho, Wei-Shio Hor, Chih-Hui Yang, View ORCID ProfileChengdao Li, Wei-Lung Wang (2020) Analysis of the

mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. bioRxiv.
59. Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, et al. (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci U S A 111: E3900-3909.
60. Wang Q, Wu J, Wang H, Gao Y, Liu Q, et al. (2020) Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. Cell.
61. Yin W, Mao C, Luan X, Shen DD, Shen Q, et al. (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science.
62. Peng Q, Peng R, Yuan B, Zhao J, Wang M, et al. (2020) Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. Cell Rep 31: 107774.
63. Warren TK, Jordan R, Lo MK, Ray AS, Mackman RL, et al. (2016) Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. Nature 531: 381-385.
64. de Wit E, Feldmann F, Cronin J, Jordan R, Okumura A, et al. (2020) Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. Proc Natl Acad Sci U S A 117: 6771-6776.
65. Wang M, Cao R, Zhang L, Yang X, Liu J, et al. (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res 30: 269-271.
66. Sheahan TP, Sims AC, Zhou S, Graham RL, Pruijssers AJ, et al. (2020) An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. Sci Transl Med 12.
67. Wan Y, Shang J, Graham R, Baric RS, Li F (2020) Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol 94.
68. Yan R, Zhang Y, Li Y, Xia L, Guo Y, et al. (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367: 1444-1448.
69. Wang Q, Zhang Y, Wu L, Niu S, Song C, et al. (2020) Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 181: 894-904 e899.
70. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, et al. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181: 281-292 e286.
71. Qiu T, Mao T, Wang Y, Zhou M, Qiu J, et al. (2020) Identification of potential cross-protective epitope between a new type of coronavirus (2019-nCoV) and severe acute respiratory syndrome virus. J Genet Genomics 47: 115-117.
72. Reguera J, Santiago C, Mudgal G, Ordono D, Enjuanes L, et al. (2012) Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLoS Pathog 8: e1002859.
73. Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, et al. (2006) Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J Biol Chem 281: 15829-15836.
74. Sui J, Li W, Murakami A, Tamin A, Matthews LJ, et al. (2004) Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. Proc Natl Acad Sci U S A 101: 2536-2541.

75. ter Meulen J, van den Brink EN, Poon LL, Marissen WE, Leung CS, et al. (2006) Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. PLoS Med 3: e237.
76. Zhu Z, Chakraborti S, He Y, Roberts A, Sheahan T, et al. (2007) Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. Proc Natl Acad Sci U S A 104: 12123-12128.
77. Coughlin M, Lou G, Martinez O, Masterman SK, Olsen OA, et al. (2007) Generation and characterization of human monoclonal neutralizing antibodies with distinct binding and sequence features against SARS coronavirus using XenoMouse. Virology 361: 93-102.
78. Yuan M, Wu NC, Zhu X, Lee CD, So RTY, et al. (2020) A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science 368: 630-633.
79. Tian X, Li C, Huang A, Xia S, Lu S, et al. (2020) Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg Microbes Infect 9: 382-385.
80. Zheng M, Song L (2020) Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. Cell Mol Immunol 17: 536-538.
81. Yu YT, Chien SC, Chen IY, Lai CT, Tsay YG, et al. (2016) Surface vimentin is critical for the cell entry of SARS-CoV. J Biomed Sci 23: 14.
82. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, et al. (2004) CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc Natl Acad Sci U S A 101: 15748-15753.
83. Zhou Y, Lu K, Pfefferle S, Bertram S, Glowacka I, et al. (2010) A single asparagine-linked glycosylation site of the severe acute respiratory syndrome coronavirus spike glycoprotein facilitates inhibition by mannose-binding lectin through multiple mechanisms. J Virol 84: 8753-8764.
84. Ip WK, Chan KH, Law HK, Tso GH, Kong EK, et al. (2005) Mannose-binding lectin in severe acute respiratory syndrome coronavirus infection. J Infect Dis 191: 1697-1704.
85. Chiodo F, Bruijns SCM, Rodriguez E, Li RJE, Molinaro A, et al. (2020) Novel ACE2-Independent Carbohydrate-Binding of SARS-CoV-2 Spike Protein to Host Lectins and Lung Microbiota. bioRxiv: 2020.2005.2013.092478.
86. Wang X, Xu W, Hu G, Xia S, Sun Z, et al. (2020) SARS-CoV-2 infects T lymphocytes through its spike protein-mediated membrane fusion. Cell Mol Immunol: 1-3.
87. Wang K, Chen W, Zhou Y-S, Lian J-Q, Zhang Z, et al. (2020) SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. bioRxiv: 2020.2003.2014.988345.
88. Bangaru S, Lang S, Schotsaert M, Vanderven HA, Zhu X, et al. (2019) A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface. Cell 177: 1136-1152 e1118.
89. Watanabe A, McCarthy KR, Kuraoka M, Schmidt AG, Adachi Y, et al. (2019) Antibodies to a Conserved Influenza Head Interface Epitope Protect by an IgG Subtype-Dependent Mechanism. Cell 177: 1124-1135 e1116.
90. Wang C, Li W, Drabek D, Okba NMA, van Haperen R, et al. (2020) A human monoclonal antibody blocking SARS-CoV-2 infection. Nat Commun 11: 2251.
91. Gao Q, Bao L, Mao H, Wang L, Xu K, et al. (2020) Rapid development of an inactivated vaccine candidate for SARS-CoV-2. Science.
92. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. Science.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 July 2020 doi:10.20944/preprints202005.0260.v2

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

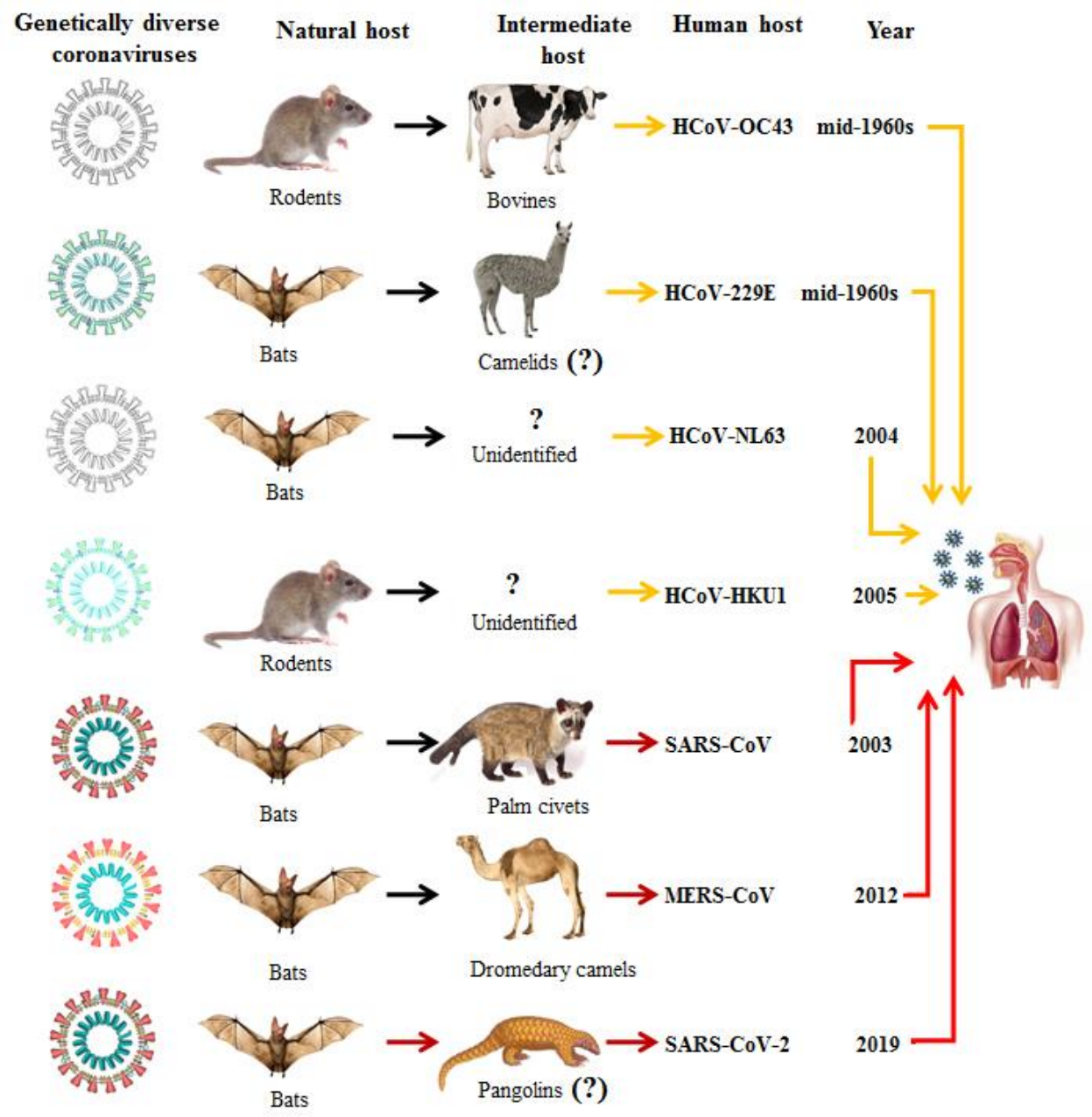


Figure 1. Origin and transmission of pathogenic human coronaviruses. Yellow and red arrows indicate mild and severe infections in humans, respectively. The figure is inspired from Jie Cui et al. [49] and the illustrations of coronaviruses (left) are adapted from "Desiree Ho, Innovative Genomics Institute", available at https://innovativegenomics.org/free-covid-19-illustrations/.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 July 2020 doi:10.20944/preprints202005.0260.v2

Peer-reviewed version available at *Plos Pathogens* **2020**, *16*; doi:10.1371/journal.ppat.1008762

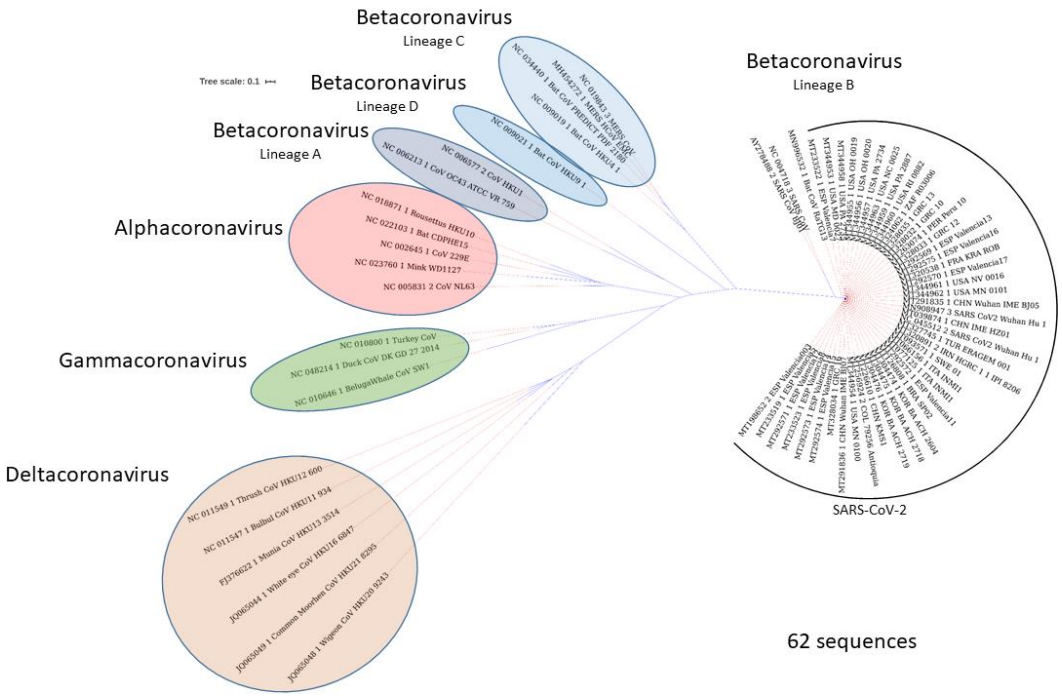
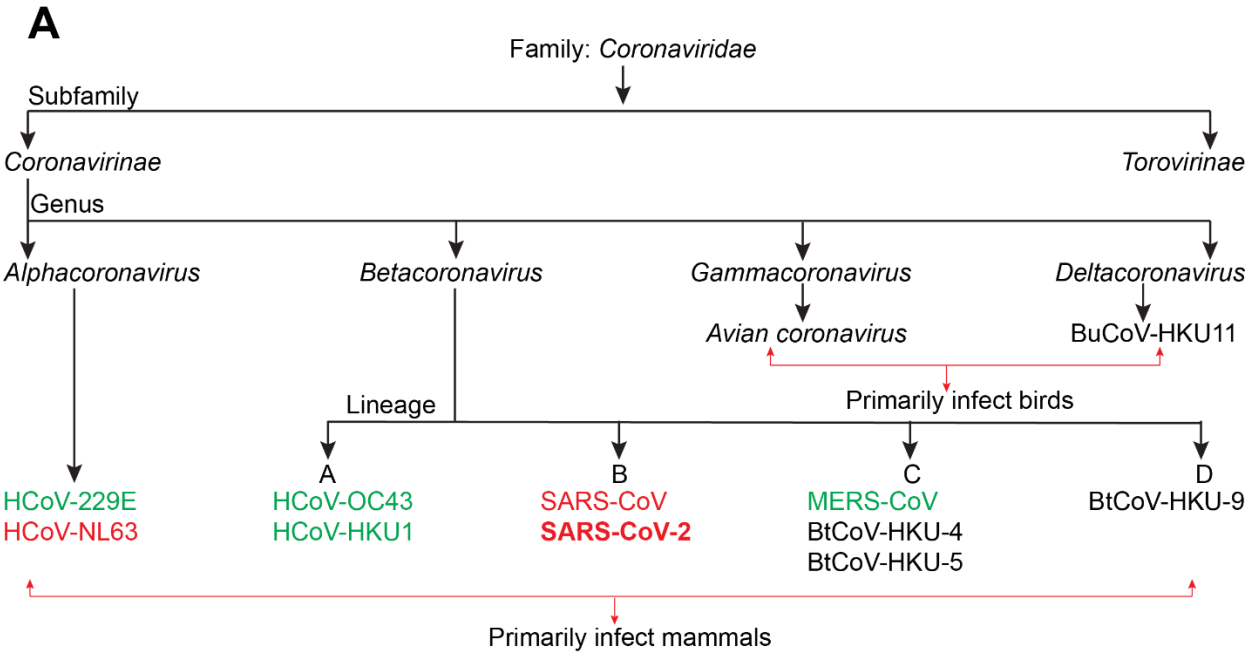
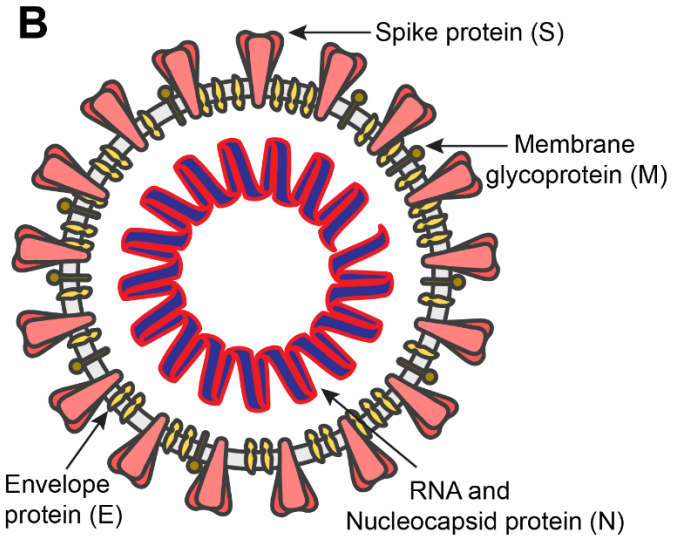


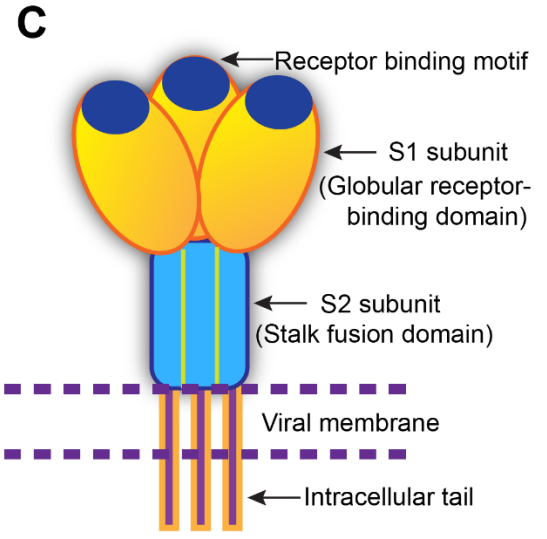
Figure 2: Phylogenetic relationships in the coronavirinae subfamily: the subfamily is formed by four genera: *Alphacoronavirus, Betacoronavirus* (linease A, B, C, and D), *Gammacoronavirus*, and *Deltacoronavirus*. We randomly picked 62 SARS-CoV-2 genome sequences, representing 15 different countries, together with other coronavirinae subfamily members. The phylogenetic tree was created using NgPhylogeny.fr tool. The analysis indicates that SARS-CoV-2 has a close relationship with Bat coronavirus RaTG13, and SARS-CoV; therefore classified as a new member of the lineage B *Betacoronavirus*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 July 2020 doi:10.20944/preprints202005.0260.v2

Peer-reviewed version available at *Plos Pathogens* **2020**, *16*; doi:10.1371/journal.ppat.1008762







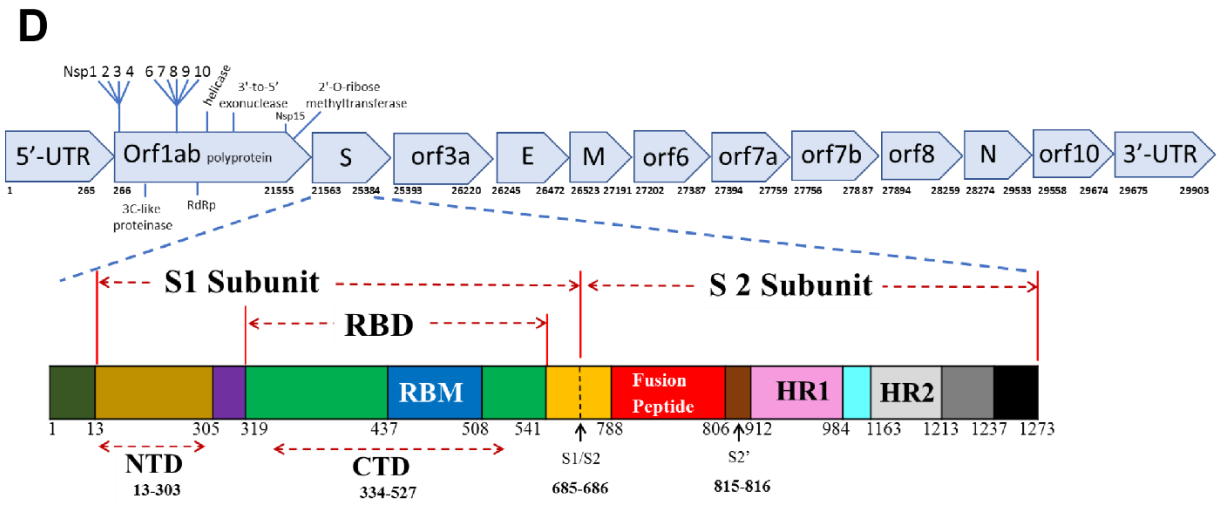


Figure 3. (A) Classification of coronaviruses: the seven known human coronaviruses (HCoVs) are shown in green and red. HCoVs in red bind the host receptor ACE2. **(B)** Schematic of the SARS-CoV-2 structure, illustration of virus is adapted from "Desiree Ho, Innovative Genomics Institute", available at https://innovativegenomics.org/free-covid-19-illustrations/. **(C)** Cartoon depicts key features and the trimeric structure of the SARS-CoV-2 spike protein. (**D**) Schematic of SARS-CoV-2 genome (top) and spike protein (bottom); annotations are adapted from the NCBI (NC_045512.2) and Expasy (https://covid-19.uniprot.org/uniprotkb/P0DTC2), respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 July 2020 doi:10.20944/preprints202005.0260.v2

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

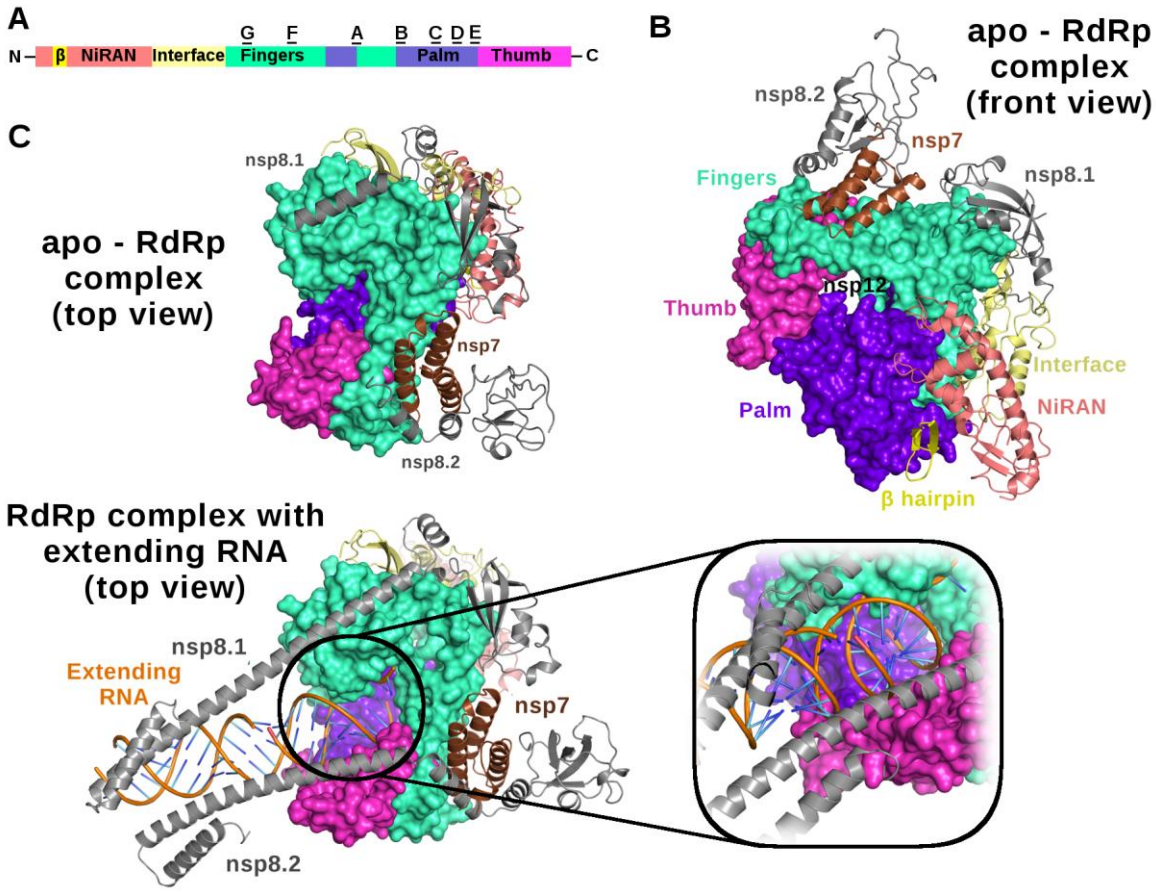


Figure 4. Cryo-EM structure of RNA dependent RNA polymerase (RdRp) of SARS-CoV-2. **(A)** The domain architecture of RdRp or nsp12 gene of SARS-CoV-2 subdivided into NiRAN, interface, finger, palm, and thumb subdomains, with A-G indicating the conserved motifs. **(B)** The cryo-EM structure of apo RdRp, PDB ID: 7BV1, complex (designated as front view) consisting of RdRp or nsp12, nsp7 (brown) and two chains of nsp8 (nsp8.1 and nsp8.2, both in grey), the subdomains of RdRp are colored according to the scheme (A) mentioned previously. A clear groove formed by the intersection of finger and thumb can be seen. While the nsp8.1 interacts directly with nsp12, nsp8.2 binds to nsp7 that interacts with nsp12, **(C)** The cryo-EM structure of RdRp complex bound to RNA, PDB ID: 6YYT, showing the two nsp8 chains stabilizing the extending RNA chain with their extended alpha helix (designated here as top view). The active site is expanded to show the RNA molecules coming out of the groove formed between the finger and the thumb.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 July 2020 doi:10.20944/preprints202005.0260.v2

Peer-reviewed version available at Plos Pathogens 2020, 16; doi:10.1371/journal.ppat.1008762

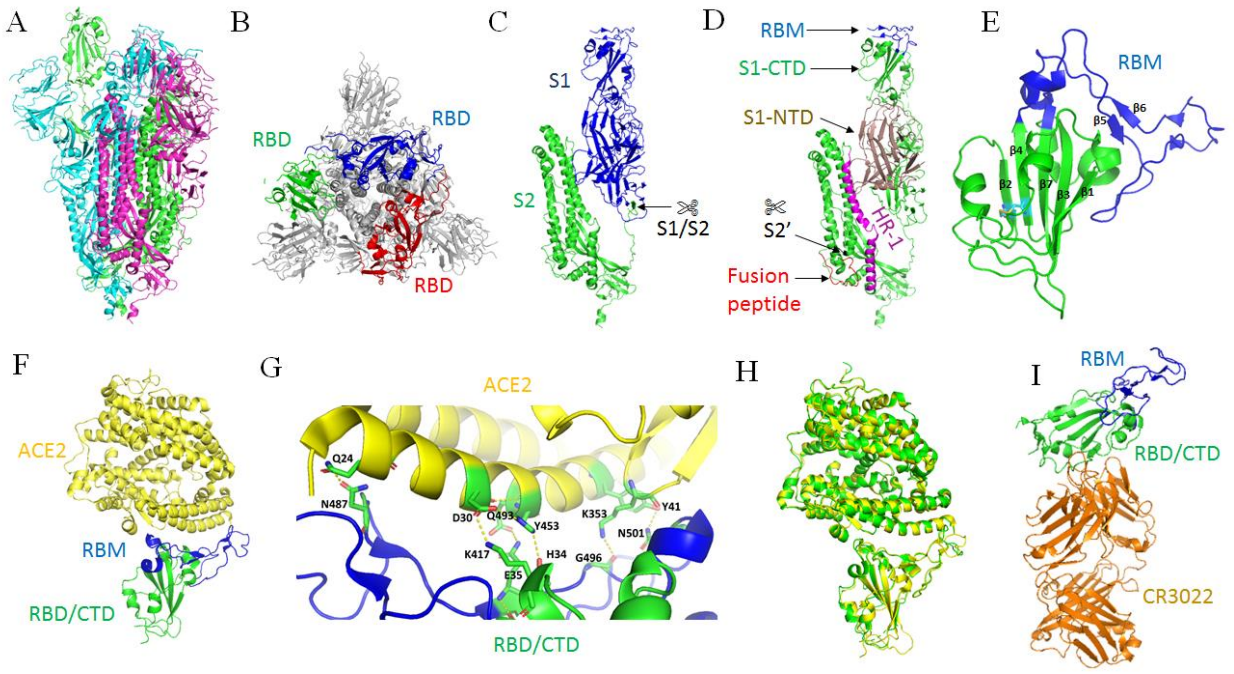


Figure 5. Structure of the SARS-CoV-2 spike protein alone and in complex with ACE2 receptor. **(A)** Side view of the trimeric SARS-CoV-2 spike ectodomain in the prefusion state (PDB ID: 6VSB). The protomer in green is in the "up" conformation and other two protomers in red and cyan are in "down" conformation. **(B)** Top view of the trimeric spike protein showing receptor binding domains (RBDs) in red, blue, and green on each protomer. **(C)** Structure of a single protomer showing the receptor-binding subunit S1 in blue and the membrane-fusion subunit S2 in green. The Furin-like protease site at the boundary of S1/S2 subunits is depicted. **(D)** The S1 subunit contain the receptor binding motif (RBM) in the CTD region in blue, and the NTD region in sand. The S2 subunit contain the fusion peptide in red, second cleavage site S2' in black, and HR1 in pink. **(E)** Structure of the RBD, core subdomain in green and RBM in blue (PDB ID: 6LZG). **(F)** Structure of the SARS-CoV-2-RBD in complex with ACE2 receptor (PDB ID: 6LZG). **(G)** SARS-CoV-2-RBD: ACE2 receptor polar interface shown by specific residues. **(H)** Structural similarity between the SARS-CoV-RBD: hACE2 (green) and SARS-CoV-2-S-CTD: hACE2 (yellow) complexes. **(I)** Crystal structure of the SARS-CoV-2-RBD (green) in complex with a monoclonal antibody CR3022 (orange). The RBM and CR3022 binding sites do not overlap and are distantly located on the RBD (PDB ID: 6W41). The figures are prepared using Pymol.

Supplementary Figure S1

Figure S1. The figure depicts the phylogenetic tree drawn for 69 coronavirus genomic sequences including SARS-Co-2 sequences. Sequences belonging to different coronavirinae subfamily have been labelled.