

Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein

Tahir S Pillay  ^{1,2}

¹Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria & National Health Laboratory Service, Pretoria, South Africa
²Division of Chemical Pathology, University of Cape Town, Cape Town, South Africa

Correspondence to

Professor Tahir S Pillay, Chemical Pathology, University of Pretoria, Pretoria, 0002, South Africa; jclinpathaic@gmail.com

Accepted 17 April 2020
 Published Online First 6 May 2020

ABSTRACT

The year 2020 has seen a major and sustained outbreak of a novel betacoronavirus (severe acute respiratory syndrome (SARS)-coronavirus (CoV)-2) infection that causes fever, severe respiratory illness and pneumonia, a disease called COVID-19. At the time of writing, the death toll was greater than 120 000 worldwide with more than 2 million documented infections. The genome of the CoV encodes a number of structural proteins that facilitate cellular entry and assembly of virions, of which the spike protein S appears to be critical for cellular entry. The spike protein guides the virus to attach to the host cell. The spike protein contains a receptor-binding domain (RBD), a fusion domain and a transmembrane domain. The RBD of spike protein S binds to Angiotensin Converting Enzyme 2 (ACE2) to initiate cellular entry. The spike protein of SARS-CoV-2 shows more than 90% amino acid similarity to the pangolin and bat CoVs and these also use ACE2 as a receptor. Binding of the spike protein to ACE2 exposes the cleavage sites to cellular proteases. Cleavage of the spike protein by transmembrane protease serine 2 and other cellular proteases initiates fusion and endocytosis. The spike protein contains an addition furin cleavage site that may allow it to be 'preactivated' and highly infectious after replication. The fundamental role of the spike protein in infectivity suggests that it is an important target for vaccine development, blocking therapy with antibodies and diagnostic antigen-based tests. This review briefly outlines the structure and function of the 2019 novel CoV/SARS-CoV-2 spike protein S.

INTRODUCTION

The pathogen severe acute respiratory syndrome (SARS)-coronavirus (CoV)-2 is a newly discovered member of the genus *Betacoronavirus* and is related to SARS-CoV but has been found to be readily transmitted between humans leading WHO to declare initially a Public Health Emergency of Immediate Concern and then later, a pandemic. In the new millennium, several CoVs have crossed species to cause severe and often fatal pneumonia: SARS-CoV; Middle-East respiratory syndrome (MERS)-CoV and the 2019 novel CoV/SARS-CoV-2. SARS-CoV-2 originated from Wuhan in the Hubei province of China and was first identified in December 2019. SARS-CoV-2 is the seventh CoV that has been shown to infect humans.¹

CoVs are lipid enveloped and spherical with a size of approximately 100–120 nm (figure 1A). The nucleocapsid houses the single-stranded non-segmented positive-sense RNA genome.² The name *Coronavirus* is derived from the appearance

of the virus—there are large petal-shaped spikes protruding from the envelope—this is the spike glycoprotein which is 20–40 nm long. In some betacoronaviruses, there is a second shorter projection which is the haemagglutinin (HA) esterase protein³ (figure 1A). The most abundant protein, the M protein, provides structural support. The E protein is a small membrane protein that is essential for the assembly and release of virions.

The genome sequence of SARS-CoV-2 was elucidated in January 2020.^{4–7} The SARS-CoV-2 genome is 96% identical to that of the bat coronavirus (BatCoV) RaTG13 and 80% identical to the SARS-CoV genome.⁷ The positive single-stranded RNA genome contains 29.3 kilobases^{4,6} (figure 2). A unique feature of the CoV family is the synthesis of a nested set of subgenomic mRNAs that are transcribed from negative strand RNAs. The subgenomic RNAs all contain a leader sequence from the 5' end of the genome. The genome is transcribed in a non-contiguous manner where the viral RNA-dependent RNA polymerase (RdRp) skips across from one part of the genome to the next to generate the RNAs.^{8,9} The basis of the most commonly used nucleic acid detection tests for SARS-CoV-2 is centred around the real-time-PCR detection of the RdRp, E (envelope) and N (nucleocapsid) genes (figure 2).^{10,11}

In common with other viruses, the genome of the virus encodes for proteins that facilitate cellular entry (figure 3). In the case of SARS-CoV-2 (and other CoVs), this is the spike protein (S), a 1273 amino acid homotrimeric class I fusion protein that allows the viral membrane to fuse with the host cell membrane.¹² (A class I viral fusion protein is a viral protein primed by cleavage of a single-chain precursor that exists as a trimer. The classical example of a type 1 fusion protein is influenza virus HA. Other examples include Ebola and paramyxovirus.) The third open reading frame in the genome encodes the spike protein⁵ (figure 2). Each monomer of the spike protein trimer is approximately 180 kDa. The S protein is the largest protein in the group of four structural proteins (including, M, E and N proteins). The spike protein of SARS-CoV-2 has 76% amino acid sequence identity with SARS-CoV suggesting that it interacts with similar protein targets. The spike protein of SARS-CoV-2 shows 93% and 97% amino acid identity with that of the BatCoV RaTG13 and Pangolin-CoV, respectively, strongly hinting at the origin of the intermediate hosts of SARS-CoV-2.^{4,7,13,14} Owing to the strong conservation of residues and binding to the receptors, immunity to SARS-CoV can confer some very limited immunity to SARS-CoV-2 based on the observation that SARS-CoV antibodies are able

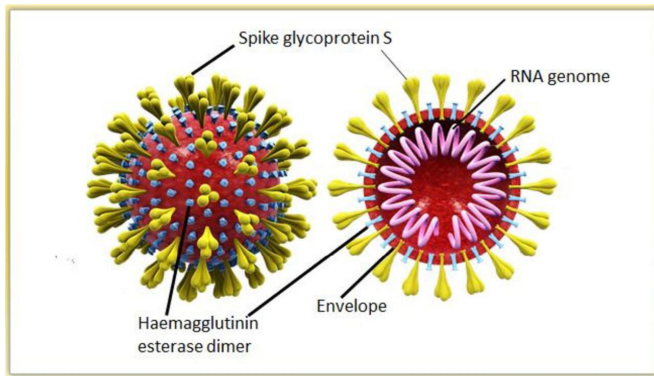


► <http://dx.doi.org/10.1136/jclinpath-2020-206685>

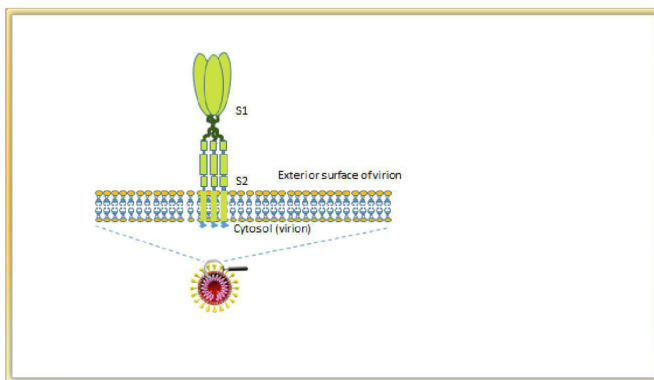


© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Pillay TS. *J Clin Pathol* 2020;**73**:366–369.



A



B

Figure 1 Structure of the (A) novel coronavirus severe acute respiratory syndrome-CoV-2 and the (B) spike protein.

to prevent the entry of SARS-CoV-2 into cells^{15 16} using either convalescent sera or immunised rabbit sera.¹⁶ However, in spite of the sequence identity, the electrostatic surface potential, antigenicity and epitopes of the spike glycoproteins from different CoV proteins are distinct, even though they use the same ACE2 receptor protein.¹⁷

The spike protein can be divided into a number of regions or functional domains or subunits. First, the globular head (or ‘petal’) contains the longer S1 N-terminal RBD. The stem contains the C-terminal membrane fusion domain (S2) followed

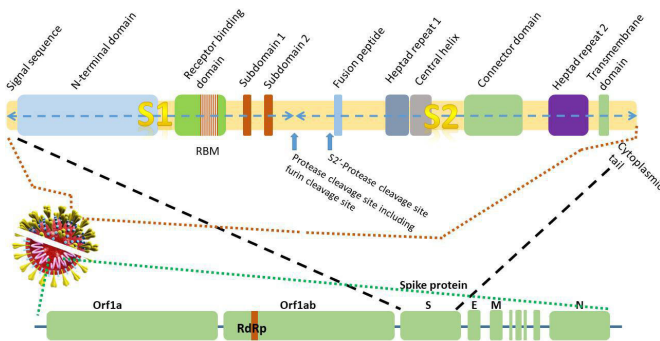


Figure 2 Figure 2 Schematic diagram of the genomic structure of the 29.3 kilobase 2019-novel coronavirus (nCoV) gene and domain structure of the 1273 amino acid spike glycoprotein S (not to scale). E, envelope protein gene; M, membrane protein gene; N, nucleocapsid protein gene; RBM, receptor-binding motif; RdRp, RNA-dependent RNA polymerase; S, spike protein gene.

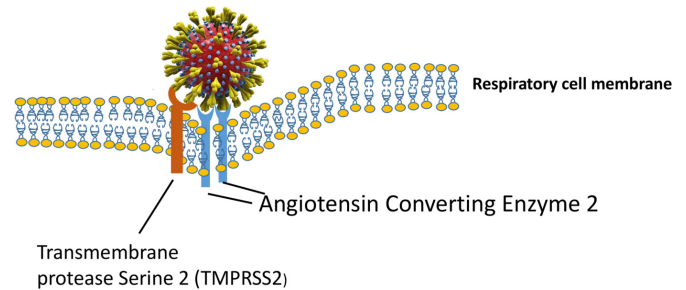


Figure 3 Cellular entry of severe acute respiratory syndrome coronavirus 2 requires ACE2 and a cellular protease such as TMPRSS2.

by the two heptad regions (HR1 and HR2), the transmembrane (TM) domain and cytosolic tail² (figure 2). The RBD of S1 is approximately 200 amino acids long. The binding of the S1 domain exposes regions within the S2 domain.¹⁸

There are significant post-translational modifications in the S protein. The spike protein ectodomain is heavily glycosylated with heterogeneous N-linked glycans¹⁹ and exists in a pre-fusion and a post-fusion conformation. The oligosaccharides could influence priming by host proteases and determine antibody recognition. Cysteine residues in the cytosolic tail undergo palmitoylation.¹⁹ The switch in conformations is triggered by binding to its cellular receptor, ACE2 (figure 3).

PROTEOLYTIC CONVERSION AND ACTIVATION OF THE SPIKE PROTEIN PRIOR TO FUSION

Entry of the virus is dependent on proteolytic activation of the spike protein. In the process of viral infection involving attachment of the virus and entry into cells, the spike protein is cleaved at the protease cleavage sites into the S1 and S2 subunits and the S2 subunit is released.^{17 18 20} There are several steps: these include binding of the virus to the cell surface; alteration of the conformation of the spike protein; proteolysis of the spike protein; release of the S2 subunit which then mediates fusion of the virion and endocytosis.^{21 22}

The first step in the process of cellular entry is the binding of the spike protein to the cell surface receptor, ACE2 mediated by the RBD of S1.^{16–18 20 23 24} The critical function of the S protein in facilitating the interaction with the cell membrane indicates that it could be a target for therapy using antibodies or chemical compounds, as well as a vaccine target. Attachment of the S1 RBD to the ACE2 enzyme exposes a cleavage site on S2 that is acted on by host cell proteases such as TMPRSS2 to initiate the process of cellular entry.^{12 16}

The RBD binds to the carboxypeptidase domain of ACE2. The RBD moves like a hinge between two conformations (‘up’ or ‘down’) to expose or hide the residues that bind the ACE2 receptor.²⁵ Within the RBD, there is a receptor binding motif (RBM) which makes the primary contact with the peptidase domain of ACE2.^{26 27} The spike protein RBM is conserved (50%) between SARS-S and SARS-CoV2-S and these residues are absent from SARS-related-CoVs which do not use ACE2.

The cleavage process of the spike protein takes place in two steps: a ‘priming’ cleavage and then an ‘activation’ cleavage^{21 22} at a cleavage site between the S1 and S2 domains as well as the S2’ site. Most CoVs spike proteins are cleaved at the junction between S1 and S2. The cleavage site between the S1/S2 domains contains multibasic arginine residues. Although SARS-CoV also uses the ACE2 receptor, SARS-CoV2 has a distinct furin cleavage site (Arg-Arg-Ala-Arg) at residues 682–685, between the S1 and S2 domains (figure 2) not found

in SARS-CoV which may explain some of the biological differences. Removal of this motif affects cellular entry. The furin cleavage site expands the versatility of SARS-CoV-2 for cleavage by cellular proteases and potentially the tropism and transmissibility owing to the wide cellular expression of furin proteases especially in the respiratory tract.¹² This means that newly synthesised virions can be secreted in a 'preactivated' state ready to fuse with and infect other cells without the need to bind to a cellular receptor, such as ACE2.

In the two-step sequential process, the activation cleavage at the S2' site is thought to activate the protein for membrane fusion. This then allows fusion of the viral membrane with the cellular membrane. Inhibition of transmembrane protease serine 2 (TMPRSS2) using camostat mesylate partially blocks the entry of SARS-2-V.¹⁶ TMPRSS2 is also critical for infection by viruses such as influenza A.¹⁶

Apart from TMPRSS2, SARS-CoV-2 spike protein can be proteolytically activated by a variety of other cellular proteases including cathepsin B and L (endosomal cysteine proteases). Inhibition of all of these prevents viral entry. Other proteases such as furin, elastase, factor X and trypsin are also capable of this 'priming' proteolysis that initiates the process of cellular entry.^{22 28}

There is some evidence that SARS-CoV-2 S protein can trigger protease-independent and receptor-dependent cellular fusion to enhance viral spreading. PIKfyve, the enzyme generating PI(3,5)P₂, also regulates early-to-late endosomal maturation and inhibition of this may block viral entry.²⁹

ACE2 AS A RECEPTOR

The role of ACE2 as a receptor for CoVs was first identified during the outbreak of the SARS-CoV using coimmunoprecipitation studies.^{30 31} A similar role for ACE2 in SARS-CoV2 infection was identified in 2020.⁷ The S1 RBD (amino acid residues 333–527) of SARS-CoV-2 binds to ACE2 with nanomolar affinity implicating it as an important component of the viral cellular interaction.¹² In comparison with SARS-CoV, the RBD of SARS-CoV-2 binds to ACE2 with more than 10-fold higher affinity.^{17 25} Within the RBD, there is a RBM at residues 438–506. Other cellular receptors that viruses use are mouse carcinoembryonic antigen-related cell adhesion molecule 1a (*Mouse hepatitis virus*); human dipeptidyl peptidase 4 (*MERS-CoV* and *SARS-CoV*).¹

ACE2 is a dipeptidyl carboxypeptidase that primarily exists as an 805 amino acid type 1 TM (single TM protein with N-terminus extracellular) protein with a zinc-binding domain and is predominantly expressed in lung, heart, kidney, testis and gastrointestinal tract.^{32 33} In the lung, ACE2 is found in alveolar epithelial type II cells (and bronchial epithelial and is a receptor for SARS-CoV, MERS-CoV and SARS-CoV-2.³⁴ MERS-CoV also uses dipeptidyl peptidase 4 (DPP4/CD26) as a receptor.¹ There is also a second circulating soluble form of ACE2. ACE2 appears to have distinct physiological functions from ACE and ACE inhibitors used clinically as antihypertensive agents do not affect ACE2 activity. The protease activity of ACE2 has no role in facilitating viral entry and it merely appears to act as a receptor to guide attachment and fusion of the spike protein. ACE2 contains two functional regions, the N-terminal M2 peptidase domain and the C-terminal collectrin domain which regulates the trafficking of amino acid transporters to the cell membrane. The physiological function of ACE2 is to convert angiotensin 1 to angiotensin(1-9) and angiotensin II to angiotensin (1-7).^{33 35} Angiotensin (1-9) binds to the Mas receptor leading to vasodilatation and

Take home messages

- ▶ The spike protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is critical for viral entry into cells and is a target for vaccine development, blocking therapy and antigen testing.
- ▶ The receptor for SARS-CoV-2 in the type II alveolar respiratory cell is ACE2. Binding is mediated by the receptor-binding domain and motif (RBM) of the S1 region of the spike protein.
- ▶ Viral attachment, fusion and entry begins with the interaction of the spike protein with ACE2 and proteolytic cleavage by cellular proteases such as transmembrane protease serine 2.
- ▶ SARS-CoV-2 binds to ACE2 with higher affinity than SARS-CoV owing to the presence of a 4-residue motif in the RBM that makes better contact with ACE2 than the spike protein of SARS-CoV.

anti-inflammatory effects. This is in contrast (antagonistic) to ACE which converts angiotensin I to angiotensin II. Angiotensin II causes vasoconstriction and proinflammatory effects via the angiotensin II receptor type 1.

The critical residues of the spike protein RBM that interact with the ACE2 domain have been identified from the cocrystal structures.^{18 26}

While both SARS-CoV and SARS-CoV-2 bind to ACE2 and the binding sites overlap, there are structural differences in both spike proteins that influence binding characteristics; for example, the SARS-CoV-2 RBM has a four residue motif (Gly-Val-Glu-Gly) at 482–485 in the RBM that allows better contact with the ACE2. ACE2 contains two lysine residue virus hotspots that appear to be critical for CoV binding.¹⁸ These create positive charges that need to be neutralised by the CoV. Two key residues in the RBM of SARS-CoV-2, Gln493 and Leu455 bind to these hotspots¹⁸ leading to considerable stabilisation of binding and a higher affinity for ACE2 than SARS-CoV-2.

Using the structural information, variations in the ACE2 gene have been studied to identify susceptibility determinants and it is postulated that two ACE2 alleles rs73635825 (S19P) and rs143936283 (E329G) may confer resistance to SARS-Cov-2 infection.³⁶

Other binding determinants that may play a role in the interaction of the spike glycoprotein include glycoproteins based on the differential susceptibility of SARS-CoV-2 relative to SARS-CoV implying that that interaction with ACE2 may not be sufficient for strong attachment to respiratory cells.³⁷ A ganglioside-binding domain has been identified in the N-terminus of the SARS-CoV-2 and this domain may be mimicked by drugs such as chloroquine and hydroxychloroquine.³⁷ Mouse studies have shown that ACE inhibitors and angiotensin receptor blockers increase the expression of ACE2. In humans, antihypertensives such as olmesartan (angiotensin receptor blocker) cause increased urinary soluble ACE2 and diabetes mellitus results in increased circulating ACE2. It is therefore speculated that the increased expression of ACE2 could potentiate infection with SARS-CoV-2. Increased soluble ACE2 may hinder binding of SARS-CoV-2 to cells and attenuate injury to cellular tissues. In engineered human tissues, administration of soluble recombinant ACE2 blocks the growth of SARS-CoV-2.³⁸

In conclusion, the spike protein of SARS-CoV-2 plays a critical and fundamental role in viral infection and in comparison with SARS-CoV the structure has evolved to exploit respiratory cell receptors and proteases to enable rapid spread.

Handling editor Runjan Chetty.

Contributors TSP is the sole author who conceived and wrote the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iD

Tahir S Pillay <http://orcid.org/0000-0002-9982-9710>

REFERENCES

- Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;17:181–92.
- Masters PS. The molecular biology of coronaviruses. *Adv Virus Res* 2006;66:193–292–292.
- de Groot RJ. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj J* 2006;23:59–72.
- Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33.
- 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;395:565–74.
- Chen L, Liu W, Zhang Q, et al. Rna based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg Microbes Infect* 2020;9:313–9.
- Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–3.
- Lai MM, Patton CD, Stohlman SA. Further characterization of mRNA's of mouse hepatitis virus: presence of common 5'-end nucleotides. *J Virol* 1982;41:557–65.
- Wu H-Y, Brian DA. Subgenomic messenger RNA amplification in coronaviruses. *Proc Natl Acad Sci U S A* 2010;107:12257–62–62.
- Chu DKW, Pan Y, Cheng SMS, et al. Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. *Clin Chem* 2020;66:55.
- Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25.
- Walls AC, Park Y-J, Tortorici MA, et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181:281–92.
- Zhang C, Zheng W, Huang X, et al. Protein structure and sequence reanalysis of 2019-nCoV genome refutes snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1. *J Proteome Res* 2020;19:1351–60.
- Zhang T, Wu Q, Zhang Z. Probable Pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol* 2020;30:1346–51.
- Zheng M, Song L. Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. *Cell Mol Immunol* 2020. doi:10.1038/s41423-020-0385-z. [Epub ahead of print: 04 Mar 2020].
- 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;181:271–80.
- Wang Q, Zhang Y, Wu L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 2020. doi:10.1016/j.cell.2020.03.045. [Epub ahead of print: 07 Apr 2020].
- Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020;382.
- Fung TS, Liu DX. Post-Translational modifications of coronavirus proteins: roles and function. *Future Virol* 2018;13:405–30.
- Yan R, Zhang Y, Li Y, et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;367:1444–8.
- Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A* 2009;106:5871–6.
- Millet JK, Whittaker GR. 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;111:15214–9.
- Zhang H, Penninger JM, Li Y, et al. Angiotensin-Converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med* 2020;46:586–90.
- Wong SK, Li W, Moore MJ, et al. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* 2004;279:3197–201.
- Wrapp D, Wang N, Corbett KS, et al. Cryo-Em structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367:1260–3.
- Li Fet al. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005;309:1864–8.
- Li W, Zhang C, Sui J, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *Embo J* 2005;24:1634–43.
- Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 2015;202:120–34.
- Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 2020;11:1620.
- Kuhn JH, Li W, Choe H, et al. Angiotensin-Converting enzyme 2: a functional receptor for SARS coronavirus. *Cell Mol Life Sci* 2004;61:2738–43–43.
- Li W, Moore MJ, Vasilieva N, et al. Angiotensin-Converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;426:450–4.
- Tipnis SR, Hooper NM, Hyde R, et al. A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase. *J Biol Chem* 2000;275:33238–43–43.
- Donoghue M, Hsieh F, Baronas E, et al. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* 2000;87:E1–9.
- 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. doi:10.1038/s41586-020-2180-5. [Epub ahead of print: 30 Mar 2020].
- Hamming I, Cooper ME, Haagmans BL, et al. The emerging role of ACE2 in physiology and disease. *J Pathol* 2007;212:1–11.
- Hussain M, Jabeen N, Raza F, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J Med Virol* 2020. doi:10.1002/jmv.25832. [Epub ahead of print: 06 Apr 2020].
- Fantini J, Di Scala C, Chahinian H, et al. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents* 2020:105960.
- Lei C, Fu W, Qian K, et al. Potent neutralization of 2019 novel coronavirus by recombinant ACE2-Ig. *bioRxiv* 2020.