

## REVIEW

# Molecular biology of the SARs-CoV-2 spike protein: A review of current knowledge

Chaogeng Zhu<sup>1</sup>  | Guiyun He<sup>2</sup> | Qinqin Yin<sup>2</sup> | Lin Zeng<sup>3</sup> | Xiangli Ye<sup>4</sup>  |  
Yongzhong Shi<sup>3</sup>  | Wei Xu<sup>3</sup> 

<sup>1</sup>Translational Medicine Laboratory of Pancreatic Diseases, Department of Hepatobiliary Surgery, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha, China

<sup>2</sup>Department of Ophthalmology, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha, China

<sup>3</sup>Institute of Translational Medicine, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha, China

<sup>4</sup>Department of Medical Laboratory Science, School of Medicine, Hunan Normal University, Changsha, China

## Correspondence

Wei Xu, Translational Medicine Research Institute, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha 410005, China.  
Email: 2982263958@hnnu.edu.cn

## Funding information

The Covid-19 Special Funding projects of Changsha Science and Technology Bureau, Grant/Award Number: (Grant No. kq2001020); Hunan Provincial Natural Science Foundation of China, Grant/Award Number: (Grant No. 2019JJ80005); The Scientific Research Foundation of Hunan Provincial Health and Family Planning Committee, Grant/Award Number: (Grant No. 20200477); The Scientific Research Fund of Hunan Provincial Education Department, Grant/Award Number: (No. 18A028)

## Abstract

The global coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to an unprecedented worldwide public health emergency. Despite the concerted efforts of the scientific field, by April 25, 2021, SARS-CoV-2 had spread to over 192 countries/regions, causing more than 146 million confirmed cases including 31 million deaths. For now, an established treatment for patients with COVID-19 remains unavailable. The key to tackling this pandemic is to understand the mechanisms underlying its infectivity and pathogenicity. As a predominant focus, the coronavirus spike (S) protein is the key determinant of host range, infectivity, and pathogenesis. Thereby comprehensive understanding of the sophisticated structure of SARS-CoV-2 S protein may provide insights into possible intervention strategies to fight this ongoing global pandemic. Herein, we summarize the current knowledge of the molecular structural and functional features of SARS-CoV-2 S protein as well as recent updates on the cell entry mechanism of the SARS-CoV-2, paving the way for exploring more structure-guided strategies against SARS-CoV-2.

## KEYWORDS

angiotensin-converting enzyme 2 (ACE2), coronavirus, COVID-19, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), spike Protein

## 1 | INTRODUCTION

The ongoing coronavirus disease 2019 (COVID-2019) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to an unprecedented worldwide public health emergency. Despite the concerted efforts of the scientific field, as of

April 25, 2021, more than 146 million confirmed cases including 31 million death were reported from around 192 countries/regions all over the world (<https://coronavirus.jhu.edu/map.html>). For now, established treatment for patients with COVID-19 remains unavailable. Thereby scientific insights and in-depth understanding of the biology and pathogenesis of the virus are imperative for deciphering

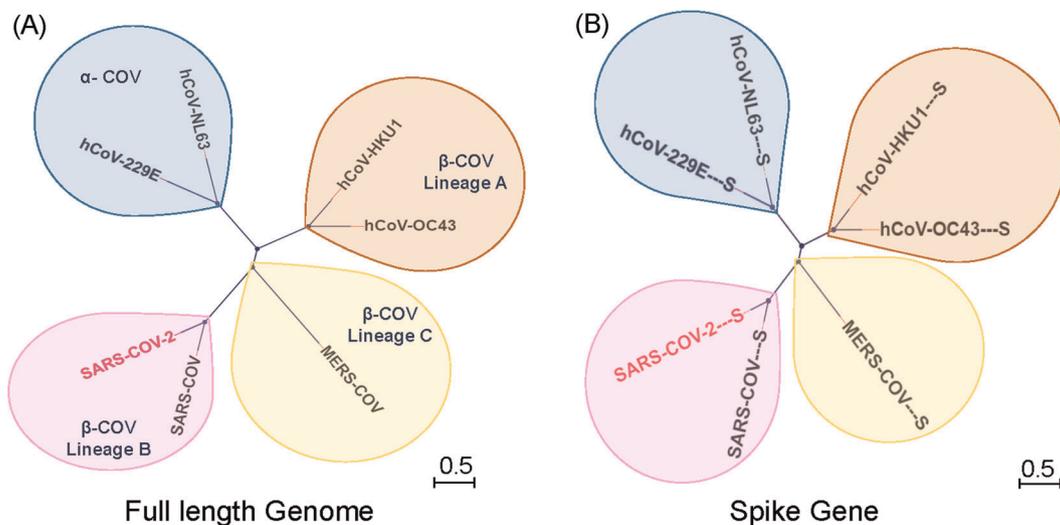
its mystery and curbing its spread. As a predominant focus, coronavirus spike (S) protein plays an essential role in coronavirus infection, pathogenicity, transmission, and evolution.<sup>1</sup> Changes in just a few residues around the spike protein can dramatically affect the infectivity, tropism, and pathogenesis of the virus.<sup>2-4</sup> Furthermore, potential adaptive mutations in the SARS-CoV-2 genome possibly make it highly pathogenic and difficult for drug or vaccine development. In this review, we summarize the current knowledge of the molecular structural and functional features of SARS-CoV-2 S protein as well as recent updates on the cell entry mechanism of the SARS-CoV-2, providing insights into virus pathogenesis, vaccine design, and drug target.

Coronaviruses (CoVs), currently the largest known genome size for RNA virus, are enveloped viruses with positive single-stranded RNA genomes ranging from 26 to 32 kb in length.<sup>5</sup> The coronaviruses are broadly classified into four genera as alpha, beta, gamma, and delta,<sup>6</sup>  $\alpha$ -CoVs and  $\beta$ -CoVs primarily infecting mammals,<sup>7,8</sup> whereas  $\gamma$ -CoVs and  $\delta$ -CoVs predominantly infect birds.<sup>9,10</sup> To date, seven coronaviruses are known to infect humans (hCoVs) including HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV, and SARS-CoV-2 have so far been identified (Table S1).<sup>11-13</sup> The former four hCoVs were known to cause just mild clinical symptoms.<sup>14</sup> However, the latter three hCoVs are highly pathogenic, causing relatively severe respiratory disease and high mortality.<sup>15,16</sup> SARS-CoV-2 exhibits 79.9% sequence identity to SARS-CoV at the whole genome level and clusters with SARS-CoV in phylogenetic analysis (Figure 1A). Further comparison and analysis of the spike (S) gene of the seven hCoVs revealed that the genetic divergence of the S genes is consistent with that of the full-length genomes (Figure 1B, Table S2). SARS-CoV-2 belongs to Beta-CoVs B

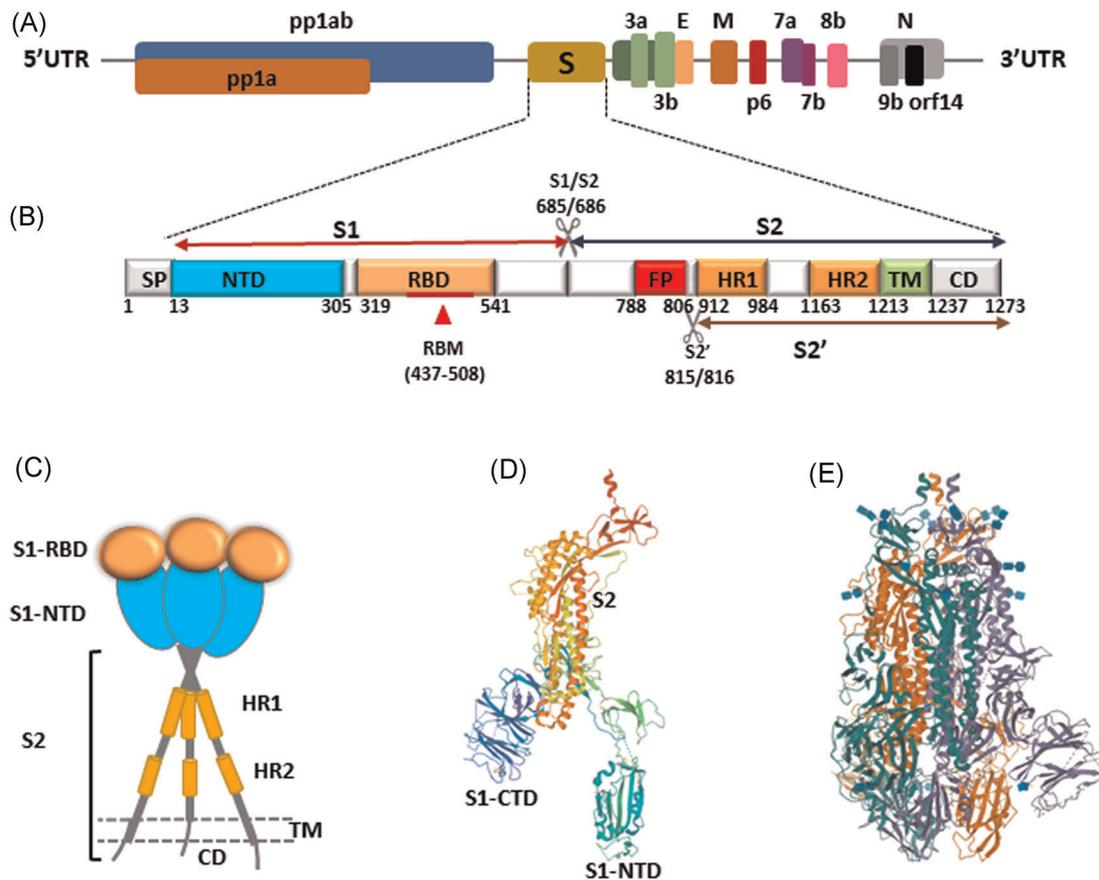
lineage closely related to SARS-CoV and SARS-related bat CoVs.<sup>17,18</sup> Although conclusive evidence is lacking, distinctive phylogenetic distances on the major clade of SARS-CoV-2 provided a clue to the evolutionary relationships among them.<sup>17</sup> Sequence analysis at full-length genome level showed that SARS-CoV-2 is much closer to the bat CoV RaTG13 (96.2% identical) than to SARS-CoV (79.5% identical). In addition, although Zhou et al.<sup>19</sup> reported a bat-derived coronavirus named RmYN02 from *Rhinolophus malayanus* sharing 93.3% identity with SARS-CoV-2 in the complete genome, RmYN02 has a low sequence identity (61.3%) with SARS-CoV-2 in the Receptor Binding Domain (RBD). However, the RaTG13 was found to have an 89% RBD identical to SARS-CoV-2, indicating a bat origin.

## 2 | MOLECULAR CHARACTERISTICS OF THE SARS-COV-2 SPIKE(S) PROTEIN

The genome of SARS-CoV-2 is approximately 29.7 kb in size, encoding four major structural proteins, including spike (S), nucleocapsid (N), membrane (M), and envelope (E), and 16 nonstructural proteins.<sup>20</sup> The full-length of SARS-CoV-2 S is 1273 amino acid long with several functional domains and multiple proteolytic cleavage sites at S1/S2 boundary and S2' site (Figure 2B). The spike protein (S) of SARS-CoV-2 is a trimeric class I fusion protein containing two subunits: S1 and S2, which remain noncovalently bound in a pre-fusion metastable conformation<sup>23,24</sup> protruding from the viral surface as a complete homotrimer and forming the distinctive surface spikes of coronaviruses.<sup>25</sup> The S protein of SARS-CoV-2 and SARS-CoV have about 74.0% identity at amino acid level. A breakdown of the functional domains of the SARS-CoV-2 S, based on



**FIGURE 1** Phylogenetic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and other hCoVs. (A) Full-length genome phylogenetic tree of 7 hCoVs; (B) phylogenetic tree of the S gene of 7 hCoVs. Nucleotide sequences of the full-length genome/S genes were aligned respectively using MEGA 7 and iTOL online software. Maximum likelihood trees were constructed by using 1000 bootstraps. Respective sequences were obtained from NCBI with indicated accession numbers (SARS-CoV-2, NC\_045512; SARS-CoV, NC\_004718.3, MERS-CoV, NC\_019843.3; hCoV-OC43, NC\_006213.1; hCoV-HKU1, NC\_006577.2; hCoV-NL63, NC\_005831.2; and hCoV-229E, NC\_002645.1). The phylogenetic trees of the full-length genome/Spike gene of SARS-Cov-2 were made using iTOL online software



**FIGURE 2** Structural features of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) protein (A) Schematic representation of the genomic organization of SARS-CoV-2; (B) Schematic representation of the SARS-CoV-2 S. The listed domain boundaries are mostly defined according to.<sup>21</sup> SP, Signal Peptide; NTD, N terminal Domain; RBD, Receptor Binding Domain; RBM, Receptor Binding Motif; RBD, Receptor Binding Domain; FP, Fusion Peptide; HR1, Heptat Repeat 1; HR2, Heptat Repeat 2; TM, Transmembrane Domain; CD, Cytoplasm Domain. (For visual clarity, the length of the boxes is not proportional to the real sequence length). The data (The listed domain boundaries of the SARS-CoV-2 S and SARS-COV S) were derived from the following resources available at (DOI: 10.1038/s41423-020-0374-2).<sup>21</sup> (C) Schematic drawing of the three-dimensional (3D) structure of SARS-CoV-2 S; (D) Single monomer from the trimeric SARS-CoV-2 S (PDB ID: 6VSB-1, Represented by Chain A).<sup>22</sup> (E) Trimeric SARS-CoV-2 spike protein (PDB ID: 6VSB),<sup>22</sup> three monomers are shown (magenta, green and cyan). These data that support the findings of this study are openly available in (PDB) at (<https://www1.rcsb.org/>)<sup>22</sup>

**TABLE 1** Functions of different subunits/domains of SARS-CoV-2 S protein

Subunit	Domain	Functions	Ref.
S1	S1	Mediate receptor recognition and viral attachment to initiate host cell entry	Walls et al. <sup>27</sup>
	NTD	Contribute to determine host range	Sironi et al. <sup>28</sup>
	RBD	Recognize and strongly bind to hACE2 receptors	Tai et al. <sup>29</sup> and Lan et al. <sup>30</sup>
	RBM	Bind and interact with hACE2	Lan et al. <sup>30</sup>
S2	S2	Mediate host-virus membrane fusion and host cell entry	Walls et al. <sup>27</sup>
	FP	Responsible for fusion of virus and target cell membrane	Gee and Freed <sup>31</sup>
	HR1/HR2	Form a 6-helix bundle (6-HB) to mediate membrane fusion between virus and target cell	Xia et al. <sup>21</sup>
	TM	Be crucial for spike protein trimerization and membrane fusion	Schroth-Diez et al. <sup>32</sup>
	CT	Contribute to anchor the trimer to the viral membrane and involve in cell-cell fusion	Petit et al. <sup>33</sup>

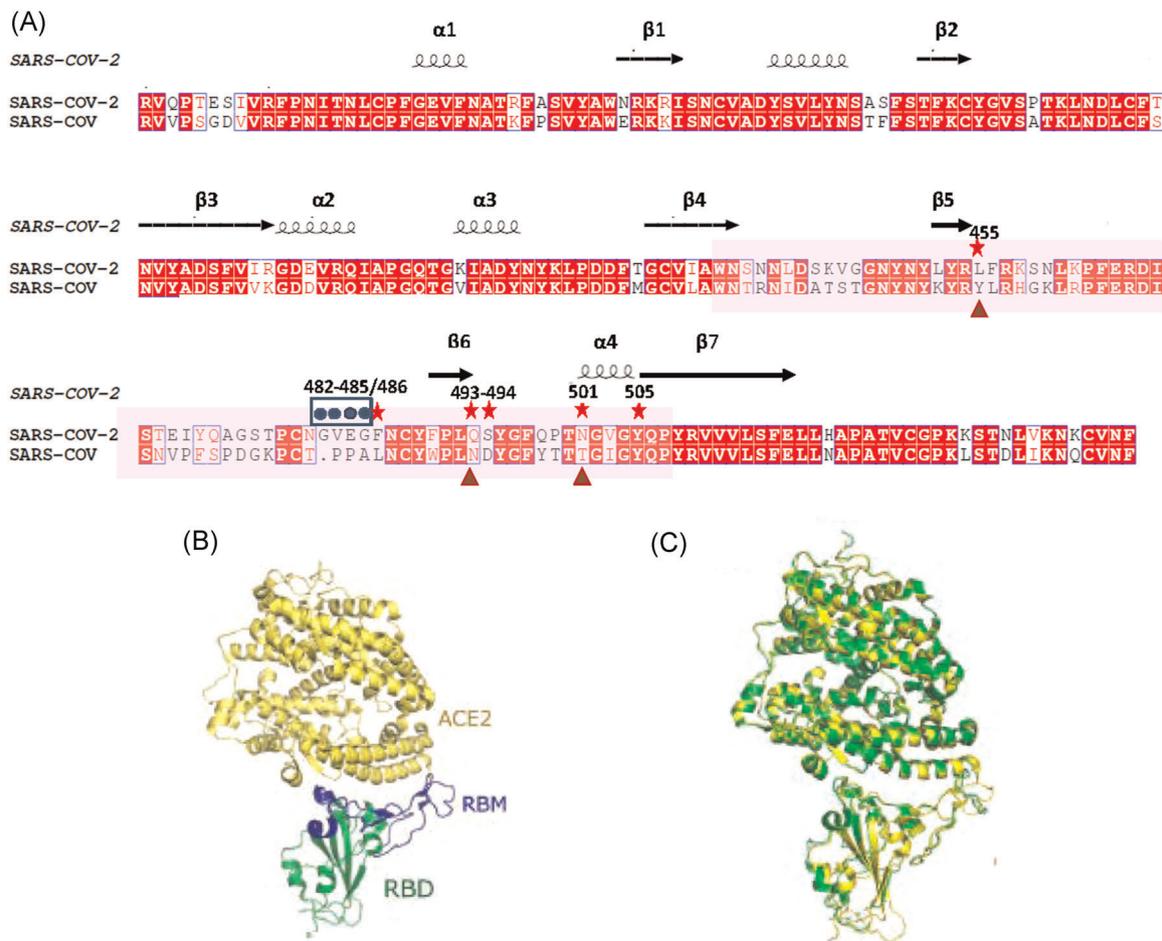
Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

SARS-CoV S sequence, reveals that the S1 subunit was quite variable (63.3% identity) (Figure S1), such observations are in line with previous reports that the S1 is the most variable region of the molecule, both across and within the four CoVs genera.<sup>26</sup> However the S2 subunit is remarkably much more conserved (90.0% identity). With such both varied and conserved domains, the spike protein plays a pivotal role in the biology and pathogenesis of SARS-CoV-2 corresponding to their sophisticated structures (Table 1).

## 2.1 | Structure and function of S1 subunit

Structurally, the S1 subunit is made up of two main domains: N-terminal domain (NTD) and C-terminal domain (CTD) both involving in RBD recognition (Figure 3). The NTDs show only about

48.9% homology to that of SARS-CoV (Figure S1), indicating its' the most diverse region in SARS-CoV-2. A recent study showed the major differences in the NTD between SARS-CoV-2 and SARS-CoV lie in three short insertions (GTNG) (GTNGTRR) (YLTPGD) in SARS-CoV-2.<sup>17</sup> Nevertheless, whether the insertions display sialic-acid-binding activity linked to glycoproteins as it does in MERS-CoV<sup>35</sup> needs to be further clarified. Among them, one insertion (YLTPGD) (248–253) is present only in SARS-CoV-2, RaTG13, and Guangxi pangolin CoVs, but is absent in other bat CoVs. This insertion formed a conformational cluster at the NTD of the spike trimer and may contribute to the host range.<sup>28</sup> Furthermore, a new type of ganglioside-binding domain (GBD), which was used as viruses receptor and/or attachment element for cell entry,<sup>35</sup> was identified at the tip of the NTD of the SARS-CoV-2 S by Fantini et al.<sup>36</sup> This GBD (111–112), enriched in aromatic and basic amino



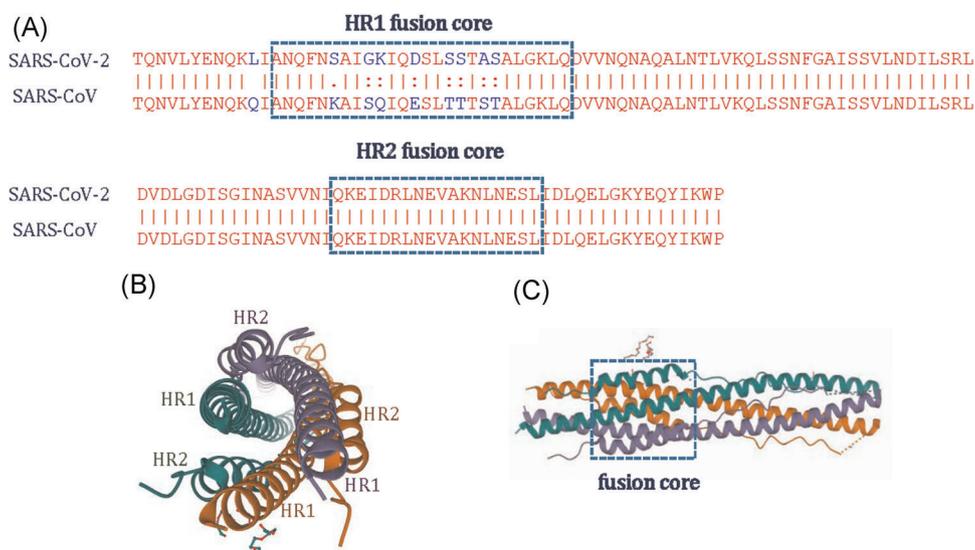
**FIGURE 3** Structure of the receptor-binding domain (RBD) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and SARS-CoV. (A) Sequence alignment of the RBDs between SARS-CoV and SARS-CoV-2. Alignment file produced by ClustalW, aligned and annotated using ESPript 3.0 online software and then adjusted for format by Photoshop. Arrows indicate the critical binding residues for interaction between RBD and receptor, stars indicate Five of these six amino acids differ between SARS-CoV-2 and SARS-CoV that are critical for hACE2 binding; dots indicate a four-residue motif that allows the ridge to become more compact and allow better contacts with hACE2; triangle marks virus-binding hotspots that lead to considerable stabilization of binding and higher affinity for ACE2. (B) Crystal structures of the SARS-CoV-2 RBD (core in green and RBM in blue) with the receptor ACE2 (in yellow) (PDB: 6LZG)<sup>34</sup>; (C) Structural similarity of the RBD-hACE2 complex between SARS-CoV-2 (green) (PDB: 6LZG) and SARS-CoV (yellow) (PDB: 2AJF).<sup>34</sup> These data that support the findings of this study are openly available in (PDB) at (<https://www.1.rcsb.org/>)<sup>34</sup>

acid residues, is longer than all linear GBD characterized to date. It's fully conserved among worldwide clinical isolates and may facilitate contact with ACE2 receptor.<sup>36</sup>

The RBD, as a key element of the CTD in SARS-CoV-2, consists of two motifs: a core structure and an extended loop, the latter also known as a receptor-binding motif (RBM). The core is a twisted five-stranded anti-parallel  $\beta$  sheet ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ ,  $\beta 4$ , and  $\beta 7$ ), with three short helices ( $\alpha 1$ ,  $\alpha 2$ , and  $\alpha 3$ ); The RBM, formed by a two-stranded  $\beta$  sheet ( $\beta 5$  and  $\beta 6$ ), lies at one edge of the core and contains most of the contacting residues binding to hACE2<sup>30</sup> (Figure 4A). Homology modeling between SARS-CoV-2 and SARS-CoV has shown that both the overall RBD structure model and the RBD-ACE2 binding model are strikingly similar<sup>5,30</sup> (Figures 4B,C). Even in the more variable RBM (45.7% identity), the overall structure is still highly similar.<sup>30</sup> In other words, most amino acid residues that are critical for hACE2 binding are highly conserved in SARS-CoV-2 (Figure 4A). RBD-ACE2 binding affinity has been shown to be a crucial element determining the infectivity of SARS-CoV.<sup>37</sup> Despite the nearly identical structure with a highly similar binding interface between SARS-CoV-2 (6LZG) and SARS-CoV (PDB: 2AJF), biophysical and structural evidence showed that the RBD-ACE2 binding affinity of SARS-CoV-2 (14.7 nM) was 10–20 fold higher than that of SARS-CoV (325.8 nM).<sup>22,38</sup> Few residues' mutations and structural changes in the RBD of SARS-CoV-2 compared with SARS-CoV may explain such subtle differences: (i) An ACE2-binding ridge in SARS-CoV-2 RBD has a much tighter conformation largely caused by a four-residue motif (residues 482–485: Gly-Val-Glu-Gly) (Figure 4A). This structural change allows the ridge to become more compact and allow better interactions with hACE2<sup>39</sup>; (ii) Previous studies have shown that two virus-binding hotspots (lysine Lys31 and Lys353) within human ACE2 are critical for SARS spike binding,<sup>40</sup> and regulate the infectivity,

pathogenesis, and cross-species transmissions of SARS-CoV.<sup>41,42</sup> The SARS-CoV-2 RBM has evolved to stabilize the two hotspots with Gln493 and Leu455 stabilizing hotspot Lys31, whereas Asn501 stabilizing hotspot Lys353 (Figure 4A), contributing to considerable stabilization and higher affinity for RBD-ACE2 binding.<sup>39</sup> (iii) Six key residues (L455, F486, Q493, S494, N501, and Y505) in RBD of SARS-CoV-2 S, which correspond to Y442, L472, N479, D480, T487, and Y491 in SARS-CoV S, have been shown to be critical for receptor binding as well as cross-species transmission of SARS-CoV-2.<sup>37</sup> Five out of these six key residues differ between them. (Figure 4A). Among them, L455, F486, Q493, and S494 in SARS-CoV-2 were revealed to provide favorable recognition and interactions with hACE2, hence enhancing viral binding, while N501 recognizes hACE2 more efficiently than that of SARS-CoV S, which increased SARS-CoV-2 infectivity.<sup>37</sup>

Paradoxically, despite the extraordinary potency of its RBD binding affinity, the overall binding ability of SARS-CoV-2 S is comparable or even lower than that of SARS-CoV S.<sup>43</sup> Cryo-electron microscopy at the atomic level revealed different conformations of the homo-trimeric SARS-CoV-2 S with both "open" (receptor-accessible) and "close" (receptor-accessible) states. In the "close" state all the three RBDs are tightly packed together, while in the "open" state, the SARS-CoV-2 S was conformed in an asymmetric conformation in which one RBD was in "up" state whereas the other two in "down" state,<sup>22,27</sup> or two RBDs in "up" while the other one in "down" state.<sup>44,45</sup> Cryo-EM structure studies revealed that the RBD in SARS-CoV S is mostly in "up" state,<sup>46,47</sup> while the RBD in SARS-CoV-2 S is mostly in "down" state.<sup>22,27</sup> Therefore, although SARS-CoV-2 RBD has stronger hACE2 binding affinity, it is less accessible, resulting in comparable or even lower overall binding ability compared to SARS-CoV.



**FIGURE 4** The fusion core structure of SARS-CoV-2 S. (A) The sequence alignment of HR1 and HR2 domains in SARS-CoV-2 and SARS-CoV, Alignment file produced by ClustalW and then adjusted for format by Photoshop. (B) Top view of the fusion core structure formed by the HR1 and HR2 in the S2 subunit of SARS-CoV-2 (PDB ID: 6LXT)<sup>52</sup>; (C) side view of the fusion core structure (PDB ID: 6LXT).<sup>52</sup> These data that support the findings of this study are openly available in (PDB) at (<https://www1.rcsb.org/>)<sup>52</sup>

## 2.2 | Structure and function of the S2 subunit

The S2 subunit of SARS-CoV-2, which plays a key role in virus-cell fusion and viral entry,<sup>27</sup> contains a fusion peptide (FP) (residues 943–982), a cleavage S2' site (residues 815/816), two heptad-repeat domains (HR1/HR2) (residues 984–1104/1246–1295), a transmembrane domain (TM) (residues 1296–1317) and a cytoplasm domain (CD) (residues 1318–1353)<sup>21</sup> (Figure 2B). Sequence comparison at amino acids level of the S2 subunit between SARS-CoV-2 S and SARS-CoV S confirmed high level of conservation (S2, 90%; FP, 78.9%; HR1, 87.7%; HR2 100%; TM, 91.7%, and CD, 97.2%) (Figure S1).

The FP (<sub>788</sub>IYKTPPIKDFGGFNFSQIL<sub>806</sub>), located at a site immediately upstream S2', is biochemically characterized by its hydrophobic nature owing to its abundance of hydrophobic residues. The potential lipids-binding residues in FP of SARS-CoV-2 is centralized amid in Lys790, Thr791, Lys795, Asp808, and Gln872 residues, contributing to lipid interaction and penetration.<sup>48</sup> Surface models predicted that the FP of SARS-CoV-2 S was organized in a more compact conformation than that of SARS-CoV S.<sup>49</sup> Cleavage at S2' exposes the FP domain, which, in turn, inserted in the host membrane triggering the viral fusion.<sup>22</sup> With strong membrane-perturbing capacities,<sup>50</sup> FP plays a crucial role in cell fusion.<sup>31</sup> Mutations in this region have been shown to block cell fusion for many viruses.<sup>51</sup>

HR1 and HR2, also known as the "fusion core region" of SARS-CoV-2<sup>21</sup> are separated by an intervening stretch of 180 amino acid residues. Both HR1 and HR2 domains in SARS-CoV-2 and SARS-CoV showed a high degree of homology, with 87.7% and 100% identity, respectively (Figure S3). However, in the fusion core of HR1, 8 of the 26 residues showed amino acid changed (~30.8% difference). These mutations were reported to enhance interaction with HR2 and imparted more stability to the six-helix bundle (6-HB) core<sup>52</sup> (Figure 4B). When the HR1 and HR2 domains are exposed to interact with each other, three long helices of HR1 in SARS-CoV-2 assemble into a coiled-coil trimer, whereas, three short helices of HR2 fold back to HR1 trimer forming an anti-parallel 6-HB fusion core (Figure 4C), which facilitates the insertion of the hydrophobic IFP into the host membrane and bridges the viral TM helix and host membranes at the proximity to promote fusion,<sup>53</sup> and ultimately result in the release of the viral RNA into the host cell. The 6-HB fusion core formed by HR1 and HR2 plays a key role in membrane fusion in SARS-CoV-2 making it one of the most important targets for vaccine and drug design.<sup>52</sup> Several studies have reported that HR1 and HR2 derived peptides can inhibit this fusion.<sup>21,52</sup>

With a stretch of 24 hydrophobic amino acid residues, the transmembrane domain (TM), which anchors the spike protein to the viral membrane, has been shown to be crucial for spike protein trimerization and membrane fusion.<sup>32</sup> It is hypothesized that the TM (anchored in the viral envelope) interacts with the FP (partitioned in the host membrane) to facilitate formation of the fusion pore.<sup>54</sup> The S2 subunit ends with a CD tail, which contains a palmitoylated cysteine-rich region (of 36 residues with 8 cysteines) and comprises the intracellular short tail part. Located at the C-terminal of the S

protein and the inner side of the cell membrane, CD is reported to be involved in viral assembly, intracellular transport, and cell-cell fusion.<sup>33</sup>

## 2.3 | Cleavage motifs and their role in priming and activating of SARS-CoV-2 spike protein

For SARS-CoV-2 to enter host cells, besides host cell receptor binding, priming and activation of the spike protein by host proteases is another crucial event modulating tropism and pathogenicity.<sup>55</sup> "Priming" occurs at S1/S2 boundary providing the SARS-CoV-2 S protein with the structural flexibility required for separating S1 and S2.<sup>56</sup> And a subsequent "activation" cleavage at S2' site generates the exposure of FP and its insertion into the membrane.<sup>57</sup> A variety of proteases including, but are not limited to, Furin, TMPSS2, and cathepsin B/L, have been shown to mediate SARS-CoV-2 for priming and activation,<sup>58–60</sup> indicating a relatively high degree of flexibility in cleavage mechanisms. Indeed, to maintain its high infectivity as well as immune surveillance evasion, SARS-CoV-2 relies on a second strategy: host protease priming and activation.

### 2.3.1 | Unique Furin-like motif at S1/S2 site of SARS-CoV-2 absent in CoVs of the same clade

Rather than the single arginine at S1/S2 observed in SARS-CoV, the SARS-CoV-2 S displays a unique feature with a polybasic motif insertion (P-R-R-A-R685↓), forming an exposed flexible loop that is easily available for protease cleavage at the S1/S2 boundary.<sup>49</sup> This Furin-like motif,<sup>61,62</sup> notably missing from SARS-CoV and other  $\beta$ -CoVs even in the closest related Bat-RaTG13 or the pangolin viruses,<sup>63</sup> but present in other human coronaviruses including hCoV-OC43, hCoV-HKU1, and MERS-CoV<sup>55,64</sup> (Figure 5). Furin recognizes the R-X-K/R-R motif and is known for its ability in the proteolytic activation of a broad range of viruses, and thus enhancing cell-to-cell fusion.<sup>65,66</sup> The presence of the polybasic Furin-like cleavage motif in SARS-CoV-2, as a hallmark of enhanced virulence,<sup>67</sup> allows effective cleavability by Furin and other proteases and thus increase the transmission efficiency in human as compared with other  $\beta$ -CoVs.<sup>63,68</sup> Cleavage of spike protein by Furin at S1/S2 site is essential for SARS-CoV-2 entry into human lung cells.<sup>59,60,63,63,69</sup> Structural and biochemical data indicated that the Furin-like cleavage motif in SARS-CoV-2 enhances the structural plasticity of the RBD and facilitates the adoption of an open conformation required for RBD-hACE2 binding.<sup>56</sup> A Furin cleavage motif insertion at the S1/S2 junction of SARS-CoV enhances spike-driven cell to cell fusion.<sup>66</sup> However blockade of the Furin cleavage motif in SARS-CoV-2 affected its entry into the TMPRSS2<sup>+</sup> human lung cell line Calu-3.<sup>60</sup> Collectively, the Furin-like motif provides SARS-CoV-2 with a gain-of-function for more efficient transmission in human, reducing its dependence on host cell proteases for entry and expanding its entry into multi-type of cells.<sup>43,63,69</sup>

	S1/S2	S2'	
* SARS-CoV-2	669- GICASYQTQTNS <b>PRRAR</b> SVA -688	808- DP----SKPS <b>KR</b> SFIED -820	
Bat-RaTG13	669- GICASYQTQTNS----- <b>R</b> SVA -684	804- DP----SKPS <b>KR</b> SFIED -816	
SARS-CoV	655- GICASYHTVSL----- <b>R</b> STS -670	790- DP----LKPT <b>KR</b> SFIED -802	
* MERS-CoV	734- SLCALPDT PSTLTP <b>RSVR</b> SVP -754	877- VSISTG <b>RSAR</b> SAIED -892	
HCoV-229E	554- GVCADGSII-----AVQ <b>PR</b> NVS-570	679- LPTSGSRVAG <b>R</b> SAIED -694	
HCoV-NL63	735- GICADGSLI-----PVR <b>PR</b> NSS-751	860- RNI RSSRIAG <b>R</b> SAIED -875	
* HCoV-HKU1	742- GFCVDYNSPSSSS <b>RRKRR</b> SI -762	895- LGPHCGS---- <b>SR</b> SFIED -909	
* HCoV-OC43	753- GYCVDYSK----- <b>NRRSR</b> GAI-768	901- LGSECSKASS <b>R</b> SAIED -916	

**FIGURE 5** Amino acid residues at the S1/S2 border and S2' site among different coronavirus spike proteins. Mono- and multibasic motifs suitable for host cell protease-mediated cleavage are highlighted with red bold; red asterisks indicate canonical Furin-like cleavage motif at the S1/S2 boundary in different coronavirus

### 2.3.2 | Dibasic S2' cleavage site

Proteolysis at a second cleavage site S2' is essential for fusion activation of all characterized CoVs spike proteins.<sup>23</sup> Multiple cellular proteases including TMPRSS2 have been implicated in the subsequent cleavage.<sup>70</sup> Proteolytic activation at the S2' site anchored spike proteins in the target cells membrane, eventually leading to early fusion at the cytoplasmic membrane.<sup>23</sup> This process has been proposed to activate the spike protein with irreversible conformational changes and trigger the membrane fusion activity.<sup>64</sup> Unlike the low pathogenic hCoVs that harbor a monobasic S2' cleavage site (R↓), the S2' cleavage site of SARS-CoV-2, in common with bat-CoVs, SARS-CoV and MERS-CoV, exhibits a markedly different feature with a dibasic cleavage site (KR↓)<sup>63</sup> (Figure 5), indicating that one or more proteases could involve in cleaving the S2' motif. TMPRSS2 cleaves at single arginine or lysine residues (R/K↓), and therefore activate viral fusion proteins. Recent studies highlighted that TMPRSS2 is also essential for SARS-CoV-2 to enter human lung cells.<sup>59,71,72</sup> TMPRSS2 expressing cell lines are highly susceptible to SARS-CoV-2.<sup>71</sup> Downregulation of TMPRSS2 activity dramatically inhibits SARS-CoV-2 replication.<sup>72</sup> Entry of SARS-CoV-2 into Calu3 cells is partially blocked by camostat mesylate, an inhibitor of TMPRSS2.<sup>59</sup>

### 2.4 | Mutations in SARS-CoV-2 S protein

RNA viruses are known to have higher mutation rates than DNA viruses.<sup>73</sup> As a typical RNA virus, coronavirus could evolve at a rate of  $10^{-4}$  substitute/bp/year.<sup>74</sup> A recent study analyzed a total of 10 022 genomes of SARS CoV-2 from 68 countries, 65 776 variants with 5775 separate variants were identified in total. Estimation of mutation rate showed a median of  $1.12 \times 10^{-3}$  mutations/site/year.<sup>75</sup> Mutations in SARS-CoV-2 are being collected each day. As of May 2021, a total of 3678 mutation sites were identified in S protein of SARS-Cov-2 (<https://bigd.big.ac.cn/ncov/variation/spike>), of which, 2694 lead to amino acid changes. Among the changed amino acids, 403 locate in the RBD, 41 are from the 17 key amino acids critical for protein interactions (Table 2).

A report from the United States over a period of 11 weeks (submitted between January 19 and April 15, 2020) revealed 16 variations in the S protein out of 579 complete SARS-CoV-2 genome sequences, and all these mutations were nonsynonymous. Out of these 16 mutations, four were in the NTD fragment, four were found in the RBD region, and the remaining eight were located in different regions within the S protein with no mutation in the FP region.<sup>76</sup> A recent analysis of 10333S protein sequences revealed 8155 proteins comprising one or more mutations. And a total of 9654 mutations were observed that correspond to 400 distinct mutation sites. Mutations are distributed in almost all regions of the S protein. The protease cleavage site (between residues 675 and 692) in the S protein is associated with the maximum mutation density.<sup>77</sup> Through in silico methods, Ahamad et al.<sup>78</sup> identified the selected mutations R408I, L455Y, F486L, Q493N, Q498Y, N501T on RBD, and A930V, D936Y on HR1 as highly deleterious, damaging the stabilization of spike protein. Furthermore, to investigate the S protein mutations and their biological importance, Qianqian Li, et al.<sup>2</sup> discovered that the D614G, along with variants containing both D614G and another amino acid change, were significantly more infectious. Sequences containing A475V, L452R, V483A, and F490L showed resistance to several neutralizing antibodies. Notably, of all variants, the D614G mutation has become the globally dominant form of SARS-CoV-2, with the frequency reaching 70.99%.<sup>79</sup> Although this mutation is located in the SD2 region, nearly all D614G mutation strains also have a replication-responsible protein mutation (ORF1ab P4715L/RdRpP323L), which can affect the speed of replication of the virus.<sup>75</sup>

Data are available in (CNCB-NGCC) at: (<https://bigd.big.ac.cn/ncov/variation/spike?lang=en>).

## 3 | CELL ENTRY MECHANISM

Coronaviruses can invade host cells through different strategies, the most important of which is receptor-mediated membrane fusion. The diversity of receptor usage is a striking feature of coronaviruses.<sup>80</sup> Different coronaviruses use different receptors to invade host cells and at the same time exhibit different tissue tropism resulting in different clinical symptoms (Supplementary Tab.1). Understanding

Mutation	Number of virus strains	Genomic location	Mutation	Number of virus strains	Genomic location
501N>Y/H/D	465 756	23063	449Y>F/C/S	21	22908
417K>T/M/R	7179	22812	456F>I/L/V	18	22928
417K>N	4579	22813	456F>L	18	22930
501N>S/I/T	2082	23064	505Y>C	17	23076
453Y>F	709	22920	496G>D/V	16	23049
446G>D/V/A	415	22899	500T>S/P/A	16	23060
455L>LX/F	389	22927	489Y>H/N	15	23027
475A>V/G	255	22986	498Q>K/X	10	23054
475A>T/P/S	137	22985	455L>S	9	22926
486F>L	137	23020	486F>S	8	23019
493Q>R/L	105	23040	487N>H/D	8	23021
496G>S/R	99	23048	498Q>R	8	23055
493Q>H	93	23041	502G>S/C	8	23066
501N>K/ NPFL	85	23065	489Y>C/S/F	7	23028
449Y>N/D/H	81	22907	487N>T/S	5	23022
493Q>*/K/E	59	23039	502G>D/V	5	23067
455L>X/V	58	22925	500T>I	4	23061
446G>R/S	57	22898	417K>E	3	22811
505Y>D/H/X	37	23075	456F>Y	3	22929
486F>L/I	34	23018	453Y>H	2	22919
498Q>H	32	23056			

Abbreviation: RBS, Receptor Binding Domain.

**TABLE 2** Mutations locate in the RBD that are from the 17 key amino acids critical for protein interaction

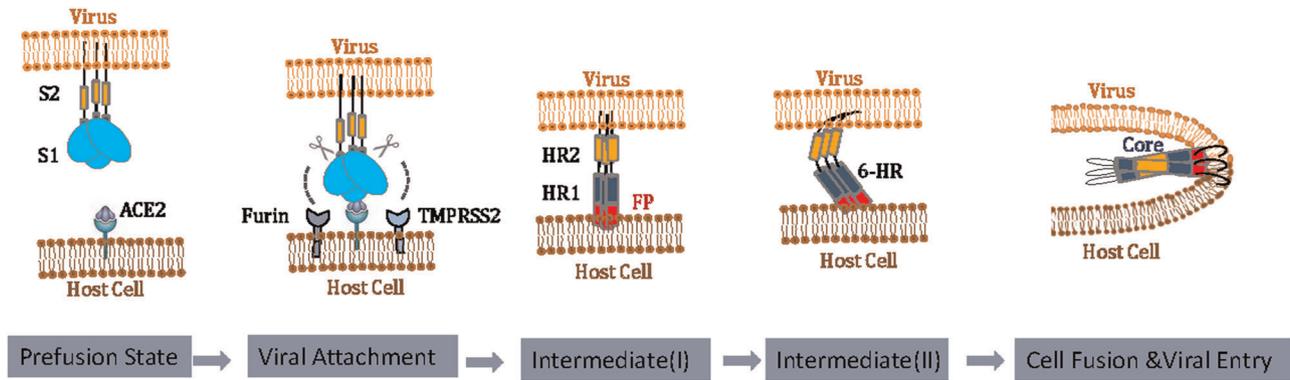
how SARS-CoV-2 enters human cells is a high priority for deciphering its mystery and curbing its spreading.

### 3.1 | ACE2-dependent receptors in viral pathogenesis

Evidence that SARS-CoV-2 utilizes host ACE2 (hACE2) as a critical receptor for cell entry is now strong and convincing,<sup>17,30,37,59,81</sup> which is strikingly similar to the mechanism exploited by SARS-CoV.<sup>82</sup> Tissues with high ACE2 expression are considered targets with potential high infection risk for SARS-CoV-2.<sup>83</sup> Although there is no evidence showing a linear correlation between the severity of the disease and the expression level of ACE2, the efficiency of ACE2 usage was reported to be a key determinant of SARS-CoV transmissibility.<sup>84,85</sup> SARS-CoV-2 S protein binds ACE2, and in concert with host proteases, principally FURIN and TMPRSS2 promote cellular entry.<sup>80</sup> Cell entry is a complex process including (i) binding of the virus to the host cell receptor brings about a conformational change in the S2 domain, followed by the proteolytic cleavage at

S1/S2 boundary<sup>60,68</sup>; (ii) a second conformational rearrangement by cleavage at S2' occurs for exposing the FP<sup>86,87</sup>; and (iii) the formation of an anti-parallel 6-HB fusion core formed by HR1 and HR2 that bridges the viral and host cell membranes into close proximity and ultimately results in cell fusion and the release of viral RNA into the host cell.<sup>88</sup> Figure 6 summarizes the spike protein-mediated cell fusion of SARS-CoV-2 in a multistep process.

ACE2 is ubiquitously expressed in human tissues with high levels in the heart, vessels, kidneys, brain, lung, and so forth.<sup>89</sup> As an endogenous counter-regulator of the renin-angiotensin system (RAS) and a cell-surface peptidase that hydrolyzes angiotensin I to angiotensin 1-9 and angiotensin II to angiotensin 1-7. ACE2 plays a vital role in the maintenance of the cardiovascular system, as well as in renal, intestinal, and respiratory systems.<sup>90,91</sup> ACE2 protein was found in two forms, namely the full-length membrane-bound ACE2 (mACE2) and circulating soluble ACE2 (sACE2). mACE2 is located on cell membranes with a transmembrane anchor and an extracellular domain. While sACE2 lacks membrane anchors and is shed from sACE2 into the circulation in low concentrations. Growing evidence suggested the protective role of sACE2 as a competitive interceptor



**FIGURE 6** Hypothesized mechanism of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral entry. The SARS-CoV-2 S engages with the hACE2 receptor and is subsequently cleaved at S1/S2 and S2' sites by proteases including, but not limited to, Furin, TMPRSS2, and cathepsin L. This leads to activation of the S2 and the forming of fusion core, ultimately drives membrane fusion and the release of viral RNA into host cell

of SARS-CoV-2 by preventing the binding of the virus to the mACE2.<sup>92,93</sup> In this line, a genetically modified sACE2, called human recombinant ACE-2 (hrsACE2) protein was developed to saturate the viral S-protein and thus prevent the cellular entry of SARS-CoV-2.<sup>94</sup> Nevertheless, Yeung et al.<sup>95</sup> reported that sACE2 could facilitate virus cell entry, *in vitro* data showed endogenous sACE2 could interact with the S of SARS-CoV-2 in the extracellular compartment. The sACE2-S complex could then enable cell entry by receptor-mediated endocytosis. Reasons for such conflicting results are yet not clear. Moreover, a decline in ACE2 expression, being a receptor for SARS-CoV-2, would prevent cellular entry of the virus thereby reducing progression of the infection. However, increased ACE2 expression have beneficial effects in the maintenance of healthy condition. The intricacy of ACE2 expression level and its regulatory mechanism in viral entry are complex, Thorough comprehension of the role of ACE2 in different pathways may help to clarify this intricate situation.

### 3.2 | ACE2-independent receptors in viral pathogenesis

It is now well established that SARS-CoV-2 utilizes the host ACE2 receptor to gain host cell entry.<sup>59,80</sup> But SARS-CoV-2 seems to infect a diverse range of cell types,<sup>96</sup> and ACE2 is widely expressed across a variety of organs. The lung, with only moderate but not the highest expression of ACE2,<sup>97</sup> however, is the major infected organ. Furthermore, the unique Furin motif at S1/S2 and dibasic cleavage S2' site may allow SARS-CoV-2 to undergo virus-cell fusion (receptor-independent entry) as well as cell-cell fusion.<sup>98</sup> Therefore, it's reasonable to speculate that there might be other co-receptors or ACE2-independent receptors for viral cell entry that are yet to be discovered.

Wang et al.<sup>99</sup> demonstrated that CD147, also known as Basigin (BSG) or EMMPRIN, functionally facilitates cell entry of SARS-CoV-2. Co-localization of CD147 and SARS-CoV-2 S was

detected by immuno-electron microscopy; the binding and interaction of the two proteins were also confirmed by Co-Immunoprecipitation and ELISA; more importantly, Meplazumab (an anti-CD147 antibody) could efficiently inhibit SARS-CoV-2 infection in a dose-dependent manner. Furthermore, a clinical trial by Bian et al.<sup>100</sup> showed confirmatory proof that treatment with Meplazumab enhanced virus clearance, and decreased lymphocytopenia and inflammation index. A recent study, however, reported that they were unable to find evidence supporting the role of basigin as a putative spike binding receptor. In their research, no proof for a direct interaction between the viral S protein to either of the two common isoforms of BSG could be found. Removing BSG from the surface of human lung epithelial cells by CRISPR/Cas9 resulted in no change in their susceptibility to SARS-CoV-2 infection.<sup>101</sup> Besides this, a genomic study investigating variants in genes involved SARS-CoV-2 infection failed to discover BSG variants in patients with COVID-19.<sup>102</sup> Although CD147/BSG could have some biological relevance through indirect routes which could indirectly influence COVID-19 clinical progression,<sup>103</sup> the hypothesis that CD147/BSG acts as a coreceptor or an equally essential new receptor remains to be well established.

Glucose Regulated Protein 78 (GRP78), also known as BiP or HSPA5, was indicated by accumulated proof to be another receptor for SARS-CoV-2 cell entry.<sup>104–108</sup> Four regions of the SARS-CoV-2 Spike protein were predicted by molecular docking to bind the host cell surface GRP78.<sup>104</sup> Palmeira et al.<sup>108</sup> found preliminary evidence that further proposed GRP78 as a possible molecular target to inhibit SARS-CoV-2 infection. The same proposition has confirmed by Allam et al.,<sup>107</sup> who identified nine compounds that could act as potential blockers of SARS-CoV-2 cell entry through GRP78. Nevertheless, although the model study confirmed that SARS-CoV-2 S could stably interact with the GRP78, the structure of SARS-CoV-2 S this study relied on differs from the previously published structure obtained by cryo-EM.<sup>22</sup> Further mechanistic studies, as well as the role of GRP78 as entry receptor in SARS-CoV-2 infection, remain to be demonstrated.

Takaharu Ichimura et al.<sup>109</sup> reported that Kidney injury molecule-1 (KIM-1), a drastically upregulated biomarker for kidney injury,<sup>110</sup> serves as an alternative receptor to ACE2 for SARS-CoV-2. KIM-1 was expressed in lung and kidney epithelial cells in patients with COVID-19, and colocalization of KIM-1 and SARS-CoV-2 nucleocapsid protein was detected in alveolar epithelium cells. Enhanced KIM-1 expression by human kidney tubuloids increased uptake of virosomes. And treatment with anti-KIM-1 antibody or TW-37 (a newly discovered inhibitor of KIM-1 small molecule inhibitor) dramatically blocked entry of the SARS-CoV-2. Using in silico simulation, co-immunoprecipitation, fluorescence resonance energy transfer Chen Yang et al.<sup>111</sup> showed confirmatory evidence that further supports KIM1 as a novel receptor for SARS-CoV-2. They demonstrated that KIM-1 binds with the receptor-binding domain of SARS-CoV-2 and facilitates its attachment to the cell membrane. The interaction between the SARS-CoV-2 receptor-binding domain and KIM-1 is potentially blocked by a rationally designed KIM-1-derived polypeptide AP2. In addition, they also found that SARS-CoV-2-RBD binds to KIM-1 with a higher affinity than that of SARS-CoV-RBD and MERS-CoV-RBD, which probably underlies the stronger contagion of SARS-CoV-2. Thus, the role of KIM1 in the cell entry of SARS-CoV-2 is worth further exploring.

## 4 | CONCLUSIONS

Emerging and re-emerging viral infections are continuous global threats to human health. Coronaviruses clearly have evolved to cross the species barrier into new hosts, making it straightforward to predict that more pandemics might outbreak in the future. The newly emerging and ongoing COVID-19 has spread all over the world, having a mortality of more than 31 million lives so far. Extensive progress in every field has been made in terms of understanding the SARS-CoV-2 as well as COVID-19. Close resemblance to other coronaviruses suggests general principles that may manage its biological activities. Nevertheless, initial testing of the drugs used against SARS-CoV and MERS-CoV has been proved to be ineffective in controlling SARS-CoV-2. And genetic mutations in the SARS-CoV-2 continue to increase, making it more challenging for vaccine design. Recent studies have demonstrated that the SARS-CoV-2 spike protein, a major target for eliciting antibodies, plays a vital role in virus invasion. Therefore, an in-depth understanding of the biological properties of the SARS-CoV-2 S is urgently important for structure-based strategies targeting SARS-CoV-2. In this view, we systematically summarized the molecular characteristics of SARS-CoV-2 S using SARS-CoV S as a comparison. The cell entry mechanism of SARS-CoV-2 has also been discussed. In particular, we offered a perspective on the notable features of the SARS-CoV-2 S that set it apart from SARS-CoV and other hCoVs. Collectively, our comprehensive understanding of the sophisticated structure and powerful function of SARS-CoV-2 S as well as its cell entry mechanism, herein, may provide insights into the mechanism of pathogenesis, interspecies

transmission as well as structure-guided intervention strategies targeting SARS-CoV-2. Notably, an intact S structure is yet still unavailable for any CoVs, and most of the structural and functional analysis of SARS-CoV-2 S are based on the discrete functional domains, although the cryo-electron microscopy structures of SARS-CoV-2 S have been well demonstrated,<sup>22,46</sup> which revealed the detailed structure of the SARS-CoV-2 RBD and how it contacts host ACE2 with striking differences compared to SARS-CoV.<sup>30</sup> The overall interaction between SARS-CoV-2 S and hACE2, however, remains to be elucidated. Future research will shed light on the biological features as well as the mechanism of SARS-CoV-2 infection.

## ACKNOWLEDGMENTS

This study was supported by funds from the COVID-19 Special Funding projects of Changsha Science and Technology Bureau (Grant No. kq2001020); the Natural Science Foundation of Hunan Province (Grant No. 2019JJ80005) and the Research Foundation of Hunan Provincial Health and Family Planning Committee (Grant No. 20200477). We thank Yun Deng for technical guidance.

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## AUTHOR CONTRIBUTIONS

Chaogeng Zhu and Guiyun He involved in data extraction, initial and final manuscript writing. Xiangli Ye and Yongzhong Shi contributed to the ideas and organization of the content. Qinqin Yin and Lin Zeng searched the literature and collected data. Wei Xu conceived the review, designed the figure, drafted and finalized the manuscript for publication. All authors read and approved the final version of the manuscript.

## ORCID

Chaogeng Zhu  <http://orcid.org/0000-0001-5522-2135>

Xiangli Ye  <http://orcid.org/0000-0001-9930-0123>

Yongzhong Shi  <http://orcid.org/0000-0001-6255-3864>

Wei Xu  <http://orcid.org/0000-0003-1549-4237>

## REFERENCES

1. Heald-Sargent T, Gallagher T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses*. 2012; 4(4):557-580.
2. Li Q, Wu J, Nie J, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*. 2020;182(5): 1284-1294.
3. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. 2020;182(4):812-827.
4. Lokman SM, Rasheduzzaman M, Salauddin A, et al. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. *Infect Genet Evol*. 2020; 84:104389.
5. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-574.

6. Wu A, Peng Y, Huang B, Ding X, Jiang T. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020;27:3-328.
7. Licitra B, Duhamel G, Whittaker G. Canine enteric coronaviruses: emerging viral pathogens with distinct recombinant spike proteins. *Viruses*. 2014;6(8):3363-3376.
8. Felten S, Hartmann K. Diagnosis of feline infectious peritonitis: a review of the current literature. *Viruses*. 2019;11:11.
9. Tang Q, Song Y, Shi M, Cheng Y, Zhang W, Xia X-Q. Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci Rep*. 2015;5(1):17155.
10. Woo PC, Lau SK, Lam CS, 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*. 2012;86(7):3995-4008.
11. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*. 2009;7(6):439-450.
12. Cunha CB, Opal SM. Middle East respiratory syndrome (MERS): a new zoonotic viral pneumonia. *Virulence*. 2014;5(6):650-654.
13. Coleman CM, Frieman MB. Coronaviruses: important emerging human pathogens. *J Virol*. 2014;88(10):5209-5212.
14. Corman V, Muth D, Niemeyer D, Drosten C. Hosts and sources of endemic human coronaviruses. *Adv Virus Res*. 2018;100:163-188.
15. Holmes, Kathryn V. SARS-associated coronavirus. *N Engl J Med*. 2003;348(20):1948-1951.
16. Zaki A, van Boheemen S, Bestebroer T, Osterhaus A, Fouchier R. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367(19):1814-1820.
17. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273.
18. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269.
19. Zhou H, Chen X, Hu T, et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr Biol*. 2020;30(11):2196-2203.
20. Nasab MG, Saghazadeh A, Rezaei N. SARS-CoV-2-A tough opponent for the immune system. *Arch Med Res*. 2020;51:589-592.
21. Xia S, Zhu Y, Liu M, et al. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol*. 2020;17(7):765-767.
22. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *bioRxiv*. 2020
23. Park JE, Li K, Barlan A, et al. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proc Natl Acad Sci U S A*. 2016;113(43):12262-12267.
24. Walls AC, Tortorici MA, Bosch BJ, et al. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature*. 2016;531(7592):114-117.
25. Tortorici MA, Veesler D. Structural insights into coronavirus entry. *Adv Virus Res*. 2019;105:93-116.
26. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res*. 2006;66:193-292.
27. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;181(2):281-292.
28. Sironi M, Hasnain SE, Rosenthal B, et al. SARS-CoV-2 and COVID-19: a genetic, epidemiological, and evolutionary perspective. *Infect Genet Evol*. 2020;84:104384.
29. Tai W, He L, Zhang X, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol*. 2020;17(6):613-620.
30. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581(7807):215-220.
31. Ge M, Freed J. Fusion peptide from influenza hemagglutinin increases membrane surface order: an electron-spin resonance study. *Biophys J*. 2009;96(12):4925-4934.
32. Schroth-Diez B, Ludwig K, Baljinnyam B, Kozerski C, Huang Q, Herrmann A. The role of the transmembrane and of the intraviral domain of glycoproteins in membrane fusion of enveloped viruses. *Biosci Rep*. 2000;20(6):571-595.
33. Petit CM, Melancon JM, Chouljenko VN, et al. Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology*. 2005;341(2):215-230.
34. Mittal A, Manjunath K, Ranjan RK, Kaushik S, Kumar S, Verma V. COVID-19 pandemic: insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. *PLoS Pathog*. 2020;16(8):e1008762.
35. Matrosovich M, Herrler G, Klenk H. Sialic acid receptors of viruses. *Top Curr Chem*. 2015;367:1-28.
36. Fantini J, Di Scala C, Chahinian H, Yahi N. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents*. 2020;55(5):105960.
37. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol*. 2020;94(7).
38. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol*. 2020;92(4):418-423.
39. Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 2020;581(7807):221-224.
40. Wu K, Chen L, Peng G, et al. A virus-binding hot spot on human angiotensin-converting enzyme 2 is critical for binding of two different coronaviruses. *J Virol*. 2011;85(11):5331-5337.
41. Li F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J Virol*. 2008;82(14):6984-6991.
42. Wu K, Peng G, Wilken M, Geraghty R, Li F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J Biol Chem*. 2012;287(12):8904-8911.
43. Shang J, Wan Y, Luo C, et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A*. 2020;117(21):11727-11734.
44. Hsieh CL, Goldsmith JA, Schaub JM, et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*. 2020;369:1501-1505.
45. McCallum M, Walls AC, Bowen JE, Corti D, Veesler D. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat Struct Mol Biol*. 2020;27(10):942-949.
46. Yuan Y, Cao D, Zhang Y, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun*. 2017;8:15092.
47. Gui M, Song W, Zhou H, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res*. 2017;27(1):119-129.
48. Shekhar N, Sarma P, Prajapat M, et al. In silico structure-based repositioning of approved drugs for spike glycoprotein S2 domain fusion peptide of SARS-CoV-2: rationale from molecular dynamics and binding free energy calculations. *mSystems*. 2020;5(5):e00382-20.
49. Jaimes JA, Andre NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic analysis and structural modeling of SARS-CoV-2 spike

- protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *J Mol Biol.* 2020;432(10):3309-3325.
50. Guillén J, Pérez-Berná A, Moreno M, Villalain J. Identification of the membrane-active regions of the severe acute respiratory syndrome coronavirus spike membrane glycoprotein using a 16/18-mer peptide scan: implications for the viral fusion mechanism. *J Virol.* 2005;79(3):1743-1752.
  51. Chakraborty H, Tarafdar P, Klapper D, Lentz B. Wild-type and mutant hemagglutinin fusion peptides alter bilayer structure as well as kinetics and activation thermodynamics of stalk and pore formation differently: mechanistic implications. *Biophys J.* 2013;105(11):2495-2506.
  52. Xia S, Liu M, Wang C, et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* 2020;30(4):343-355.
  53. Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol.* 2009;7:3-236.
  54. Reuven E, Dadon Y, Viard M, Manukovsky N, Blumenthal R, Shai Y. HIV-1 gp41 transmembrane domain interacts with the fusion peptide: implication in lipid mixing and inhibition of virus-cell fusion. *Biochemistry.* 2012;51(13):2867-2878.
  55. Millet J, Whittaker G. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A.* 2014;111(42):15214-15219.
  56. Antoni Wrobel DB, Pengqi Xu, Chloë R, et al. Evolution of SARS-CoV-2 spike glycoprotein. *Research Square.* 2020:1-10.
  57. Belouzard S, Chu V, Whittaker G. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A.* 2009;106(14):5871-5876.
  58. Belouzard S, Millet J, Licitra B, Whittaker G. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses.* 2012;4(6):1011-1033.
  59. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 2020;181(2):271-280.
  60. Hoffmann M, Kleine-Weber H, Pohlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell.* 2020;78(4):779-784.
  61. Braun E, Sauter D. Furin-mediated protein processing in infectious diseases and cancer. *Clin Transl Immunol.* 2019;8(8):e1073.
  62. Izaguirre G. The proteolytic regulation of virus cell entry by furin and other proprotein convertases. *Viruses.* 2019;11(9).
  63. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* 2020;176:104742.
  64. Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* 2015;202:120-134.
  65. Dahms S, Arciniega M, Steinmetzer T, Huber R, Than M. Structure of the unliganded form of the proprotein convertase furin suggests activation by a substrate-induced mechanism. *Proc Natl Acad Sci U S A.* 2016;113(40):11196-11201.
  66. Follis KE, York J, Nunberg JH. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology.* 2006;350(2):358-369.
  67. Steinhauer D. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology.* 1999;258(1):1-20.
  68. Lemmin T, Kalbermatter D, Harder D, Plattet P, Fotiadis D. Structures and dynamics of the novel S1/S2 protease cleavage site loop of the SARS-CoV-2 spike glycoprotein. *J Struct Biol X.* 2020;4:100038.
  69. Ali A, Rabaan SHA-A S, Haque R, et al. SARS-CoV-2, SARS-CoV, and MERS-CoV: a comparative overview. *Infez Med.* 2020;2:174-184.
  70. Shirato K, Kawase M, Matsuyama S. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol.* 2013;87(23):12552-12561.
  71. Matsuyama S, Nao N, Shirato K, et al. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc Natl Acad Sci U S A.* 2020;117(13):7001-7003.
  72. Dorothea B, Heindl MR, Limburg H, et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Allianc.* 2020;3(9):1-14.
  73. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* 2018;16(8):e3000003.
  74. Su S, Wong G, Shi W, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *TIM.* 2016;24(6):490-502.
  75. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ.* 2020;98(7):495-504.
  76. Kaushal N, Gupta Y, Goyal M, Khaiboullina SF, Baranwal M, Verma SC. Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. *Pathogens.* 2020;9(7):565.
  77. Guruprasad L. Human SARS CoV-2 spike protein mutations. *Proteins.* 2021;89(5):569-576.
  78. Ahamad S, Kanipakam H, Gupta D. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding. *J Biomol Struct Dyn.* 2020:1-13.
  79. Ugurel OM, Ata O, Turgut-Balik D. An updated analysis of variations in SARS-CoV-2 genome. *Turk J Biol.* 2020;44(3):157-167.
  80. Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol.* 2016;3(1):237-261.
  81. Wang Q, Zhang Y, Wu L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell.* 2020;181(4):894-904.
  82. Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature.* 2003;426(6965):450-454.
  83. Ziegler CGK, Allon SJ, Nyquist SK, et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell.* 2020;181(5):1016-1035.
  84. Li F, Li W, Farzan M, Harrison S. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science.* 2005;309(5742):1864-1868.
  85. Hoffmann M, Hofmann-Winkler H, Pöhlmann S. Priming time: how cellular proteases arm coronavirus spike proteins. *Activation of Viruses by Host Proteases.* 2018:71-98.
  86. Walls AC, Tortorici MA, Snijder J, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc Natl Acad Sci U S A.* 2017;114(42):11157-11162.
  87. Harrison S. Viral membrane fusion. *Nat Struct Mol Biol.* 2008;15(7):690-698.
  88. Guillen J, Kinnunen PK, Villalain J. Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein. *Biochim Biophys Acta.* 2008;1778(12):2765-2774.
  89. Tipnis SR, Hooper NM, Hyde R, Karran E, Christie G, Turner AJ. A human homolog of angiotensin-converting enzyme cloning and functional expression as a captopril-insensitive carboxypeptidase. *J Biol Chem.* 2000;275(43):33238-33243.
  90. Oliveira Andrade J, de Farias Lelis D, Mafra V, Cota J. The angiotensin converting enzyme 2 (ACE2), gut microbiota, and cardiovascular health. *Protein Pept Lett.* 2017;24(9):827-832.
  91. Jia H. Pulmonary angiotensin-converting enzyme 2 (ACE2) and inflammatory lung disease. *Shock.* 2016;46(3):239-248.

92. Monteil V, Kwon H, Prado P, et al. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell*. 2020;181(4):905-913.
93. Lei C, Qian K, Li T, et al. Neutralization of SARS-CoV-2 spike pseudotyped virus by recombinant ACE2-Ig. *Nat Commun*. 2020;11(1):2070.
94. Batlle D, Wysocki J, Satchell K. Soluble angiotensin-converting enzyme 2: a potential approach for coronavirus infection therapy. *Clin Sci*. 2020;134(5):543-545.
95. Yeung ML, Teng JLL, Jia L, et al. Soluble ACE2-mediated cell entry of SARS-CoV-2 via interaction with proteins related to the renin-angiotensin system. *Cell*. 2021;184(8):2212-2228.
96. Chiodo F, Bruijns SCM, Rodriguez E, et al. Novel ACE2-independent carbohydrate-binding of SARS-CoV-2 spike protein to host lectins and lung microbiota [published online ahead of print May 14, 2020]. *BioRxiv*. 2020.
97. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med*. 2020;14(2):185-192.
98. Menachery VD, Dinnon KH, Yount BL, et al. Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection. *J Virol*. 2020;94(5):e01774-19.
99. Wang K, Chen W, Zhou Y-S, et al. SARS-CoV-2 invades host cells via a novel route: CD147-spike protein [published online ahead of print March 14, 2020]. *BioRxiv*. 2020.
100. Bian H, Zheng Z-H, Wei D, et al. Meplazumab treats COVID-19 pneumonia: an open-labelled, concurrent controlled add-on clinical trial [published online ahead of print March 24, 2020]. *medRxiv*. 2020.
101. Shilts J, Crozier TWM, Greenwood EJD, Lehner PJ, Wright GJ. No evidence for basigin/CD147 as a direct SARS-CoV-2 spike binding receptor. *Sci Rep*. 2021;11(1):413.
102. Latini A, Agolini E, Novelli A, et al. COVID-19 and genetic variants of protein Involved in the SARS-CoV-2 entry into the host cells. *Genes*. 11, 2020;9.
103. Lucas C, Wong P, Klein J, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*. 2020;584(7821):463-469.
104. Ibrahim IM, Abdelmalek DH, Elshahat ME, Elfiky AA. COVID-19 spike-host cell receptor GRP78 binding site prediction. *J Infect*. 2020;80(5):554-562.
105. Ha DP, Van Krieken R, Carlos AJ, Lee AS. The stress-inducible molecular chaperone GRP78 as potential therapeutic target for coronavirus infection. *J Infect*. 2020;81(3):452-482.
106. Sudeep HV, Gouthamchandra K, Shyamprasad K. Molecular docking analysis of Withaferin A from *Withania somnifera* with the Glucose regulated protein 78 (GRP78) receptor and the SARS-CoV-2 main protease. *Bioinformation*. 2020;16(5):411-417.
107. Allam L, Ghriji F, Mohammed H, et al. Targeting the GRP78-dependant SARS-CoV-2 cell entry by peptides and small molecules. *Bioinform Biol Insights*. 2020;14:1177932220965505.
108. Palmeira A, Sousa E, Koseler A, et al. Preliminary virtual screening studies to identify GRP78 inhibitors which may interfere with SARS-CoV-2 infection. *Pharmaceuticals*. 2020;13(6):132.
109. Ichimura T, Mori Y, Aschauer P, et al. KIM-1/TIM-1 is a receptor for SARS-CoV-2 in lung and kidney. *medRxiv*. 2020.
110. Yang L, Brooks CR, Xiao S, et al. KIM-1-mediated phagocytosis reduces acute injury to the kidney. *J Clin Invest*. 2015;125(4):1620-1636.
111. Chen Yang YZ, Zeng Xia, Chen H, et al. Kidney injury molecule-1 is a potential receptor for SARS-CoV-2. *J Mol Cell Biol*. 2021.

#### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Zhu C, He G, Yin Q, et al. Molecular biology of the SARS-CoV-2 spike protein: A review of current knowledge. *J Med Virol*. 2021;93:5729-5741.

<https://doi.org/10.1002/jmv.27132>