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

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

The human pandemic coronaviruses on the show: The spike glycoprotein as the main actor in the coronaviruses play

Pedro F.N. Souza a,⁎,1, Felipe P. Mesquita b,1, Jackson L. Amaral a, Patrícia G.C. Landim a, Karollyny R.P. Lima a, Marília B. Costa b, Izabelle R. Farias b, Luina B. Lima b, Raquel C. Montenegro b

a Department of Biochemistry and Molecular Biology, Federal University of Ceara, Brazil
b Drug research and Development Center, Department of Medicine, Federal University of Ceara, Brazil

Article history:
Received 23 January 2021
Received in revised form 25 February 2021
Accepted 26 February 2021
Available online 2 March 2021

Abstract

Three coronaviruses (CoVs) have threatened the world population by causing outbreaks in the last two decades. In late 2019, the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) emerged and caused the coronaviruses to disease 2019 (COVID-19), leading to the ongoing global outbreak. The other pandemic coronaviruses, SARS-CoV and Middle East respiratory syndrome CoV (MERS-CoV), share a considerable level of similarities at genomic and protein levels. However, the differences between them lead to distinct behaviors. These differences result from the accumulation of mutations in the sequence and structure of spike (S) glycoprotein, which plays an essential role in coronavirus infection, pathogenicity, transmission, and evolution. In this review, we brought together many studies narrating a sequence of events and highlighting the differences among S proteins from SARS-CoV, MERS-CoV, and SARS-CoV-2. It was performed here, analysis of S protein sequences and structures from the three pandemic coronaviruses pointing out the mutations among them and what they come through. Additionally, we investigated the receptor-binding domain (RBD) from all S proteins explaining the mutation and biological importance of all of them. Finally, we discuss the mutation in the S protein from several new isolates of SARS-CoV-2, reporting their difference and importance. This review brings into detail how the variations in S protein that make SARS-CoV-2 more aggressive than its relatives coronaviruses and other differences between coronaviruses.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction .................................................................................................................. 2
2. Coronaviruses taxonomy and linking to bats ................................................................. 2
3. The pandemic coronaviruses ....................................................................................... 3
3.1. SARS-CoV .............................................................................................................. 3
3.1.1. Origin and evolution of SARS-CoV-1 ................................................................. 3
3.1.2. Pathogenicity .................................................................................................... 4
3.2. MERS-CoV ............................................................................................................. 4
3.2.1. Origin and evolution of MERS-CoV .................................................................. 5
3.2.2. Pathogenicity .................................................................................................... 6
3.3. SARS-CoV-2 ......................................................................................................... 6
3.3.1. Origin and evolution of SARS-CoV-2 ................................................................. 6
3.3.2. Pathogenicity .................................................................................................... 7
4. Molecular biology of coronaviruses ............................................................................ 7
4.1. Coronaviruses genome organization ...................................................................... 7
4.2. Virion particles ........................................................................................................ 7

⁎ Corresponding author at: Biochemistry and Molecular Biology Department, Federal University of Ceará, CE, Brazil, Laboratory of Plant Defense Proteins, Av. Mister Hull, P.O. Box: 60451, Fortaleza, CE, Brazil.
E-mail address: pedrofihobio@gmail.com (P.F.N. Souza).
1 Both authors contributed equally to this article.
1. Introduction

Coronaviruses (CoVs) belong to a family of positive-sense, single-stranded, RNA viruses that are lipid-enveloped [1]. The classification by virologists classified coronaviruses into four genera as alpha, beta, gamma, and delta. The most famous are the α- and β-coronaviruses, given their ability to pass through the animal-human barriers, thus becoming clinically relevant to humans [2]. Nowadays, virologists reported seven coronaviruses able to infect humans and named as human coronaviruses (hCoVs). Of these, human coronavirus OC43 (hCoV-OC43), Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), Human coronavirus HKU1 (hCoV-HKU1), MERS-CoV are classified into the beta-genera, and human coronavirus NL-63 (hCoV-NL63) and Human coronavirus 229E (hCoV-229E) into the α-genera [1,3–6].

The hCoVs hCoV-HKU1, hCoV-OC43, hCoV-NL63, and hCoV-229E are not that harmful to humans as their infection results in non-symptomatic infection or, worse case, mild respiratory and less common gastrointestinal infection. Today, 5–30% of common cold cases are attributed to these hCoVs. Therefore, humans did not give the necessary attention to the hCoVs. However, in the last two decades, three outbreaks caused by hCoVs made humans pay more attention to them. The ongoing outbreak showed the human population how devastating hCoVs could be and driven many researchers worldwide to find either a vaccine or a treatment [1–6].

In December 2019, China warned the World Health Organization (WHO) about pneumonia caused by a new virus in Wuhan [7–9]. Later, the new virus was grouped into the coronavirus family and named severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). SARS-CoV-2 spreads quickly worldwide, and in a few months, more precisely in March 2020, the situation became worse and established a pandemic state [9–11]. The pandemic situation was followed by lockdown strategies adopted by the entire world. Despite that, the health care systems around the world went through high pressure, followed by the shuttering down of the economic situation in many countries [7].

Unlike the 2002–2003 SARS-CoV-1 outbreak, SARS-CoV-2 infection reached all continents quickly, proving to be more contagious. Comparing both viruses, the infection caused by SARS-CoV-2 is recognized by a broader clinical spectrum that comes from asymptomatic infection to severe viral pneumonia with respiratory failure and death [8,12,13]. In contrast to SARS-CoV and MERS-CoV, many SARS-CoV-2-infected patients developed low-titer of neutralizing antibody, leading them to suffer with prolonged severe symptoms and illness, suggesting a more effective SARS-CoV-2 immune surveillance evasion than SARS-CoV and MERS-CoV [14–18].

The ongoing outbreak caused by SARS-CoV-2 has taken many lives, threatened thousands more, and destroyed entire families worldwide. Indeed, SARS-CoV-2 (Coronavirus Disease 2019) is by far more transmissible than SARS-CoV-1 and MERS-CoV but is less lethal than they are (Table 1). However, SARS-CoV-2 higher transmissibility has resulted in 113,299,920 million of infected people with 2,512,823 million of death as of 25 February 2021, by far a larger number compared to other outbreaks [19,20].

The coronaviruses outbreak presented different spreads worldwide. Starting in November 2002 in Foshan-China, until March 2003, the SARS-CoV-1 outbreak has spread all over 29 countries (Fig. 1) [21,22]. The WHO status for SARS-CoV-1 today is controlled [21]. The MERS-CoV outbreak started in April 2012 in Zarqa-Jordan, and by September 2012, spread to 27 countries (Fig. 1) [22–25]. The status of MERS-CoV by WHO is sporadic continuous [24]. For SARS-CoV-2, as we know, the status of the outbreak is ongoing and has already spread to 213 countries (Fig. 1) [26].

The higher transmissibility of the SARS-CoV-2 is probable due to the higher number of accumulated positive mutations in the spike glycoprotein (S), which led this protein to be 20 times more effective in recognized human receptors than the spike from SARS-CoV-1 and MERS-CoV [27,28]. Based on that, the review is focused on the discussion and tracking the mutations in the S protein in three recent coronaviruses that posed outbreaks of SARS-CoV-1, MERS-CoV, and SARS-CoV-2. Additionally, we intend to understand the contribution of the S protein to the transmissibility of SARS-CoV-2 compared to SARS-CoV-1 and MERS-CoV.

2. Coronaviruses taxonomy and linking to bats

The coronaviruses are one of the two genera belonging to the Coronaviridae family. The other one is the toroviruses [29–31], both included in the Nidovirales order. Coronaviruses are highly disseminated among mammals in general, causing mild infections such as cold and gastrointestinal. However, in some cases, they might cause severe respiratory infection. In the past, coronavirus was known the most by an infection caused in animals such as bronchitis virus (IBV) in chickens and pigs and Feline infectious peritonitis (FIP) in cats [7,11,32].

Nowadays, at least 60 coronaviruses have been isolated and described, mainly from bats (BtCoVs). Most of the coronaviruses identified belong to the betacoronavirus group. It has been accepted that the power to flight of bats associated with the ability to migrate to regions far from the original spot has imposed a strong selective pressure for the coexistence with viruses [33–36]. Additionally, bats’ immune system is known as permissive, which means that bats can act as a reservoir to coronaviruses without developing the disease [37–40]. Bats, at the same time, could act as a carrier to 10 up to 17 α-coronaviruses and 7 up to 12 β-coronaviruses with the potential to jump to humans passing interspecies barrier causing disease [41]. To corroborate that,
long-term studies by genome sequencing revealed nucleotide similarity between coronaviruses found in humans and bats using the same cell entry, the Angiotensin-Converting Enzyme 2 receptor (ACE2) [38]. These studies [33–41] helped to track back the hCoVs to bats. For instance, it is hypothesized that the MERS-CoV evolved from a progenitor hosted in bats to infect dromedary and camels with the ability to directly infect humans [6].

| hCoV       | Case fatality rate | Symptoms                          | Refs          |
|------------|--------------------|-----------------------------------|---------------|
| SARS-CoV   | 9.6%               | Fever, Myalgia, Headache, Malaise, Chills, Nonproductive cough, Dyspnea, Respiratory distress, Diarrhea | [42,56,113]   |
| MERS-CoV   | 35.5%              | Fever, Cough, Chills, Sore throat, Myalgia, Arthralgia, Dyspnea, Pneumonia, Diarrhea and vomiting | [9,111,113]   |
| SARS-CoV-2 | 6.8%               | Fever, Cough, Chills, Myalgia, Arthralgia, Dyspnea, Pneumonia, Diarrhea and vomiting | [9,111,113]   |

Table 1: Human coronaviruses (hCoV) related-symptoms and their case fatality rates.

3. The pandemic coronaviruses

3.1. SARS-CoV

In 2002, SARS-CoV-1 emerged in Guangdong Province, China, that spread over five continents 29 countries, leading to 8098 cases and 774 deaths in nearly September 2003 [42–44]. The SARS-CoV was the first threat imposed by hCoVs to humankind that outbreaks around the world. During the outbreak, the estimated case fatality rate (CRF) was 9.6%, and the human-to-human transmission of SARS-CoV was confirmed. On 30 January, WHO, warned by Chinese scientific council, declared the SARS-CoV outbreak as an international public health emergency [21]. The first SARS-CoV cases occurred in employees who worked in a restaurant that handled wild mammals served as exotic food. Studies found that Chinese horseshoe bats had sequences of SARS-related CoVs and hosted a virus that shares similarities with SARS-CoV. Based on that, the origin of SARS-CoV is believed to have occurred in Chinese horseshoe bats [44,45].

The first suspected source of SARS-CoV was civets, a small mammal, due to the detection of the virus in those animals. Nonetheless, these animals are only transient hosts and found no evidence in wild civets [46]. Meanwhile, the evidence points to bats as hosts for SARS-CoV since they are permissive to SARS-CoV-like viruses [46].

3.1.1. Origin and evolution of SARS-CoV-1

SARS-CoV belongs to β-coronaviruses (Fig. S1) of the Coronaviridae family and is involved in zoonotic transmission and spread among humans through close contact [47]. It was later reported the first patient with SARS-CoV had prior contact with animals before developing the symptoms. The causative agent of SARS was later found in palm civets. However, strains isolated from handled market civets were transmitted from other animals. This suggests civets only as an intermediate host. Based on that, the hypothesis that the bat was the natural host of SARS-CoV came back to the spotlight [48,49].
Hu et al. [50] provided new information suggesting that the horseshoe bat (*Rhinolophus sinicus*) is the natural host of SARS-CoV. The authors reached that conclusion after the isolation of SARS-like CoVs that was homologous to SARS-CoV. Moreover, evidence shows the possible origin of SARS-CoV was based on the recombination of different SARS-like coronavirus in bats since some potential recombination sites were identified around the S gene [50,51]. Rest and Mindell [52] tried to elucidate the phylogenetic origin of SARS-CoV by comparing RNA dependent RNA polymerase (RdRp) sequences between SARS-CoV and other coronaviruses. The results showed that the SARS-CoV sequence is a recombinant virus. Recently, phylogenetic analyses showed a high similarity (80%) between SARS-CoV and SARS-CoV-2 [53]. Regarding the S gene, the similarity between SARS–CoV-2, and SARS-CoV is about 76% [47,54].

Coronaviruses often undergo mutations and genetic recombination, as they have error-prone RdRp, which results in a bigger diversity, adaptive evolution, and capacity to cause disease. Previous studies have shown that SARS-CoV mutated over the 2002 and 2004 epidemic to better bind to its cellular receptor (ACE2) and replication in human cells, enhancing virulence [55]. The S1 subunit of S protein has a receptor-binding domain (RBD) involved in the direct interaction with the ACE2 receptor [47,53–55].

3.1.2. Pathogenicity

During the SARS-CoV epidemic in 2002, patients usually presented fever, myalgia, headache, malaise, chills, cough, dyspnea, and respiratory distress generally 5 to 7 days later, resulting in death. In some cases, the gastrointestinal tract, liver, kidney, and brain [56]. SARS-CoV infection of the lungs leads to diffuse alveolar damage, epithelial cell proliferation, and an increase in macrophages [57,58].

The high transmission efficiency of the SARS-CoV occurs because it binds to a target (ACE2) on cells that are abundantly expressed, including pneumocytes in the respiratory system. The virus enters and replicates in these cells. Thus, mature virions are then released to infect new target cells [47,52]. SARS-CoV-1 has an incubation period of ~5 days, and 95% of patients develop the disease within 13 days of exposure [46]. SARS-CoV was reported to cause the respiratory system and the gastrointestinal and other organ systems [56]. This is because the SARS-COV human receptor ACE2 is abundantly expressed in the lungs and small intestine [55,57,58].

SARS-CoV has unique pathogenesis because it causes upper and lower respiratory tract infections [55]. Common early symptoms are fever, chills, coughing, malaise, myalgia, headache, and less common symptoms, including diarrhea, vomiting, and nausea. In some cases, not common, SARS-CoV is associated with thrombotic complications and hematologic manifestations [42]. Abnormal chest X-rays are detected in roughly 60% of patients infected with SARS-CoV. It was reported that 20–30% of patients infected with SARS-CoV require intensive care and mechanical ventilation [59–61]. In addition, at that time, it was noticed the SARS-CoV had developed high stability in aerosol and other surfaces, which had improved its transmissibility [62,63].

3.2. MERS-CoV

The MERS-CoV was primarily identified in Arabian Peninsula in 2012 as the major agent of a severe respiratory condition with a high case fatality rate (CFR) of ~35% (Table 1) [24]. Saudi Arabia was the first country to report MERS-CoV and the hotspot to its outbreak. MERS-CoV has officially 2279 laboratory-confirmed cases with 806 associated deaths in 27 countries (Fig. 1) [23,24,42,48,64,65]. Nowadays, there are yet reported cases of MERS-CoV infection. For example, WHO [24] has reported nine confirmed infection cases by MERS-CoV with five deaths from April to September 2020. It is proposed that during replicating the genome, the mutation rate of MERS-CoV is $4.81 \times 10^{-4}$ substitutions per site per year [65,66]. MERS-CoV has the same genome size and produces the same proteins (Fig. 2) of SARS-CoV, which will be in-depth, discussed later [67].

During the infection process, the S glycoprotein plays a central role in cell recognition, attachment, cell entry, and infection [27,28]. Structurally, the S protein is a trimetric protein with two subunits, S1 and
S2, in each monomer (Fig. 3A). The S protein from MERS-CoV binds to the cellular receptor dipeptidyl peptidase 4 (DPP4/CD26) driven by S1. At this point, the MERS-CoV infection differs from that presented by SARS-CoV and SARS-CoV-2, which employ interaction with ACE2 receptor (Fig. 3B). SARS-CoV-2 is the only known CoV that could use both ACE2 and DPP4/CD26 receptors. The second step of infection is the same for the three pandemic coronaviruses, which involve the viral membrane and host membrane fusion and virus arrivals into the cytoplasm [64]. The replication process inside the cytoplasm will be later discussed in this review.

3.2.1. Origin and evolution of MERS-CoV

Since the MERS-CoV outbreak, research worldwide has put together pieces of information to decipher the puzzle about MERS-CoV origin, pathogenicity, and transmissibility. The zoonotic event played a non-trivial part in MERS-CoV evolution and transmission [66]. Because a serosurvey study revealed that dromedary camels have antibodies against MERS-CoV, indicating that MERS-CoV has circulated in that area a long time before the outbreak. Harbor viruses that are closely genetically related to humans, the dromedary camels, constitute a source of pathogens that can be harmful to humans [68–72]. However, as to other coronaviruses, bats are believed to be the primary source of MERS-CoV origin and still works as a reservoir host of MERS-CoV. Genome sequence data revealed a similarity of 99.2–99.5% between bat and human MERS-CoV [23,48,68,69,71].

By analyzing 5030 fecal specimens from bats, Annan et al. [73] strongly suggested MERS-CoV likely comes from bats. Studies employing phylogenetic analysis showed two main MERS-CoV clades: A and B hosting camel and human MERS-CoV. MERS-CoV strains from humans from Jordan and Saudi Arabia belong to clade A. In contrast, clade B, divided into five groups (I to V), hosts MERS-CoV collected of humans from other regions and camels. The MERS-CoV from bats and hedgehogs forms a basal paraphyletic group to all camel and human MERS-CoV clades. This result suggests that, in addition to bats, hedgehog could also be the ancestor of camel and human MERS-CoV [36,66].

Additionally, it is known that recombination events can create new viral strains capable of infecting new hosts and evading immune responses from hosts. Overall, those mutations affect the S protein involved in both processes. There is evidence that recombination events on clade B groups, involving camel MERS-CoV and human MERS-CoV of different groups (I to V), produced different recombinant virus types (1 to 7) [66]. It was reported these recombination events might happen frequently and involving group V might happen broadly. Additionally, multiple recombination events indicate double infection, persistent infection, and superinfection during the transmission history [66].

The most recent common MERS-CoV ancestor was estimated as of March 2012, consistent with initial case detection and the beginning of the outbreak [74]. The genetic mechanisms underlying cross-species jumping remain poorly understood. However, it was reported that mutations on MERS-CoV S protein changed its surface charge,
enhancing cross-species jumping and viral entry on cells through sub-optimal DPP4 receptor [75–78].

The higher number of positive selections in the S protein is completely comprehensive given the central role played by S protein in MERS-CoV infection [66]. Most of the positive selection sites were in the RBD of S protein. RBD has two portions, one binding region, and one core region. Both are crucial to the virus recognizing and entering host cells and it. It was observed at two sites of mutations in the binding region of RDB, suggesting these amino acid substitutions might improve MERS-CoV binding capability to bind to different host cells and thus facilitate its cross-species transmission [66]. Chen et al. [79] state that most of the recombination and sequence diversity is in the S gene. Those changes may affect the structural conformation of RBD and the interactions with cognate human receptors.

As discussed before, the receptor used by MERS-CoV in the cells is different from that employed by SARS-CoV and SARS-CoV-2. SARS-CoV and SARS-CoV-2 employ the ACE2 receptor, MERS-CoV picked DPP4 [80,81]. As S-CoVs, the S1 C-terminal (CTD) of MERS-CoV acts as RBD, with two domains: the core and receptor-binding motif (RBM) [79,82–84]. S1-CTD from MERS-CoV is structurally similar to that from SARS-CoVs, composed of a major β-sheet scaffold. However, while in SARS-CoVs, the RBM has many unorderd structures forming loops, the MERS-CoV RDM mainly contains stranded β-sheets. It is postulated these structural differences were responsible for receptor-specificity between viruses [85].

3.2.2. Pathogenicity

The high similarities among the DPP4 from humans, camels, horses, and bats allow MERS-CoV to infect them. However, mutations in the DPP4 receptor from mice, hamsters, and ferrets prevent MERS-CoV infection [86–88]. MERS-CoV isolates sampled from humans and camels are highly similar to each other and use the human DPP4 receptor efficiently [89]. Studies revealed that these mutations change critical residues in mouse DPP4 compared to humans DPP4 affecting host specificity of MERS-CoV. Some of these residues mapped in mouse and hamster are different from humans; two residues 288 and 330 in mouse DPP4, and five in 291, 295, 336, 341, and 346 in hamsters DPP4 lead to the incompatibility of DPP4 from mouse and hamster with MERS-CoV S proteins thus preventing the infection [89].

Interaction between the residues 330 of DPP4 with MERS-CoV RBD Y499 has been suggested as a key interaction [82,83]. Besides that, Peck et al. [87] observed that the residue 330 mutation knocks out an NXT glycosylation motif in mouse DPP4. These results show that glycosylation acts as a determinant of DPP4-mediated host range and inhibits infection by MERS-CoV. Other nonpermissive hosts (ferret, hamster, guinea pig) also have a nonconserved glycosylation site in the region of RBD of S protein. Rhinolophus (RBM) [79,82–84], S1-CTD from MERS-CoV is structurally similar to that from SARS-CoVs, composed of a major β-sheet scaffold. However, while in SARS-CoVs, the RBM has many unorderd structures forming loops, the MERS-CoV RDM mainly contains stranded β-sheets. It is postulated these structural differences were responsible for receptor-specificity between viruses [85].

Overall, people infected and confirmed by a laboratory test for MERS-CoV were around 49 years old; 65% are males. The hospitalization time for patients to come back healthy was about 41 days [90,91]. MERS-CoV has a median incubation period of 5.2 days, reaching 12 days in normal people conditions, and a more extended period in patients immunocompromised and comorbidities [92,93]. The onset of symptoms to death 11.5 days. Even nowadays, the morbidity is still 36% (WHO, 2020), with more than 50% of patients showing viral accumulation on lungs revealed by radiography of the chest. Generally, the high viral loads on the lungs lead to acute respiratory distress syndrome [92,94,95].

The clinical manifestations of MERS-CoV could be in three forms. First, asymptomatic. Second, flu-like symptoms, cold, fever, cough, dyspnea, pneumonia, myalgia, diarrhea, vomiting, abdominal pain, chills or rigors, and malaise (Table 1). Third, acute progressive infection leading to acute respiratory distress syndrome, septic shock followed by multiorgan failure, and death [96–101]. Nearly 50% of patients developing either form 2 or 3 of MERS-CoV infection require intensive medical care in the intensive care unit (ICU). From those, up to 70% typically go to mechanical ventilation [100,101].

It is reported 50% of patients infected by MERS-CoV develop an atypical symptom called Acute Kidney Injury (AKI), which is a sudden kidney failure or damage. From those, up to 70% will pass by renal replacement therapy given the damage caused by MERS-CoV to kidneys [102–105]. MERS-CoV has been detected in upper and lower respiratory secretions at relatively high virus load and fecal samples [106,107].
CoV-2 progenitors [120]. Mutations, insertions, and deletions can occur near the S1–S2 junction of S protein from SARS-CoV-2, increasing the probability of polybasic cleavage site acquisition, enhancing transmision crossing the mammalian–human line, and improving human-to-human [121] (Fig. 3). Once acquired, adaptation was fixed as a new feature in the SARS-CoV-2 genome and enabled the rapid spread of SARS-CoV-2 to pandemic status.

3.3.2. Pathogenicity

Respiratory air droplets, aerosol, direct contact with contaminated surfaces, and fecal–oral transmission drive human coronaviruses. SARS-CoV-2 reaches the host via the respiratory tract, alveolar epithelial cells, vascular endothelial cells, and alveolar macrophages are the primary targets of the viral entry (Supplementary Fig. 2) [122].

Cell entry of SARS-CoV-1 and SARS-CoV-2 depends on the S protein binding with specific cellular receptors, ACE2 and TMPRSS2, playing a crucial role in the entry of both viruses into the host cells [28]. Briefly, the S protein of SARS-CoV-2 consists of two subunits, the S1 domain, and the S2 domain. SARS-CoV-2 utilizes RBD of the S1 domain to bind to the cellular receptor ACE2, which stimulates the TMPRSS2 to cleavages protein S at the S1 and S2 sites, allowing the cell membrane fusion and viral entry [27] (Fig. 3). Later, the role of S protein from three pandemic viruses in cell entry will be more in-depth discussed.

The clinical symptoms of SARS-CoV-2 infection are similar to SARS-CoV-1 and MERS-CoV infections described, being pneumonia the most described in the studies with abnormal chest CT examinations. However, patients infected with SARS-CoV-2 rarely have significant upper respiratory signs indicating the targeted cell may exist in the lower respiratory tract. Indeed, it has been reported SARS-CoV-2 infection mostly triggers deep airway inflammatory reactions and alveolar damage [123]. For SARS-CoV-2, the most common symptoms are cough, dyspnea, chest pain, myalgia/arthritis, diarrhea, nausea, vomiting, and common systemic symptoms observed: fever, chills, fatigue [124]. Also, SARS-CoV-2 infects several human tissues, such as the lung, intestinal tract, pharynx, heart, kidney, liver, brain, and blood [125,126]. Nevertheless, further studies are necessary to evaluate the effects of SARS-CoV-2 in extra-pulmonary infection sites.

The immune system may play a critical role in the severity of COVID-19. SARS-CoV-2 infection of pneumocytes induces a “cytokine storm”, which is an activation cascade of auto-amplifying cytokine production [116]. Local inflammatory responses promote the release of cytokines, including transforming growth factor-β1 (TGF-β1), tumor necrosis factor-α (TNF-α), interleukin-1β (IL-1β), IL-6, IL-2, IL-7, IL-10, granulocyte colony-stimulating factor (G-CSF), monocyte chemotactic protein (MCP), TNF-α and other chemokines to recruit leukocytes for the inflammation region leading to a multiple organ functional failure [116].

A recent study showed a plausible answer for this event: the direct SARS-CoV-2 infection of vascular endothelial cells with concomitant accumulation of inflammatory mononuclear cells in multiple organs in patients with severe COVID-19 [127]. Another plausible point is the involvement of SARS-CoV-2 infection with multiorgan failure due to the ACE2 and TMPRSS2 expression distribution in several human organs [125].

Another aspect is the higher proportion of macrophage and neutrophils observed in the patients with severe symptoms than those with mild symptoms and the chemokine related to these cells [128]. Interestingly, Zuo et al. [129] described the formation of neutrophil extracellular traps (NET) inside the microvessels of severe disease patients that could be a potential factor for severity.

4. Molecular biology of coronaviruses

4.1. Coronaviruses genome organization

Viruses classified as coronaviruses possess a single-stranded, non-segmented, large, and positive-sense RNA. The positive-sense RNA is an advantage acquired over the evolution process that mimics the cellular mRNA making its translation easier. Like cellular mRNAs, coronaviruses RNA possesses both 5’ cap and a tail at 3’ end [130,131] (Fig. 2A). The viral RNA is way larger than cellular mRNA and even compared to other viruses such as picornavirus (four times) and flavivirus (three times). The larger size of the coronaviruses genome varies from 27.3 to 31.3 kb, classified as the bigger RNA molecules are known until now [132].

Coronaviruses genome holds many ORFs (Fig. 2A), another difference compared to cellular mRNA. Precisely, the coronaviruses genome has 10 ORFs. The two largest ORF1a and ORF1b are responsible for producing up to 16 non-structural proteins (nsp) (Fig. 2A). Among those are PL proteinase, 3CL protease, RNA-dependent RNA polymerase (RdRp), and Helicase. The PL proteinase and 3CL protease are involved in polyprotein post-translational processes (Fig. 2A) [67]. The other ORFs are responsible for producing structural and accessory spike glycoprotein (S), an envelope protein (E), membrane attached protein (M), and nucleocapsid protein (N) (Fig. 2B). The genome of coronaviruses produces five canonical proteins. The RdRp at the 5’ end and S, E, M, and N at the 3’ end. This is a common sequence that RNA is translated following the sequence 5’–RdRp–S–E–M–N–3′ (Fig. 2A) [130–132].

Holding a positive-sense RNA is an excellent advantage to any viruses. For coronaviruses, it is not different. If only the RNA entry in a permissive cell the infective process starts. These processes have been shown in many studies over the years [131,133–135]. The most recent example of that is the vaccine produced by companies Pfizer and BioNTech (BNT162b2). BNT162b2 is a lipid nanoparticle–formulated, nucleoside-modified positive sense-RNA encoding membrane-anchored SARS-CoV-2 full-length S protein. Given the positive sense, when RNA is absorbed by the cell is promptly translated to viral S protein, which is exposed or externalized by the cells and recognized by the immune system producing antibodies against it [136].

During coronavirus infection, the genome acts as cellular mRNA producing the larger replica protein. The RNA from coronaviruses also possesses an important site dedicated to a ribosomal frameshifting event. After that, there is a ribosomal frameshifting site [137] (Fig. 2A). Ribosomal frameshifting in the displacement of ribosomal frames, also known as translation recording, is a biological phenomenon that results in multiple protein production from a single mRNA molecule (Fig. 2A). Although it is common in viruses, it goes beyond them. Organisms from all three kingdoms employ frameshifting to regulate gene expression [137–142]. Next, as usual, the genome is a model for protein synthesis, RNA replication, and assembly, producing new virus particles [143].

4.2. Virion particles

Coronavirus particles are spherical, pleomorphic (Fig. 2B) with diameters varying from 80 to 120 nm [144,145]. Attached to it membranes are found two types of the spike. One larger with projection 17–20 nm long produced S protein (Fig. 2B), and a smaller one 5–10 nm long today known as the hemagglutinin–esterase (HE) found in a subset of coronaviruses [144,145]. Recently, the development of advanced microscopy techniques has clarified coronavirus morphology. Ng and collaborators [146] catch three-dimensional (3D) images from SARS-CoV moving out from the Vero infected cell. By employing scanning electron (SEM) and atomic force (AFM) microscopies, it was possible to notice the small cauliflowers-like structures assumed by viral particles. To see the outside of the virus is not that hard. The problem is to evaluate the inside of a viral particle. To go that far, virion particles have been treated to nonionic detergents, allowing them to look inside the coronavirus particles. It was revealed that coronaviruses possess helical symmetric nucleocapsids, which is not common to positive-sense RNA viruses but for negative-sense RNA viruses. This is a great point of discordance among virologists. However, a few published studies are reporting that, but only we know is that more studies are required to obtain a clear picture of coronavirus virion particles.
5. The coronaviruses proteins

5.1. Membrane protein (M)

The M protein (Fig. 2B) is one of the three main structural proteins in coronaviruses. It is responsible for the viral particle conformation after assembling [147–150]. The M protein holds a small N-terminal domain extended to the outside of the viral particle, which is important to interact with the endoplasmic reticulum during infection [151]. Right after the N-terminal domain, there are three large, highly hydrophobic transmembrane domains followed by an even larger C-terminus domain, which stands inside the viral particle.

The M protein has regular conservation into the coronavirus family. The same is not applied to the coronaviruses groups. A consensus in all groups is that M protein has multiple-glycosylated stages. The preglycosylated stage of M proteins ranges in size from 25 to 30 kDa. However, multiple glycosylations could lead to a high molecular Mass of M protein [152]. Neuman et al., [147] reported the ability of M protein to oligomerizes to form larger structures. In the same study, the authors revealed two stages: the larger and compact. In the larger stage, M protein forms a structure 8 nm long. In contrast, the compact form has structures 6 nm long. The oligomerization stages of M protein are involved in many virion particle assemblies and even genome protection [147,153–155]. The interaction between transmembrane domains drives the interaction between M proteins; this interaction is important to exclude any leftover host membrane during the release of viral particles [147,156].

5.2. Envelope protein (E)

The E (Fig. 2B) is the smallest, membrane-integral, and less known structural protein from coronaviruses. One quite intriguing point is that E protein accumulates the most in the infected cells during viral replication. Still, only a very small portion is inserted in the new viral particle. The function of the excess of E protein in replication is obscure [157,158]. Most of E protein is localized outside the viral particle. This is because it is involved in the viral attachment to the ER and Golgi apparatus controlling virus intracellular traffic [158–160]. Although not known how the E protein is involved in the viral titers, viral particle production, and maturation [159,161–168].

5.3. Nucleocapsid protein (N)

The N (Fig. 2B) protein holds a molecular weight varying from 43 to 50 kDa, larger than M and E proteins but far smaller than S protein. It has a quite essential function to coronaviruses. Interacting with the RNA genome forms a helical structure like a necklace with beads providing stability to the RNA genome [169]. N protein sequence is divided into three domains. The highest part of the molecule has great amounts of Lys and Arg residues leading to a positive charge, which is involved in the interaction with the negative charge of viral RNA. This portion comprehends domains 1 and 2. For example, it was identified that the RNA bind ability of MHV N is attributed to domain 2 [169–171].

An unusual feature of N proteins in the phosphorylation. N protein from some but not all coronaviruses are phosphorylated. Some phosphorylation happens in a Ser and others in the N residues [172,173]. The biological function of phosphorylation in the N until today is obscure. However, the hypothesis is that the phosphorylation helps N protein to recognize only viral RNA. Chen et al. [172], employing Mass spectroscopic analysis and surface plasmon resonance, brought some information about the biological importance of N phosphorylation. The nonphosphorylated and phosphorylated N protein has the same ability to bind to viral RNA. However, the phosphorylated N protein binds only to the viral RNA. In contrast, the nonphosphorylated binds to both viral and non-viral RNA. These experiments by Chen et al. [172] revealed one biological effect of phosphorylation of N protein, which is the recognition exclusively of viral RNA.

5.4. Spike protein (S)

Now, after a brief explanation about coronaviruses and their proteins. It will be deeply discussed about the protein that is the focus of the review and plays essential roles during coronaviruses infection, Spike glycoprotein (S). The S protein is the third protein component and most abundant in the viral envelope (Fig. 2) [174]. The S protein is the larger protein from coronaviruses with about 180–200 kDa [174]. It has a larger N-terminal region (90% of protein) exposed outside the viral envelope and a shorter C-terminal region inside the viral envelope. Between N- and C-terminal regions, there is a transmembrane section liking both N- and C-regions (Fig. 3A) [174].

The S protein possesses high-level glycosylation on its structure. Before post-translational processing, the monomeric structure of S protein that inserts carbohydrates on its structure is about 128–160 kDa. After the glycosylation process, the size of S protein ranges between 180 and 200, which indicates a high amount of carbohydrates are inserted in its structure, most of them N-linked [175]. It is proposed that the biological function of glycosylation in the S proteins is to evade the host immune system during infection [176,177]. For example, a study from Delmas and Laude [177] with the S protein from porcine transmissible gastroenteritis virus revealed the glycosylation process coincides with the translation. The authors also showed that the glycosylation precedes the S protein trimerization. The glycosylation of TGEV S is essential to correct the folding of monomers [177].

S protein is still a huge protein even in the monomeric stage with 1273 amino acid residues. The first 13 amino acids form the signal peptide in the N-terminal site targeting the protein from the endoplasmic reticulum to the viral envelope. The S1 subunit goes from 14 to 685 amino acid residues. The S2 subunit starts at the amino acid 686 until the 1273 residue. Both regions are involved in cell recognition and entry (Fig. 3A and Supplementary 3) [178,179]. Like other proteins, the S protein is synthesized as an inactive precursor present in the viral envelope. During cell infection, the S protein is cleaved by a cellular protease releasing the S1 and S2 subunits activating it and allowing the membrane fusion and virus entry (Fig. 3A) [28,180,181]. This will be discussed later in this review.

5.4.1. The S1 subunit from S protein

The S1 subunit is smaller than the S2 subunit but has a quite relevant coronaviruses function (Fig. 3A and Supplementary 3). The S1 subunit is responsible for recognizing the cell receptor to allow virus entry into the cell cytoplasm. The S1 subunit hosts the RBD responsible for binding to ACE2 in the portion where the aminopeptidase cleavage point is present [181–183]. The S1 region also contains the N-terminal (NTD) and C-terminal (CTD) domains, both involved in the RBD recognition (Fig. 3A and Supplementary 3). For comparison purposes, the SARS-CoV-2 S1 CTD has more amino acid residues (21 aa) directly involved with the ACE2 interaction that SARS-CoV S1 CDT does (17 aa). Additionally, the mutational analysis revealed amino acid substitution from I472 in SARS-CoV S1 CTD for F486 SARS-CoV-2 S1 CTD strength interaction [181–183]. The S1 region also contains the N-terminal (NTD) and C-terminal (CTD) domains, both involved in the RBD recognition (Fig. 3A and Supplementary 3). For comparison purposes, the SARS-CoV-2 S1 CTD has more amino acid residues (21 aa) directly involved with the ACE2 interaction that SARS-CoV S1 CDT does (17 aa). Additionally, the mutational analysis revealed amino acid substitution from I472 in SARS-CoV S1 CTD for F486 SARS-CoV-2 S1 CTD strength interaction [181–183].

The RBD hosted by the S1 domain is critical for recognizing the ACE2 receptor by S protein. There are nine residues fully conserved in all coronaviruses involved in the ACE2-RBD interaction. Given the importance of RBD for the coronavirus, it becomes a great target for the action of neutralizing antibodies [27,28,180,182,183]. Our research group has recently performed a molecular docking with synthetic peptides against S protein from SARS-CoV-2 [184]. Out of eight, two peptides showed up as promising the most to bind toward S protein. Indeed, the peptides do not interact with the RBD domain. However, the interaction with S protein led to conformational changes in the S protein structures. The docking simulation led to a wrong interaction with the ACE2 receptor,
hence inhibiting virus entry in the cell [184]. In another study, Souza et al. [185] designed four peptides from the amino acid sequence of the ACE2 receptor. Those peptides specifically targeted the RBD domain from S protein. After interaction with peptides, RBD presented changes in the conformational structure impairing the correct interaction with ACE2 protein, suggesting that peptides could prevent cell infection by SARS-CoV-2.

5.4.2. The S2 subunit from S protein

The S2 subunit from the S protein hosts many important domains for S protein function. The first domain is the fusion peptide (FP) from 788 to 806 amino acid residues (Fig. 3A and Supplementary 3). FP is a short fragment highly conserved in the coronavirus family. Its composition is essentially made by apolar residues such as Gly and Ala, all-important during the membrane fusion activity of the S2 subunit during viral infection (Fig. 3A and Supplementary 3) [186,113,23].

Right after the FP domain, two sequences called heptapeptide repeated sequence 1 (HR1) (912–984 residues) and heptapeptide repeated sequence 2 (HR2) (1163–1213) with, respectively, 72 and 50% (Fig. 3A and Supplementary 3) [187]. Both HR1 and HR2 are located at the N-terminus portion of the transmembrane domain (TM) and follow the same structure HPPHCPC where H designates a hydrophobic residue, P regards a polar or hydrophilic, and C is any other charged residue. The HR1 and HR2 are essential to the viral fusion and entry function of the subunit in the cell [187].

The HR1 domains assemble to form a homotrimeric structure with highly conserved hydrophobic surfaces. These hydrophobic surfaces are exposed to the outside of the 3D structure of the S2 subunit to interact with the HR2 domain. In turn, HR2 forms a rigid helix followed by a highly flexible loop in the face that interacts with the HR1 domain. The interaction between HR1 and HR2 domains is strong and supported by hydrophobic and aromatic interactions forming a bigger domain fusion core region [188].

The HR1 and HR2 are essential to infection establishment by hCoVs and are thus highly conserved [188]. Given that, these domains are targeted continuously by potential antiviral molecules. That happens because another critical target, the RBD from the S1 domain of S protein, is highly variable, so it is hard to develop a drug targeting RBD. But the high conservation of HR domains makes them a great target [186–189]. For instance, Xia et al. [189] have developed an HR2-derived peptide that targets the HR1 region and inhibits cell infection by human coronaviruses.

6. Biological roles of the S protein

The S protein is the most important factor involved in viral infection. Because of that, it stands exposed on the viral envelope surface. It is a trimetric protein essential to receptor recognition, cell attachment, membrane fusion and virus entry [28,38,54,80,83,85,126,179,182,184,191]. Given S protein roles, it is present in all kinds of hCoVs and other viruses such HIV and Ebola [191] but with other names.

The S protein possesses two conformational states (Fig. 3B). The conformational dynamics of S protein is due to the flexibility of NTD and RBD. The first is the closed state where the RBD domain cannot recognize the ACE2 receptor (Fig. 3B). The closed state is also called fusion state because it precedes membrane fusion [192]. The closed state (prefusion) of beta coronaviruses was determined by cryo-electron microscopy [193]. In general, the closed state of S protein from coronavirus is similar to the, considering higher complexity, hemagglutinin from influenza virus. In the closed state conformation, the RBD domain is trapped into a pocket formed by the NTD region and RBD itself, inhibiting the interaction with the ACE2 receptor [85,193].

Additionally, in the closed state, the S1 heads stand on top of S2 subunits to avoid the S2 conformational transitions. In the S2 domain, the HRs domain assumes the helices form to stabilize the S2 subunit. In the coronavirus fusion HR-1 domain, the hydrophobic residues support the formation of a small loop and helix, which are buried inside the prefusion structure. The proteolysis is essential for the conformational transition of S2, one at the S1/S2 border and the other at the N-terminal region [194,195]. The S protein moves to the second stage called an open state (Fig. 3B). In the open state, the RBD is now exposed outside the S protein structure and can recognize and bind to the cellular receptor (Fig. 3B) [27,28,67]. In this stage, the RBD dynamics is modulated by proteolysis allowing the interaction with the ACE2 receptor. After interaction by RBD-ACE2, the virus will start the process to enter within the cell, and the S protein starts to change to another state called post-fusion, leading to the viral and cellular membrane fusion [27,28,67,189,193–195]. Recently, Wang et al. [196] reported the S protein could have an alternative route to enter cells. Rather than interact with RBD, S protein from hCoVs could interact with a CD147, which is a transmembrane glycoprotein from the immunoglobin superfamily. It is known for the CD147 involvement in tumor development, bacterial, virus, and Plasmodium invasion and infection [196–198].

Wang et al. [196] revealed vero E6 and BEAS-2B cell lines either lacking or blocked to CD147 or blocking presented no signal of SARS-CoV-2 amplification. Additionally, the expression of CD147 in SARS-CoV-2 non-permissive human cells allowed virus entry and infection. The authors also found high viral loads in the mice lungs expressing human CD147, but not in wild type mice. Indeed, the results presented by Wang and co-authors [196] are quite interesting.

In the same way, since November 2020, a new cellular receptor came up as hosts in facilitating SARS-CoV-2 entry. Those receptors either enhance the ACE2-mediated entry of SARS-CoV-2 or act as alternative receptors for SARS-CoV-2 entry. SARS-CoV-2 could employ these alternative receptors to enter the cell with a low expression of ACE2. For instance, the tyrosine-protein kinase UFO (AXL) acts as a receptor by interacting with the NDT domain of S protein, demonstrated by in vitro cell culture model and COVID-19 patients samples [201]. Another in vitro model study showed the high-density lipoprotein (HDL) scavenger receptor B type 1 (SR-B1) as a facilitator of ACE2-dependent entry of SARS-CoV-2 through the interaction of the S1 subunit. Blocking the S protein interaction with a specific monoclonal antibody against cholesterol/HDL inhibits the HDL-enhanced SARS-CoV-2 infection [202].

Cantuti-Castelvetri et al. [203] and Daly et al. [204] revealed that neuropilin-1 (NRP1), which is known to bind furin-cleaved substrates, potentiate SARS-CoV-2 infectivity. The NRP1 protein is highly expressed in the respiratory and olfactory epithelium, with the highest expression in endothelial and epithelial cells. Those cells present a low expression of ACE2 receptors, which is why SARS-CoV-2 employs this protein to enter the cell [203,204]. Amraei et al. [205] reported that CD209L and CD209 are members of the C-type lectin superfamily and could act as mediators of SARS-CoV-2 entry cell and increase viral pathogenesis. The CD209L is highly expressed in human type II alveolar cells, lungs, and liver, whereas CD209 protein is expressed in alveolar macrophages. Both CD209L and CD209 interact with the RBD of S protein from SARS-CoV-2 to enter cells. The mechanism of these proteins is quite similar to ACE2-mediated cell entry by SARS-CoV-2.

Recently, in an elegant experiment, Tang et al. [206] revealed a new candidate molecule used by SARS-CoV-2 to enter cells. ACE2-knockout mice are not protected from SARS-CoV-2 infection. The authors showed that the transferrin receptor (TIR) is the target by S protein to allow SARS-CoV-2 entry on cells. Even though new finding supports alternative molecule employment to invade cells by SARS-CoV-2, little is known about the role of these receptors for SARS-CoV-2 entry on cells. Based on that, the discussion below will be mainly focused on the ACE2-S protein interaction, which there more information allowing a deeper discussion has been supported by the literature.

Most of the CoVs enter cells by interacting with the RBD domain from S protein with the ACE2 receptor [199]. The ACE2 receptor is a widely expressed protein in the cell membrane of many tissues such as lungs, heart, kidney, renal, cardiovascular, and intestine [200].
Discovered in 2000, ACE2 holds 61% of similarity with ACE. Structural analysis of ACE2 protein revealed an N- and C-terminal followed by a unique transmembrane α-helix domain followed by an intracellular portion [113]. The difference between ACE and ACE2 is the most related to their activity. At the time, ACE catalyzed the conversion of angiotensin I in angiotensin II. ACE2 acts in angiotensin II conversion in two forms: angiotensin I-1 and angiotensin I-9 [201]. Until the pandemic situation imposed by the coronaviruses, ACE2 moves from the shadow to light as an important protein to viral infection. Now, it is quite clear that SARS-CoV-2 and other coronaviruses employ the ACE2 to come into a cell and start an infection.

The infectious process starts when a coronavirus-infected person sneezes, releasing air droplets containing virus particles into the environment (Fig. 4). Without a social distancing, a healthy person intake, by natural breathing, the contaminated air allowing the air droplets infected with the virus to reach the lungs (Fig. S2). At that point, the S protein becomes the principal actor in the coronavirus infection by driving viral recognition and entry on cells. The S1 subunit from the S protein starts the process by promoting RBD interaction with ACE2 (Fig. 4 - 1). After the attachment of the virus supported by S protein and low pH conditions act as targets to the viral fusion of membranes [178,179,182].

After the attachment, the S protein, driven by the S2 subunit, promotes the viral fusion, which is the fusion of the viral membrane with the host cell membrane. The trigger to initiate the membrane fusion is the proteolytic action suffered by S protein after the attachment to the receptor. As a result of the cleavage, the S protein is separated into two parts, S1 and S2, by cellular proteases. Even after the cleavage, the subunit S1 and S2 remain non-covalently bonded to ensure the membrane fusion follows through [202,203].

Hasan et al. [203] and Millet et al. [202] revealed that in the SARS-like coronaviruses such as SARS-CoV and SARS-CoV-2, the S protein has an uncleaved state. In contrast, other coronaviruses possess S protein mainly in the cleaved state. Additionally, the mutational analysis revealed that SARS-CoV-2 had accumulated mutations increasing the number of cleavage sites on S protein, leading to higher cleavage rates and cell entry and infectivity [202,203]. Many studies over the years [28,180] revealed host proteases involved in the most S protein cleavage are TMPRSS2 and trypsin, even for SARS-CoV and SARS-CoV-2. The S protein from SARS-CoV and SARS-CoV-2 are essentially the same. However, the high number of cleavage sites in the SARS-CoV-2 S protein might be involved in higher contagious levels of SARS-CoV-2 compared to SARS-CoV.

Back to viral fusion, the domains FP and both HR on S2 is critical to viral fusion. The cleavage of S protein exposes the FP domain, which is, in turn, inserted in the host membrane triggering the viral fusion [114,178,191]. Once anchored in the host membrane by the FP domain, the distances between viral and host membrane are dramatically reduced, and HR1 stands close to the cell membrane, and HR2 is yet anchored in the viral membrane. Now starts the last and crucial process in membrane fusion. The HR2 domain turns up toward the HR1. Both domains form a helix structure to produce a fusion core, and the viral membrane is pushed over toward the cell membrane tightly, and both membranes fuse. The virus reaches the cytoplasm. Now it is just a matter of making the cell work in virus replication. Once inside the cytoplasm, the virus hijacks cellular proteins and supplies to replicate the genome and produce thousands of new viral particles [204,205].

Inside the cell, the positive-sense RNA from coronaviruses, which works as mRNA, is released, and translated by the cellular ribosomes (Fig. 4 - 2). The first protein produced is the huge replicase polyprotein complex (Fig. 4 - 3). The positive-sense RNA is then used as a model by the replicase to produce the negative-sense RNA, the subgenomic RNAs, and thus new positive-sense RNA molecules (Fig. 4 - 4). The subgenomic RNAs were used to produce S, E, M, N proteins (Fig. 4 - 5 and 6). The proteins S, E, and M move to the ER and attach to its membrane. At the same time, the N protein interacts with the RNA molecules to form the nucleocapsid particles (Fig. 4 - 7). Now, induced by structural membrane-bounded protein, the membrane of ER suffers conformation changes to assume a spherical format with the nucleocapsid particle inside, which is the new virion particle. The new viral particle moves now out to infect new cells (Fig. 4 – 9). This is the common way of CoVs replication. S protein is not called here as a principal actor without a motive. In some cases, during CoVs, the infection can lead to the formation of syncytia [206–208].

Syncytia are giant infected cells induced by some viruses, like coronaviruses. All pandemic coronaviruses, SARS-CoV-1, MERS-CoV and SARS-CoV-2 [206–208] have the ability to induce syncytia formation. Somehow, in coronavirus infection, the host cell showed the S protein on the membrane, which is a quite important function of S protein. The S protein interacts with the ACE2 with neighboring healthy cells and thus induces membrane fusion of many cells forming giant known syncytia. By inducing the syncytia formation, the coronaviruses can continuously infect cells without moving out cells. This is an incredible and essential ability acquired by S protein over evolution. The syncytia formation induced by S protein allows the virus to keep spreading without extracellular virus particles working as an evasion mechanism of the immune system.

7. S protein in the pandemic coronaviruses

As discussed above, in the last two decades, the human king faced three outbreaks caused by hCoVs. These coronaviruses have threatened and taken many lives worldwide with different spreading rates, fatality cases, and regions affected (Fig. 1 and Table 1). The pandemic hCoVs had present different spread and human-to-human transmissible rates. That factor could be attributed to the mutation in the S protein from all of them. Over the years, hCoVs have accumulated many mutations, some negative, some positive, which culminate in the current pandemic scenario. This section will discuss the characteristics and differences of S protein from the pandemic hCoVs. Most of the comparisons are made between SARS-CoV and SARS-CoV-2 when pertinent MERS-CoV will be included.

The mutations that lead to the change in amino acid in the S protein could be important information to track the natural sources of CoVs, allowing intra-species or inter-species jumping of the virus. For example, genomic sequence experiments revealed that SARS-CoV-2 is closer to bat CoVs than other hCoVs, which indicates its origin [37,51,62,69,115,117,121,190,209]. The RatG13, CoV from bats, has a similarity of 93.1% in the gene and 98% in the protein sequence of S protein with SARS-CoV-2. In contrast, SARS-CoV-2 has only 80% similarity in the S protein with other hCoVs [37,51,62,69,115,117,121,190,209].

In contrast to other CoVs, SARS-CoV-2 has accumulated many mutations in S protein conserved regions (Fig. S3). Because of that, even with some similarities, SARS-CoV-2 is much sufficiently different from SARS-CoV and SARS-like coronaviruses [40,47,51,111]. In comparison to pandemic hCoVs, SARS-CoV-2 is, indeed, closer to SARS-CoV than MERS-CoV (Fig. S1A) [45,93,98,178]. Even though evolutionary studies put SARS-CoV-2 and SARS-CoV in the same group, they still hold remarkable variations between them [111]. Wu et al. [111] revealed high similarities between SARS-CoV-2 and SARS-CoV at the polyprotein level, but many significant alterations exist. For instance, the most remarkable is the absence of protein 8a in SARS-CoV-2 and its presence in the SARS-CoV, which indicates a mutation that deletes or produces a truncated inactive 8a protein in SARS-CoV-2. For sure, the absence of 8a did not affect SARS-CoV-2 fitness [111].

To have a better picture of the S protein changes of the pandemic hCoVs, we performed an alignment comparing the entire sequence of S protein from MERS-CoV (accession number K9NSQ8), SARS-CoV-1 (accession number P59595), and SARS-CoV-2 (accession number P0DT2C) deposited in the UniProt database (Supplementary Fig. 3). By analyzing the results, the first point noticed was the length of S protein in all viruses. The MERS-CoV S protein has a higher number of amino acid residues, 1345. In contrast, S protein from SARS-CoV and SARS-
CoV-2 has, respectively, 1243 and 1261 amino acid residues (Supplementary Fig. 3). That might be a result of many deletions in S protein from both SARS-CoV and SARS-CoV-2.

The S protein alignment from the pandemic hCoVs revealed that compared to the MERS-CoV S protein, the S protein from SARS-CoV-1 and SARS-CoV-2 has 14 points of deletions. Probably, because of those deletions, the S protein from SARS-CoV-1 and SARS-CoV-2 are smaller than the MERS-CoV S protein (Supplementary Fig. 3). Of these, 12 points of deletions are present in the S1 subunit and 2 on the S2 subunit. Most of the point mutations are situated in regions that have no important domains for the infection process. But somehow alter the 3D of S protein, making it more compact in both SARS-CoV and SARS-CoV-2 than MERS-CoV.

Another great difference between S protein from SARS-CoV and SARS-CoV-2 is the pattern of glycosylation [210–212]. The presence of an additional glycosylation spot in Asn370 in SARS-CoV-2 compared to S protein from SARS-CoV provides an additional glycosylation state in S protein from SARS-CoV-2. This spot belongs to a domain involved in the membrane attachment and thereby enhances the interaction with the receptor and membrane fusion [211]. Point mutations in SARS-CoV-2 S protein are said to increase virulence through the instability of the viral machinery and altered viral-to-cell-membrane fusion (Supplementary Fig. 3).

It is already known that the amino acid composition change does not reflect protein structure changes and thereby function. To check that, we provide a root mean square deviation (RMSD) calculation of the atomic position on S protein from the pandemic hCoVs (Fig. 5). The RMSD provides a value: If the value is 0, the atoms from both structures are in the same position, with no alteration in the 3D structure. However, if the value is anyone higher than zero, it indicates the atoms are in different positions, and this value is high, the structures are very different.

To perform the structural alignments was used the 3D structures from S proteins MERS-CoV (PDB ID: 5X5C), SARS-CoV-1 (PDB ID: 5X58), and SARS-CoV-2 (PDB ID: 6Z97) deposited in the UniProt database deposited in the Protein Data Bank (PDB, https://www.rcsb.org/). By performing a structural alignment, it is possible to see structural changes. Fig. 5 shows the structural alignment between the S proteins from MERS-CoV and SARS-CoV. It is possible to analyze in many positions the differences in both proteins. The RMSD value for this alignment was 12.93 Å, indicating both atomic positions are quite different. The comparison between S protein from MERS-CoV and SARS-CoV-2 is more interesting (Fig. 5). In some regions such as NDT, the arm of S1 and HR is possible to evaluate that the S protein structure from MERS-CoV is more prominent than S protein than SARS-CoV-2. It seems the S protein from SARS-CoV-2 is more compact with a lower atomic distance than MERS-CoV (Fig. 5). The RMSD value for the structural alignment of S protein from MERS-CoV and SARS-CoV-2 is 15.90 Å, which indicates both structures are quite different. Despite the differences in S protein it is important to remember both viruses employed different receptor to come inside the cell. Together these differences are responsible for complete behaviors regarding the spread, infectivity, and severity of the disease.

The RMSD analysis of S protein from SARS-CoV and SARS-CoV-2 (Fig. 5) revealed a lower value, 9.21 Å, than that obtained in the comparisons of SARS-CoV x MERS-CoV and SARS-CoV-2 x MERS-CoV. It is important to notice that even the S protein from SARS-CoV and SARS-CoV-2 are similar, they share many differences. By looking at amino acid sequence alignment (Supplementary Fig. 3) it is possible to notice many changes in the amino acid residues at the same position.
These changes could be responsible for the altered structures revealed by RMSD analysis.

It is important to notice that some point mutations occur on the RBD site (Supplementary Figs. 3 and 4). To have a better overview it was performed an alignment comparing the sequence of RBD extracted from the sequence of S protein from MERS-CoV (accession number K0NSQ8), SARS-CoV-1 (accession number P95959), and SARS-CoV-2 (accession number P0DT12) deposited in the UniProt database (Supplementary Fig. 4). For example, the sequences 538EDCDYRRKLSS546, 575VQYV578, and 587KLEFAN592 are present in the S protein from MERS-CoV absent in that one from SARS-CoV and SARS-CoV-2 (Supplementary Fig. 4). Deletion points were not an exclusivity from SARS-CoV and SARS-CoV-2. The S protein from MERS-CoV presents 4 points of deletion than SARS-CoV and SARS-CoV-2 (Supplementary Fig. 3).

The RBD portion has nine cysteine residues able to form four disulfide bonds. Comparing the location of these disulfide bonds in the SARS-CoV, MERS-CoV, and SARS-CoV-2, it was possible to see if the positions of disulfide bonds are similar SARS-CoV-2 and SARS-CoV, but in a different position in comparison both with MERS-CoV (Fig. 5 and Supplementary Fig. 3) [183]. This result suggests a closed mode of action between RBD from SARS-CoV and SARS-CoV-2 but not too close to the action of RBD from MERS-CoV.

Regarding S protein, most of the amino acid changes were in RBD, affecting the conformational flexibility of the protein or binding interactions with ACE2 [213]. The mutations accumulated by SARS-CoV-2 lead to five amino acid residues that are different than those in SARS-CoV. In SARS-CoV, the residues are Tyr455, Leu486, Asn494, Asp495, Glu494, and Ser495, in contrast, in SARS-CoV-2, the residues are Leu455, Phe486, Asp495, Ser495, and Tyr495 [214] (Supplementary Fig. 4). These differences in the SARS-CoV-2 RBD domain allow it to bind to ACE2 with an affinity 20 times higher than SARS-CoV [214].

A key mutation in the RBD from SARS-CoV-2 compared to SARS-CoV is the presence of a Lys417. The core region of RBD from SARS-CoV-2 has a unique residue of Lys417, forming a salt-bridge interaction with a residue of Asp at position 420. This interaction strengthens the energy bind of SARS-CoV-2 RBD with ACE2. In contrast, this position of RBD from SARS-CoV holds a Val residue that did not allow such type of interaction (Supplementary Fig. 4) [183]. Indeed, the presence of Val residue interrupts the interaction, thus weakens the interaction with ACE2. Besides, the presence of Lys417 provides a positively charged patch providing an electrostatic surface, which is not present in the RBD from SARS-CoV [183]. Besides strengthening the SARS-CoV-RBD::ACE2 interaction, the Lys417 is also important to 5 protein immunogenicity (Supplementary Fig. 4). The substitution of a Val to Lys in the RBD from SARS-CoV-2 hinders the interaction of anti-SARS-CoV antibodies and thus the neutralization of SARS-CoV-2 by anti-SARS-CoV antibodies (Supplementary Fig. 4) [183].

The RMSD analysis revealed some important information about the structure of all S proteins from pandemic hCoVs. However, the S protein is a huge protein with a complex structure (Fig. 5). To obtain a better view of the alteration in RBD, it was performed an RMSD analysis of only the RBD portion, comparing in all pandemic hCoVs. The RMSD allowed a more in-depth evaluation of differences (Fig. 6). To perform the structural alignments was used the 3D structures of RBD from MERS-CoV (PDB ID: 6L8Q), SARS-CoV-1 (PDB ID: 2AJF), and SARS-CoV-2 (PDB ID: 6YW1) deposited in the Protein Data Bank (PDB, https://www.rcsb.org/). The comparison between RBD from MERS-CoV and SARS-CoV revealed at least three positions with a high variation in the structure (Fig. 6 arrows). The RMSD value was 10.34 A. Corroborating the results from S protein RMSD analysis of MERS-CoV and SARS-CoV-2 (Fig. 6), the RBD analysis domain presented a high RMSD value of 11.6 A, indicating the structures were quite different. The arrows highlighted the most prominent differences (Fig. 6).

Regarding the RMSD analysis of RBD follows the same pattern of S protein for SARS-CoV and SARS-CoV-2 (Fig. 6). Although there are many differences in the amino acid sequence of RBD showed by the alignment (Fig. 54). The conformational structure of RBD from SARS-CoV and SARS-CoV-2 is quite similar as represented by the RMSD analysis with a value of 2.09, which is very low compared to other RMSD analysis (Fig. 6). The similarity between RBD sequences from SARS-CoV and SARS-CoV-2 is around 74% allowing both viruses to use ACE2 receptors to a different extent in the cell [118]. Indeed, RBD from SARS-CoV-2 can form a higher number of interactions with ACE2 than RBD from SARS-CoV, supporting the high affinity of SARS-CoV-2 RBD to ACE2 than SARS-CoV. This is possible because RBD from SARS-CoV-2 performs more atomic interaction with the ACE2 compared to SARS-CoV. Experiments have shown the RBD from SARS-CoV-2 binds to ACE2 in an nM scale concentration, with dissociation at 4.7 nM. In contrast, the RBD from SARS-CoV has a constant of dissociation about 31 nM [28,178,183].

All these differences in the atomic arrangement led to suitable differences in S protein are directly involved in the interactions with the ACE2 receptor and thus binding affinities. Based on that, it is feasible to suggest that those alterations are responsible for the higher affinity of SARS-CoV-2 RBD than SARS-CoV RBD by the receptor, and thereby higher transmissibility [28,178,183]. Experimental analysis using animals is an important source of information if the mutations on RBD affect COVID-19 symptomatology and may provide information about the origin of the pandemic situation [215]. Based on RBD sequences’ similarities, there have been discussions of convergent evolution between SARS-CoV and SARS-CoV-2 RBD structures enhancing the affinity of SARS-CoV-2 by ACE2 [183,215].
Albeit everybody looks for a mutation in the RBD given its interaction with ACE2 receptor, other regions. For example, it was reported that many mutations changed amino acid sequences in the HR1 domain of SARS-CoV-2 compared to SARS-CoV, which might be associated with improved and more robust interaction with the HR2. Thereby enhancing the membrane fusion ability of S protein, which increased SARS-CoV-2 infectivity (Fig. S3) [181,189,193,195]. Comparing SARS-CoV and SARS-CoV-2, the S2 domain is well preserved in both N- and C-terminal domains. All regions involved in viral particle fusion such as FP, HR1, HR2, TM, cytoplasmic portion have similarities, respectively, of 93%, 88%, 100%, 93%, and 97% (Fig. S4) [216].

It was already described how mutations favoring SARS-CoV-2 S protein cleavage improve viral loads and the higher efficiency of spread compared to other hCoVs [27,181,216]. These mutations increase the number of proteases that could cleavage the extracellular portion of S protein from SARS-CoV-2, increasing its infectivity and virulence [181,183]. Mutational studies showed that compared to other coronaviruses S protein, one from SARS-CoV-2 has a higher number of hotspots for glycosylation. It is quite possible that SARS-CoV-2 is using these new sites of glycosylation to evade immune system surveillance, making it S protein less antigenic than those from other coronaviruses do and thereby contributing to its pandemic spreading [118,181,207,216,218].

8. Mutations in S protein from SARS-CoV-2

Based on all discussion made above, it is quite clear the high mutation rates in hCoVs. Most of the discussion focused on the mutations in the pandemic hCoVs SARS-CoV, MERS-CoV, and SARS-CoV-2. In this topic, the focus will be the mutations that happened in the S protein from SARS-CoV-2 since the beginning of the outbreak.

Mutations are a process intrinsic to all RNA viruses. There are some results from mutation: (1) mutations are a natural process resulting from the errors made by the RNA polymerase during RNA replication; (2) Mutations could be induced by RNA-editing system as a natural defense system from the host; (3) genetic variation from recombination of two viral lineages 8,9,11,12,13. To viruses, mutations could be neutral, which is the majority, harmful or to a less extent represent some advantages to virus [219]. In SARS-CoV-2, the mutations have a lower rate, an estimated 1–2 mutations per month [220], because the RNA polymerase has an unusual proof-reading mechanism [221]. Even so, many mutations have been related to SARS-CoV-2 since the outbreak has started [222,223]. The mutations in SARS-CoV-2 are the most listed in S protein and associated with higher transmissibility and virulence [222,223].

At the beginning of the outbreak, in January of 2020, WHO recognized the circulation of a SARS-CoV-2 variant with a D614G substitution in the S protein (Fig. 7). [7,26,109]. According to WHO [7,26,109], the D614G variant was first identified in China and became the dominant variant worldwide by June 2020. The D614G variant, compared to other SARS-CoV-2, has enhancements in infectivity and transmission [222] but does not lead to higher illness.

The D614G mutation changes a Gly residue by Asp residue at 614 positions located at the C-terminal region of S1 subunits from S protein, a region directly associated with S2 subunit [222,223]. The advantage of D614G mutation is the high transmission, infectivity, and viral loads in patients with COVID-19 [222,223]. Experiments performed by Plante et al. [224] showed that sera from hamsters infected with the wildtype slightly neutralized the D614G variant. This is a concerning result because it has implications in the vaccine efficacy and thus in antibody therapy.

As happens to other mutations discussed in SARS-CoV-2, the D614G is not in the RBD domain and does not improve the affinity of S protein by ACE2. Indeed, the D614G results in a more stable S protein than wild type S protein and increases the assembly of more functional S protein outside viral particles. Both these advantages increase SARS-CoV-2 infectivity [222–224].

Last December 2020, WHO recognized a new SARS-CoV-2 mutant discovered in the United Kingdom from samples collected 20-Sept-2020 in the county of Kent and another on 21-Sept-2020 from London [225]. This mutant was named SARS-CoV-2 lineage B.1.1.7 (SARS-B117). The mutant SARS-B117 has 17 mutations. Of these, eight mutations were found in the gene responsible for S protein production B117). The mutant SARS-B117 has 17 mutations. Of these, eight mutations were found in the gene responsible for S protein production [225]. The mutations in the S protein were: two deletions: HV 69–70 and Y144 (Fig. 7). The deletions in the 69 and 70 position of the S protein are related to the evasion of S protein from the immune response.
The rapid increase in cases of COVID-19 in Amazonas, Brazil, has led to the collapse of the public health system and the loss of many lives. The emergence of a new variant, SARS-B117, has been identified as the source of the epidemic. This variant, which carries the N501Y mutation in the spike protein, is more transmissible than its predecessors. The N501Y mutation has also been found in variants circulating in other parts of the world, such as P.1 in Brazil and B.1.1.7 in the UK. These variants have been shown to be more resistant to neutralization by monoclonal antibodies, which could affect the effectiveness of vaccines. As a result, there is concern that the landscape of COVID-19 variants may be changing, and vaccine manufacturers are exploring the possibility of producing vaccines that are more effective against these new variants.

The N501Y mutation is located in the receptor-binding domain (RBD) of the spike protein, which is responsible for binding to the ACE2 receptor on human cells. This mutation changes an amino acid on the spike protein, allowing the virus to bind more tightly to its receptor. Other mutations, such as E484K, D614G, and P681H, have also been identified in SARS-B117. These mutations may have a synergistic effect, allowing the variant to spread more easily.

The emergence of these variants raises questions about the future of the COVID-19 pandemic. While vaccines have been effective in preventing severe illness and death, they may not be as effective against the new variants. This could lead to a resurgence of the virus, particularly in regions with low vaccination rates. As a result, there is a need for continued monitoring of the virus and rapid response to new variants.

In conclusion, the COVID-19 pandemic is far from over, and new variants of the virus continue to emerge. The N501Y mutation in the spike protein is a key factor in the increased transmissibility of SARS-B117. Other mutations, such as E484K, D614G, and P681H, may also contribute to the spread of the virus. As a result, there is a need for continued research into the properties of these new variants and the development of vaccines that are effective against them.
These differences made S protein from SARS-CoV-2 more efficient than that one from SARS-CoV and MERS-CoV RBD. Here we discuss some mutations that change the RBD domain, which is noticeably different from SARS-CoV and MERS-CoV RBD. These differences made S protein from SARS-CoV-2 more efficient to bind host receptors than SARS-CoV and MERS-CoV. Mutations also provided an additional cleavage in S1/S2 exclusively from SARS-CoV-2 improving membrane fusion. The unique glycosylation pattern is another advantage accumulated by SARS-CoV-2 S protein playing essential roles in viral tropism, pathogenesis, transmissibility, and evasion of the immune system. Therefore, S protein from pandemic coronaviruses is an essential factor in the show presented by them. Based on that, in-depth studies are required to investigate and understand more about S protein to develop potent drugs and vaccines to eliminate these threats.

CRediT authorship contribution statement

Conceptualization: Pedro F. N. Souza.

Data curation: Pedro F.N. Souza Felipe P. Mesquita, Jackson L. Amaral, Patricia G. C. Landim, Karolllyn R.P. Lima, Marília B. Costa, Izabelle R. Farias, Luina B. Lima.

Formal analysis: Pedro F.N. Souza, Jackson L. Amaral and Felipe P. Mesquita.

Funding acquisition: Pedro F.N. Souza, Felipe P. Mesquita and Raquel C. Montenegro.

Resources: Pedro F.N. Souza, Felipe P. Mesquita and Raquel C. Montenegro.

Supervision: Pedro F.N. Souza.

Writing – original draft: Pedro F.N. Souza, Felipe P. Mesquita, Jackson L. Amaral, Patricia G. C. Landim, Karolllyn R.P. Lima, Marília B. Costa, Izabelle R. Farias, Luina B. Lima.

Writing, review and editing: Pedro F.N. Souza, Felipe P. Mesquita and Raquel C. Montenegro.

Final approval of manuscript and submission: Pedro F.N. Souza.

Ethical approval

Not applicable.

Funding and acknowledgments

Grants from the following Brazilian agencies supported this work: The National Council for Scientific and Technological Development (CNPq), with a doctoral and master grant to JLA and FPM a research grant; the Office to Coordinate Improvement of University Personnel (CAPES) sponsored PFFS with a postdoctoral fellowship.

Declaration of competing interest

All authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbiomac.2021.02.203.
S. Leopardi, E.C. Holmes, M. Gastaldelli, L. Tassoni, P. Priori, D. Scaravelli, G. W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J.H. Epstein, H. Wang, G. Crameri, Z. Hu, H. X.Y. Ge, J.L. Li, X. Lou Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. S.K.P. Lau, P.C.Y. Woo, K.S.M. Li, Y. Huang, H.W. Tsoi, B.H.L. Wong, S.S.Y. Wong, S.Y. W. Wang, X.D. Lin, W.P. Guo, R.H. Zhou, M.R. Wang, C.Q. Wang, S. Ge, S.H. Mei, M.H. E.C. Teeling, M.S. Springer, O. Madsen, P. Bates, S.J. O

L.E. Gralinski, V.D. Menachery, Return of the coronavirus: 2019-nCoV, Viruses 12 (2020) 104362, https://doi.org/10.3201/eid1903.121503.

V.M. Corman, J. Jores, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M.A. Müller, V.M. Corman, J.B. Bosch, J. F. Drexler, et al., Antibodies against MERS coronavirus in dromedary camels from Dubai, mSphere 5 (2020)https://doi.org/10.1128/msphere.00898-19.

M.A. Müller, V.M. Corman, J. Meyer, B. Meyer, M. Younan, A. Liljander, B.J. Bosch, E. Lattwein, S. Kallies, A. Siemens, J. van Beek, et al., Antibodies against MERS corona-

M. Cotten, S.J. Watson, A.I. Zumla, H.Q. Makhdoom, A.L. Palser, S.H. Ong, A.A. Al Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotta, A.K. Dimitrov, M. Grada, G. Pirotin, et al., Viral persistence in the respiratory tract of COVID-19 patients, J. Infect. 82 (2020) e1–e9,https://doi.org/10.1016/j.jinf.2020.03.017.

A. Bonanni, H.J. Baldwin, V.M. Corman, S.M. Klose, M. Owusu, E.E. Nkrunmah, E.K. Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotten, S.J. Watson, A.I. Zumla, H.Q. Makhdoom, A.L. Palser, S.H. Ong, A.A. Al Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotta, A.K. Dimitrov, M. Grada, G. Pirotin, et al., Viral persistence in the respiratory tract of COVID-19 patients, J. Infect. 82 (2020) e1–e9,https://doi.org/10.1016/j.jinf.2020.03.017.

A. Bonanni, H.J. Baldwin, V.M. Corman, S.M. Klose, M. Owusu, E.E. Nkrunmah, E.K. Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotten, S.J. Watson, A.I. Zumla, H.Q. Makhdoom, A.L. Palser, S.H. Ong, A.A. Al Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotta, A.K. Dimitrov, M. Grada, G. Pirotin, et al., Viral persistence in the respiratory tract of COVID-19 patients, J. Infect. 82 (2020) e1–e9,https://doi.org/10.1016/j.jinf.2020.03.017.

A. Bonanni, H.J. Baldwin, V.M. Corman, S.M. Klose, M. Owusu, E.E. Nkrunmah, E.K. Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotta, A.K. Dimitrov, M. Grada, G. Pirotin, et al., Viral persistence in the respiratory tract of COVID-19 patients, J. Infect. 82 (2020) e1–e9,https://doi.org/10.1016/j.jinf.2020.03.017.

A. Bonanni, H.J. Baldwin, V.M. Corman, S.M. Klose, M. Owusu, E.E. Nkrunmah, E.K. Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotta, A.K. Dimitrov, M. Grada, G. Pirotin, et al., Viral persistence in the respiratory tract of COVID-19 patients, J. Infect. 82 (2020) e1–e9,https://doi.org/10.1016/j.jinf.2020.03.017.

A. Bonanni, H.J. Baldwin, V.M. Corman, S.M. Klose, M. Owusu, E.E. Nkrunmah, E.K. Badu, P. Anti, O. Aghenyeza, B. Meyer, et al., Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe, Emerg. Infect. Dis. 19 (2013) 233.

M. Cotta, A.K. Dimitrov, M. Grada, G. Pirotin, et al., Viral persistence in the respiratory tract of COVID-19 patients, J. Infect. 82 (2020) e1–e9,https://doi.org/10.1016/j.jinf.2020.03.017.
M. Alghamdi, F. Mushtaq, N. Awn, S. Shalhoub, MERS CoV infection in two renal transplant recipients: a descriptive study, Lancet Infect. Dis. 13 (2013) 752–761, https://doi.org/10.1016/S1473-3099(13)70204-4.

A. Assiri, J.A. Al-Tawfiq, A.A. Al-Rabeah, F.A. Al-Rabiah, S.A. Al-Hajjar, A. Al-Barrak, H. Flemban, W.N. Al-Nassir, H.H. Balkhy, R.F. Al-Hakeem, et al., Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus from Saudi Arabia: a descriptive study, Lancet Infect. Dis. 13 (2013) 574–583, https://doi.org/10.1016/S1473-3099(13)70245-2.

C. Wang, W.P. Horby, G.F. Hayden, G.E. Fowlkes, A novel coronavirus outbreak of global health concern, Lancet 395 (2020) 470–473, https://doi.org/10.1016/S0140-6736(20)30154-9.

WHO Coronavirus Disease (COVID-19). Dashboard | WHO coronavirus disease (COVID-19) dashboard, Available online:https://covid19.who.int/.

M. Bassetti, A. Vena, D.R. Giaccober, The novel Chinese coronavirus (2019-nCoV) infections: challenges for fighting the storm, Eur. J. Clin. Investig. (2020) 50.

A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. ZHU, Z. JIANG, J. WANG, W. FUXIANG, A. SHENQIANG, M. KAIJUN, et al., First report of 2019 novel coronavirus infection in China, J. Virol. 88 (2014) 723–732, https://doi.org/10.1128/JVI.03764-13.

F. Li, Receptor recognition mechanisms of coronaviruses: a decade of structural studies, J. Virol. 89 (2015) 1954–1964, https://doi.org/10.1128/JVI.02615-14.

K.M. Peck, T. Scoley, J. Swanstrom, K.L. Jensen, C.L. Burch, R.S. Baric, M.T. Heise, Permissivity of dipetidyl peptidase 4 expression is required for Middle East respiratory syndrome coronavirus infection, J. Virol. 91 (2017) 20170345. 10.1128/JVI.02051-16.

K.M. Peck, A.S. Cockrell, B.L. Yount, T. Scoley, R.S. Baric, M.T. Heise, Glycosylation of mouse DPP4 helps innate immune response to inhibiting Middle East respiratory syndrome coronavirus infection, J. Virol. 89 (2015) 4696–4699, https://doi.org/10.1128/JVI.01445-14.

A.S. Cockrell, K.M. Peck, B.L. Yount, S.s. Agnihothram, T. Scobey, N.R. Curnes, R.S. Baric, M.T. Heise, Mouse Dipetidyl peptidase 4 is a non-functional receptor for Middle East respiratory syndrome coronavirus infection, J. Virol. 88 (2014) 5199–5205, https://doi.org/10.1128/JVI.03764-13.

A. Barlan, J. Zhao, M.K. Sarkar, K, l. P.B. McCray, S. Perlman, T. Gallagher, Receptor variation and susceptibility to Middle East respiratory syndrome coronavirus infection, J. Virol. 88 (2014) 4831–4836, https://doi.org/10.1128/JVI.01611-14.

Y.M. Arabi, A. Omairi, Y. Mhandouz, A. Haimed, A.A. Siddi, B. Aliardadi, S. Shalhoub, A. Almotairi, K. Al Khatib, A. Abdulmomen, et al., Critically ill patients with the middle eastern respiratory syndrome: a multicenter retrospective cohort study, Crit. Care Med. 45 (2017) 1683–1695, https://doi.org/10.1097/CCM.0000000000003262.

Y.M. Arabi, A.A. Arifil, H.H. Balkhy, N. Hajim, A.S. Aldawood, A. Ghabashi, H. Hawa, A. Allmeiß, A. Khaldi, B. Al Ray, Clinical course and outcomes of critically ill patients with middle east respiratory syndrome coronavirus infection, Ann. Intern. Med. 160 (2014), 389–397, https://doi.org/10.7326/m3-2486.

A.M. Al-Osaili, M.J. Al-wazzah, The history and epidemiology of Middle East respiratory syndrome coronavirus, Multidiscip. Respir. Med. 12 (2017).

N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, R. Huang, W.S. Li, et al., A novel coronavirus from patients with pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733, https://doi.org/10.1056/NEJMoa2001017.

E.I. Azhar, S. Lanini, G. Ippolito, A. Zumla, The middle east respiratory syndrome coronavirus – a cocktail of opportunities, Environ. Health Perspect., vol. 125, New York, Springer New York LLC 2017, pp. 49–60.

A. Zumla, J.W. Chan, E. Fan, S.S. Agnihothram, T. Scobey, A. Almotairi, K. Al Khatib, A. Abdulmomen, et al., Critically ill patients with middle east respiratory syndrome coronavirus infection, Crit. Care Med. 20 (2016).

K. Kupferschmidt, Camel vaccine offers hope to stop MERS: vaccinated animals.

[93] A. Zumla, J.F.W. Chan, E.I. Azhar, D.S.C. Hui, K.Y. Yuen, Coronaviruses-drug discovery, Lancet 388 (2016) 267–270, https://doi.org/10.1016/S0140-6736(16)30999-0.

[94] A. Zumla, J.F.W. Chan, E.I. Azhar, D.S.C. Hui, K.Y. Yuen, Coronaviruses-drug discovery, J. Virol. 89 (2015) 4696–4699, https://doi.org/10.1128/JVI.01445-14.

[95] K.Z. Alkhaldi, E.L. Almohammadi, B.M. Alraddadi, S.I. Gerber, et al., 2014 MERS-coronavirus infection, J. Virol. 88 (2014) 4953–4961, https://doi.org/10.1128/JVI.01756-13.

[96] A. Sieberg, B.J. Bosch, E. Lattwein, et al., Transmission of MERS-coronavirus in KSA: a monoclonal antibody and a potential vaccine, PLoS Med. 10 (2013) e1001346. 10.1371/journal.pmed.1001346.

[97] A.K. Eitel, D. Debatin, J. Berkefeld, et al., A novel coronaviridae, Cell Res. 23 (2013) 986–994, https://doi.org/10.1038/cr.2013.92.

[98] N. Wang, X. Shi, L. Jiang, Z. Shang, D. Wang, P. Tong, D. Guo, L. Fu, Y. Cui, X. Liu, et al., Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4, Cell Res. 23 (2013) 586–593, https://doi.org/10.1038/cr.2013.52.

[99] H.M. Al-Dosly, S. Saleh, Y.M. Arabi, Critically ill patients with Middle East respiratory syndrome coronavirus infection, Crit. Care Med. 20 (2016).

[100] I.K. Oboho, S.M. Tomczyk, A.M. Al-Nasair, B. Banjar, H. Al-Mugti, M.S. Aloraini, K.Z. Alkhaldi, E.L. Almohammadi, B.M. Alraddadi, S.I. Gerber, et al., 2014 MERS-CoV outbreak in Jeddah – a link to health care facilities, N. Engl. J. Med. 372 (2015) 846–854, https://doi.org/10.1056/NEJMoa1408152.

[101] M. Senga, Y.M. Arabi, R.A. Fowler, Clinical spectrum of the Middle East respiratory syndrome coronavirus (MERS-CoV), J. Infect. Public Health 10 (2017) 191–194, https://doi.org/10.1016/j.jiph.2016.04.008.

[102] M. Alghamdi, F. Mushtaq, N. Awn, S. Shalhoub, MERS CoV infection in two renal transplant recipients: case report, Am. J. Transplant. 15 (2015) 1101–1104, https://doi.org/10.1111/ajt.13085.

[103] A.M. Huang, S. Li, N. Wang, S. Zhang, D. Wang, J. Liu, J. Wang, Y. Li, et al., A probable case of MERS-CoV infection due to the high level of transmission in a hospital, J. Infect. 71 (2015) 524–528, https://doi.org/10.1016/j.jinf.2015.07.027.
R.S. Baric, B. Yount, Subgenomic negative-strand RNA function during mouse hepatitis virus infection, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 5821–5826, https://doi.org/10.1073/pnas.0811739106.

J.L. Jacobs, A.T. Belew, R. Rakauskaite, J.D. Dinman, Identification of putative membrane protein interaction localizes to the carboxy-terminal domain of the coronavirus membrane protein, J. Virol. 75 (2001) 12228–12236, https://doi.org/10.1128/jvi.75.19.12228-12236.2001.

J.K. Locker, G. Grigoriev, E. Corse, C.E. Machamer, Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles, J. Virol. 74 (2000) 4967–4973, https://doi.org/10.1128/JVI.74.10.4967-4973.2000.

M.L. DeDio, E. Alvarez, F. Almazán, M.T. Rejas, E. Lamariande, A. Roberts, W.-J. He, S.R. Zaki, K. Subbarao, L. Enjuanes, A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo, J. Virol. 81 (2007) 8127–8134, https://doi.org/10.1128/JVI.01467-06.

L. Kuo, K.R. Hurst, P.S. Masters, Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function, J. Virol. 81 (2007) 2249–2262, https://doi.org/10.1128/JVI.01577-06.

E. Corse, C.E. Machamer, Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles, J. Virol. 74 (2000) 4967–4973, https://doi.org/10.1128/JVI.74.10.4967-4973.2000.

M. DeDio, E. Alvarez, F. Almazán, M.T. Rejas, E. Lamariande, A. Roberts, W.-J. He, S.R. Zaki, K. Subbarao, L. Enjuanes, A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo, J. Virol. 81 (2007) 8127–8134, https://doi.org/10.1128/JVI.01467-06.

L. Kuo, K.R. Hurst, P.S. Masters, Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function, J. Virol. 81 (2007) 2249–2262, https://doi.org/10.1128/JVI.01577-06.

E. Corse, C.E. Machamer, Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles, J. Virol. 74 (2000) 4967–4973, https://doi.org/10.1128/JVI.74.10.4967-4973.2000.

M. DeDio, E. Alvarez, F. Almazán, M.T. Rejas, E. Lamariande, A. Roberts, W.-J. He, S.R. Zaki, K. Subbarao, L. Enjuanes, A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo, J. Virol. 81 (2007) 8127–8134, https://doi.org/10.1128/JVI.01467-06.

L. Kuo, K.R. Hurst, P.S. Masters, Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function, J. Virol. 81 (2007) 2249–2262, https://doi.org/10.1128/JVI.01577-06.
associated with viral infectivity, Biochem. Biophys. Res. Commun. 359 (2007) 174–179, https://doi.org/10.1016/j.bbrc.2007.05.092.

[182] J. Wang, Y. Zhang, L. Wu, S. Song, Z. Zhang, C. Liu, C. Qiao, Y. Hu, K.-Y. Yuan, et al., Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 181 (2020) 894–904, e9, https://doi.org/10.1016/j.cell.2020.03.045.

[183] J. Fan, J. Ge, Y. Yuan, S. Han, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, et al., Structure of the spike of the SARS-CoV-2 spike receptor-binding domain. Nature 586 (2020) 215–220, https://doi.org/10.1038/s41586-020-2180-5.

[184] P.F.N. Souza, F.E.S. Lopes, J.L. Amaral, C.T. Frestedt, J.T.A. Oliveira, International journal of biological macromolecules a molecular docking study revealed that synthetic peptides induced conformational changes in the structure of SARS-CoV-2 spike glycoprotein, disrupting the interaction with human ACE2 receptor, Int. J. Biol. Macromol. 164 (2020) 66–76, https://doi.org/10.1016/j.ijbiomac.2020.07.174.

[185] P.F.N. Souza, J.L. Amaral, L.P. Bezerra, F.E.S. Lopes, V.N. Freire, J.T.A. Oliveira, C.T. Frestedt, ACE2-derived peptides interact with the RBD domain of SARS-CoV-2 spike glycoprotein, disrupting the interaction with the human ACE2 receptor, J. Biomol. Struct. Dyn. (2021) 1–14, https://doi.org/10.1080/07391102.2021.1871145.

[186] J.K. Millet, G.R. Whittaker, Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells, Virology 517 (2018) 3–4, https://doi.org/10.1016/j.virol.2017.12.015.

[187] Chambers, P., Pinrige, C.R., Easton, A.J. Hepatid repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. J. Gen. Virol. 1990, 71 (12), 3075–3081, doi:https://doi.org/10.1099/0022-1317-71-12-3075.

[188] S. Liu, J. Xiao, Y. Chen, Y. He, Y. Jiu, N. Cui, S. Eschlanite, X. Huang, J. Farman, K.A. Debnath, P. Tien, et al., Interaction between hepatid repeat 1 and 2 regions in spike glycoproteins of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of the most effective inhibitors. Lancet Infect. Dis. 363 (2004) 938–947, https://doi.org/10.1016/S0140-6736(04)61578-7.

[189] S.X. Li, Y. Yan, W. Xu, A.S. Agrawal, A. Algaissi, C.-T.K. Tseng, Q. Wang, L. Dou, W. Tan, W. Widagdo, N.M.A. Okba, W. Li, A. de Jong, R.L. de Swart, L. Begeman, J.M.A. van Weissenhorn, A. Dessen, L.J. Calder, S.C. Harrison, J.J. Skehel, D.C. Wiley, Structural basis for the recognition of transmembrane protease TMPRSS2, J. Virol. 84 (2020) 12568–12564, https://doi.org/10.1128/jvi.00661-20.

[190] P.M. Kamau, Z. Zhang, L. Yang, X. Zhao, X. Peng, R. Lai, Transferrin receptor is another receptor for SARS-CoV-2 entry, BioRxiv 2020 (2020) 10.35304/biorxiv.2020.23.2053347.

[191] R. Amraei, Y. Wim, M.A. Napoleon, E.L. Suter, J. Berrigan, Q. Zhao, J. Olejnik, K.B. Cheng, J.A. Hiscott, C. Xia, J. Helman, K.M. Hauser, T.T. Caradonna, A.C. Schmidt, M. Ubbelohde, C. Mühltal, A. Taguchi, Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2, J. Virol. 84 (2020) 12658–12664, https://doi.org/10.1128/jvi.01542-20.

[192] P.F.N. Souza, F.E.L. Carvalho, Killing two birds with one stone: how do Plant viruses break down plant defenses and manipulate cellular processes to replicate themselves? Plant Sci. 62 (2019) 170–180, https://doi.org/10.1016/j.plantsci.2019.11.010.

[193] Y. Fan, K. Zhao, Z.L. Shi, P. Zhou, Bat coronaviruses in China, Viruses 11 (2019).

[194] Y. Mao, N.C. Xu, X. Zhu, C.-C.D. Lee, S. Ho, L.V. C.K.P. Mok, J.A. Wilson, A highly conserved cryo-electron tomograms of the receptor binding domain of SARS-CoV-2 spike protein, J. Virol. 84 (2020) 630–633, https://doi.org/10.1128/jvi.00226-19.

[195] S. Matsuyma, N. Nagata, K. Shira, M. Kawaue, M. Takeda, F. Taguchi, Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2, J. Virol. 84 (2020) 12658–12664, https://doi.org/10.1128/jvi.01542-20.

[196] S. Duchene, L. Featherstone, M. Haritopoulou-Sinanidou, A. Rambaut, P. Lemey, G. Gascuel, The origin of SARS-CoV-2, Nature 581 (2020) 220, https://doi.org/10.1038/s41586-020-2180-5.

[197] E. Minskaia, T. Hertzig, A.E. Gorbalenya, V. Campanacci, C. Cambillau, B. Canard, J.A. Plante, Y. Liu, J. Liu, H. Xia, B.A. Johnson, K.G. Lokugamage, X. Zhang, A.E. Wimmer, et al., SARS-CoV-2 infection in human cells, Signal Transduct. Target. Ther. 5 (2020) 1–12, https://doi.org/10.1038/s41392-020-00426-x.

[198] R.N. Kirchdoerfer, C.A. Cottrell, N. Wang, J. Pallesen, H.M. Yassine, H.L. Turner, K.S. J.K. Millet, G.R. Whittaker, Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells, Virology 517 (2018) 3–4, https://doi.org/10.1016/j.virol.2017.12.015.

[199] Chambers, P., Pinrige, C.R., Easton, A.J. Hepatid repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. J. Gen. Virol. 1990, 71 (12), 3075–3081, doi:https://doi.org/10.1099/0022-1317-71-12-3075.