The Yale Undergraduate Research Journal

Volume 1
Issue 1 *Fall 2020*

Article 25

2020

A Comparison of the Evolution, Structure, and Function of SARS-CoV and SARS-CoV-2 Spike Proteins

Tai Michaels
Yale University

A Comparison of the Evolution, Structure, and Function of SARS-CoV and SARS-CoV-2 Spike Proteins

Cover Page Footnote

Thank you to Amelia Hallworth, Will Oles, Dr. Iain Dawson, and the YURJ team for their feedback and guidance.

STEM I Structural Biology VOL. 1.1 I Nov. 2020

A Comparison of the Evolution, Structure, and Function of SARS-CoV and SARS-CoV-2 Spike Proteins

Tai Michaels[1]

[1]*Yale University*

Abstract

As the COVID-19 pandemic has developed into the largest pandemic of the twenty-first century, it has become apparent that this disease, caused by the SARS-CoV-2 virus, is unlike anything the modern world has faced before. Not only has the disease infected more than 16 million people worldwide, but its rapid spread has drawn global attention to the gaps in our understanding of its pathogenesis and the development of vaccines and treatments. One of the most important topics of research in the disease is the viral spike (S) protein which facilitates binding and entering host cells and plays a key role in host specificity and pathogenicity among others. This review attempts to capture the importance of S protein in this new disease through evolutionary, structural, and functional lenses while drawing parallels to the recent SARS-CoV outbreak to identify what has made COVID-19 so different.

1. INTRODUCTION

The recent COVID-19 pandemic has exploded across the globe and slowed much of the world to a standstill as it forced the largest quarantines in history[1]. Although other coronaviruses (CoVs) are present in human and other mammal populations[2], the causative coronavirus pathogen, named SARS-CoV-2[3], is now the third novel coronavirus (nCoV) to attain epidemic proportions within the past two decades along with SARS-CoV (2002) and MERS-CoV (2012)[4]. Similar to past outbreaks, COVID-19 has been characterized by respiratory complications and has limited treatment options with vaccines still under development[5,6].

Unlike prior nCoV's, however, SARS-CoV-2 has infected multiple orders of magnitude more people and has a much lower case fatality rate. As of July 27, 2020, COVID-19 has caused over 16 million confirmed cases globally with over 600 thousand confirmed deaths[7] potentially due to a higher reproductive number (which is contested)[8,9] or a higher asymptomatic case proportion[10,11]. However, most research findings point towards a case fatality rate of 1-2%[12] which indicates that reported cases may be as little as 20% of the true number of infections[13,14]. This stands in stark contrast to SARS-CoV which had 8,500+ cases and a 6-10% case fatality rate[15]. SARS-CoV-2 evidently has a vastly different epidemiological character for unclear reasons, and understanding why could be critical for targeting response efforts and preparing for the future pandemics of the modern age. As the scientific community scrambles to understand this virus, it is important to recognize both its similarities with past outbreaks and what makes COVID-19 fundamentally different.

One crucial, conserved component of all CoV's is the spike (S) protein - a structural protein of the viral capsid. The S protein is a portion of the viral capsid which binds the host cell receptor and initiates the introduction of the viral contents into the cell. Within the S protein, one of the most important portions

is the receptor binding domain (RBD) which is the portion of the protein which initiates binding with the cell receptor[16]. In both SARS-CoV and SARS-CoV2, the RBD primarily binds angiotensin converting enzyme 2 (ACE2)[17,18]. In this literature review, I aim to capture the scope of what is known of the evolution and function of the RBD in CoV's with specific focus on SARS-CoV and SARS-CoV-2, their similarities and differences.

2. EVOLUTION

2.1 Classification

Coronaviruses are members of the family coronaviridae which is composed of enveloped +ssRNA viruses which infect mammals and birds and includes genuses alphacoronavirus, betacoronavirus, gammacoronavirus, and deltacoronavirus which are generally delineated by sequence homology[19]. Genus alphacoronavirus includes diverse bat coronaviruses as well as two notable human coronaviruses (HCoVs), HCoV-NL63 and HCoV-229E, which both cause relatively mild influenza-like symptoms in infected individuals. Genus betacoronavirus, which includes the three recent nCoVs and is the most thoroughly studied genus, has been subdivided into 4 subgenus lineages. Genuses gammacoronavirus and deltacoronavirus are not currently known to contain HCoVs and are primarily composed of avian coronaviruses[20]. The most recent common ancestor (MRCA) of coronaviruses was 10ka (kiloannum), and the MRCAs of the four genuses were between 4-5ka[21].

The four subgenuses of betacoronavirus have been termed embecoronavirus (subgroup A), sarbecoronavirus (subgroup B), merbecoronavirus (subgroup C), and nobecoronavirus (subgroup D). Embecoronavirus contains several bat CoVs along with HCoV-OC43 which (similar to HCoV-NL63 and HCoV-229E) causes mild influenza-like symptoms. Sarbecoronavirus includes the severe acute respiratory syndrome coronavirus (SARS-CoV), the more recent SARS-CoV-2 (causal agent of COVID-19), and a range of primarily bat CoVs termed severe acute respiratory syndrome related coronaviruses (SARSr-CoV). Merbecoronavirus contains the middle east respiratory syndrome coronavirus (MERS-CoV) along with related (primarily bat) coronaviruses. MERS-CoV has been excluded from the study due to its much lower genetic relatedness and distinct epidemiologic character[4]. Although it merits further study in light of ongoing cases and much higher case fatality rates[4], it is beyond the scope of this review. Nobecoronavirus is a lineage of mostly bat coronaviruses with no known HCoVs[19–21].

Of these lineages, the one of most interest to the study is subgenus sarbecoronavirus since it contains SARS-CoV-2 (the causal agent of COVID-19) along with SARS-CoV, a relatively closely related nCoV[22]. Genetically, have genomes of 27.9kb and 29.9kb respectively[13], both trace their origins to bat CoVs[2], and share primarily droplet and fomite[23] transmission (although aerosol transmission is debated[24,25]). The RBDs of SARS-CoV and SARS-CoV-2 are very distinct compared to other sarbecoronaviruses[26] but bear strong resemblance to one another[19]. The overall genetic similarity of SARS-CoV-2 to SARS-CoV is 79.5%[19] (with its most closely related identified relative at 96.2%[27]). Both viruses have generally similar S protein structures[19] and retain the asymmetric homotrimer with two RBDs "down" and one "up"[28]. However, their most recent common ancestor (MRCA) has been estimated to have occured 1,400 years ago[27], and their accumulated differences have clearly made them very different epidemiologically.

2.2 Lifecycle

CoVs generally follow a pattern of evolution and diversification in reservoir organisms - typically birds for gammacoronavirus and deltacoronavirus, and bats for alphacoronavirus and betacoronavirus. Rarely, CoV strains spill over into other mammal populations through contact with bats

or birds[21]. HCoVs tend to result from secondary transmission of viral strains from intermediate mammal hosts. For example, MERS-CoV and HCoV-229E are believed to have originated from bats with an intermediate host of dromedary camels (*Camelus dromedarius*)[29]. Many other HCoVs have been demonstrated to have been transmitted to humans from bats through various mammalian intermediate hosts although not all have been studied[20].

In the case of SARS-CoV, the virus most likely originated in bats and then was transmitted to palm civets (*Paradoxurus hermaphroditus*) and raccoon dogs (*Nyctereutes procyonoides*) which then transmitted the virus to humans[20]. The most closely related strain (WIV16) has 96% genetic similarity and 97% amino acid similarity in the S protein and was isolated from horseshoe bats further supporting the bat origin theory[30]. Broader study found a diverse group of SARSr-CoVs in these and other bats from which SARS-CoV and SARS-CoV-2 are believed to be descendents[2].

The origin of SARS-CoV-2 is still up to some debate. The most closely related CoV isolated so far is RaTG13 isolated from horseshoe bats in 2013 with 96.2% genetic similarity[19,31], but since early cases had no clear exposure to bats, the existence of an unknown intermediate host is likely[2,19]. One study of a related pangolin CoV found "conclusive" evidence that the virus was transmitted from bats to a pangolin reservoir population from which RaTG13 and SARS-CoV-2 are descended[32]. However, other studies noted the pangolin CoV is likely an outgroup rather than a direct ancestor of SARS-CoV-2 and that the pangolin CoV and SARS-CoV likely both diverged from a horseshoe bat CoV[2,27]. The most closely related bat coronaviruses sequenced were three times more closely related for SARS-CoV than for SARS-CoV-2[2], so with further sequencing of related SARSr-CoVs in bats and other organisms, a clearer picture may emerge in the coming. The most closely related bat coronaviruses sequenced were three times more closely related for SARS-CoV than for SARS-CoV-22, so with further sequencing of related SARSr-CoVs in bats and other organisms, a clearer picture may emerge in the future[2].

2.3 Evolution of the S Protein

More specific to the role of the S protein in evolution, it is believed that the ability of the S protein to bind ACE2 originates from the ability to bind ACE2 orthologs in bat species, but it is not a conserved trait in the common ancestors of SARS-CoV and SARS-CoV-2[18]. One other HCoV (HCoV-NL63) associated with more mild respiratory disease also binds ACE2 in humans[19], but is not closely related at all to SARSr-CoVs since it is in the alphacoronavirus genus[33] and has a distinct binding interaction[34]. In general, the S protein is highly mutable[35] and a large degree of the RBD similarities between SARS-CoV and SARS-CoV-2 are likely a product of convergent evolution.

In terms of proximal evolution, related viruses identified in pangolin populations share several key similarities to SARS-CoV-2 in the RBD (while RaTG13 does not), but do not possess a key furin cleavage site (discussed later) which may have derived from evolution in human populations from repeated introductions as in MERS[36]. Further evolution of the spike protein post transition to human hosts is evidenced by the increasing predominance of the G614 mutant form of the spike protein which has been correlated with higher case-fatality rates, viral loads, and potentially transmissibility[37,38].

3. STRUCTURE

CoV genomes all contain variants of four key structural proteins - nucleocapsid (N), matrix (M), envelope (E), and spike (S) proteins - in addition to highly variable numbers of non-structural proteins (nsp) often in overlapping open reading frames (ORFs)[39]. In SARS-CoV and SARS-CoV-2, the genome has two majors ORFs (ORF1a and ORF1b) which contain 15 nsp. The final third of the genome contains four structural

proteins[13,22,40] which are separated by accessory proteins which are not believed to be essential to viral function[41]. The gene encoding the S protein is located after ORF1ab and nonstructural protein 2 (ns2)[40] and is about 3800nt[13] and 1270 amino acids[22].

3.1 Spike Protein

Structurally, the S protein is expressed as a homotrimer spike on the viral surface with each monomer noncovalently linked[42]. This trimer form has also been described in related CoVs[43]. Each S protein monomer is made up of three domains. The extracellular domain (EC), the transmembrane anchor domain, and a small intracellular tail domain. Of these, the EC is of greatest interest due to its role in host cell binding and as a potential antibody target. The EC contains two domains: the S1 domain which is primarily involved with host cell binding and the S2 domain which is primarily involved in fusion with the host cell[44,45].

The S1 domain has been described as having four subdomains: the N-terminal domain (NTD) and three C-terminal domains numbered CTD1, CTD2, and CTD3. Of these, CTD1 (the closest to the NTD) is of greatest interest to this review since it contains the RBD[42]. CTD2 consists primarily of beta sheets extending out from CTD3 which is bound to S2[46]. As compared to the general genetic stability of coronaviruses and overarching similarities of major S protein structures, the RBD is one of the most mutable regions of the virus[39]. For example, just three point mutations were found to be key to the transition of SARS-CoV to humans, and the RBD had very little similarity to other sarbecoronaviruses[26,35].

Studies using cryo-EM have identified four distinct conformations of the trimer. One of these is a symmetric form of the trimer which is unable to bind to ACE2 and thus unable to fuse the host cell. The CTD1 domains in this conformation are said to be in a "down" state and are folded in towards the rest of the spike protein which blocks the RBD from being able to interact with ACE2[42,46]. The other three each have one of the RBD's in an "up" conformation in which the CTD1 inverts exposing the RBD allowing binding[28,42]. Each of these three "active" conformations binds ACE2 at a different angle to the S protein[46]. Another study identified only two binding conformations[47], and the resolution of binding states so far described is in the range of 5-10Å, so further research is likely necessary.

The S2 subunit is a class I viral fusion protein which facilitates fusion and viral entry to the cell[16,39]. The protein contains two heptad repeats (HR1 and HR2) which integrate into the membrane and facilitate endosome formation[42].

3.2 Host Receptor

ACE2, the RBD's binding target, is a homodimer with three domains. Each monomer has a single pass transmembrane domain, a collectrin-like domain (CLD) just outside the cell membrane, and a peptidase domain (PD) which extends further into the extracellular space. The RBD binds ACE2 on one limb of the binding pocket near the N-terminal domain[17]. It primarily interacts with one of the PDs alpha helices α1 and α2 as well as beta sheets β3 and β4[48].

One trait of SARS-CoV that has not yet been conclusively demonstrated in SARS-CoV-2 is the ability to utilize CD209L (also called DC-SIGN) as a host cell receptor instead of ACE2[49]. CD209L is a C-type lectin receptor present in many immune cells as well as alveolar cells and is targeted by other viruses including HIV, hepatitis C, ebola, and HCoV-229E[49,50]. One study has made a model of potential SARS-CoV-2 CD209L binding with a heavily glycosylated NTD, but it has yet to be shown experimentally[50]. While it is widely recognized that ACE2 is the primary receptor for both viruses, CD209L deserves further consideration in the face of the many unknowns surrounding COVID-19 infections.

3.3 Receptor Binding Domain

The RBD's primary role is binding the host cell through the receptor angiotensin converting enzyme 2 (ACE2)[16,17]. The RBD has a core complex of about 170 amino acids forming disulfide linked beta sheets which forms a projection from the rest of S1[26]. Within the RBD is a section called the receptor binding motif (RBM) of about 70 residues. The RBM is the only part of the S protein which directly interacts with ACE2 and is thus critical to binding and host specificity[26]. The RBM forms a concave structure of two antiparallel beta sheets linked on one end by a loop and on the other by strands connecting to the RBD core complex[26,42,46].

Despite macro-scale similarities, there are many key differences in the region of the S protein. According to one study, the genetic sequence of SARS-CoV-2 only indicates 6 amino acid substitutions in the RBD - P348A, E354N, V417K, K430N, T438S, N519H - out of 17 substitutions in S1 and 27 in all of S (out of 1273 residues)[22]. Also worth noting are significant changes in many accessory proteins including 3ab and 8a, but these are not the focus of this review[22].

Another study comparing an X-ray crystallography structure of SARS-CoV S[45] and a cryo-EM structure of SARS-CoV-2 S, the overall similarity was strong in the RBD with only a 0.64Å root mean square deviation in the core sequence of 120 alpha carbons[46]. In terms of the specific residues interacting with ACE2, however, there were many notable changes. Three substitutions in the region interacting with the α1 N-terminal region and one substitution towards the C-terminal end were noted. The middle portion contained five substitutions including a key $\mathbf{V^{404} \rightarrow K^{417}}$ substitution that may increase binding affinity to $\mathbf{D^{30}}$ on α1 by forming a new salt bridge. Another notable substitution is $\mathbf{R^{426} \rightarrow N^{439}}$ weakening a salt bridge to $\mathbf{D^{329}}$ on a helix near β4. The study also notes, however, that overall change in binding affinity could not be determined from the structure alone[46]. One study has found similar ACE2 binding affinities for the two viruses[48] while others have found that SARS-CoV-2 S protein has as much as 10 times greater affinity[47,61].

It is worth noting that the comparison of genetic sequences yielding 6 substitutions in the RBD surveyed a wide range of SARS-CoV and SARS-CoV-2 sequences to ensure that these were polymorphisms representative of true differences between the populations[22]. The structural comparison which noted 10 substitutions compared a single representative structure of each virus, thus the difference in approach may account for some of the differences in results of these and other analyses.

3.4 Proteolytic Activation

After binding ACE2, another critical function of the S protein is proteolytic activation. Cleavage at one or more sites in the S protein are necessary to activate S2 activity and initiate endosome fusion[34,42,51,52]. The primary cleavage region is between the S1 and S2 subunits allowing them to separate and S2 to activate, but there is increasing evidence for an additional cleavage site within S2. This site, called the S2' cleavage site, is located between the L segment of S2 and the heptad repeats and is believed to be similarly crucial to viral function[34,51]. Unlike related viruses such as MERS-CoV, the S protein is not cleaved during development and is instead cleaved by host cell proteases upon binding[53]. Additionally, unlike some viruses including influenza, the S1 and S2 domains are not linked by disulfide bonds and are therefore bound only by noncovalent bonds after S1/S2 cleavage[34]. While some studies on related strains of murine CoV have found that proteolytic cleavage is not necessary for viral fusion[54,55], it increases the rate of fusion by 2-3 orders of magnitude[55].

A variety of enzymes have diverse involvements with the proteolytic activation of the S protein. The primary enzyme involved in S protein cleavage is a type II transmembrane serine protease (TTSP) called transmembrane protease serine 2 (TMPRSS2) which cleaves the protein in multiple places

including between S1 and S2[45] and at S2'[51] leading to activation of S2 fusion activity[52,56–58]. TMPRSS2 is notably present in similar cell types to ACE2 including alveolar cells which SARS-CoV and SARS-CoV-2 notably infect[56,58]. This protease activity also produces fragments of the S protein which may bind antibodies helping evade humoral immunity[34,56]. A different type of serine protease, elastase, has been found to cleave only the S2' cleavage region, but it may be important in some severe cases since elastase is released by neutrophils in response to many infections. This would appear to indicate that the body's inflammatory response increases the proteolytic activation of the S protein[55,59].

Another enzyme known to cleave the S protein is cathepsin L, a lysosomal peptidase, which cleaves the protein between the S1 and S2 subunits thereby activating fusion activities[60,61]. However, unlike TMPRSS2, cathepsin L is an intracellular protein and can only affect S protein activity after endocytosis has begun[56,57]. Though it is sufficient for fusion[61], TMPRSS2 alone is more efficient than cathepsin L[62]. Another TTSP, human airway trypsin-like protease (HAT), is also able to cleave between S1 and S2 in both a cis and trans state. However, HAT cannot induce fusion activity in the absence of cathepsin L[52]. There are still significant gaps in our understanding of how the necessity of cleavage at the cell surface, in late endosomes, or lack thereof affects viral and epidemiologic characteristics.

Proteolytic cleavage is also one of the significant differences in the S protein between SARS-CoV-2 and SARS-CoV. While MERS-CoV and some other HCoVs such as HCoV-OC43 and HCoV-HKU1 have one or more furin cleavage sites, SARS-CoV has no significant furin protease activity[34], and furin cleavage sites in the S protein have not been identified in any of closely related SARSr-CoVs such as RaTG13 and the pangolin CoVs[36]. However, recent studies have predicted a furin cleavage site along with some associated O-linked glycans in SARS-CoV-2 in the S1/S2 cleavage region[44,47]. Comparison of fusion activity in the presence of TMPRSS2 and cathepsin L inhibitors found that SARS-CoV-2 maintained a low but significantly higher rate of fusion than SARS-CoV which was interpreted as a result of additional furin proteolytic activation[62] It has been shown in a porcine CoV that a single point mutation creating a furin cleavage site rendered trypsin proteases unnecessary for fusion[63], and thus, the furin cleavage site may serve to increase the ability of SARS-CoV-2 to fuse with host cell membranes. Crucially, furin is expressed in far more cell types than TMPRSS2 and may thus contribute to the virus' capability to infect a wide range of cells including intestinal and pancreatic cell[64].

4. SIGNIFICANCE

4.1 Epidemiology

It is still unclear how mutations in the SARS-CoV-2 RBD have affected its binding affinity for ACE2, but the general consensus is that SARS-CoV-2 has a greater binding affinity which may play a role in its pathology. Studies in related viruses have found that the patterns of S protein binding are correlated with pathogenicity[65] indicating that the changes in binding interaction observed between the viruses could be key to epidemiologic properties. A study of HCoV-NL63 has also suggested that lower observed binding affinity for ACE2 compared to SARS-CoV may have contributed to lower pathogenicity[33].

While there is both evidence for enhanced binding affinity and for the role of binding affinity in pathogenicity, the extent to which this may explain the dramatic epidemiological differences between the viruses is unclear. It is likely that the altered RBD increases infectivity and enhances transmission[66], but elucidating the magnitude of the effect from the wide range of factors has and will likely continue to prove extremely challenging.

Differential cleavage certainly is a significant difference between the viruses, but conclusive evidence for any epidemiological significance is still lacking. It has been predicted that the addition of a furin cleavage site in the SARS-CoV-2 genome will enable the virus to infect a greater range of human cell types due to the more widespread presence of furins and therefore increase the ease of transmission[44]. However, studies have shown that both viruses infect the same cell types when tested in vitro suggesting that this effect may not be present[62]. Furthermore, in a single-cell RNA-seq study of a wide range of cell types, ACE2 (not TMPRSS2 or other proteolytic activators) was found to be the limiting factor on capacity for viral infection[64].

On a final note, the importance of S protein to host transitions and treatments make it an important determinant of the likelihood of future nCoV outbreaks. In the case of SARS-CoV, the transition from palm civet to human hosts required only two amino acid substitutions in the RBD domain to yield a virus capable of infecting humans and with 3-4x the binding affinity it had in civets[35]. For SARS-CoV-2, just six substitutions in the RBD separate it from SARS-CoV, but it has already proved to be a much more devastating epidemic[22,67], and the more recent G614 mutation may have further increase its pathogenicity and transmissibility[37,38]. Given the relatively high mutability of this protein[39], it seems that further zoonosis or recurrence of past nCoVs seem likely within the foreseeable future.

"Given the relatively high mutability of this protein, it seems that further zoonosis or recurrence of past nCoVs seem likely within the foreseeable future.

4.2 Treatment

Since there are still no approved vaccines or treatments for SARS-CoV-2, it is crucial to understand the ways in which S1 and its interactions may be exploited for treatments to prevent viral binding and fusion. The S protein is widely regarded as an important target for antibody, protease inhibitor, and vaccine development[62,68].

Much of the ongoing research on treatments for SARS-CoV focuses on developing antibody therapies. One of the most promising targets for monoclonal antibodies is the S protein which has seen significant research, but these therapies are susceptible to small mutations in the crucial RBD yielding resistance[26,28]. On the other hand, multiple studies have demonstrated antibody cross neutralization of SARS-CoV and SARS-CoV-2 indicating that there may be some conserved regions[47,62,68].

Some studies have examined the effects of TMPRSS2 inhibitors due to the importance of it and other trypsin-like proteases to viral fusion. One study found strong fusion inhibition by using camostat mesylate (although the SARS-CoV-2 notably had more residual activity than SARS-CoV indicating residual fusion activity)[62]. Also notable is that studies in TMPRSS2 -/- mutant mice have found no increased fatality or other significant changes in phenotype implying that TMPRSS2 inhibitors may be safely used although confirmation for humans has yet to occur[69]. Other studies have confirmed that inhibitors of cathepsin L reduce SARS-CoV-2 fusion[61,62], but not nearly as much as TMPRSS2 inhibitors and thus are of less therapeutic importance[62]. Inhibitors of furins exist, but furins are widely expressed and play many critical functions in the body. However, it may be important to note that while proteases dramatically increase the rate of viral fusion, they are not strictly necessary for it[54,55], and while studies of protease inhibitors

have demonstrated efficacy in vitro, the effects of these are unknown in human systems as of yet[58].

Additionally, most current front-line treatments are more general, focusing on dampening the cytokine storm such as tocilizumab or acting as broad spectrum antivirals such as the nucleoside analog remdesivir. These treatments have proven useful, but it is quite possible that specific treatments such as S-protein targeted antibodies could prove more effective with less nontarget effects than broad spectrum treatments. Targeted treatments also have the ability to be administered along with broad spectrum or other targeted therapeutics since they can act constructively to achieve more complete effectiveness.

4.3 Vaccines

Thus far, many of the vaccine candidates for SARS-CoV with the most promise are whole virus or S protein isolates[68]. These are also the most common targets for SARS-CoV-2 vaccines by number of vaccines under investigation, and even the frontrunner nucleic acid vaccines primarily target mRNA and DNA sequences which encode the S protein[6].

Furthermore, there is strong evidence for conserved regions from high degrees of cross reactivity of antibodies between SARS-CoV and SARS-CoV-2[48,60,62] but not other related CoVs[62]. Since 70% of SARS-CoV antibodies target structural proteins including the prominent S protein[62], it seems probable that changes in SARS-CoV-2 S have not directly yielded immune evasion (although indirect evasion including additional cleavage is under investigation[32,54]). Therefore, it may be possible to create vaccines that induce vigorous immune responses to conserved regions of the S protein to confer lasting immunity against COVID-19 and future related pandemics.

However, it is also worth noting that the S1 domain, while a valuable target for treatments and vaccines, is also among the most mutable regions of the virus and thus may be more susceptible to the development of resistance[39]. Although clearly important, S protein is not the only factor in pathogenesis. Recent studies have implicated nsp2 and nsp3 as key to pathogenicity[70] and potentially nsp1 in related CoVs[71]. Interestingly, currently identified broad spectrum antiviral treatments which have demonstrated efficacy in vitro including remdesivir and chloroquine are not believed to affect the S protein or its function[72].

5. DISCLAIMER

COVID-19 research is a rapidly changing landscape, and while this review aims to be as up to date as possible at the time of its creation, some information may grow outdated as research progresses.

6. ACKNOWLEDGEMENTS

Thank you to Amelia Hallworth, Will Oles, Dr. Iain Dawson, and the YURJ team for their feedback and guidance.

7. WORKS CITED

1. Begley, S. Wuhan quarantine bought the world time to prepare for Covid-19. *STAT* https://www.statnews.com/2020/02/21/coronavirus-wuhan-quarantine-bought-world-time-to-prepare/ (2020).

2. Zhang, L., Shen, F., Chen, F. & Lin, Z. Origin and Evolution of the 2019 Novel Coronavirus. *Clin. Infect. Dis.* doi:10.1093/cid/ciaa112.

3. Gorbalenya, A. E. *et al.* The species Severe acute respiratory syndrome-related coronavirus : classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).

4. Eastern Mediterranean Regional Office. *MERS Situation Update November 2019*. 1 (2019).

5. CDC. Information for Clinicians on Investigational Therapeutics for Patients with COVID-19. *Centers for Disease*

Control and Prevention https://www.cdc.gov/coronavirus/2019-ncov/hcp/therapeutic-options.html (2020).

6. Thanh Le, T. *et al.* The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.* **19**, 305–306 (2020).

7. World Health Organization. *COVID-19 Situation Report*. 1–18 https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200727-covid-19-sitrep-189.pdf?sfvrsn=b93a6913_2 (2020).

8. Choi, S. & Ki, M. Estimating the reproductive number and the outbreak size of COVID-19 in Korea. *Epidemiol. Health* **42**, (2020).

9. Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, (2020).

10. Nishiura, H. *et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)*. http://medrxiv.org/lookup/doi/10.1101/2020.02.03.20020248 (2020) doi:10.1101/2020.02.03.20020248.

11. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **25**, (2020).

12. Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).

13. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

14. Pedersen, M. G. & Meneghini, M. Quantifying undetected COVID-19 cases and effects of containment measures in Italy: Predicting phase 2 dynamics. (2020) doi:10.13140/RG.2.2.11753.85600.

15. Update: Severe Acute Respiratory Syndrome --- Worldwide and United States, 2003. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5228a4.htm.

16. Bosch, B. J., Zee, R. van der, Haan, C. A. M. de & Rottier, P. J. M. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *J. Virol.* **77**, 8801–8811 (2003).

17. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).

18. Demogines, A., Farzan, M. & Sawyer, S. L. Evidence for ACE2-Utilizing Coronaviruses (CoVs) Related to Severe Acute Respiratory Syndrome CoV in Bats. *J. Virol.* **86**, 6350–6353 (2012).

19. Guo, Y.-R. *et al.* The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Mil. Med. Res.* **7**, 11 (2020).

20. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).

21. Woo, P. C. Y. *et al.* Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. *J. Virol.* **86**, 3995–4008 (2012).

22. Wu, A. *et al.* Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **27**, 325–328 (2020).

23. Rabenau, H. F. *et al.* Stability and inactivation of SARS coronavirus. *Med. Microbiol. Immunol. (Berl.)* **194**, 1–6 (2005).

24. Wang, J. & Du, G. COVID-19 may transmit through aerosol. *Ir. J. Med. Sci. 1971* - (2020) doi:10.1007/s11845-020-02218-2.

25. Asadi, S., Bouvier, N., Wexler, A. S. & Ristenpart, W. D. The coronavirus pandemic and aerosols: Does COVID-19 transmit via expiratory particles? *Aerosol Sci. Technol.* **54**, 635–638 (2020).

26. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **309**, 1864–1868 (2005).

27. Boni, M. F. *et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic*. http://biorxiv.org/lookup/doi/10.1101/2020.03.30.015008 (2020) doi:10.1101/2020.03.30.015008.

28. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).

29. Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses. in *Advances in Virus Research* (eds. Kielian, M., Mettenleiter, T. C. & Roossinck, M. J.) vol. 100 163–188 (Academic Press, 2018).

30. Yang, X.-L. *et al.* Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **90**, 3253–3256 (2016).

31. Li, X., Song, Y., Wong, G. & Cui, J. Bat origin of a new human coronavirus: there and back again. *Sci. China Life Sci.* **63**, 461–462 (2020).

32. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr. Biol.* **30**, 1346-1351.e2 (2020).

33. Abdul-Rasool, S. & Fielding, B. C. Understanding Human Coronavirus HCoV-NL63. *Open Virol. J.* **4**, 76–84 (2010).

34. Millet, J. K. & Whittaker, G. R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120–134 (2015).

35. Song, H.-D. *et al.* Cross-Host Evolution of Severe Acute Respiratory Syndrome Coronavirus in Palm Civet and Human. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2430–2435 (2005).

36. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).

37. Becerra-Flores, M. & Cardozo, T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int. J. Clin. Pract.* **74**, e13525 (2020).

38. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* (2020) doi:10.1016/j.cell.2020.06.043.

39. Tortorici, M. A. & Veesler, D. Structural insights into coronavirus entry. in *Advances in Virus Research* vol. 105 93–116 (Elsevier, 2019).

40. Sawicki, S. G., Sawicki, D. L. & Siddell, S. G. A Contemporary View of Coronavirus Transcription. *J. Virol.* **81**, 20–29 (2007).

41. Schwarz, B., Routledge, E. & Siddell, S. G. Murine coronavirus nonstructural protein ns2 is not essential for virus replication in transformed cells. *J. Virol.* **64**, 4784–4791 (1990).

42. Gui, M. *et al.* Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **27**, 119–129 (2017).

43. Kirchdoerfer, R. N. *et al.* Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118–121 (2016).

44. Wang, Q. *et al.* A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral

Transmissibility. *Virol. Sin.* (2020) doi:10.1007/s12250-020-00212-7.

45. Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643 (2005).

46. Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLOS Pathog.* **14**, e1007236 (2018).

47. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).

48. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).

49. Jeffers, S. A. *et al.* CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci.* **101**, 15748–15753 (2004).

50. Uslupehlivan, M. & Şener, E. Computational analysis of SARS-CoV-2 S1 protein O-glycosylation and phosphorylation modifications and identifying potential target positions against CD209L-mannose interaction to inhibit initial binding of the virus. *bioRxiv* 2020.03.25.007898 (2020) doi:10.1101/2020.03.25.007898.

51. Whittaker, G. R., Chu, V. C. & Belouzard, S. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *PNAS* **106**, 5871–5876 (2009).

52. Bertram, S. *et al.* Cleavage and Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by Human Airway Trypsin-Like Protease. *J. Virol.* **85**, 13363–13372 (2011).

53. Song, H. C. *et al.* Synthesis and Characterization of a Native, Oligomeric Form of Recombinant Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein. *J. Virol.* **78**, 10328–10335 (2004).

54. Taguchi, F. Fusion formation by the uncleaved spike protein of murine coronavirus JHMV variant cl-2. *J. Virol.* **67**, 1195–1202 (1993).

55. Matsuyama, S., Ujike, M., Morikawa, S., Tashiro, M. & Taguchi, F. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc. Natl. Acad. Sci.* **102**, 12543–12547 (2005).

56. Glowacka, I. *et al.* Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol.* **85**, 4122–4134 (2011).

57. Iwata-Yoshikawa, N. *et al.* TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J. Virol.* **93**, (2019).

58. Wu, C. & Zheng, M. Single-cell RNA expression profiling shows that ACE2, the putative receptor of COVID-2019, has significant expression in nasal and mouth tissue, and is co-expressed with TMPRSS2 and not co-expressed with SLC6A19 in the tissues. *Prepr. Httpswww Res. Comarticlers-16992v1 Accessed* **29**, (2020).

59. Belouzard, S., Madu, I. & Whittaker, G. R. Elastase-mediated Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein at Discrete Sites within the S2 Domain. *J. Biol. Chem.* **285**, 22758–22763 (2010).

60. Bosch, B. J., Bartelink, W. & Rottier, P. J. M. Cathepsin L Functionally Cleaves the Severe Acute Respiratory Syndrome Coronavirus Class I Fusion Protein Upstream of Rather than Adjacent to the Fusion Peptide. *J. Virol.* **82**, 8887–8890 (2008).

61. Simmons, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl. Acad. Sci.* **102**, 11876–11881 (2005).

62. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically

Proven Protease Inhibitor. *Cell* (2020) doi:10.1016/j.cell.2020.02.052.

63. Li, W. *et al.* A Single Point Mutation Creating a Furin Cleavage Site in the Spike Protein Renders Porcine Epidemic Diarrhea Coronavirus Trypsin Independent for Cell Entry and Fusion. *J. Virol.* **89**, 8077–8081 (2015).

64. Yang, L. *et al.* A Human Pluripotent Stem Cell-based Platform to Study SARS-CoV-2 Tropism and Model Virus Infection in Human Cells and Organoids. *Cell Stem Cell* **27**, 125-136.e7 (2020).

65. Wickramasinghe, I. N. A., Vries, R. P. de, Gröne, A., Haan, C. A. M. de & Verheije, M. H. Binding of Avian Coronavirus Spike Proteins to Host Factors Reflects Virus Tropism and Pathogenicity. *J. Virol.* **85**, 8903–8912 (2011).

66. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).

67. *Coronavirus disease 2019 (COVID-19) Situation Report - 85*. 1–12 (2020).

68. Prompetchara, E., Ketloy, C. & Palaga, T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pac. J. Allergy Immunol.* (2020) doi:10.12932/AP-200220-0772.

69. Kim, T. S., Heinlein, C., Hackman, R. C. & Nelson, P. S. Phenotypic analysis of mice lacking the Tmprss2-encoded protease. *Mol. Cell. Biol.* **26**, 965–975 (2006).

70. Angeletti, S. *et al.* COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J. Med. Virol.* **92**, 584–588 (2020).

71. Züst, R. *et al.* Coronavirus Non-Structural Protein 1 Is a Major Pathogenicity Factor: Implications for the Rational Design of Coronavirus Vaccines. *PLoS Pathog.* **3**, (2007).

72. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* **30**, 269–271 (2020).