



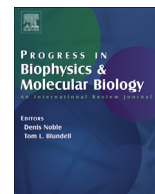
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Human coronavirus spike protein–host receptor recognition

Lalitha Guruprasad

School of Chemistry, University of Hyderabad, Hyderabad, 500046, India



ARTICLE INFO

Article history:

Received 27 July 2020

Received in revised form

12 October 2020

Accepted 23 October 2020

Available online 31 October 2020

Keywords:

Human coronavirus

SARS-CoV

SARS-CoV-2

MERS-CoV

HCoV-HKU1

HCoV-NL63

HCoV-OC43

HCoV-229E

Spike protein

Angiotensin-converting enzyme 2

Dipeptidyl peptidase 4

Amino peptidase N

Sialic acid

Receptor binding domain

ABSTRACT

A variety of coronaviruses (CoVs) have infected humans and caused mild to severe respiratory diseases that could result in mortality. The human CoVs (HCoVs) belong to the genera of α - and β -CoVs that originate in rodents and bats and are transmitted to humans via zoonotic contacts. The binding of viral spike proteins to the host cell receptors is essential for mediating fusion of viral and host cell membranes to cause infection. The SARS-CoV-2 originated in bats (RaTG13 SARS-CoV) and is transmitted to humans via pangolins. The presence of 'PRRA' sequence motif in SARS-CoV-2 spike proteins from human, dog, cat, mink, tiger and lion suggests a common viral entry mechanism into host cells. In this review, we discuss structural features of HCoV spike proteins and recognition of host protein and carbohydrate receptors.

© 2020 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	40
2. Coronavirus	40
3. SARS-CoV and SARS-CoV-2	44
4. MERS-CoV	46
5. HCoV-OC43	48
6. HCoV-HKU1	48
7. HCoV-NL63	49
8. HCoV-229E	49
9. Conclusions and perspectives	50
Author contribution	51
Declaration of competing interest	51
Acknowledgements	51
Supplementary data	51
References	51

E-mail address: lalitha.guruprasad@uohyd.ac.in.

1. Introduction

During the past two decades, life-threatening diseases caused by human coronaviruses (HCoVs); severe acute respiratory syndrome coronavirus (SARS-CoV) in China (2002) (Drosten et al., 2003; Ksiazek et al., 2003; Peiris et al., 2003) and Middle East respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia (2012) (Zaki et al., 2012) have resulted in major epidemics in several countries leading to a number of deaths. In addition, HCoV-229E (named after the specimen code 229E), HCoV-OC43 (organ culture 43), HCoV-NL63 (Netherlands 63) and HCoV-HKU1 (Hong Kong University 1) responsible for causing common colds and mild to acute endemic respiratory illnesses have been reported among the human populations from time to time. The HCoV-229E (McIntosh et al., 1967) and HCoV-OC43 (Kaye et al., 1971; Hendley et al., 1972) first identified during 1960s as human respiratory pathogens were isolated from Salisbury, United Kingdom. The HCoV-NL63 was isolated during 2002–2003 (van Der Hoek et al., 2004; Fouchier et al., 2004) and HCoV-HKU1 was isolated during early 2005 (Woo et al., 2005) from patients with pneumonia. Currently the world is worst hit by a deadly infectious pandemic coronavirus disease 2019 (COVID-19) caused by novel SARS coronavirus-2 (SARS-CoV-2). COVID-19 was first reported from Wuhan City, Hubei-1 Province, China, during December 2019 and in less than 7 months spread throughout the world. The SARS-CoV-2 was first reported from individuals to have been in contact with wildlife animals at the live animal and seafood market in Jiangnan District, Wuhan (Zhu et al., 2020). The SARS-CoV-2 is observed to spread more rapidly compared to other HCoVs. The transmission of SARS-CoV-2 in humans takes place via contact with respiratory droplets during coughing and sneezing, or surface contacts. The World Health Organization (WHO) has acknowledged evidence of the emergence of airborne transmission of SARS-CoV-2. The prominent symptoms of this viral infection are fever, dry cough, tiredness, loss of smell and taste, severe respiratory illness, and when compounded with conditions of comorbidity, may lead to blood clots in lungs, or respiratory failure, severe hypoxaemia, renal failure, septic shock, multiple organ failure, cardiogenic shock, resulting in death (Vincent and Taccone, 2020). Despite the re-emergence of HCoV infections in altered forms resulting in mortality, there have been no drugs or vaccines to treat them. Some drugs are being evaluated for use as repurposed drugs for COVID-19 (Thorlund et al., 2020; Shaffer 2020) and a list of these drugs is mentioned in Table 1. Convalescent plasma therapy is another

intervention that is being used to treat COVID-19 patients. Design and development of vaccines are underway for combating COVID-19 in several countries. Some of these vaccines have made considerable progress and are in phase III clinical trials; Johnson & Johnson Janssen (Country: USA, Name: JNJ-78436735), Novavax (USA, NVX-CoV2373), Moderna Therapeutics (USA, mRNA-1273), Pfizer (USA, BNT162b2), University of Oxford–AstraZeneca–IQVIA–Serum Institute of India (UK, AZD1222), Sinovac (China, CoronaVac), CanSino Biologics (China, Ad5-nCoV), The Gamaleya National Center of Epidemiology and Microbiology (Russia, Sputnik V) and Sinopharm (China). The methods currently observed for preventing spread of the disease include, use of proper masks to cover nose and mouth, maintenance of social distancing and personal hygiene. COVID-19 has made severe indentation to the socio-economic status in all countries. As of October 11, 2020, SARS-COV-2 has spread to 214 countries with ~37,471,071 confirmed COVID-19 cases and ~1,077,509 deaths. There is an urgent need to discover specific drugs and vaccines to combat this deadly infectious disease.

2. Coronavirus

Orthocoronavirinae is a subfamily of the virus superkingdom; Riboviria (clade); Orthornavirae (kingdom); Pisuviricota (phylum); Pisoniviricetes (class); Nidovirales (order); Coronavirineae (suborder), and Coronaviridae (family) according to the Coronaviridae Study Group (CSG) of the International Committee on Taxonomy of Viruses (ICTV) (Gorbalenya et al., 2020). Orthocoronavirinae subfamily is classified into four genera, α -, β -, γ - and δ -CoVs (Woo et al., 2009). Phylogenetic evidence has shown, that rodents and bats serve as the genetic source of most α -CoVs and β -CoVs, while birds are the main reservoir of γ -CoVs and δ -CoVs (Su et al., 2016). HCoV-OC43 and HCoV-HKU1 have their origins in rodents (Su et al., 2016; Forni et al., 2017; Cui et al., 2019) and HCoV-NL63, HCoV-229E, SARS-CoV, MERS-CoV and SARS-CoV-2 are considered to have originated from bats (Donaldson et al., 2010; Huynh et al., 2012; Pfeifferle et al., 2009; Hu et al., 2015; Zhou et al., 2020). The CoVs of zoonotic origin during evolution of the virus acquire sufficient genomic mutations to be able to adapt to a new host. Zoonotic contacts compounded with genomic recombination, facilitate cross-species transmission of the virus that result in the spread of CoVs among mammals eventually causing threat to human health and life. In order to gain the ability to infect a human host, CoV genomes undergo significant mutations during evolution, importantly in spike proteins (Song et al., 2005; Menachery et al., 2016).

Table 1
Repurposed drugs in trials for COVID-19.

Drug	Original disease condition	Mechanism of action in COVID-19
Chloroquine, Hydroxychloroquine	Malaria	Increases pH in endosomes and prevents fusion and entry of virus into the host cell. Inhibits terminal glycosylation of ACE-2
Remdesivir	Ebola viral infection	SARS-CoV-2 RNA dependent RNA polymerase inhibitor
Favipiravir	Influenza viral infection	SARS-CoV-2 RNA dependent RNA polymerase inhibitor
Azithromycin	Broad-spectrum antibiotic to treat respiratory, enteric and urinary bacterial infections	Antiviral effects in bronchial epithelial cells
Lopinavir, Ritonavir	Human immunodeficiency virus	SARS-CoV-2 protease inhibitors
Nafamostat, Camostat	Human pancreatitis	SARS-CoV-2 protease inhibitors
Famotidine	Acidity	Potential inhibitor of the 3-chymotrypsin-like protease
Umifenovir	Influenza and other respiratory infections	Blocks entry of virus into host target cells. Inhibits synthesis of viral RNA
Nitazoxanide	Broad-spectrum drug for parasitic, bacterial and viral infections	Inhibits replication of the virus and protein expression
Fluvoxamine	Anti-depressant drug	Inhibits cytokine production
Dexamethasone	Bacterial infections with inflammation	Decreases cytokine storm and hyperinflammation
Tocilizumab, Sarilumab, Itolizumab, Bevacizumab	Inflammatory and autoimmune conditions	Alleviates inflammation of the lungs by suppressing cytokine storm

The classification of HCoVs implicated in human respiratory diseases, along with representative reference genomes, mode of viral transmission, carbohydrates and proteins that act as receptors of HCoVs required for human infection are shown in Table 2.

HCoVs are enveloped spherical shaped virions with 100–120 nm diameter. These are Baltimore class IV (Baltimore 1971) positive-sense single-stranded RNA viruses with 27–31 kb genome size that are translated into structural and non-structural proteins (NSPs). From the 5' proximal direction, ~two-thirds of the HCoV genomes code for two large open reading frames (ORFs), ORFs 1a and 1b genes that utilize one overlapping base and code for NSPs. The latter one-third of the genome encodes for the structural proteins; spike, envelope, membrane and nucleocapsid and genome dependent ORFs. In HCoV-OC43 and HCoV-HKU1 genomes, an additional protein, hemagglutinin esterase (HE) is also present. The name coronavirus stems from the characteristic club-shaped corona-like projections of the spike proteins as seen in the electron microscopy structures of the virus surface (Almeida et al., 1968; Lai and Cavanagh 1997). The spike protein is required for attachment to the host cell-specific receptors and subsequent catalysis of the virus-host cell membrane fusion. The envelope protein is a small integral membrane protein comprising 75 amino acid residues that plays a role in the virus life cycle. The membrane protein is a type III glycoprotein consisting of a short amino-terminal ecto-domain, a triple-spanning transmembrane domain, and a long carboxy-terminal endo-domain. The nucleocapsid protein is a highly basic phosphoprotein that modulates viral RNA synthesis. The short projections of HE are present on the surface of HCoV-OC43 and HCoV-HKU1 virions. The proteins encoded by the genomes of HCoVs are shown in Table 3.

CoVs enter the human cells through the fusion of viral and host cellular membranes, mediated by the interaction between the viral spike glycoprotein and host receptor protein and/or carbohydrate (Belouzard et al., 2012, Li, 2015). The spike proteins also called fusion proteins are heavily glycosylated, single pass integral membrane proteins that assemble as ~500 kDa homo-trimers projecting ~18–20 nm from viral envelopes. The multiple sequence alignment of representative HCoV spike proteins is shown in Supplementary Fig. 1. As per the phylogenetic tree in Fig. 1A, the α -CoV spike proteins (HCoV-NL63 and HCoV-229E) are grouped into one clade, the β -CoVs (SARS-CoV and SARS-CoV-2) and (HCoV-OC43 and HCoV-HKU1) into distinct clades and the MERS-CoV spike protein as yet another clade.

The spike protein comprises S1 subunit towards the N-terminus followed by membrane-proximal S2 subunit that consists of hydrophobic fusion peptide and heptad repeat regions. Towards the

C-terminus, there is a single membrane-spanning α -helix and an intracellular segment rich in cysteines. The S1 subunit is divided into four domains; S1^A, S1^B, S1^C and S1^D. The larger S1^A and S1^B domains are termed N-terminal domain (NTD) and C-terminal domain (CTD), respectively. The S1^A domain is characterized by a galectin-like fold. The S1^B domain is characterized either by a five-stranded anti-parallel β -sheet as in β -HCoVs or a six-stranded β -sandwich as in α -HCoVs. The S1^C and S1^D domains are associated with a five-stranded β -sheet structure.

Among HCoVs, the S1 subunit varies in length and amino acid sequence, while the S2 subunits share relatively high sequence homology (Supplementary Fig. 1). The 'linker' region connecting the S1^A and S1^B domains comprises ~32 amino acid residues and contributes one β -strand each to the S1^C and S1^D domains. The S2 subunit comprises three long α -helices, multiple α -helical segments and extended twisted β -sheets that span up to the viral membrane proximal end folding into a β -sheet domain. The amino acid sequences in the region connecting the S1 and S2 subunits are variable. The sequence variation among HCoVs in this region provides the flexibility for spike proteins to accommodate the cleavage sites for recognition by different proteases that is necessary for cellular entry of virion into the host cells. For instance, it has been reported that the 'PRRA' sequence motif is gained by SARS-CoV-2 providing a furin cleavage site (Ou et al., 2020). The loops connecting the domains within the S1 subunit and between the S1 and S2 subunits provide the flexibility for structural rearrangements of the protein in its pre-fusion, post-fusion states, binding to host cell receptors and carbohydrates, and for presentation of the cleavage sites at the S1/S2 and S2' sites during infection. The linear sequence and cartoon representation of the three-dimensional structure of human SARS-CoV-2 spike glycoprotein obtained from the Protein Data Bank (PDB) (Berman et al., 2000) (PDB code: 6VSB A-chain) are shown in Fig. 1B and Fig. 1C, respectively.

The HCoV spike proteins undergo structural alterations during various stages of host receptor recognition and viral-host cell membrane fusion. The different conformational states of the SARS-CoV spike glycoproteins during virus entry are reported (Song et al., 2018). Initially, the spike glycoprotein adopts the pre-fusion trimer structure with the S1^B in "Down" (or closed) conformation. Upon binding with host cell receptor during infection, the spike protein undergoes a large conformational transition from pre-fusion (closed) state to a post-fusion (open) state such that the S1^B domain is converted to the "Up" conformational state. This promotes the hydrophobic fusion peptide to be exposed thereby facilitating the viral and cellular membranes fusion by bringing them closer. The conformational changes present in the SARS-CoV

Table 2
Human coronaviruses.

Coronavirus	Transmission	Disease outcome	Human receptor	Classification	Reference genome NCBI ID (size)
HCoV-229E	Bats to humans through alpacas, camelids	Mild respiratory symptoms in immunocompromised patients	Amino-peptidase N	α -CoV	NC_002645.1 (27,317 bp)
HCoV-NL63	Bats to humans through an unknown intermediate	Mild respiratory symptoms in immunocompromised patients	Angiotensin- converting enzyme 2, Heparan sulfate	α -CoV	NC_005831 (27,553 bp)
HCoV-OC43	Rodents to humans through cattle	Mild respiratory symptoms in immunocompromised patients	9-O-acetylsialic acids	β -CoV	NC_006213.1 (30,741 bp)
SARS-CoV	Bats to humans through wild animals, palm civets	Acute pneumonia and respiratory disease	Angiotensin- converting enzyme 2, C-type lectin, Pulmonary surfactant protein D	β -CoV	NC_004718.3 (29,751 bp)
MERS-CoV	Bats to humans through dromedary camels	Acute pneumonia and respiratory disease	Dipeptidyl-peptidase 4, Sialic acid	β -CoV	NC_038,294 (30,111 bp)
HCoV-HKU1	Rodents to humans through an unknown intermediate	Mild respiratory symptoms in immunocompromised patients	9-O-acetylsialic acid	β -CoV	NC_006577.2 (29,926 bp)
SARS-CoV-2	Bats to humans possibly through pangolins	Acute pneumonia and respiratory disease	Angiotensin- converting enzyme 2	β -CoV	NC_045512.2 (29,903 bp)

Table-3
UniProtKB IDs of human coronavirus proteins.

HCoV-229E	ORF1ab P0C6X1; 1–111 (NSP1), 112–897 (NSP2), 898–2484 (NSP3), 2485–2965 (NSP4), 2966–3267 (3C-like proteinase), 3268–3546 (NSP6), 3547–3629 (NSP7), 3630–3824 (NSP8), 3825–3933 (NSP9), 3934–4068 (NSP10), 4069–4995 (RdRp), 4996–5592 (helicase), 5593–6110 (Exoribonuclease), 6111–6458 (Uridylate-specific endoribonuclease), 6459–6758 (Putative 2'-O-methyl transferase), P15423 (spike glycoprotein, 1173 aa), P19739 (NSP4a, 133 aa), P19740 (NSP4b, 88 aa), P19741 (Envelope small membrane protein, 133 aa), P15422 (membrane protein, 225 aa), P15130 (Nucleoprotein, 389 aa).
HCoV-NL63	ORF1ab P0C6X5; 1–110 (NSP1), 111–898 (NSP2), 899–2462 (NSP3), 2463–2939 (NSP4), 2940–3242 (3C-like proteinase), 3243–3521 (NSP6), 3522–3604 (NSP7), 3605–3799 (NSP8), 3800–3908 (NSP9), 3909–4043 (NSP10), 4044–4970 (RdRp), 4971–5567 (helicase), 5568–6085 (Exoribonuclease), 6086–6429 (Uridylate-specific endoribonuclease), 6430–6729 (Putative 2'-O-methyl transferase), Q6Q1S2 (Spike glycoprotein, 1356 aa), Q6Q1S1 (NSP3, 225 aa), Q6Q1S0 (Envelope small membrane protein, 77 aa), Q6Q1R9 (Membrane protein, 226 aa), Q6Q1R8 (Nucleoprotein, 377 aa).
SARS-CoV	ORF1ab P0C6X7; 1–180 (Host translation inhibitor NSP1), 181–818 (NSP2), 819–2740 (NSP3), 2741–3240 (NSP4), 3241–3546 (3C-like proteinase), 3547–3836 (NSP6), 3837–3919 (NSP7), 3920–4117 (NSP8), 4118–4230 (NSP9), 4231–4369 (NSP10), 4370–5301 (RdRp), 5302–5902 (helicase), 5903–6429 (Proofreading exoribonuclease), 6430–6775 (Uridylate-specific endoribonuclease), 6776–7073 (2'-O-methyl transferase), P59594 (Spike glycoprotein, 1255 aa), P59632 (Protein 3a, 274 aa), P59633 (Non-structural protein 3 b, 154 aa), P59637 (Envelope small membrane protein, 76 aa), P59596 (Membrane protein 221 aa), P59634 (NSP6, 63 aa), P59635 (Protein 7a, 122 aa), Q7TFA1 (Protein non-structural 7 b, 44 aa), Q7TFA0 (Protein non-structural 8a, 39 aa), Q80H93 (Non-structural protein 8 b, 84 aa), P59595 (Nucleoprotein, 9a, 422 aa), P59636 (Protein 9 b, 98 aa), Q7TLC7 (ORF14, 70 aa).
MERS-CoV	ORF1ab K9N7C7; 1–193 (Host translation inhibitor NSP1), 194–853 (NSP2), 854–2740 (Papain-like proteinase), 2741–3247 (NSP4), 3248–3533 (3C-like proteinase), 3554–3845 (NSP6), 3846–3928 (NSP7), 3929–4127 (NSP8), 4128–4237 (NSP9), 4238–4377 (NSP10), 4378–5310 (RdRp), 5311–5908 (helicase), 5909–6432 (Guanine-N7 methyltransferase), 6433–6775 (Uridylate-specific endoribonuclease), 6776–7078 (2'-O-methyl transferase), K9N5Q8 (Spike glycoprotein, 1353 aa), K9N796 (NSP ORF3, 103 aa), K9N5R3 (Envelope small membrane protein, 82 aa), K9N4V0 (NSP ORF4a, 109 aa), K9N643 (Non-structural protein ORF4b, 246 aa), K9N7D2 (ORF5, 224), K9N7A1 (Membrane protein, 219 aa), K9N4V7 (Nucleoprotein, 411 aa).
SARS-CoV-2	ORF1ab P0D7D1; 1–180 (Host translation inhibitor NSP1), 181–818 (NSP2), 819–2763 (NSP3), 2764–3263 (NSP4), 3264–3569 (3C-like proteinase), 3570–3859 (NSP6), 3860–3942 (NSP7), 3943–4140 (NSP8), 4141–4253 (NSP9), 4254–4392 (NSP10), 4393–5324 (RdRp), 5325–5925 (helicase), 5926–6452 (Proofreading exoribonuclease), 6453–6798 (Uridylate-specific endoribonuclease), 6799–7096 (2'-O-methyl transferase), P0D7C2 (Spike glycoprotein, 1273 aa), P0D7C3 (Protein 3a, 275 aa), P59633 (Protein 3 b, 154 aa), P0D7C4 (Envelope small membrane protein, 75 aa), P0D7C5 (Membrane protein 222 aa), P0D7C6 (NSP6, 61 aa), P0D7C7 (Protein 7a, 121 aa), P0D7D8 (Protein non-structural 7 b, 43 aa), P0D7C8 (Protein non-structural 8, 121 aa), P0D7C9 (Nucleoprotein, 419 aa), P0D7D2 (Protein 9 b, 97 aa), P0D7D3 (ORF14, 73 aa).
HCoV-OC43	ORF1ab P0C6X6; 1–246 (NSP1), 247–851 (NSP2), 852–2750 (Papain-like proteinase, NSP3), 2751–3246 (NSP4), 3247–3549 (3C-like proteinase), 3550–3836 (NSP6), 3837–3925 (NSP7), 3926–4122 (NSP8), 4123–4232 (NSP9), 4233–4369 (NSP10), 4370–5297 (RdRp), 5298–5900 (helicase), 5901–6421 (Guanine-N7 methyltransferase), 6422–6796 (Uridylate-specific endoribonuclease), 6797–7095 (2'-O-methyl transferase), Q80872 (NSP2a, 278aa), Q66165 (Hemagglutinin-esterase, 424 aa), P36334 (Spike glycoprotein, 1353 aa), Q04853 (NSP12.9 kDa, 109 aa), Q04854 (Envelope small membrane protein, 84 aa), Q01455 (Membrane protein, 230 aa), P33469 (Nucleoprotein 7a, 448 aa), Q4VID0 (Protein I, 7 b, 207 aa).
HCoV-HKU1	ORF1ab P0C6X2; 1–222 (Host translation inhibitor NSP1), 223–809 (NSP2), 810–2838 (Papain-like proteinase, NSP3), 2839–3334 (NSP4), 3335–3637 (3C-like proteinase), 3638–3924 (NSP6), 3925–4016 (NSP7), 4017–4210 (NSP8), 4211–4320 (NSP9), 4321–4457 (NSP10), 4458–5385 (RdRp), 5386–5988 (helicase), 5989–6509 (Guanine-N7 methyltransferase), 6510–6883 (Uridylate-specific endoribonuclease), 6884–7182 (2'-O-methyl transferase), Q5MQD1 (Hemagglutinin-esterase, 386 aa), Q5MQD0 (Spike glycoprotein, 1356 aa), Q5MQC9 (NSP4, 109 aa), Q5MQC8 (Envelope small membrane protein, 82 aa), Q5MQC7 (Membrane protein, 223 aa), Q5MQC6 (Nucleoprotein 7a, 441 aa), Q5MQC5 (Protein I, 7 b, 205 aa).

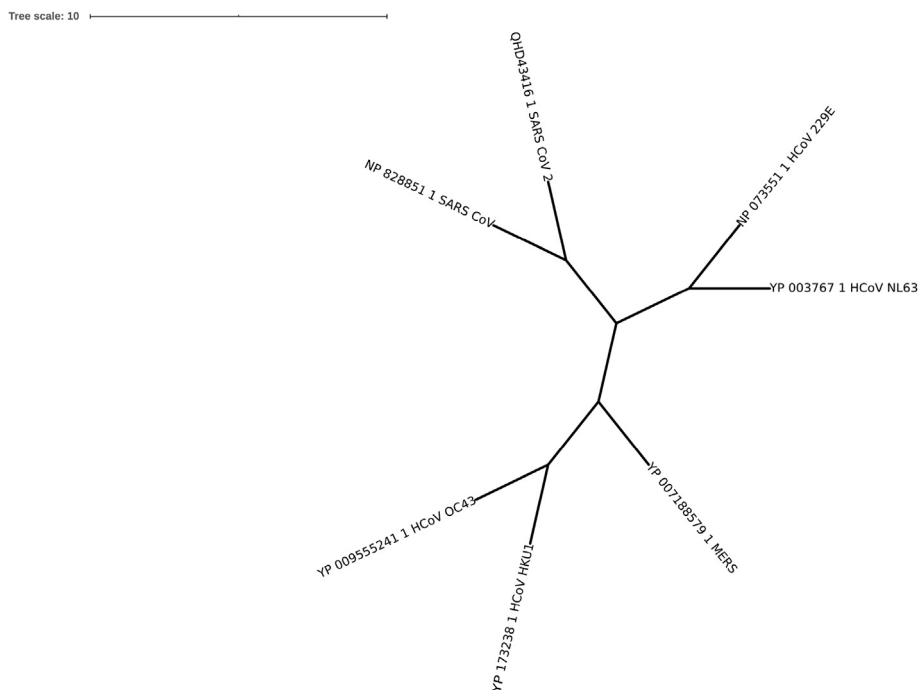


Fig. 1A. Phylogenetic tree of representative spike proteins from human coronaviruses. SARS-CoV, SARS-CoV-2, MERS-CoV, HCoV-HKU1, HCoV-NL63, NCoV-229E and HCoV-OC43.

spike protein reported by (Song et al., 2018) are shown in Fig. 1D. After the virion uptake by target host cells, the proteolytic cleavage between S1/S2 subunits is aided by host proteolytic enzymes in order to release the S1 subunit and trigger the pre-fusion to post-fusion conformational transition (Burkard et al., 2014, Millet and Whittaker, 2014, Millet and Whittaker, 2015). A second cleavage site S2', upstream of the fusion peptide in the S2 subunit becomes available during the onset of membrane fusion. This second cleavage step occurs for all CoVs and is believed to activate the protein for membrane fusion which takes place via irreversible conformational changes (Belouzard et al., 2009; Walls et al., 2017). The cleavage sites present near S1/S2 and S2' participate in the viral entry and modulate host range and cell tropism. The structure of SARS-CoV-2 spike protein was analysed *in situ* by combing cryo-electron tomography, sub-tomogram averaging and molecular dynamics simulations (Turoňová et al., 2020). Molecular dynamics simulations of the spike proteins reveal the conformational alterations in the protein required for the mechanistic function. The SARS-CoV-2 spike protein contains three hinge regions that provide the S1^A and S1^B domains orientational freedom allowing them to scan the host cell surface for potential receptors (Turoňová et al., 2020).

The attachment of HCoV to host cells is mediated via viral spike protein and host cell surface receptor interactions. The host receptors are carbohydrates that function as “Attachment receptors” and membrane proteins that function as “Entry receptors” thus mediating protein-carbohydrate and protein-protein interactions, respectively. Either the S1^A or S1^B or both domains are referred to as the receptor binding domain (RBD) that bind a carbohydrate for host cellular attachment or a cell surface receptor protein for host cellular entry. For instance, the S1^A domain of the spike proteins in

HCoV-OC43 and HCoV-HKU1 bind to 9-*O* acetyl sialic acid host receptor (Hulswit et al., 2019), and the MERS-CoV spike protein S1^A domain selectively binds to sialic acid (Li et al., 2017; Park et al., 2019). These studies demonstrate that the spike proteins recognize the cell-surface sialic acid glycoconjugates and can therefore serve as attachment factor on host cells. Sialic acid is a negatively charged ubiquitous monosaccharide that is terminally linked to oligosaccharides and forms glycoconjugates of proteins and lipids (Vlasak et al., 1988; Huang et al., 2015). The sialic acid derived glycoconjugates, glycoproteins and gangliosides decorate on the host cell surface and are therefore recognized by the HCoV spike proteins to cause the initial attachment to the host cells. The host cell selectivity of the virus can be hampered since such glycoconjugates are differentially expressed in several cells and tissues. In this context, some viruses, such as, HCoV-OC43 and HCoV-HKU1 express HE protein that functions as sialate-*O*-acetyltransferases that can reverse sialic acid binding of spike proteins to non-targeted host cells (De Groot, 2006). Likewise, HCoV-NL63 spike protein binding to heparan sulfate is required for viral attachment and infection of target cells (Milewska et al., 2014). These studies indicate that the different HCoV spike protein and host-receptor carbohydrate-binding partners that mediate host cell recognition remain to be explored.

The spike protein S1^B domain is reported to bind various host receptor transmembrane proteins; SARS-CoV (Li et al., 2003, 2005), SARS-CoV-2 (Shang et al., 2020b; Zhang et al., 2020) and HCoV-NL63 (Hofmann et al., 2005) bind the human angiotensin-converting enzyme 2 (ACE-2) receptor, MERS-CoV binds the human dipeptidyl peptidase 4 (DPP4) (Wang et al., 2013) and the HCoV-229E binds human amino peptidase N (APN) (Delmas et al., 1992; Yeager et al., 1992). The SARS-CoV also binds to human

```

1 MFVFLVLLPL VSSQCVNLT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS
61 NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNIRGWIFGTTL DSKTQSLIV
121 NMATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFMLDLE
181 GKQGNFKNLR EEFVKRIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT
241 LLALHRSYLT PGDSSSGWTA GAAAYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK
301 CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN
361 CVADYSVLYN SASFSTFKCY GVSPTKLNLD CFTNVYADSF VIRGDEVQR I APGQTGKIAD
421 YNYKLPDDFT GCVIAWNSNN LDSKVGNNY YLYRFLRKS LKPFERDIST EIYQAGSTPC
481 NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN
541 FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP
601 GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY
661 ECDIPIGAGI CASYQTQNS PRRARSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI
721 SVTTEILPVS MTKTSVDCTM YICGDSTEC NLLQYGSFC TQLNRALTGI AVEQDKNTQE
781 VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLENKVTLLAD AGFIKQYGC
841 LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM
901 QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN
961 TLVKQLSSNF GAISSVLNDI LSRLDKVEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA
1021 SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTVVPA QEKNFTTAPA
1081 ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITDNT FVSGNCDVVI GIVNNTVYDP
1141 LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL
1201 QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCCMTSC CSCLKGCCSC GSKCFDEDD
1261 SEPVLKGVKL HYT

```

Fig. 1B. Amino acid sequence and structural annotation of human SARS-CoV-2 spike protein (NCBI code: QHD43416). The N-glycosylation sites are shown in bold and italics. Structural regions are labelled below the sequence.

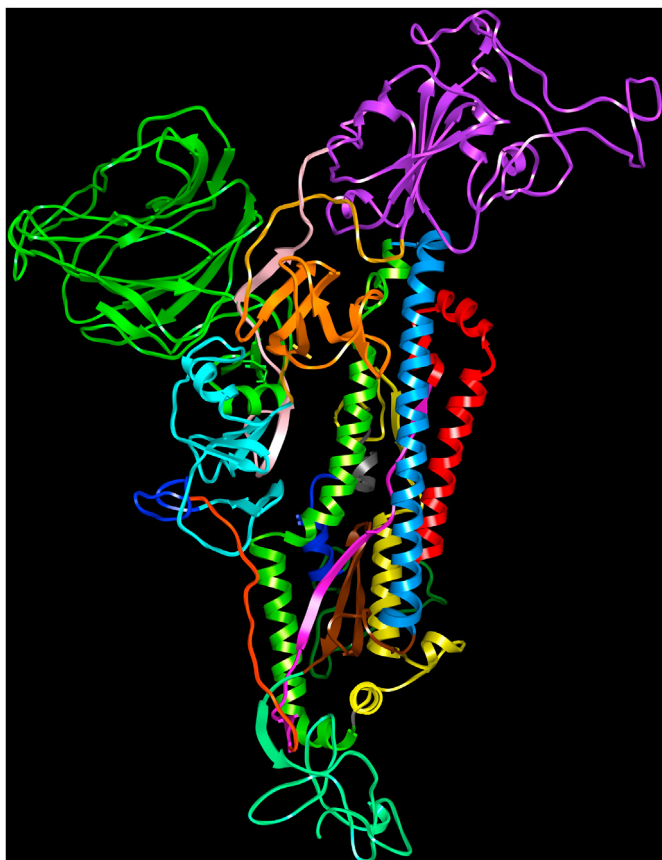


Fig. 1C. Human SARS-CoV-2 spike protein (PDB code: 6VSB A-chain). S1^A domain (27–300, green), S1^A-S1^B linker (301–335, pink), S1^B domain (336–516, purple), 517–533 (linker, golden rod), S1^C domain (534–589, orange), 590–593 (linker, golden rod), S1^D domain (594–674, cyan), protease cleavage site (675–688, blue), S1–S2 subunits linker (689–710, orange), central β -strand (711–737, magenta), downward helix (738–782, red), S2' cleavage site (783–815, sea green), fusion peptide (816–833, navy blue), connecting region (834–910, yellow), heptad repeats (912–983, chartreuse), central helix (984–1034, dodger blue), β -hairpin (1035–1069, brown), connecting β -sheet domain (1070–1134, spring green).

CD209L, a C-type lectin expressed in human lung type II alveolar cells and endothelial cells (Jeffers et al., 2004). The purified SARS-CoV spike protein was shown to bind pulmonary surfactant protein D, a collectin found in the lung alveoli (Leth-Larsen et al., 2007). The human host cell receptor proteins for HCoV-OC43 and HCoV-HKU1 have not been identified so far. The structure, function, and evolution of CoV spike proteins from all four genera and the structural basis for fusion of viral and host membranes is reviewed (Li, 2016).

The manifestation of three major HCoV epidemics; (SARS-CoV, MERS-CoV and SARS-CoV-2) in less than two decades along with instances of other less pathogenic HCoV infections is good reason, to study the molecular mechanisms underlying HCoV-host receptor interactions, in order to understand viral entry into host cells. The three-dimensional structures of HCoV spike proteins and their S1^A and S1^B domains in complex with host receptors provide the atomistic details of the virus and host inter-molecular interactions. The structures aid the development of therapeutics and vaccines against these infectious diseases. This manuscript reviews HCoV spike proteins, their structure and function, and their binding to receptors based on the three-dimensional structures.

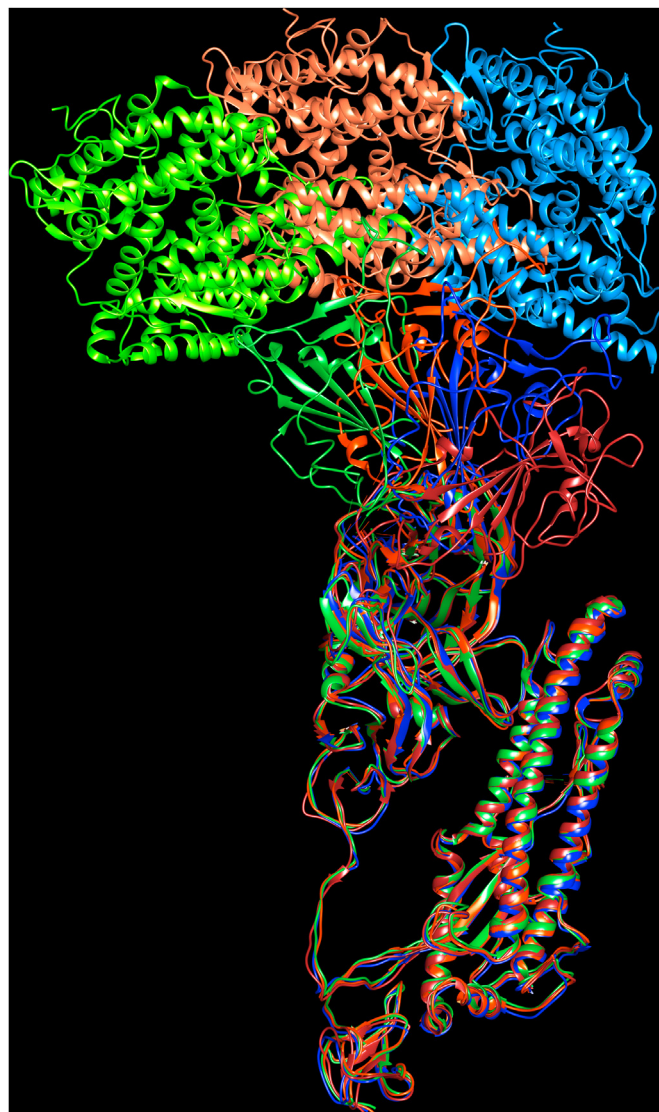


Fig. 1D. Structural superposition of apo SARS-CoV (PDB code: 6ACC) and human ACE-2 receptor complexes (6ACC, 6ACJ, 6ACK). The pre-fusion to post-fusion conformations are indicated. 6ACC: brown (S1^B downward), 6ACG: conformation 1, spike (blue), ACE-2 (dodger blue), 6ACJ: conformation 2, spike (red), ACE-2 (orange), 6ACK: conformation 3, spike (green), ACE-2 (chartreuse).

3. SARS-CoV and SARS-CoV-2

The human SARS-CoV infection in Guangdong province of southern China in 2002 was transmitted from bats and civets (Guan et al., 2003, Marra et al., 2003; Rota et al., 2003; Ksiazek et al., 2003). The disease outbreak then spread over Asia, Europe and North America. A total of 8,096 cases and 774 deaths were recorded. The civet SARS-CoV (Wang et al., 2005; Li et al., 2006) and SARS-like CoVs from some bats and civets were predicted to result in human infections (Menachery et al., 2015; Wang et al., 2018). Based on the comparative analyses of SARS-CoV genomes, it has been proposed that they have evolved from civet SARS-CoV and that their spike proteins are highly homologous (Shi and Hu, 2008). SARS-CoV uses its spike protein for the initial recognition of ACE-2 as host receptor for entry into human epithelial cells (Li et al., 2003) to cause infection. ACE-2 is an angiotensin-converting enzyme related carboxypeptidase, a type I integral membrane protein that comprises extracellular (18–740 amino acids), transmembrane

(741–761) and cytoplasmic (762–840) regions. The extracellular region is composed of two domains, zinc metallopeptidase domain (19–615) and C-terminal collectrin-like domain (616–740). The three-dimensional structures of human SARS-CoV spike protein (PDB code; 6ACG; Song et al., 2018, 6CRW, 5WRG) and its S1^B domain complexed with human ACE-2 are known (2AJF, 3SCI, 3SCJ). The S1^B domain of SARS-CoV comprises a five-stranded anti-parallel β -sheet and the structure is stabilised by three disulfide bridges. Between strands β 4 and β 5, a \sim 65 amino acid residues region folds into an extended loop comprising short stretches of two α -helices and two anti-parallel β -strands in a β -sheet. The SARS-CoV spike protein binds the human ACE-2 receptor via this extended loop (3SCI) (Fig. 2A). The interaction of virus with host cell receptor is mediated via the receptor binding motifs (RBMs) in the extended loop of spike protein S1^B domain and the virus binding motifs (VBMs) on the ACE-2 receptor comprising helices; H1, H2, H15, H17 and β -hairpin 'B' (labelled according to the PDBsum for PDB code: 3SCI). The amino acid residues that stabilise the virus-host recognition via protein-protein interactions are shown in Table 4. The spike protein binding site and the carboxypeptidase catalytic site are two different sites in the ACE-2 receptor.

During mid-December 2019, SARS-CoV-2 infection was reported in Wuhan, China. The nucleotide sequences of SARS-CoV and SARS-CoV-2 share 79.6% sequence identity at the genomic level (Zhou et al., 2020). The SARS-CoV-2 uses ACE-2 as receptor for cellular entry (Zhou et al., 2020). Recombinant overexpressed SARS-CoV-2 spike proteins form non-aggregated homotrimers that specifically bind only to human ACE-2 (Herrera et al., 2020). The complete genome of bat SARS-CoV isolated in 2013 (RaTG13) was sequenced during 2020 and shown to be similar to the novel SARS-CoV-2. This bat SARS-CoV has been proposed to have recently crossed species and caused infection in humans (Wu et al., 2020). The pangolin SARS-CoV genomes isolated during 2014–2018 and sequenced in 2020 share 85.5%–92.4% sequence similarity to SARS-CoV-2 (Han 2020; Lam et al., 2020). The SARS-CoV-2 spike protein shares \sim 97.5% and \sim 92% sequence identity with spike proteins of RaTG13 bat SARS-CoV and pangolin SARS-CoV, respectively. The three-dimensional structures of SARS-CoV-2 spike proteins (6VSB; Wrapp et al., 2020, 6VXX, 6VYB; Walls et al., 2020, 6X6P; Herrera et al., 2020, 6X29; Henderson et al., 2020) and crystal structures of the S1^B domain complexed with ACE-2 (6M0J; Lan et al., 2020, 6LZG; Wang et al., 2020) are available in the PDB. The S1^B domain comprising five-stranded anti-parallel β -sheet in SARS-CoV-2 (6LZG; Wang et al., 2020) interacts with the human ACE-2 receptor as shown in Fig. 2B. The three-dimensional structures of human SARS-CoV (3SCI) and SARS-CoV-2 (6LZG) S1^B domains complexed with ACE-2 are highly superimposable. The protein-protein interaction sites between the SARS-CoV-2 virus and host receptor are similar to the interactions with SARS-CoV although there are differences in their amino acid sequences. The residues involved in the protein-protein interactions between SARS-CoV-2 spike protein and human ACE-2 are shown in Table 4.

The multiple sequence alignment of the bat, civet, pangolin, human SARS-CoV spike proteins and the dog, cat, mink, lion, tiger, human SARS-CoV-2 spike proteins is shown in Supplementary Fig. 2 and the corresponding phylogenetic tree is shown in Fig. 2C. The phylogenetic tree suggests, unlike the SARS-CoV spike proteins, the SARS-CoV-2 spike proteins representing the different host sources are not associated with distinct clades. The RaTG13 bat SARS-CoV (NCBI code: QHR63300) and pangolin SARS-CoV spike proteins are close to human SARS-CoV-2 as seen from the phylogenetic tree. The SARS-CoV and SARS-CoV-2 S1^B domain region contain three insertion loops according to the sequence and structure comparisons. A 5-residue loop "V₄₄₅GGNY₄₄₉", 8-residue loop "Y₄₇₃QAGSTPC₄₈₀" and 6-residue loop "E₄₈₄GFNCY₄₈₉"

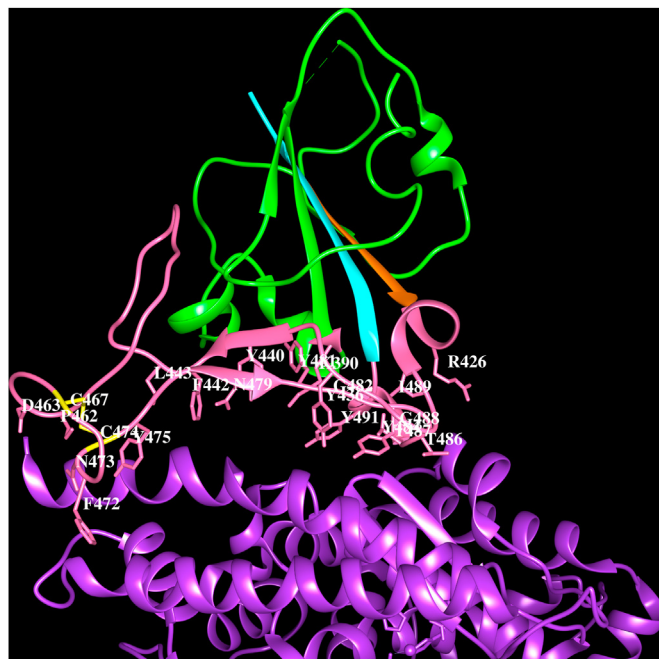


Fig. 2A. Cartoon representation of SARS-CoV spike protein S1^B domain (PDB code: 3SCI) (green) interacting with human ACE-2 (purple). Strand β 4 (orange), strand β 5 (cyan), extended loop (pink) and the side chains of residues within 4.5 Å from ACE-2 (K390, **R426, **Y436**, Y440, F442, L443, P463, D463, F472, N473, Y475, N479, Y481, G482, **Y484**, **T486**, **T487**, G488, I489, Y491) are indicated. The disulfide bridge C467–C474 (yellow). The residues involved in the hydrogen bonds with ACE-2 are shown in bold and italics.**

recognize the host receptor (numbering is according to PDB code: 6LZG). The location of these three insertion loops are same in human SARS-CoV and SARS-CoV-2 and match the ACE-2 binding region. While similar sequences are also present in the RaTG13 bat SARS-CoV and pangolin SARS-CoV, all other SARS-CoVs comprise only 5-residues (instead of 6-residues) in the third loop. The absence of the three loop regions in some bat SARS-CoV at the equivalent positions may be responsible for their inability to bind human ACE-2 receptor (Guruprasad 2020). The second and third loops are stabilised by a disulfide bridge between half-cystines at positions 480 and 488 (6LZG). This disulfide bridge that tethers the 2nd and 3rd insertion loops is conserved in all spike proteins that recognize ACE-2. The three insertion loops and the tethered disulfide bridge in SARS-CoV and SARS-CoV-2 S1^B domain of spike proteins represent important structural features required for recognition of human ACE-2 receptor.

As shown in Supplementary Fig. 2, SARS-CoV-2 spike proteins have two insertion sequences M₁₅₃ESEFR₁₅₈, S₂₄₇YLTPG₂₅₂ in the S1^A domain. These sequences are also present in bat RaTG13 and pangolin SARS-CoV spike proteins. The crystal structure of the spike protein S1^B domain bound to ACE-2 (6LZG) shows inter-molecular hydrogen-bond interactions of Tyr449 with Asp38 and Gln42. In the pangolin SARS-CoV spike protein, the equivalent Tyr449 residue is present, whereas, in the bat RaTG13 SARS-CoV spike protein it is mutated to Phe449, indicating the common ACE-2 interactions between human SARS-CoV-2 and pangolin SARS-CoV. This raises the possibility of the virus transmission to have occurred from pangolin to humans. Further, based on the virus isolation timelines for bat and pangolin SARS-CoV and spike protein sequence comparisons, we propose that the virus transmission may have occurred from RaTG13 bat SARS-CoV to pangolin SARS-CoV that may have ultimately caused the virus transmission to humans as

Table 4
HCoV spike protein-receptor interactions (≤ 4.5 Å).

PDB: 3SCI
SARS-CoV: K390, R426, Y436, Y440, F442, L443, P462, D463, F472, N473, Y475, N479, Y481, G482, Y484, T486, T487, G488, I489, Y491
hACE-2: S19, Q24, T27, F28, D30, K31, H34, E37, D38, Y41, Q42, L45, M82, Y83, Q325, E329, N330, K353, G354, D355, R357, R393
Hydrogen bonds between SARS-CoV and hACE-2: T486–Y41, Y484–Q42, R426–E329, Y436–D38, C488–K353
PDB: 6LZG
SARS-CoV-2: K417, G446, Y449, Y453, L455, F456, Y473, A475, G476, E484, F486, N487, Y489, F490, Q493, G496, Q498, T500, N501, G502, Y505
hACE-2: S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45, L79, M82, Y83, N330, K353, G354, D355, R357, R393
Hydrogen bonds between SARS-CoV-2 and hACE-2: A475–S19, N487–Q24, T500–Y41, K417–D30, Y449–D38, Y449–Q42, N487–Y83, Q498–Q42, T500–Y41, N501–Y41, G502–K353
PDB: 6Q04
MERS-CoV: Q36, F39, H91, A92, F101, I132, S133, P134, S135, Q304, R307
Hydrogen bonds between MERS-CoV with sialic acid: A92–O9, R307–O8, R307–O9, I132–N5, A92–O9
PDB: 4L72
MERS-CoV: S454, D455, P463, Y499, N501, K502, S504, L506, D510, R511, E513, P515, E536, D537, G538, D539, Y540, R542, W553, V555, A556, S557, S559
hDPP4: T188, T265, K267, F269, Q286, T288, A289, P290, A291, S292, L294, I295, G296, H298, R317, Y322, S333, S334, R336, V341, Q344, I346, K392
Hydrogen bonds between MERS-CoV and hDPP4: D359–K267, E513–A291, D510–Y322, Y499–R336, E513–Q344
PDB: 6NZK
HCoV-OC43: N27, K29, T31, L80, K81, G82, S83, L85, L86, S87, W90
Hydrogen bonds between HCoV-OC43 with sialic acid: N27–OA9, S83–O3, K81–N5
PDB: 3KBH
HCoV-NL63: G494, G495, S496, C497, Y498, Y499, C500, H503, G534, S535, P536, G537, S539, S540, W585, H586
hACE-2: D30, N33, H34, E37, Y41, P321, N322, M323, T324, Q325, G326, N330, K353, G354, D355, F356, R357, M383, A386, A387, R393
Hydrogen bonds between HCoV-NL63 and hACE-2: S540–T324, Y498–E37, S535–K353, S540–T324
PDB: 6ATK
HCoV-229E: S312, G313, G314, G315, K316, C317, F318, N319, C320, R359, W404, S407, K408
hAPN: T244, E286, F287, D288, Y289, V290, E291, K292, W303, I309, A310, D315, L318
Hydrogen bonds between HCoV-229E and hAPN: N319–E291, G314–D288, G315–F287, G315–D288, F318–Y289, N319–E291, K408–E291



Fig. 2B. Cartoon representation of the SARS-CoV-2 spike protein $S1^B$ domain (green) interacting with human ACE-2 (purple) (PDB code: 6LZG). Strand $\beta 4$ (orange) and $\beta 5$ (cyan), the extended loop (pink) and the side chains of residues within 4.5 Å from ACE-2 (***K417***, G446, ***Y449***, Y453, L455, F456, Y473, ***A475***, G476, E484, F486, ***N487***, Y489, F490, Q493, G496, ***Q498***, ***T500***, ***N501***, ***G502***, Y505) are indicated within 4.5 Å from ACE-2. The disulfide bridge C480–C488 (yellow). The residues that form hydrogen bonds with ACE-2 are shown in bold and italics.

the novel SARS-CoV-2. The multiple sequence alignment in [Supplementary Fig. 2](#) was edited to generate [Fig. 2D](#) in order to depict the RBMs discussed above in the representative spike proteins from bat, civet, pangolin, human SARS-CoV, and dog, cat, mink, tiger, lion and human SARS-CoV-2.

A motif “P₆₈₁RRA₆₈₄”, (amino acid numbering according to human SARS-CoV-2, NCBI code: QHD43416) gained in the human SARS-CoV-2 spike protein is referred as the furin cleavage site ([Ou](#)

[et al., 2020](#)). This gain of the ‘PRRA’ sequence motif in human SARS-CoV-2 distinguishes it from its closest bat homologue, RaTG13 SARS-CoV. To facilitate viral entry into host cell, SARS-CoV-2 is preactivated by furin and its spike protein is cleaved during viral packaging thereby reducing its dependence on host cell proteases for viral entry ([Shang et al., 2020a](#)). The SARS-CoV-2 is also activated by transmembrane protease serine 2 and cathepsin L, and both enzymes show cumulative effects with furin on activating SARS-CoV-2 entry ([Shang et al., 2020a](#)). The spike proteins from *Canis lupus familiaris* (NCBI code: QIT08292.1), *Felis catus* (QLG96797.1), *Mustela lutreola* (QJS39496.1), *Neovison vison* (QNJ45106.1), *Panthera leo* (QLC48407.1), *Panthera tigris* (QLC48443.1) also comprise the ‘PRRA’ sequence motif, indicating that these mammals employ a similar SARS-CoV-2 entry mechanism.

4. MERS-CoV

MERS-CoV was first identified in humans in the Middle East during 2012. This virus originated in bats and was transmitted from dromedary camels to humans in Saudi Arabia with fatality rate ~35% ([Zaki et al., 2012](#); [Bermingham et al., 2012](#); [Azhar et al., 2014a, 2014b](#); [Chan et al., 2015](#); [Sabir et al., 2016](#); [Alagaili et al., 2014](#); [Hemida et al., 2013](#)). Later, it spread to other Middle East countries; Jordan, United Arab Emirates, Qatar and to France, Germany, United Kingdom and Italy in Europe and Tunisia in North Africa via human to human transmission. MERS-CoV outbreak was later also reported in South Korea during 2015 ([Ki, 2015](#)). MERS-CoV virions cause agglutination of human erythrocytes. Hemagglutination by MERS-CoV is mediated through simultaneous low-affinity binding of multiple spike proteins of the virus with multiple receptors on human erythrocytes surface. The MERS-CoV hemagglutinating activity is sialic acid dependent and the binding site of sialic acid is located in the $S1^A$ domain ([Li et al., 2017](#)). The pathogenicity of MERS-CoV is caused by the specific binding of its $S1^B$ domain to the human DPP4 receptor. DPP4 is a transmembrane serine protease that comprises cytoplasmic (1–6 amino acids), transmembrane (7–28) and extracellular (29–766) regions. Amino acid residues

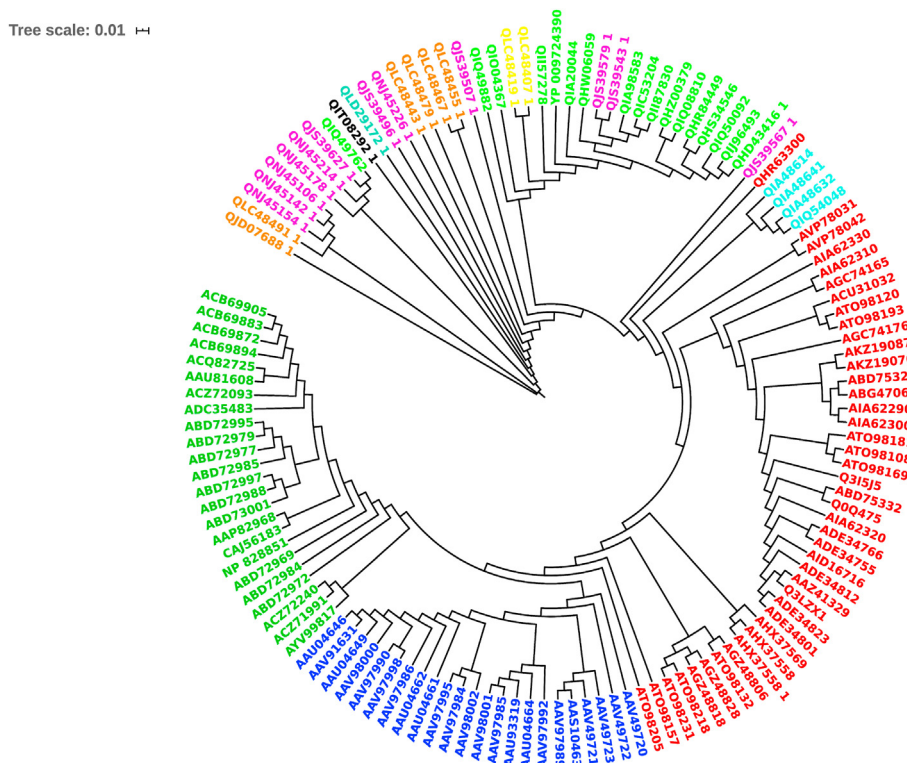


Fig. 2C. Phylogenetic tree of SARS-CoV and SARS-CoV-2 spike proteins. Human SARS-CoV-2 (light green), dog SARS-CoV-2 (black), cat SARS-CoV-2 (aqua green), mink SARS-CoV-2 (magenta), tiger SARS-CoV-2 (orange), lion SARS-CoV-2 (yellow), pangolin SARS-CoV (cyan), human SARS-CoV (dark green), bat SARS-CoV (red), civet SARS-CoV (violet).

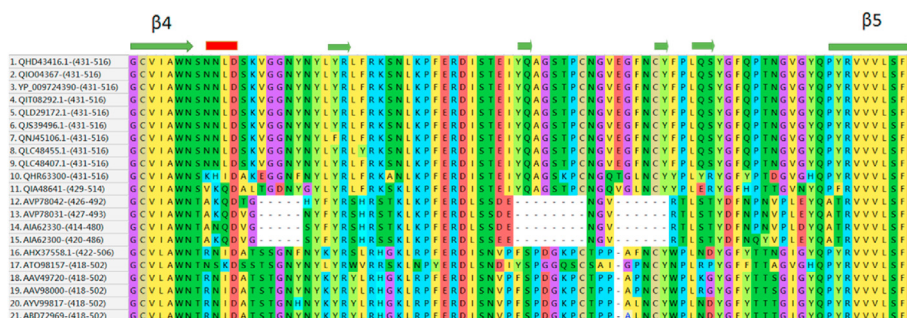


Fig. 2D. Portion of the alignment of spike proteins. Extracted from the multiple sequence alignment (Supplementary Fig. S2) showing the insertion sequences that form receptor binding motifs within the RBD ($S1^B$ domain) for human SARS-CoV-2 (1–3), *Canis lupus familiaris* SARS-CoV-2 (4), *Felis catus* SARS-CoV-2 (5), *Mustela lutreola* SARS-CoV-2 (6), *Neovison vison* SARS-CoV-2 (7), *Panthera tigris* SARS-CoV-2 (8), *Panthera leo* SARS-CoV-2 (9), bat SARS-CoV RaTG13 (10), pangolin SARS-CoV (11), bat SARS-CoV (12–17), civet SARS-CoV (18–19), human SARS-CoV (20–21). The secondary structure conformations; β -strands (olive green arrows) and α -helices (red bars) are indicated. The starting and ending amino acid numbers of the regions are indicated after the NCBI code within brackets.

39–508 fold into the N-terminal eight-bladed β -propeller domain and 509–766 fold into the C-terminal α/β hydrolase domain. DPP4 has ecto-peptidase activity and expressed on the surface of several cell types including cells found in human airways (Gallagher and Perlman 2013). The $S1^B$ domain of MERS-CoV spike protein binds to soluble DPP4 protein *in vitro* and specifically binds to DPP4 expressing cells *in-vivo* thereby inducing significant neutralizing antibody responses (Du et al., 2014).

The cryo-electron microscopy structures of the MERS-CoV spike proteins (5X5C, 5X5F, 5X59; Yuan et al., 2017), the crystal structure of $S1^A$ domain (5X4R, Yuan et al., 2017) and its complex with *O*-sialic acid (6Q04, Park et al., 2019), the $S1^B$ domain with DPP4 receptor (4L72, Wang et al., 2013) are available in the PDB. The binding of MERS-CoV spike protein to the β -propeller domain does not interfere with hydrolase activity of DPP4. *O*-sialic acid binds to

$S1^A$ domain and is exposed to solvent as shown in Fig. 3A, residues on the strands $\beta 1$, $\beta 4$ and $\beta 5$, adjacent loop regions and helix H1 are involved in the recognition (6Q04). The MERS-CoV spike protein $S1^B$ domain binds to the β -propeller of DPP4 (4L72, Wang et al., 2013) and comprises five-stranded anti-parallel β -sheet stabilised by three disulfide bridges. A long insertion region between the $\beta 4$ and $\beta 5$ strands comprising ~80 amino acid residues forms a four-stranded anti-parallel β -sheet. The insertion region and the loop connecting $\beta 3$ and $\beta 4$ strands are involved in interacting with the human receptor DPP4 (4L72, Wang et al., 2013). Each blade of the β -propeller in DPP4 comprises four anti-parallel β -strands and the MERS-CoV $S1^B$ domain recognizes amino acid residues on the 4th and 5th blades in the β -propeller. The region of interactions between MERS $S1^B$ domain of spike protein and DPP4 is shown in Fig. 3B. The residues involved in the protein-carbohydrate and

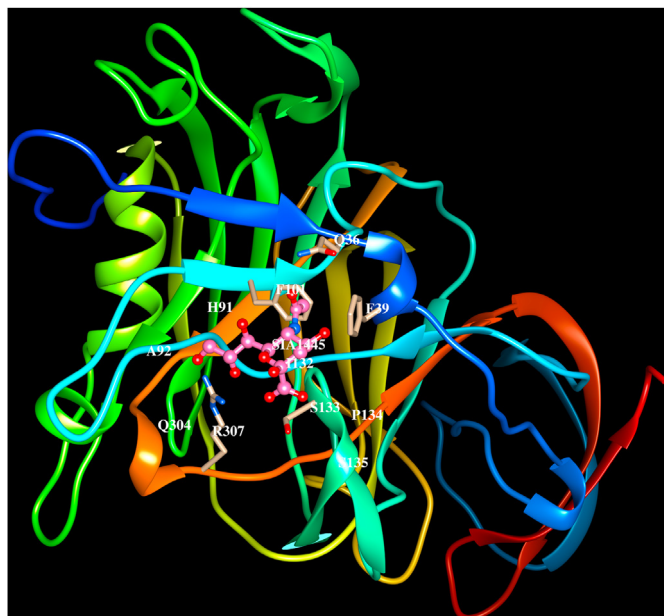


Fig. 3A. Cartoon representation of the MERS-CoV spike protein (PDB code: 6Q04) S1^A domain binding to O-sialic acid. The side chains of residues (Q36, F39, H91, **A92**, F101, **I132**, S133, P134, S135, Q304, **R307**) that lie within 4.5 Å from sialic acid are indicated. The residues that form hydrogen bonds with O-sialic acid are shown in bold and italics.

protein-protein interactions are listed in [Table 4](#).

5. HCoV-OC43

HCoV-OC43 was isolated for the first time in 1967 from volunteers at the Common Cold Unit in Salisbury, United Kingdom. Modified sialic acid, such as 9-*O*-acetyl-N-acetylneuraminic acid is a major receptor for HCoV-OC43 ([Vlasak et al., 1988](#)). Also, HCoV-OC43 uses 9-*O*-acetyl-sialic acid as a receptor ([Tortorici et al., 2019](#)). The HE protein reverses attachment of sialic acid by activating sialate-*O*-acetyl-esterase activity. Site-directed mutagenesis, binding experiments, and the three-dimensional structures showed that residues involved in sialic acid binding are essential for HCoV-OC43 spike protein mediated entry into host cells. Further, the HCoV-OC43 spike protein does not bind free sialic acid and/or acidic pH conditions do not induce conformational changes in the spike protein, suggesting that multivalent interactions with sialoglycans and/or attachment to a protein receptor are essential to promote membrane fusion ([Tortorici et al., 2019](#)).

The cryo-electron microscopy structures of HCoV-OC43 spike protein trimer in apo form and in complex with 9-*O*-acetyl-sialoglycan at 2.9 Å and 2.8 Å resolution, respectively, are available in the PDB (6OHW, 6NZK, [Tortorici et al., 2019](#)). The HCoV-OC43 uses its S1^A domain to bind sialic acid and the ligand interacts with a groove at the periphery of S1^A domain. The 9-*O*-acetylated sialic acid binds to a large exposed surface in the S1^A domain involving β1, β4 and β5 strands, adjacent loop regions and helix H1 as shown in [Fig. 4A](#) (6NZK) similar to MERS-CoV. These protein-carbohydrate interactions are mediated by several non-bonding interactions listed in [Table 4](#). The S1^B domain of HCoV-OC43 shown in [Fig. 4B](#) (6NZK) forms a five-stranded anti-parallel β-sheet. The long insertion between the 4th and 5th β-strands comprising ~130 amino acid residues folds into five β-strands and five short α-helices connected by loops. Seven of the nine disulfide bridges in S1^B domain are associated with the 130 amino acids long insertion. The human receptor for the spike protein of HCoV-OC43 is not yet known.



Fig. 3B. Cartoon representation of the MERS-CoV spike protein (PDB code: 4L72) S1^B domain (green) interacting with human DPP4 (purple). The strand β4 (orange) and β5 (cyan), the extended loop (pink) and the side chains of residues within 4.5 Å from DPP4 (S454, D455, P463, **Y499**, N501, K502, S504, L506, **D510**, R511, **E513**, P515, E536, D537, G538, **D539**, Y540, R542, W553, V555, A556, S557, S559) are indicated. The C503–C526 disulfide bridge (yellow). The residues that form hydrogen bonds with DPP4 are shown in bold and italics.

6. HCoV-HKU1

The HCoV-HKU1 was first reported in Hong Kong in 2005 ([Woo et al., 2005](#)) and known to spread all over the world subsequently. The HCoV-HKU1 causes mild upper respiratory tract disease among young children but sometimes can lead to severe respiratory diseases in young children and in immunocompromised elderly patients and accounts for ~3% acute respiratory infections. The attachment receptor for HCoV-HKU1 was identified as *O*-acetylated sialic acid and the binding of this carbohydrate to S1^A domain is required but not sufficient to cause the infection ([Huang et al., 2015](#)). The HCoV-HKU1 HE protein demonstrates sialate-*O*-acetyl-esterase receptor destroying enzyme activity specific to the *O*-acetylated sialic acids recognized by the spike protein ([Huang et al., 2015](#)). Although a protein receptor has not been identified for HKU1, antibodies against the S1^B domain but not those against the S1^A domain blocked HKU1 infection of cells ([Qian et al., 2015](#)). These data suggest that the S1^B domain is the primary HCoV-HKU1 receptor-binding site, whereas, the S1^A domain mediates initial attachment via glycan binding ([Kirchdoerfer et al., 2016](#)). Most HCoV-HKU1 spike monoclonal antibodies recognized epitopes in the region between amino acids 535 and 673 (i.e., S1^C domain and some region from S1^B domain) indicating that this region is immunodominant ([Qian et al., 2015](#)). The electron microscopy structure of HCoV-HKU1 spike protein in pre-fusion conformation (5I08, [Kirchdoerfer et al., 2016](#)) and the crystal structure of HCoV-HKU1 spike glycoprotein S1^B domain required for host receptor binding and S1^C domain at 1.9 Å resolution are available (5KWV, [Ou et al., 2017](#)). The HCoV-HKU1 S1^B (310–612 amino acid residues) that comprises a ~155 amino acid residues insertion and located between the β4 and β5 strands is made up of several β-strands, small helices and eight disulfide bridges as shown in [Fig. 4C](#). This region with an extended surface is capable of interactions with an unknown human receptor ([Ou et al., 2017](#)).



Fig. 4A. Cartoon representation of the HCoV-OC43 spike protein (PDB code: 6NZK) S1^A domain binding to O-sialic acid. The side chains of residues within 4.5 Å from sialic acid (N27, K29, T31, L80, **K81**, G82, **S83**, L85, L86, S87, W90) are indicated. The residues that form hydrogen bonds with sialic acid are shown in bold and italics.

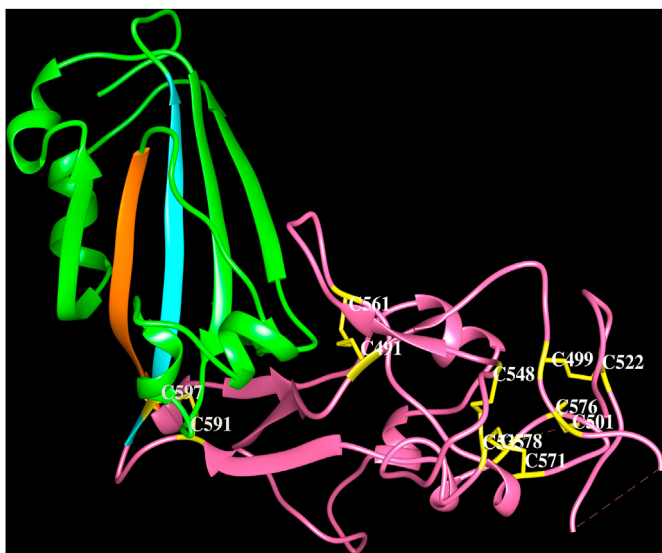


Fig. 4B. Cartoon representation of the HCoV-OC43 spike protein (PDB code: 6NZK) S1^B domain (green). The strands β 4 (orange) and β 5 (cyan) and the extended loop (pink) are indicated. The Cys-Cys pairs; 591–597, 491–561, 535–548, 571–578, 535–548, 499–522, 501–576 form disulfide bridges (yellow).

7. HCoV-NL63

HCoV-NL63 of the α -HCoV genera was first detected in 2002–2003 soon after SARS-CoV epidemic. HCoV-NL63 is a major cause of bronchiolitis and pneumonia in newborns worldwide and can cause severe lower respiratory tract infections among young children and immune-compromised adults (Chiu et al., 2005). HCoV-NL63 infections have been reported in countries across Europe, Asia and North America. Gene duplication events contribute to an additional N-terminal domain within the α -CoVs of HCoV-NL63 spike protein as observed in the cryo-electron microscopy structure (5SZS, Walls et al., 2016). This region referred as “0” domain was shown to adopt a galectin-like β -sandwich fold similar



Fig. 4C. Cartoon representation of the HCoV-HKU1 spike protein (PDB code: 5KWB) S1^B domain (green). The strands β 4 (orange) and β 5 (cyan) and the extended loop (pink) are indicated. The Cys-Cys pairs; 466–546, 474–495, 476–567, 520–533, 504–518, 485–516, 582–588, 556–569 form disulfide bridges (yellow).

to S1^A domain of HCoV-NL63 with an additional three-stranded β -sheet. The host-cell heparan sulfate proteoglycans participate in HCoV-NL63 anchoring and infection (Milewska et al., 2014). The binding of heparan sulfate to the HCoV-NL63 spike protein using surface plasmon resonance (Walls et al., 2016) was hypothesized to be mediated either via the “0” domain or S1^A domain that exhibit several positively charged patches on their surface.

The crystal structure of HCoV-NL63 S1^B domain in complex with human ACE-2 (3KBH, Wu et al., 2009) reveals a β -sandwich fold comprising two layers of three stranded β -sheets stacked against each other through extensive hydrophobic interactions. Three loops connecting the strands; β 1– β 2, β 3– β 4 and β 5– β 6 form the RBMs and are responsible for recognising human ACE-2. Among these, residues in the loop connecting the β 1 and β 2 strands that are stabilised by a disulfide bridge makes extensive interactions with the receptor as shown in Fig. 4D. The HCoV-NL63 spike protein RBMs recognize residues on the helices; H1, H16, H17 and residues in the loops comprising turns connecting helices; H15–H16, H17–H18 and H18–H19 and the β -hairpin ‘C’ on the ACE-2 receptor termed as VBMs. The protein-protein interactions are mediated by several non-bonding interactions listed in Table 4.

8. HCoV-229E

HCoV-229E of α -HCoV genera isolated in 1966 causes severe lower respiratory tract infections among young children. This virus is proposed to have originated in African hipposiderid bats and transferred to camelids and alpacas as intermediate hosts (Corman et al., 2015). The phylogenetic tree of the spike protein sequences from bat, camel and human CoV-229E suggests that the camel CoV-229E is intermediate to the bat CoV-229E and HCoV-229E. The bat CoV-229E comprises the N-terminal domain “0” as in HCoV-NL63, but this domain is absent in the camel and human CoV-229E, suggesting the evolutionary changes across species in CoV-229E.

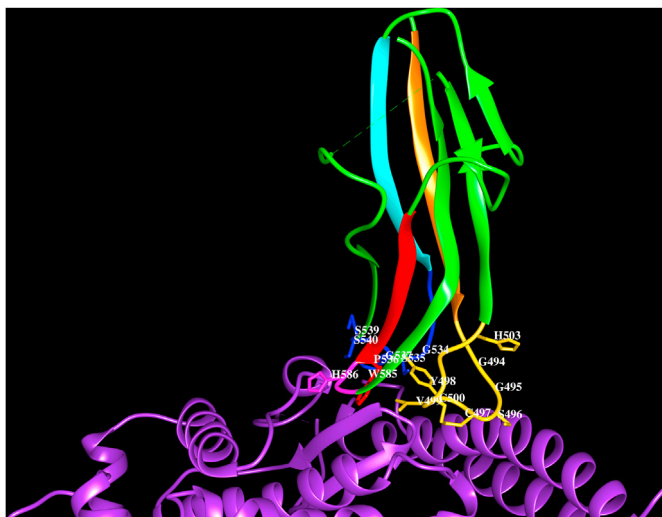


Fig. 4D. Cartoon representation of the HCoV-NL63 spike protein (PDB code: 3KBH) S1^B domain (green) interacting with human ACE-2. The three loops (gold, blue, magenta) connecting the strands β 1 (orange) to β 2 (green), β 3 (cyan) to β 4 (green) and β 5 (red) to β 6 (green), along with side chains of residues lie within 4.5 Å from hACE-2 (G494, G495, S496, C497, **Y498**, V499, C500, H503, G534, **S535**, P536, G537, S539, **S540**, W585, H586) are shown. The C497–C500 forms disulfide bridge (yellow). The residues that form hydrogen bonds with ACE-2 are shown in bold and italics.

The HCoV-229E uses hAPN as receptor for cellular entry (Yeager et al., 1992). The hAPN comprises a cytoplasmic tail (1–8 amino acids), transmembrane region (9–32) and extracellular region (33–967). This protein contains four domains and domain II corresponds to the catalytic site with metalloprotease activity. The electron microscopy structure of the spike proteins HCoV-229E (6U7H, Li et al., 2019) and the crystal structures of the S1^B domain in complex with hAPN (6ATK, Wong et al., 2017; Li et al., 2019) reveal a β -sandwich fold as in HCoV-NL63. The two layers of three-stranded β -sheets are tightly packed against each other through extensive hydrophobic interactions. Three loops form the RBMs and are presented by the β -sandwich core to bind hAPN. The loop connecting strands β 1– β 2 stabilised by a disulfide bridge makes extensive interactions with the receptor. Several amino acid residues mediate the interactions between the S1^B domain of HCoV-229E spike protein and the human APN shown in Fig. 4E. The HCoV-229E spike protein RBMs recognize residues on helices; H4, H5 and on the β -strands in β -sheets ‘D’ and ‘E’ and in the loop preceding β -strand ‘E’ on domain II. The sites of HCoV-229E S1^B domain interactions and the active site of the enzyme are different. The residue interactions between hAPN and the spike protein S1^B domain are listed in Table 4.

9. Conclusions and perspectives

Seven HCoVs of α - and β -genera identified during the last six decades are known to cause human infections that could lead to mortality. The viral spike protein uses a double receptor mechanism, i.e., carbohydrate and protein binding for attachment and entry into host cells, respectively, to cause the infection. The dynamical conformational states of the spike protein structure facilitate its binding to host cell receptors and viral entry. The spike protein’s S1^A domain from MERS-CoV, HCoV-HKU1, HCoV-OC43 has evolved to bind carbohydrates, such as, sialic acid, and to heparan sulfate in HCoV-NL63, in order to mediate virus-host interactions. The spike protein’s S1^B domain display significant sequence variability across the HCoVs and is therefore able to

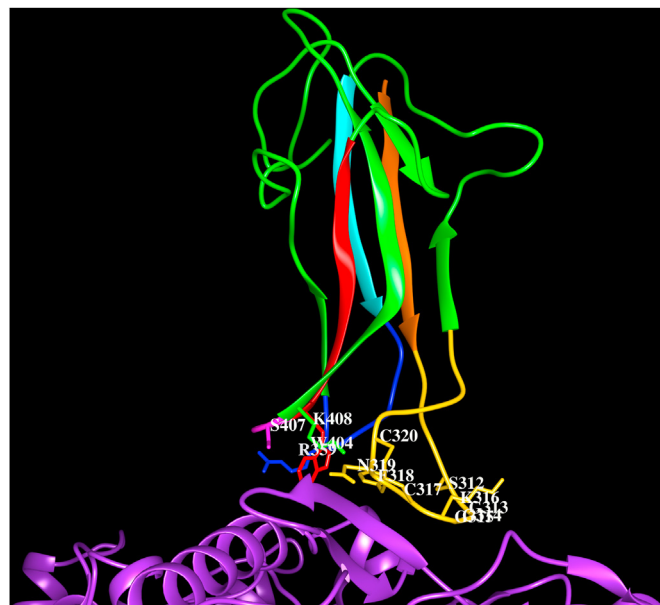


Fig. 4E. Cartoon representation of the HCoV-229E spike protein (PDB code: 6ATK) S1^B domain (green) interacting with human APN. The three loops (gold, blue, magenta) connecting strands β 1 (orange) to β 2 (green), β 3 (cyan) to β 4 (green) and β 5 (red) to β 6 (green), respectively, along with side chains of residues that lie within 4.5 Å from hAPN (S312, G313, **G314**, **G315**, K316, C317, **F318**, **N319**, C320, R359, W404, S407, **K408**) are shown. The C317–C320 forms disulfide bridge (yellow). The residues that form hydrogen bonds with APN are shown in bold and italics.

specifically recognize different host receptors. The S1^B domain in α -HCoVs comprises a six-stranded β -sandwich fold and in the β -HCoVs a five-stranded anti-parallel β -sheet. According to the phylogenetic tree, the α -HCoVs; HCoV-229E and HCoV-NL63 are closely related in sequence and have similar S1^B domain structure. However, these proteins have evolved to recognize different receptors; the HCoV-229E binds to the human APN receptor, whereas, the HCoV-NL63 binds to human ACE-2 receptor. Among, the β -HCoVs, the SARS-CoV and SARS-CoV-2 bind human ACE-2 receptor, whereas, MERS-CoV binds DPP4 receptor. The protein receptors for HCoV-OC43 and HCoV-HKU1 that facilitate entry into host are currently not known. In β -HCoVs, the extended loop connecting the β 4 and β 5 strands in the spike protein S1^B domain is of variable length and amino acid sequence that have evolved to recognize different receptors. The extended disordered loops are stabilised by disulfide bridges. The extended loop connecting the β 4 and β 5 strands in the S1^B domain of human SARS-CoV-2 spike protein comprises three loops and a tethered disulfide bridge that are important structural determinants for ACE-2 recognition. The sequence regions that specifically interact with human receptors serve as potential candidate epitopes for design of antibodies. The sites of protein-protein interactions between the HCoV spike proteins and host receptors serve as potential sites for design of small molecule inhibitors. The human SARS-CoV-2 that has evolved from bat RaTG13 SARS-CoV has gained the ‘PRRA’ sequence motif that is absent in human and bat SARS-CoV. The gain of ‘PRRA’ sequence motif in SARS-CoV-2 involved in furin cleavage is known to enhance efficiency of the virus entry into the host target cells.

In perspective, the availability of complete HCoV genome sequences from different host sources at different timelines of virus isolation will be helpful to trace their evolutionary mutations and transmission pathways. The identification of host entry receptors for HCoV-OC43 and HCoV-HKU1 and attachment receptors for the HCoVs discussed in this review will provide further insights into

the mechanism of virus-host interactions.

Author contribution

Lalitha Guruprasad: carried out the work and wrote the manuscript.

Declaration of competing interest

The author declares that there is no potential conflict of interest.

Acknowledgements

LGP thanks School of Chemistry and CAS, UGC for providing research facilities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pbiomolbio.2020.10.006>.

References

- Alagaili, A.N., Briese, T., Mishra, N., Kapoor, V., Sameroff, S.C., de Wit, E., Munster, V.J., Hensley, L.E., Zalmout, I.S., Kapoor, A., Epstein, J.H., 2014. Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *mBio* 5 (2) e00884–14.
- Almeida, J.D., Berry, D.M., Cunningham, C.H., Hamre, D., Hofstad, M.S., Mallucci, L., McIntosh, K., Tyrrell, D.A.J., 1968. Coronaviruses. *Nature* 220 (5168), 650.
- Azhar, E.I., El-Kafrawy, S.A., Farraj, S.A., Hassan, A.M., Al-Saeed, M.S., Hashem, A.M., Madani, T.A., 2014b. Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* 370 (26), 2499–2505.
- Azhar, E.I., Hashem, A.M., El-Kafrawy, S.A., Sohrab, S.S., Aburizaiza, A.S., Farraj, S.A., Hassan, A.M., Al-Saeed, M.S., Jamjoom, G.A., Madani, T.A., 2014a. Detection of the Middle East respiratory syndrome coronavirus genome in an air sample originating from a camel barn owned by an infected patient. *mBio* 5 (4), e01450–14.
- Baltimore, D., 1971. Expression of animal virus genomes. *Bacteriol. Rev.* 35 (3), 235–241.
- Belouzard, S., Chu, V.C., Whittaker, G.R., 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. Unit. States Am.* 106 (14), 5871–5876.
- Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* 4 (6), 1011–1033.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242.
- Bermingham, A., Chand, M.A., Brown, C.S., Aarons, E., Tong, C., Langrish, C., Hoschler, K., Brown, K., Galiano, M., Myers, R., Pebody, R.G., 2012. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill.* 17 (40), 20290.
- Burkard, C., Verheije, M.H., Wicht, O., van Kasteren, S.I., van Kuppeveld, F.J., Haagmans, B.L., Pelkmans, L., Rottier, P.J., Bosch, B.J., de Haan, C.A., 2014. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog.* 10 (11), e1004502.
- Chan, J.F., Lau, S.K., To, K.K., Cheng, V.C., Woo, P.C., Yuen, K.Y., 2015. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin. Microbiol. Rev.* 28 (2), 465–522.
- Chiu, S.S., Hung Chan, K., Wing Chu, K., Kwan, S.W., Guan, Y., Man Poon, L.L., Peiris, J.S.M., 2005. Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China. *Clin. Infect. Dis.* 40 (12), 1721–1729.
- Corman, V.M., Baldwin, H.J., Tateno, A.F., Zerbini, R.M., Annan, A., Owusu, M., Nkrumah, E.E., Maganga, G.D., Oppong, S., Adu-Sarkodie, Y., Vallo, P., 2015. Evidence for an ancestral association of human coronavirus 229E with bats. *J. Virol.* 89 (23), 11858–11870.
- Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17 (3), 181–192.
- De Groot, R.J., 2006. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconjugate J.* 23 (1–2), 59–72.
- Delmas, B., Gelfi, J., L'Haridon, R., Sjöström, H., Laude, H., 1992. Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV. *Nature* 357 (6377), 417–420.
- Donaldson, E.F., Haskew, A.N., Gates, J.E., Huynh, J., Moore, C.J., Frieman, M.B., 2010. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J. Virol.* 84 (24), 13004–13018.
- Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., Berger, A., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348 (20), 1967–1976.
- Du, L., Zhao, G., Yang, Y., Qiu, H., Wang, L., Kou, Z., Tao, X., Yu, H., Sun, S., Tseng, C.T.K., Jiang, S., 2014. A conformation-dependent neutralizing monoclonal antibody specifically targeting receptor-binding domain in Middle East respiratory syndrome coronavirus spike protein. *J. Virol.* 88 (12), 7045–7053.
- Forni, D., Cagliani, R., Clerici, M., Sironi, M., 2017. Molecular evolution of human coronavirus genomes. *Trends Microbiol.* 25 (1), 35–48.
- Fouchier, R.A., Hartwig, N.G., Bestebroer, T.M., Niemeyer, B., De Jong, J.C., Simon, J.H., Osterhaus, A.D., 2004. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc. Natl. Acad. Sci. Unit. States Am.* 101 (16), 6212–6216.
- Gallagher, T., Perlman, S., 2013. Public health: broad reception for coronavirus. *Nature* 495 (7440), 176–177.
- Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., 2020. The species severe acute respiratory syndrome related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 5, 536–544.
- Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302 (5643), 276–278.
- Guruprasad, L., 2020. Evolutionary relationships and sequence-structure determinants in human SARS coronavirus-2 spike proteins for host receptor recognition. *Proteins* 88, 1387–1393.
- Han, G.Z., 2020. Pangolins harbor SARS-CoV-2-related coronaviruses. *Trends Microbiol.* 28 (7), 515–517.
- Hemida, M.G., Perera, R.A., Wang, P., Alhammedi, M.A., Siu, L.Y., Li, M., Poon, L.L., Saif, I., Alnaeem, A., Peiris, M., 2013. Middle East Respiratory Syndrome (MERS) coronavirus seroprevalence in domestic livestock in Saudi Arabia, 2010 to 2013. *Euro Surveill.* 18 (50), 20659.
- Henderson, R., Edwards, R.J., Mansouri, K., Janowska, K., Stalls, V., Gobeil, S., Kopp, M., Hsu, A., Borgnia, M., Parks, R., Haynes, B.F., 2020. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* 27, 925–933.
- Hendley, J.O., Fishburne, H.B., Gwaltney Jr., J.M., 1972. Coronavirus infections in working adults: eight-year study with 229 E and OC 43. *Am. Rev. Respir. Dis.* 105 (5), 805–811.
- Herrera, N.G., Morano, N.C., Celikgil, A., Georgiev, G.I., Malonis, R.J., Lee, J.H., Tong, K., Vergnolle, O., Massimi, A.B., Yen, L.Y., Noble, A.J., 2020. Characterization of the SARS-CoV-2 S protein: biophysical, biochemical, structural, and antigenic analysis. *bioRxiv*. <https://doi.org/10.1101/2020.06.14.150607>.
- Hofmann, H., Pyrc, K., Van Der Hoek, L., Geier, M., Berkhout, B., Pöhlmann, S., 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. Unit. States Am.* 102 (22), 7988–7993.
- Hu, B., Ge, X., Wang, L.F., Shi, Z., 2015. Bat origin of human coronaviruses. *Virol. J.* 12 (1), 1–10.
- Huang, X., Dong, W., Milewska, A., Golda, A., Qi, Y., Zhu, Q.K., Marasco, W.A., Baric, R.S., Sims, A.C., Pyrc, K., Li, W., 2015. Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. *J. Virol.* 89 (14), 7202–7213.
- Hulswit, R.J., Lang, Y., Bakkers, M.J., Li, W., Li, Z., Schouten, A., Ophorst, B., Van Kuppeveld, F.J., Boons, G.J., Bosch, B.J., Huizinga, E.G., 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. Unit. States Am.* 116 (7), 2681–2690.
- Huynh, J., Li, S., Yount, B., Smith, A., Sturges, L., Olsen, J.C., Nagel, J., Johnson, J.B., Agnihotram, S., Gates, J.E., Frieman, M.B., 2012. Evidence supporting a zoonotic origin of human coronavirus strain NL63. *J. Virol.* 86 (23), 12816–12825.
- Jeffers, S.A., Tusell, S.M., Gillim-Ross, L., Hemmila, E.M., Achenbach, J.E., Babcock, G.J., Thomas, W.D., Thackray, L.B., Young, M.D., Mason, R.J., Ambrosino, D.M., 2004. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. Unit. States Am.* 101 (44), 15748–15753.
- Kaye, H.S., Marsh, H.B., Dowdle, W.R., 1971. Seroepidemiologic survey of coronavirus (strain OC 43) related infections in a children's population. *Am. J. Epidemiol.* 94 (1), 43–49.
- Ki, M., 2015. 2015 MERS outbreak in Korea: hospital-to-hospital transmission. *Epidemiology and Health* 37, e2015033.
- Kirchdoerfer, R.N., Cottrell, C.A., Wang, N., Pallesen, J., Yassine, H.M., Turner, H.L., Corbett, K.S., Graham, B.S., McLellan, J.S., Ward, A.B., 2016. Pre-fusion structure of a human coronavirus spike protein. *Nature* 531 (7592), 118–121.
- Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* 348 (20), 1953–1966.
- Lai, M.M., Cavanagh, D., 1997. The molecular biology of coronaviruses. In: *Advances in Virus Research*, vol. 48. Academic Press, pp. 1–100.
- Lam, T.T.Y., Jia, N., Zhang, Y.W., Shum, M.H.H., Jiang, J.F., Zhu, H.C., Tong, Y.G., Shi, Y.X., Ni, X.B., Liao, Y.S., Li, W.J., 2020. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 583, 282–285.

- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., Wang, X., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581 (7807), 215–220.
- Leth-Larsen, R., Zhong, F., Chow, V.T., Holmskov, U., Lu, J., 2007. The SARS coronavirus spike glycoprotein is selectively recognized by lung surfactant protein D and activates macrophages. *Immunobiology* 212 (3), 201–211.
- Li, F., Li, W., Farzan, M., Harrison, S.C., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309 (5742), 1864–1868.
- Li, F., 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* 89 (4), 954–1964.
- Li, F., 2016. Structure, function, and evolution of coronavirus spike proteins. *Annual Review of Virology* 3, 237–261.
- Li, W., Hulswit, R.J., Widjaja, I., Raj, V.S., McBride, R., Peng, W., Widagdo, W., Tortorici, M.A., Van Dieren, B., Lang, Y., Van Lent, J.W., 2017. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci. Unit. States Am.* 114 (40), E8508–E8517.
- Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., Choe, H., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426 (6965), 450–454.
- Li, W., Wong, S.K., Li, F., Kuhn, J.H., Huang, I.C., Choe, H., Farzan, M., 2006. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. *J. Virol.* 80 (9), 4211–4219.
- Li, Z., Tomlinson, A.C., Wong, A.H., Zhou, D., DesForges, M., Talbot, P.J., Benlekbir, S., Rubinstein, J.L., Rini, J.M., 2019. The human coronavirus HCoV-229E S-protein structure and receptor binding. *Elife* 8, e51230.
- Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattri, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., 2003. The genome sequence of the SARS-associated coronavirus. *Science* 300 (5624), 1399–1404.
- McIntosh, K., Dees, J.H., Becker, W.B., Kapikian, A.Z., Chanock, R.M., 1967. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci. U.S.A.* 57 (4), 933–940.
- Menachery, V.D., Yount, B.L., Debink, K., Agnihothram, S., Gralinski, L.E., Plante, J.A., Graham, R.L., Scobey, T., Ge, X.Y., Donaldson, E.F., Randell, S.H., 2015. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* 21 (12), 1508–1513.
- Menachery, V.D., Yount, B.L., Sims, A.C., Debink, K., Agnihothram, S.S., Gralinski, L.E., Graham, R.L., Scobey, T., Plante, J.A., Royal, S.R., Swanstrom, J., 2016. SARS-like WIV1-CoV poised for human emergence. *Proc. Natl. Acad. Sci. Unit. States Am.* 113 (11), 3048–3053.
- Milewska, A., Zarebski, M., Nowak, P., Stozek, K., Potempa, J., Pyrc, K., 2014. Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. *J. Virol.* 88 (22), 13221–13230.
- Millet, J.K., Whittaker, G.R., 2014. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl. Acad. Sci. Unit. States Am.* 111 (42), 15214–15219.
- Millet, J.K., Whittaker, G.R., 2015. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* 202, 120–134.
- Ou, X., Guan, H., Qin, B., Mu, Z., Woydyla, J.A., Wang, M., Dominguez, S.R., Qian, Z., Cui, S., 2017. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat. Commun.* 8, 15216.
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 11 (1), 1–12.
- Park, Y.J., Walls, A.C., Wang, Z., Sauer, M.M., Li, W., Tortorici, M.A., Bosch, B.J., DiMaio, F., Veesler, D., 2019. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nat. Struct. Mol. Biol.* 26 (12), 1151–1157.
- Peiris, J.S.M., Lai, S.T., Poon, L.L.M., Guan, Y., Yam, L.Y.C., Lim, W., Nicholls, J., Yee, W.K.S., Yan, W.W., Cheung, M.T., Cheng, V.C.C., 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361 (9366), 1319–1325.
- Pfefferle, S., Oppong, S., Drexler, J.F., Gloza-Rausch, F., Ipsen, A., Seebens, A., Müller, M.A., Annan, A., Vallo, P., Adu-Sarkodie, Y., Kruppa, T.F., 2009. Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerg. Infect. Dis.* 15 (9), 1377.
- Qian, Z., Ou, X., Góes, L.G.B., Osborne, C., Castano, A., Holmes, K.V., Dominguez, S.R., 2015. Identification of the receptor-binding domain of the spike glycoprotein of human betacoronavirus HKU1. *J. Virol.* 89 (17), 8816–8827.
- Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300 (5624), 1394–1399.
- Sabir, J.S., Lam, T.T.Y., Ahmed, M.M., Li, L., Shen, Y., Abo-Aba, S.E., Qureshi, M.I., Abu-Zeid, M., Zhang, Y., Khiyami, M.A., Alharbi, N.S., 2016. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* 351 (6268), 81–84.
- Shaffer, L., 2020. 15 drugs being tested to treat COVID-19 and how they would work. *Nat. Med.* <https://www.nature.com/articles/d41591-020-00019-9>.
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., Li, F., 2020a. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. Unit. States Am.* 117 (21), 11727–11734.
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., Li, F., 2020b. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581 (7807), 221–224.
- Shi, Z., Hu, Z., 2008. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res.* 133 (1), 74–87.
- Song, H.D., Tu, C.C., Zhang, G.W., Wang, S.Y., Zheng, K., Lei, L.C., Chen, Q.X., Gao, Y.W., Zhou, H.Q., Xiang, H., Zheng, H.J., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. Unit. States Am.* 102 (7), 2430–2435.
- Song, W., Gui, M., Wang, X., Xiang, Y., 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* 14 (8), e1007236.
- Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C., Zhou, J., Liu, W., Bi, Y., Gao, G.F., 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24 (6), 490–502.
- Thorlund, K., Dron, L., Park, J., Hsu, G., Forrest, J.I., Mills, E.J., 2020. A real-time dashboard of clinical trials for COVID-19. *The Lancet Digital Health* 2 (6), e286–e287.
- Tortorici, M.A., Walls, A.C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G.J., Bosch, B.J., Rey, F.A., de Groot, R.J., Veesler, D., 2019. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* 26 (6), 481–489.
- Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J., Welsch, S., Blanc, F.E., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., van Zandbergen, G., 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* 370 (6513), 203–208.
- Van Der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J., Berkhout, B., 2004. Identification of a new human coronavirus. *Nat. Med.* 10 (4), 368–373.
- Vincent, J.L., Taccone, F.S., 2020. Understanding pathways to death in patients with COVID-19. *The Lancet Respiratory Medicine* 8 (5), 430–432.
- Vlasak, R., Luytjes, W., Spaan, W., Palese, P., 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc. Natl. Acad. Sci. Unit. States Am.* 85 (12), 4526–4529.
- Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 180, 281–292.
- Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., DiMaio, F., Bosch, B.J., Veesler, D., 2016. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 23 (10), 899.
- Walls, A.C., Tortorici, M.A., Snijder, J., Xiong, X., Bosch, B.J., Rey, F.A., Veesler, D., 2017. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. Unit. States Am.* 114 (42), 11157–11162.
- Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., Wang, H., 2005. SARS-CoV infection in a restaurant from palm civet. *Emerg. Infect. Dis.* 11 (12), 1860.
- Wang, N., Li, S.Y., Huang, X.L., Huang, H.M., Zhang, Y.J., Guo, H., Luo, C.M., Miller, M., Zhu, G., Chmura, A.A., Hagan, E., 2018. Serological evidence of bat SARS-related coronavirus infection in humans, China. *Virology* 513 (1), 104–107.
- Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K.C., 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* 23 (8), 986–993.
- Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K.Y., Wang, Q., Zhou, H., Yan, J., Xi, J., 2020. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 181 (4), 894–904 e9.
- Wong, A.H., Tomlinson, A.C., Zhou, D., Satkunarajah, M., Chen, K., Sharon, C., DesForges, M., Talbot, P.J., Rini, J.M., 2017. Receptor-binding loops in alphacoronavirus adaptation and evolution. *Nat. Commun.* 8 (1), 1–10.
- Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367 (6483), 1260–1263.
- Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* 79 (2), 884–895.
- Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y., 2009. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med.* 234 (10), 1117–1127.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269.
- Wu, K., Li, W., Peng, G., Li, F., 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl. Acad. Sci. Unit. States Am.* 106 (47), 19970–19974.
- Yeager, C.L., Ashmun, R.A., Williams, R.K., Cardellicchio, C.B., Shapiro, L.H., Look, A.T., Holmes, K.V., 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* 357 (6377), 420–422.
- Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., Zhang, X., 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* 8 (1), 1–9.
- Zaki, A.M., Van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367 (19), 1814–1820.
- Zhang, H., Penninger, J.M., Li, Y., Zhong, N., Slutsky, A.S., 2020. Angiotensin-

converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.* 46 (4), 586–590.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798), 270–273.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733.