Check for updates

Scrutinizing Coronaviruses Using Publicly Available Bioinformatic Tools: The Viral Structural Proteins as a Case Study

Sonia Beeckmans* and Edilbert Van Driessche

Research Unit Protein Chemistry, Vrije Universiteit Brussel, Brussels, Belgium

Since early 2020, the world suffers from a new beta-coronavirus, called SARS-CoV-2, that has devastating effects globally due to its associated disease, Covid-19. Until today, Covid-19, which not only causes life-threatening lung infections but also impairs various other organs and tissues, has killed hundreds of thousands of people and caused irreparable damage to many others. Since the very onset of the pandemic, huge efforts were made worldwide to fully understand this virus and numerous studies were, and still are, published. Many of these deal with structural analyses of the viral spike glycoprotein and with vaccine development, antibodies and antiviral molecules or immunomodulators that are assumed to become essential tools in the struggle against the virus. This paper summarizes knowledge on the properties of the four structural proteins (spike protein S, membrane protein M, envelope protein E and nucleocapsid protein N) of the SARS-CoV-2 virus and its relatives, SARS-CoV and MERS-CoV, that emerged few years earlier. Moreover, attention is paid to ways to analyze such proteins using freely available bioinformatic tools and, more importantly, to bring these proteins alive by looking at them on a computer/laptop screen with the easy-to-use but highly performant and interactive molecular graphics program DeepView. It is hoped that this paper will stimulate non-bioinformaticians and non-specialists in structural biology to scrutinize these and other macromolecules and as such will contribute to establishing procedures to fight these and maybe other forthcoming viruses.

Keywords: beta-coronavirus, COVID-19, SARS-CoV-2, SARS-CoV, MERS-CoV, spike protein, pandemic, DeepView molecular graphics program

INTRODUCTION

The year 2020 will always be remembered as "*the year of the pandemic.*" A new type of virus causing severe respiratory illness emerged in December 2019 in Wuhan, China. Since then, it has rapidly spread throughout the entire world, leaving a trail of destruction with high mortality. The pathogen was soon identified to belong to the *Coronaviridae* family, subfamily of the *Coronavirinae*, which is further subdivided in four genera called alpha, beta, gamma and delta (Belouzard et al., 2012; Fehr and Perlman, 2015; Fung and Liu, 2019; Li et al., 2020a). The new virus could be classified as a beta-coronavirus and was found to be closely related to other human beta-coronaviruses that

emerged early in the 21-st century, i.e., SARS-CoV that died out after about one year and MERS-CoV that is still lingering. The new virus was officially named SARS-CoV-2 and the disease it causes is known as Covid-19 (Gorbalenya et al., 2020). The original SARS-CoV-2 virus, which evolved in bats (Andersen et al., 2020; Shereen et al., 2020) like many other alpha- and beta-coronaviruses (Li et al., 2005; Chan et al., 2013; Wang and Anderson, 2019), is easily transmitted from human to human, has an appreciably high reproductive number when no containment measures are taken (R_o = 2–4), a high infection fatality rate (IFR = 0.3–1.3%), and it remains infective for extensive periods of time outside the human body (Bar-On et al., 2020; Rabi et al., 2020). Moreover, it spreads already for several days before an infected person notices the first symptoms of disease because the virus developed several ways to thwart the immune system's response (Astuti and Ysrafil, 2020; Banerjee et al., 2020; Kikkert, 2020). SARS-CoV-2, together with SARS-CoV and MERS-CoV (hereafter referred to as the SARS-CoV-s), are still of great concern because of their worldwide health threat to humans. For this reason, since its first appearance, numerous studies have been conducted to understand its structure, organization, ways of infection, multiplication and pathogenesis. These studies are anticipated to continue guiding us in the development of strategies using antivirals and/or immunomodulators to attenuate the severity of illness in case of infection, and/or to prevent infection through the development of vaccines (Abd Ellah et al., 2020; Capell et al., 2020; Callaway, 2020a; Dai et al., 2020; Dong et al., 2020; Graham, 2020; Hu et al., 2020a; Kaslow, 2020; Krammer, 2020; Li et al., 2020c; Liu et al., 2020; Poland et al., 2020; Riva et al., 2020; Tay et al., 2020), thereby trying to avoid serious problems that may show up such as cytokine storm development, suboptimal antibody response or immune enhancement (Tisoncik et al., 2012; de Alwis et al., 2020; Hotez et al., 2020; Iwasaki and Yang, 2020; Moore and June, 2020; Pedersen and Ho, 2020). Another point of attention should be the prevention of mutational escape of viral proteins that seems to occur following administration of single antibody species (Baum et al., 2020). Such studies are all the more important as it is envisaged that many more SARS-CoV/MERS-CoV-like coronaviruses might be lurking around the corner, ready to jump to and thereafter spread amongst humans following interspecies transmission in the years or decennia to come (Wang and Anderson, 2019; Valitutto et al., 2020).

Abbreviations: ACE, Angiotensin-I-Converting Enzyme; ACE2, Angiotensin-Converting Enzyme-2; ADE, Antibody-Dependent Enhancement; BLI, Bio-Layer Interferometry; CD, Connector Domain; CD26, Cluster of Differentiation-26; CendR, C-end Rule; CH, Central Helix; CLR, C-type Lectin Receptor; Covid-19, Coronavirus disease-2019; CPK, "Corey-Pauling-Koltun", type of atom coloring; CT, Cytoplasmic Tail; CUB, C1r/C1s/Uegf/Bmp1 kind of protein domains; DC/L-SIGN, Dendritic Cell/Liver-Specific ICAM, *inter-cellular adhesion molecule)*-3-Grabbing Non-integrin; DMV, Double Membrane Vesicle; DPP4, DiPeptidyl-Peptidase-4; dsRNA, double-stranded RNA; ELISA, Enzyme-Linked Immuno-Sorbent Assay; EM, Electron Microscopy; EMBL-EBI, the European Molecular Biology Laboratory–European Bioinformatics Institute; ER, Endoplasmic Reticulum; ERGIC, ER-Golgi Intermediate Compartment; ExPASy, Expert Protein Analysis System, the bioinformatics resource portal operated by the Swiss Institute of Bioinformatics; Fab, antibody's antigen-binding Fragment;

AIM OF THE PAPER

This paper presents an overview of the current knowledge of the SARS-CoV-s' structural proteins, on their spatial organization and functional properties, with emphasis on the spike protein. Also, the involvement of the host's own proteins in the development of Covid-19 is considered. Moreover, attention is paid to how antibodies and peptides may help to overcome infections. In the **Supplementary Material** to this paper, we will explore and demonstrate how bioinformatic tools that are freely available on the internet may help students and researchers who are neither trained bioinformaticians nor structural biologists, to understand and visualize macromolecules such as those from the beta-coronaviruses. In view of the overwhelming numbers of structural studies and the continuous increasing availability of data in the protein databank (wwPDB consortium, 2019; Berman et al., 2000), it is definitely an asset to be able to visualize (macro)molecules on a personal computer screen. Although authors do their utmost to present structures they show in publications in optimal orientations and with the most instructive coloring, it is essential to be able to walk around in these structures yourself to gain a much better understanding of these molecules and appreciate their 3D structure and flexibility. Therefore, the **Supplementary Material** will guide the reader within this exciting area, which steadily continues to grow in importance. Playing around with the structural data that are amply available nowadays is becoming a prerequisite to understand complex particles such as the SARS-CoV-s and helps us to deal with them. In fact, looking in detail to the structures of the respective viral components allows us to understand the whole sequence of events that occur during infection and pathogenesis of the virus.

THE GENOME OF THE SARS-COV-S

Coronaviruses are (+)ssRNA (positive-sense single-stranded RNA) viruses with a very large RNA genome, typically

---

Fc, antibody's crystallizable Fragment; FP, Fusion Peptide; GAG, Glycos-Amino-Glycan; Glc, D-Glucose; GlcNAc, N-Acetyl-D-Glucosamine; GRP78/BiP, Glucose-Regulated Protein, 78 kDa/Binding immunoglobulin Protein: two alternative names for an ER-Hsp70 chaperone; HR1, HR2, first, second Heptad Repeat; HS, Heparan-Sulfate; IDR, Intrinsically Disordered Region; K_D, Dissociation constant; mAb, monoclonal Antibody; MAM, Meprin/A5-antigen/ptp-Mu kind of protein domains; Man, D-Mannose; MERS-CoV, Middle-east respiratory syndrome Corona Virus; MGL, Macrophage Galactose-type Lectin; NCBI, National Center for Biotechnology Information, U.S. resource for bioinformatics tools and services; NRP1, Neuropilin-1; Nsp, viral Non-structural protein; NTD, N-Terminal Domain; NxT/S, consensus sequence for N-glycosylation: Asn-x-Thr or Asn-x-Ser; ORF, Open Reading Frame; PBMC, Peripheral Blood Mononuclear Cells; PDB, Protein Data Bank; PDZ, PSD-95/Dlg/ZO-1 kind of protein domains; RBD, Receptor-Binding Domain; RdRp, RNA-dependent RNA polymerase; RMS, Root Mean Square; RNP, Ribonucleoprotein; S1, N-terminal half of Spike protein; S1/S2, first cleavage site; S2, C-terminal half of Spike protein; S2', second cleavage site; SARS-CoV, Severe acute respiratory syndrome Corona Virus; SARS-CoV-2, Severe acute respiratory syndrome Corona Virus-2; SD1, SD2, Structural sub-Domains 1 and 2; Siglec, Sialic acid-binding ImmunoGlobulin-type LECtins; SPR, Surface Plasmon Resonance; SS, Signal Sequence; ssRNA, single-stranded RNA; TM, Trans-Membrane domain; TMPRSS2, Trans-Membrane Protease-Serine-2), UniProtKB, Universal Protein Knowledge Base, database of protein sequences and functional information).

around 30 kb. The viral RNA is packaged inside a spherical membrane (roughly 100–125 nm in diameter) with the help of soluble nucleocapsid proteins (N). Other structural proteins comprise three membrane proteins, i.e., the spike protein (S), the membrane protein (M) (occasionally also called matrix protein) and the envelope protein (E). These four structural proteins occur at a ratio of roughly (E:S:N:M) = (1:5:50:100), according to estimations done for SARS-CoV (Bar-On et al., 2020).

At the 5′ end, the viral RNA contains two large so-called replicase genes (*rep1a*, *rep1b*) organized as two extended open reading frames (5′-ORF-1a/1b), followed by genes that code for the four structural and some accessory proteins (Fehr and Perlman, 2015; Tang et al., 2020; **Figure 1**). As soon as the viral RNA enters a host cell, it acts as an mRNA molecule and hijacks not only the host translational machinery to make all its encoded proteins, but also the host's intricate post-translational modification systems. The open reading frames, 5′-ORF-1a/1b, encode a series of non-structural proteins (Nsps), some of which are enzymes, while others have yet unknown functions (Fehr and Perlman, 2015; Chen et al., 2020). Translation of these two ORFs in the cytoplasm of the host cell results in the synthesis of two long polyproteins (pp1a, pp1b), which are autoproteolytically cleaved by viral proteases. One of the enzymes (Nsp12) is a unique RNA-dependent RNA polymerase (RdRp), responsible for replication of the viral RNA genome (Jiang et al., 2020). Another one (Nsp14) also has an essential role in replication and transcription: it is a bifunctional enzyme with an exoribonuclease domain (ExoN) that, extraordinary for a virus, has proofreading activity and, as such, limits the occurrence of lethal mutations in the viral RNA (Ferron et al., 2017; Romano et al., 2020). Moreover, two important viral proteases are produced from its RNA, i.e., the main protease (M^{Pro}, also called $3CL^{Pro}$) and a papain-like protease (PL^{Pro}), both implicated in processing of the polyproteins (Zhang et al., 2020a). Besides other components, numerous copies of the nucleocapsid protein N are also made in the cytoplasm.
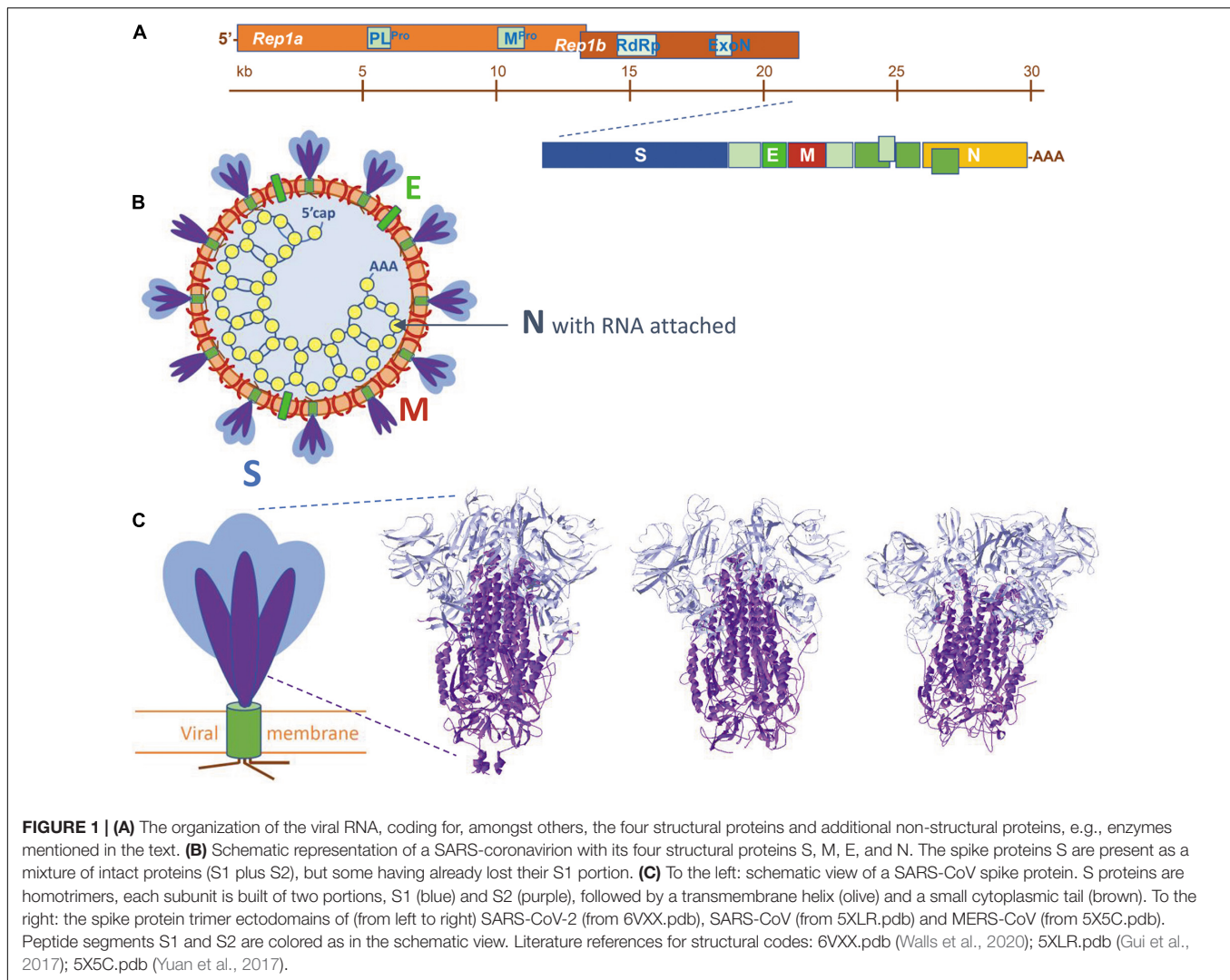
FROM RNA TO MATURE VIRIONS

Upon infection, (+)ssRNA viruses usurp host cell membranes from certain organelles. In SARS-CoV-s, ER membranes are captured from which a complex reticulovesicular network forms that contains double membrane vesicles (DMVs, 200–300 nm) and which remains connected to the ER (Snijder et al., 2020). DMVs act as little factories in which ssRNA is first transformed into dsRNA (an intermediate in multiplication), from which new viral (+)ssRNA molecules are generated. In this way, viral RNA is isolated and shielded from innate immune sensing. This entire process also relies on several viral Nsps that form complexes, many of which have yet to be unraveled. Some Nsps assemble to form pores in the DMV membranes, through which newly made ssRNA molecules are exported into the host cytoplasm (Wolff et al., 2020). During this process they collect N proteins that bind in a "beads-on-a-string" fashion to stabilize the RNA

(Yao et al., 2020). These RNP complexes then travel to virus assembly sites at the ERGIC and/or Golgi complex.

Meanwhile, the viral structural membrane proteins (S, M, E) follow the export route. Only protein S is equipped with a signal peptide to access the ER in the classical way. How the other two reach the ER is not yet clear. However, it is known that some membrane proteins also face the same problem in various organisms (Ott and Lingappa, 2002). The three viral membrane proteins follow the normal flow from ER towards the Golgi apparatus. In this process they are decorated with N-linked glycans, which is essential for proper folding and maturation of the molecules (Zhao et al., 2015). The three proteins assemble in the ERGIC/Golgi membrane and leading to its invagination. RNA-(protein N) complexes enter the pro-virions (Wolff et al., 2020), driven by N-M protein interactions, after which new virions containing S, M, and E proteins, as well as RNA-N complexes, pinch off by ERGIC or Golgi membrane fission. Finally, mature virions are released from the host cell in a non-classical manner. Instead of using the secretory exocytosis pathway, the virus utilizes lysosomes that are deacidified (possibly through the action of the viral protein ORF3a), concomitantly inactivating lysosomal degradative enzymes and disturbing cellular processes including autophagy, pathogen degradation and antigen presentation (Ghosh et al., 2020). During trafficking, the virions are continuously accompanied by KDEL-containing ER-chaperones GRP78/BiP and calreticulin and by the KDEL-receptor, which are also co-released. Hundreds of new virions may be excreted from an infected lung cell, which dies from exhaustion or is eliminated by the host's immune system. The new virions can then infect other cells of the same host or be expelled in the air in droplets or aerosols that may invade another human host.

THE SPIKE PROTEIN (S) ATTACHES THE VIRUS TO HOST CELLS AND MEDIATES INTERNALIZATION

The spike protein is an integral single-pass type-I membrane protein that protrudes in many copies from the outer surface of the virus, contributing its characteristic appearance. Spikes are responsible for binding the virus to a human or animal cell by recognizing specific receptors and, thereafter, for entry of the virus into the host cell. Spike proteins are the major antigenic determinants of the virus and the main targets in numerous active and passive immunization studies (Amanat and Krammer, 2020; de Alwis et al., 2020; Wrapp et al., 2020a). Indeed, there is an urgent need to generate neutralizing antibodies to fight Covid-19 (as Klasse (2014) explains: "*neutralization*" is defined as "*the reduction in viral infectivity by the binding of antibodies to the surface of virions, thereby blocking a step in the viral replication cycle that precedes virally encoded transcription or synthesis*"). Detailed structural studies on the SARS-CoV-s have paved the way to understanding the complexity of the spike proteins and their way of action. Schematically, the spikes are built up as shown in **Figure 1**. Spike proteins are homotrimers.

FIGURE 1 | (A) The organization of the viral RNA, coding for, amongst others, the four structural proteins and additional non-structural proteins, e.g., enzymes mentioned in the text. (B) Schematic representation of a SARS-coronavirion with its four structural proteins S, M, E, and N. The spike proteins S are present as a mixture of intact proteins (S1 plus S2), but some having already lost their S1 portion. (C) To the left: schematic view of a SARS-CoV spike protein. S proteins are homotrimers, each subunit is built of two portions, S1 (blue) and S2 (purple), followed by a transmembrane helix (olive) and a small cytoplasmic tail (brown). To the right: the spike protein trimer ectodomains of (from left to right) SARS-CoV-2 (from 6VXX.pdb), SARS-CoV (from 5XLR.pdb) and MERS-CoV (from 5X5C.pdb). Peptide segments S1 and S2 are colored as in the schematic view. Literature references for structural codes: 6VXX.pdb (Walls et al., 2020); 5XLR.pdb (Gui et al., 2017); 5X5C.pdb (Yuan et al., 2017).
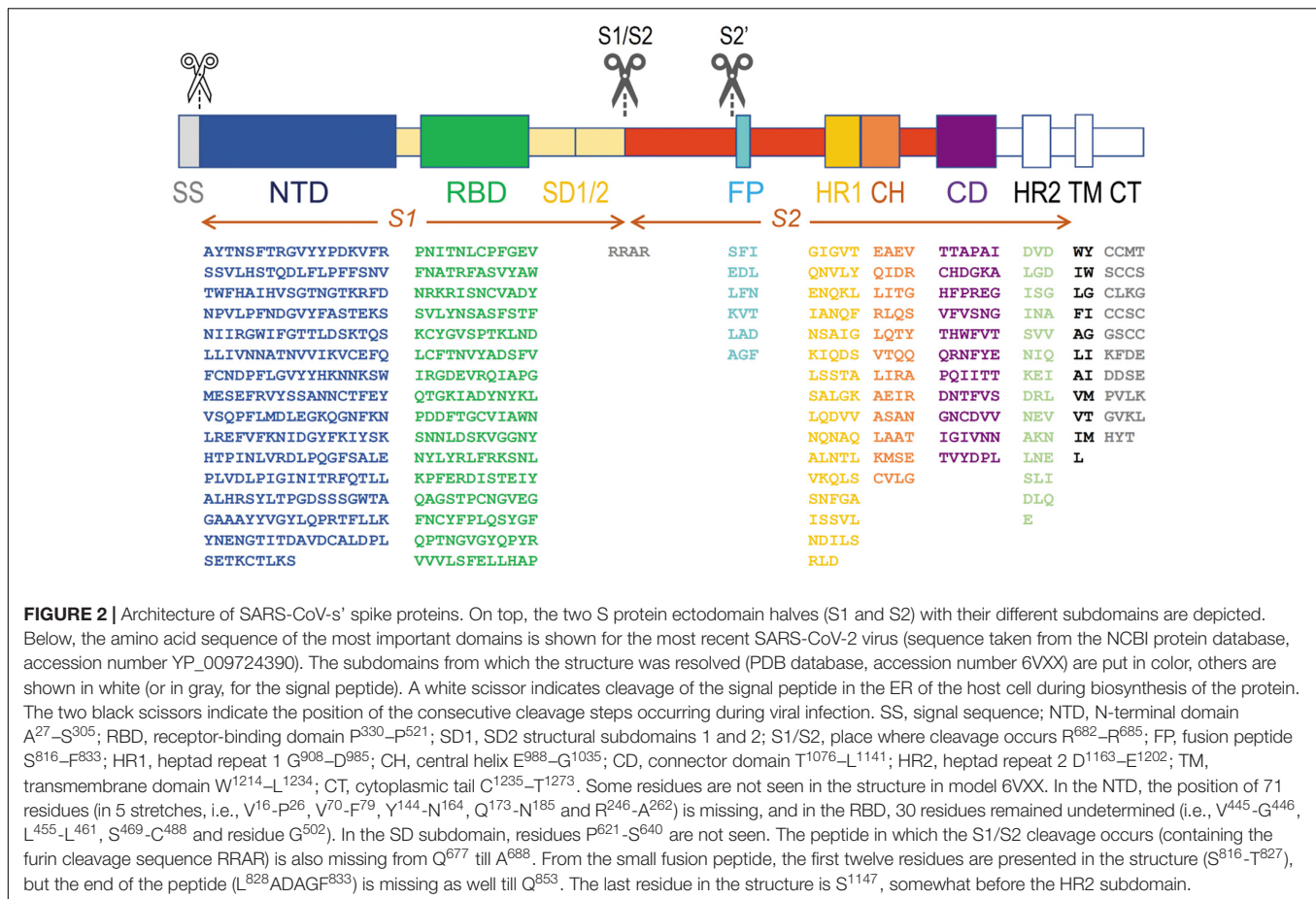
Each monomer has a large ectodomain that consists of several subdomains, followed by a transmembrane domain and an endodomain that contains a series of cysteine residues with palmitoyl chains attached (Petit et al., 2007; McBride and Machamer, 2010; Veit, 2012). S-palmitoylation is a well-known reversible protein post-translational modification (Charollais and van der Goot, 2009; Blaskovic et al., 2013). The ectodomain as well undergoes extensive post-translational modification and becomes heavily glycosylated by the host's N-/O-glycosylation machinery (Fung and Liu, 2018; Watanabe et al., 2020a,b; Yao et al., 2020; Shajahan et al., 2020).

The overall architecture of the SARS-CoV-s is well-understood and has been conserved during evolution of the viruses. It is summarized in **Figure 2**. Although the spike glycoproteins of the three SARS-CoV-s are structurally quite similar, their primary structures differ substantially (**Table 1**, and **Supplementary Material**: **Supplementary Table 1**, **Supplementary Figure 2**). Both the S1 and the S2 domains of the spike protein consist of a series of subdomains, each having a well-defined function. The N-terminal S1 is responsible for receptor binding, while the

C-terminal S2 mediates membrane fusion to facilitate entry of the virus into a host cell.

Overall Appearance of the SARS-CoV-2 Spike Glycoprotein Trimer: The "Pre-Fusion State"

A model of the complete SARS-CoV-2 spike glycoprotein using coordinates from 6VXX.pdb is shown in **Figure 3**. This structure starts at the twelfth residue of the mature polypeptide chain (A^{27}) and ends at residue S^{1147}. A spike protein subunit comprises 1173 residues (including the signal peptide of 15 residues). In the S1 half of the spike protein ectodomain we mostly find beta-strands, but the S2 part mainly consists of long alpha-helices. This figure (**Figure 3**) also indicates the end with which a spike protein trimer is attached to the virion (schematically drawn on top and decorated with multiple spike proteins). The organization of each of the spike protein subunits in different subdomains can be appreciated in **Figure 4**. In this figure, the B-subunit in model 6VXX is colored following the color code used in **Figure 2**. The

FIGURE 2 | Architecture of SARS-CoV-s' spike proteins. On top, the two S protein ectodomain halves (S1 and S2) with their different subdomains are depicted. Below, the amino acid sequence of the most important domains is shown for the most recent SARS-CoV-2 virus (sequence taken from the NCBI protein database, accession number YP_009724390). The subdomains from which the structure was resolved (PDB database, accession number 6VXX) are put in color, others are shown in white (or in gray, for the signal peptide). A white scissor indicates cleavage of the signal peptide in the ER of the host cell during biosynthesis of the protein. The two black scissors indicate the position of the consecutive cleavage steps occurring during viral infection. SS, signal sequence; NTD, N-terminal domain A^{27}–S^{305}; RBD, receptor-binding domain P^{330}–P^{521}; SD1, SD2 structural subdomains 1 and 2; S1/S2, place where cleavage occurs R^{682}–R^{685}; FP, fusion peptide S^{816}–F^{833}; HR1, heptad repeat 1 G^{908}–D^{985}; CH, central helix E^{988}–G^{1035}; CD, connector domain T^{1076}–L^{1141}; HR2, heptad repeat 2 D^{1163}–E^{1202}; TM, transmembrane domain W^{1214}–L^{1234}; CT, cytoplasmic tail C^{1235}–T^{1273}. Some residues are not seen in the structure in model 6VXX. In the NTD, the position of 71 residues (in 5 stretches, i.e., V^{16}-P^{26}, V^{70}-F^{79}, Y^{144}-N^{164}, Q^{173}-N^{185} and R^{246}-A^{262}) is missing, and in the RBD, 30 residues remained undetermined (i.e., V^{445}-G^{446}, L^{455}-L^{461}, S^{469}-C^{488} and residue G^{502}). In the SD subdomain, residues P^{621}-S^{640} are not seen. The peptide in which the S1/S2 cleavage occurs (containing the furin cleavage sequence RRAR) is also missing from Q^{677} till A^{688}. From the small fusion peptide, the first twelve residues are presented in the structure (S^{816}-T^{827}), but the end of the peptide (L^{828}ADAGF^{833}) is missing as well till Q^{853}. The last residue in the structure is S^{1147}, somewhat before the HR2 subdomain.

last residue seen in the structure (S^{1147}) lies just before the second heptad repeat (HR2). The position of an aspartate residue (D^{614}) that very early spontaneously mutated to glycine (see section "The Much-d ebated lucrative spike protein mutant D^{614}G") is indicated as well in chain B.

The Closed (Down) and Open (Up) Conformation of the Spike Protein

The viral spike protein responsible for binding to the host cell is initially in a "pre-fusion" conformation, in search for a receptor on a host cell. Thereby, each of the three subunits exists a certain period of time in a closed (or "down") configuration, a state that is more stable but unable to bind the receptor (see section "A Viral Spike Protein Cannot Bind to the Ace2 Receptor When All Its Subunits Are in Closed Conformation"), and some time in a more unstable open (or "up") configuration, which is receptor accessible. This hinge-like open ↔ closed transition is described in a publication by Wrapp et al. (2020b) and two very instructive videos of the internal movements in the SARS-CoV-2 spike trimer are included in the same publication on-line. When S1 successfully binds to its host cell receptor, the S protein structure becomes unstable and proteolysis may easily occur, resulting in shedding of the N-terminal S1 half of the molecule (S1/S2 cleavage). A second proteolytic cleavage (S2') then takes place

in the SARS-CoV-s, which further removes a long peptide up to just before the fusion peptide (FP), thereby fully exposing this small peptide. The virus is now ready to fuse its own membrane with the host cell membrane (as will be described in section "Events Causing Virus Entry Into Host Cells: The Spike Protein "Post-Fusion" State").

In **Figures 3**, **4**, all spike protein monomers are in the closed conformation. For the SARS-CoV-2 spike protein, a structure is available in which chain B exists in the open form. The two models 6VXX.pdb (all chains closed) and 6VYB.pdb (B-chain open) were uploaded in the database exactly in the same orientation, allowing direct comparison of both structures. Such a superposition of the B-chain in both models is seen in the **Supplementary Material** (**Supplementary Figure 3**). Similar events occur in the SARS-CoV and the MERS-CoV spike proteins. Coordinates may be found as well for MERS-CoV (5X5C.pdb: all chains down, and 5X5F.pdb: B-chain in up conformation).

Comparison of the SARS-CoV-2, SARS-CoV and MERS-CoV Spike Protein Structures

High quality structural details of all three SARS-CoV spike proteins are available. However, each of them was uploaded

TABLE 1 | Properties of the three human Sars-CoV-s' structural proteins S (ectodomain), M, E and N.

	SARS-CoV-2	SARS-CoV	MERS-CoV
Membrane spike protein S			
Identity (NCBI accession number)	YP_009724390	P59594	ASY99778
Ectodomain	V^{16}–P^{1213}	S^{14}–V^{1198}	Y^{18}–W^{1300}
Molecular mass (kDa)	132.924	131.577	141.544
Theoretical pI	6.30	5.56	5.65
Aliphatic index	83.23	82.63	81.61
Protein sequence SARS-CoV-2	–	76.0% *Id* – 17.0% *Si*	26.7% *Id* – 34.3% *Si*
SARS-CoV	76.0% *Id* – 17.0% *Si*	–	27.3% *Id* – 33.9% *Si*
MERS-CoV	26.7% *Id* – 34.3% *Si*	27.3% *Id* – 33.9% *Si*	–
Membrane protein M			
Identity (NCBI accession number)	QIC53216	AAP13444	AGH58718
Molecular mass (kDa)	25.146	25.070	24.552
Theoretical pI	9.51	9.63	9.27
Aliphatic index	120.86	115.06	103.70
Protein sequence SARS-CoV-2	–	89.2% *Id* – 8.1% *Si*	39.1% *Id* – 31.8% *Si*
SARS-CoV	89.2% *Id* – 8.1% *Si*	–	41.8% *Id* – 30.9% *Si*
MERS-CoV	39.1% *Id* – 31.8% *Si*	41.8% *Id* – 30.9% *Si*	–
Membrane envelope protein E			
Identity (NCBI accession number)	P0DTC4	AAP13443	AGH58723
Molecular mass (kDa)	8.365	8.361	9.354
Theoretical pI	8.57	7.01	7.64
Aliphatic index	144.00	145.92	111.59
Protein sequence SARS-CoV-2	–	96.0% *Id* – 4.0% *Si*	34.1% *Id* – 30.5% *Si*
SARS-CoV	96.0% *Id* – 4.0% *Si*	–	34.1% *Id* – 32.9% *Si*
MERS-CoV	34.1% *Id* – 30.5% *Si*	34.1% *Id* – 32.9% *Si*	–
Soluble nucleocapsid protein N			
Identity (NCBI accession number)	P0DTC9	AAP13445	AGG22549
Molecular mass (kDa)	45.625	46.025	44.857
Theoretical pI	10.07	10.11	10.05
Aliphatic index	53.52	49.81	56.08
Protein sequence SARS-CoV-2	–	89.3% *Id* – 8.1% *Si*	46.1% *Id* – 26.3% *Si*
SARS-CoV	89.3% *Id* – 8.1% *Si*	–	44.8% *Id* – 25.8% *Si*
MERS-CoV	46.1% *Id* – 26.3% *Si*	44.8% *Id* – 25.8% *Si*	–

Values on the molecular properties of the viral structural proteins in this table were calculated using the program ProtParam (Gasteiger et al., 2005). For the spike proteins, only the ectodomains are considered. The aliphatic index is calculated as: $X(Ala) + a*X(Val) + b*[X(Ile) + X(Leu)]$, in which X is the mole fraction of each of the 3 amino acids, and a = 2.9 and b = 3.9, i.e., the relative volumes with respect to Ala of Val (a) and Leu/Ile (b), respectively (Ikai, 1980). The primary structures of the proteins were compared using Clustal omega (Sievers et al., 2011) and resemblance between them is indicated under "protein sequence" as percentages identity (Id) and similarity (Si).

in the PDB database in a different orientation, so they first need to be superimposed for comparison. After superposition of the individual RBDs we see that, especially for SARS-CoV-2 and SARS-CoV, the structures are very similar (**Figure 5B**). But, also from SARS-CoV-2 and MERS-CoV, the RBDs can be nicely superposed over an extended part of the domain (**Figure 5C**). Towards the end of the RBD, the protein fold in MERS-CoV starts diverting from the structure seen in SARS-CoV-2. This becomes also understandable when looking at a multiple sequence alignment of the three RBDs (**Figure 5D**). In the first part of the alignment (S^{364} till V^{484}, MERS-CoV numbering), identities and similarities in the sequences are 19.5% and 37.4%, while in the second part (P^{485} till M^{569}) the value for identities and similarities drops to 9.4% and 27.2%, respectively. Moreover, long gaps needed to be introduced to optimize the alignment. However, most of the disulfide bonds in

the RBD are conserved in all three SARS-CoV-s (**Figures 5B,C**). In SARS-CoV-2, the stretch I^{468} till Y^{489} is missing, so the disulfide bond 4 is not evident from this structure. However, it is visualized in other structures of the SARS-CoV-2 RBD (see **Supplementary Material**, **Supplementary Figures 4**, **5**). In MERS-CoV, the corresponding cysteine residue (in between T^{533} and V^{534}) is absent in the sequence. But in this spike protein RBD, C^{526} forms an alternative disulfide bond with C^{503} instead.

Spike Protein Glycosylation

N-glycosylation is a very ancient process that is fully conserved in all eukaryotes. The N-glycosylation process starts in the ER, during biosynthesis of the protein and its co-translational import in the ER lumen, by the covalent attachment of a pre-synthesized $GlcNAc_2$-Man_9-Glc_3 precursor chain (Marth and Grewal, 2008; Aebi et al., 2009, Aebi, 2013;
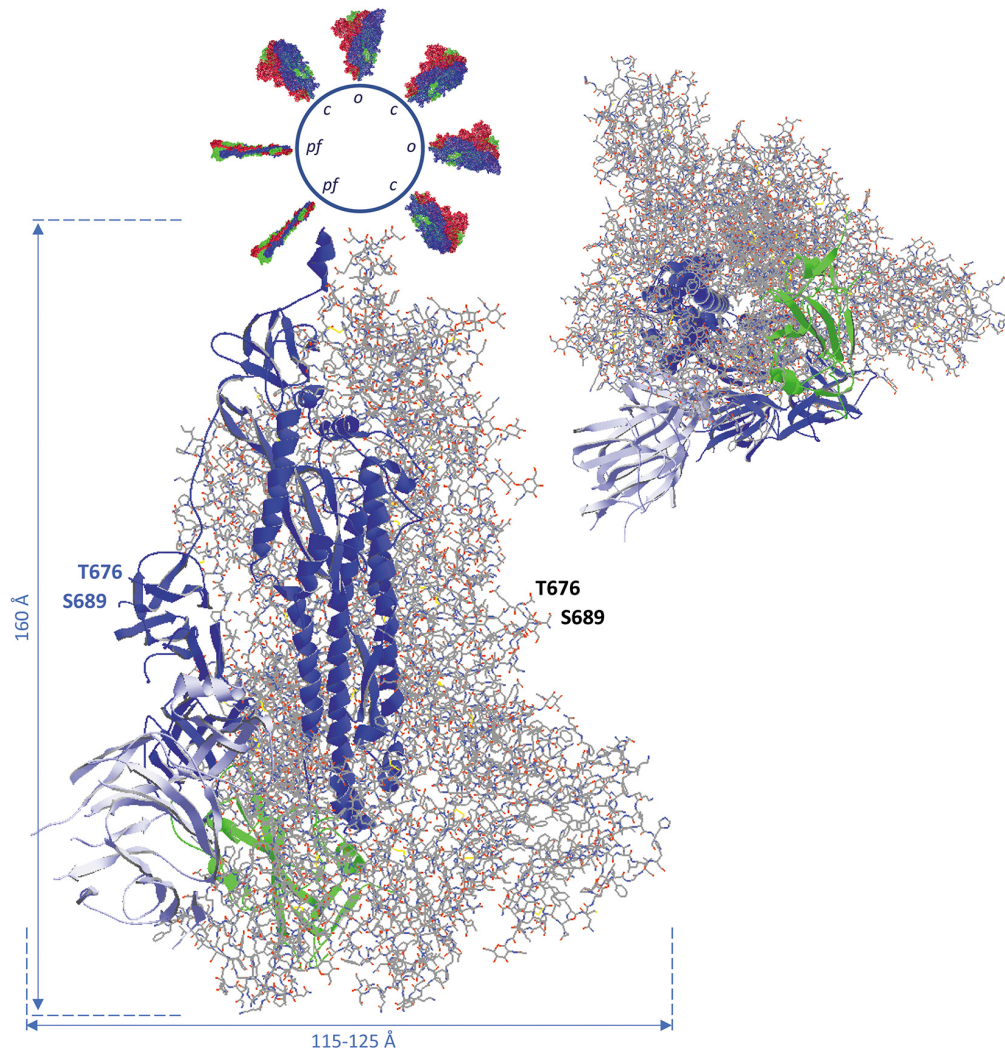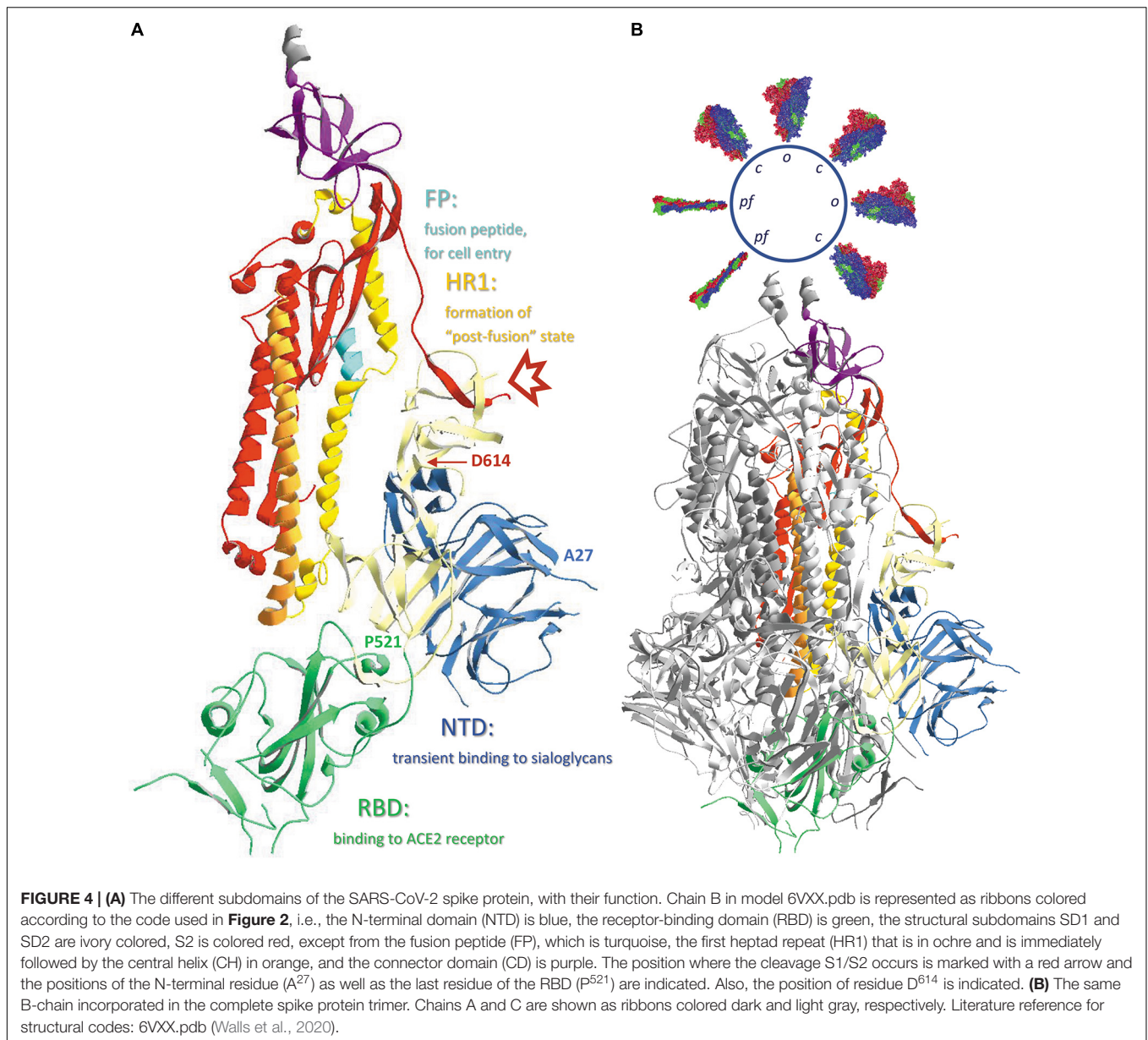
FIGURE 3 | Side view (left) and bottom-to-top view with the virus particle being behind the spike protein (right) of the complete SARS-CoV-2 spike glycoprotein trimer, using coordinates from 6VXX.pdb. All subunits are in the closed (down) position. The B-chain is represented as ribbons, while A and C-chains are shown as backbone with side chains, in CPK colors (explained in the **Supplementary Material**). The B-chain ribbons are colored blue, with the NTD light blue and the RBD green. The location of the position where the activating proteolytic cleavage in the spike protein occurs is indicated for the subunit B (left, in blue) and for subunit A (to the right). The peptide of 12 residues in which cleavage takes place to generate spike protein molecules S1 and S2 (i.e., Q^{677}TNSPRRARSVA^{688}) is missing in the structure, but the flanking residues T^{676} and S^{689} are shown with their side chains and (manually) labeled. In the side view, the virion is at the top. It is represented as a sphere from which other spike proteins emerge. The latter are represented with their surface, either in the pre-fusion state with all subunits in closed state (c) and some of them with one subunit in open configuration (o), each subunit colored differently (chain A red, chain B blue, chain C green), or in the post-fusion state (pf; colored in the same way). Approximate spike dimensions were measured on the model and are indicated. In the open state, the spike protein length increases from 160 to about 175 Å. Literature reference for structural codes: 6VXX.pdb (Walls et al., 2020).

Stanley et al., 2017; Watanabe et al., 2019). After an elaborate quality control procedure in the ER (Ruddock and Molinari, 2006; Słomińska-Wojewódzka and Sandvig, 2015), during which the three glucose residues are removed, the protein that is now decorated with high-mannose chains is transported to the Golgi apparatus by ERGIC (Cop-II vesicles called 'ER-Golgi Intermediate Compartment'). In the subsequent Golgi stacks, which move on by cisternal progression (Luini, 2011), all or some of the high-mannose chains may be enzymatically modified into complex-type or hybrid-type glycans (Strasser et al., 2014; Stanley et al., 2017; Wang et al., 2017).

For SARS-CoV-2 (Watanabe et al., 2020a), as well as for SARS-CoV and MERS-CoV (Watanabe et al., 2020b) the extent of N-glycosylation, as well as the maturation of the glycans in each of the NxT/S glycosylation sites, was thoroughly investigated. All three SARS-CoV-s' spike proteins are heavily glycosylated over the whole length of their subunits. A SARS-CoV-2/SARS-CoV subunit has 22 potential N-glycosylation sites, while a MERS-CoV subunit has 23. All potential sites

FIGURE 4 | (A) The different subdomains of the SARS-CoV-2 spike protein, with their function. Chain B in model 6VXX.pdb is represented as ribbons colored according to the code used in **Figure 2**, i.e., the N-terminal domain (NTD) is blue, the receptor-binding domain (RBD) is green, the structural subdomains SD1 and SD2 are ivory colored, S2 is colored red, except from the fusion peptide (FP), which is turquoise, the first heptad repeat (HR1) that is in ochre and is immediately followed by the central helix (CH) in orange, and the connector domain (CD) is purple. The position where the cleavage S1/S2 occurs is marked with a red arrow and the positions of the N-terminal residue (A^{27}) as well as the last residue of the RBD (P^{521}) are indicated. Also, the position of residue D^{614} is indicated. **(B)** The same B-chain incorporated in the complete spike protein trimer. Chains A and C are shown as ribbons colored dark and light gray, respectively. Literature reference for structural codes: 6VXX.pdb (Walls et al., 2020).

are also effectively glycosylated (though some scarcely not to their full extent), as is described in two publications by Watanabe et al. (2020a,b). In these two papers, the authors investigated the extent of glycan maturation in all these sites by analyzing glycopeptides with mass spectrometry. In each of the glycosylation positions, the glycan chains turned out to have been enzymatically modified in the Golgi apparatus to different extents, leading to very heterogeneous combinations of high-mannose-type ($Man_9GlcNAc_2$ to $Man_5GlcNAc_2$) and hybrid- and complex-type of glycan chains containing different numbers of antennae (A1 to A4), some of which are sialylated, with or without core fucosylation. It is worthwhile mentioning that most of the structural data available on the SARS-CoV-s' spike trimers used proteins that were expressed in insect cells (details in **Supplementary Material Supplementary Table 3**), which

produce glycan structures that differ from those in mammalian cells (Marth and Grewal, 2008). However, none of the pdb files shows any trace of covalently linked glycans. Besides being very heterogeneous, N-glycans are also extremely flexible structures, so that, except in a very few exceptional cases, they can only be modeled into a protein structure. Such modeling studies have been done for the SARS-CoV-2, SARS-CoV and MERS-CoV spike glycoproteins (Walls et al., 2019, 2020; Casalino et al., 2020; Grant et al., 2020; Vankadari and Wilce, 2020; Watanabe et al., 2020a,b; Zhou et al., 2020).

In order to at least visualize where the N-glycans are attached to the spike proteins, the same subunits from SARS-CoV-2, SARS-CoV and MERS-CoV that were superposed and displayed in **Figure 5A** were used, to make sure that we are looking to the three subunits in the same orientation for comparison. **Figure 6**

A

S2

S1

B

C

D
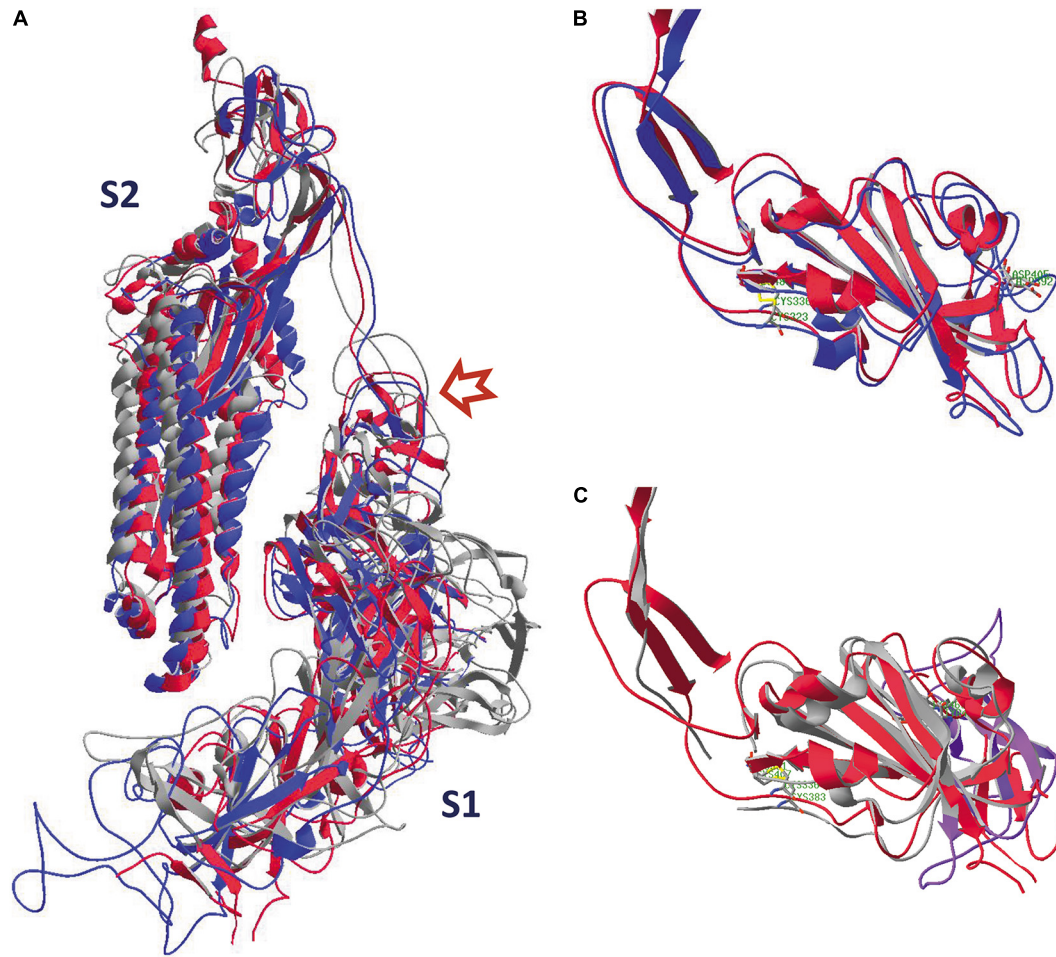
```
          1                            1
Sars_CoV_2  SNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFS   375
Sars_CoV    SNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFS   362
Mers_CoV    SSFEAKPSGSVVEQAE-GVECDFSPLLSG-TPPQVYNFKRLVFTNCNYNLTKLLSLFSVN   421
            *.*.. *: .:*. :    * *. ::..   .** ::* ::**  : : *. ..

            2         3                                   2
Sars_CoV_2  TFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIA   435
Sars_CoV    TFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLA   422
Mers_CoV    DFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNPTCLIL   481
            *.*  :* : : . *::.:  * *   .   ::: ..:* *:::***  .   *::

                          4b                       4            4a
Sars_CoV_2  WNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFP----   491
Sars_CoV    WNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTP-PALNCYWP----   477
Mers_CoV    ATVPHNLTTITKPLKYSYINKCSRLLSDDRTEVPQLVNANQYSPCVSIVPST-VWEDGDY   540
            . : :. :* * *    .  .::   .** .   :

                                                  3
Sars_CoV_2  -------LQSYGFQPTNGVGYQPYRVVVLSFELL----HAPATVCGPKKSTNLVKNKCVNFN   542
Sars_CoV    -------LNDYGFYTTTGIGYQPYRVVVLSFELL----NAPATVCGPKLSTDLIKNQCVNFN   528
Mers_CoV    YRKQLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVC-PKLEFA           591
            *:. *:  :.*    . : :.* :    *  :** **  :.
```

FIGURE 5 | **(A)** One single chain is displayed of each of the spike protein trimer ectodomains from SARS-CoV-2 (6VXX.pdb), SARS-CoV (5XLR.pdb) and MERS-CoV (5X5C.pdb), after superposition of the three models (as explained in the **Supplementary Material**). They have been given different colors, i.e. red (SARS-CoV-2), blue (SARS-CoV) and gray (MERS-CoV). The upper left part shows the S2 halves of the chains, at the lower right are the S1 halves. The red arrow points to the place where the cleavage S1/S2 occurs. **(B)** The RBDs of one subunit of the SARS-CoV-2 (red) and SARS-CoV (blue) spike proteins were superposed, starting from the previous picture (**Figure 5A**). Side chains of residues C^{336} and C^{361} in SARS-CoV-2 (forming a disulfide bridge) and of residue D^{405} are also shown. They correspond with residues C^{323} and C^{348} (also forming a disulfide bridge) and D^{392} in SARS-CoV. These residues are labeled in green. **(C)** The RBDs of one subunit

(Continued)

FIGURE 5 | Continued

of the SARS-CoV-2 (red) and MERS-CoV (gray) spike proteins were superposed. Side chains of residues C^{336} and C^{361} in SARS-CoV-2 (forming a disulfide bridge) and of residue Y^{423} are shown as well. They correspond with residues C^{383} and C^{407} (also forming a disulfide bridge) and Y^{469} in MERS-CoV. These residues are labeled in green. The ribbons in the RBD stretch where major structural differences are observed with the SARS-CoV-2 protein (i.e., from V^{484} till M^{569} in the MERS-CoV sequence) are colored pink. The SARS-CoV-2 RBD is shown in figures B and C in roughly the same orientation. The SARS-CoV-2 RBD is shown from residue S^{316} till N^{544} in both figures, the SARS-CoV RBD from N^{304} till N^{528}, and the MERS-CoV RBD from S^{364} till C^{585}. A limited number of residues is missing in the three structures. **(D)** Multiple sequence alignment of the RBDs of the three spike protein sequences using Clustal omega (Sievers et al., 2011). The sequence where the structure of the MERS-CoV RBD diverts from that of the SARS-CoV and SARS-CoV-2 RBDs is shown in pink letters. The tyrosine that is conserved in the three RBDs, as well as the aspartate residue conserved in SARS-CoV and SARS-CoV-2, are highlighted in green. Disulfide bond cysteine residues are highlighted in yellow and numbers above them indicate their covalent interaction; 4a and 4b refer to different disulfide bond formation in SARS-CoV/SARS-CoV-2 and in MERS-CoV, respectively. Residues that are missing in the structures are in gray and underlined. Literature references for structural codes: 6VXX.pdb (Walls et al., 2020); 5XLR.pdb (Gui et al., 2017); 5X5C.pdb (Yuan et al., 2017).

---

shows that the glycosylation sites are nicely spread over the whole surface of the three SARS-CoV-s. Keeping in mind that N-glycans are very voluminous but also very flexible structures, it is obvious that the spike glycoproteins will be extremely well covered by an extensive glycan coat that may act as a real shield. Besides being important for protein folding and/or stability, this glycan coat might also hide epitopes and prevent antibodies from binding. Therefore, it may interfere with the host's immune defense mechanisms. Finally, the published observations (Watanabe et al., 2020a,b) that many of the N-glycan chains are complex-type in nature tell us that the spike glycoproteins should have passed through the different stacks of the Golgi apparatus, where the glycan modifying enzymes are located in the correct order.

SARS-CoV/SARS-CoV-2 and MERS-CoV Bind to Different Host Cell Receptors

Upon infection with SARS-CoV-s, the skin and mucosal membranes form the first layer of defense. Through the nose, eyes or mouth, the virus reaches the respiratory system where it may recognize a receptor on the surface of lung cells. The spike protein RBD (receptor-binding domain) is responsible for recognition and attachment to a host cell. The receptor for both the SARS-CoV-2 and the SARS-CoV spike proteins was identified to be ACE2 (angiotensin-converting enzyme-2), while the receptor for MERS-CoV is DPP4 (dipeptidyl-peptidase-4) (Belouzard et al., 2012; Fehr and Perlman, 2015; Li, 2016; Skariyachan et al., 2019; Hoffmann et al., 2020; Letko et al., 2020; Matheson and Lehner, 2020; Walls et al., 2020). DPP4, also known as CD26, is a single-pass type-II transmembrane protein of 766 residues with a very extended (738 residues) C-terminal ectodomain that is N-glycosylated. Thanks to its dipeptidyl-peptidase activity, it acts as a regulator of numerous physiological processes.

Angiotensin-Converting Enzyme-2 (ACE2) on the other hand is a proteolytic enzyme acting on angiotensin-I and -II, as well as on some other vasoactive peptides, and it is a regulator of blood pressure. It is a single-pass type-I membrane protein (805 residues) with an extended N-terminal ectodomain of 723 residues. Human ACE2 has six potential N-glycosylation sites and is heavily N-glycosylated (Warner et al., 2004). ACE2 is expressed in many different organs (Xu et al., 2020a), which might contribute to the damage to organs other than lungs in some Covid-19 patients. For SARS-CoV-s, it was observed that co-expression of a plasma cell membrane-anchored surface protease, TMPRSS2 (transmembrane protease-serine-2), highly
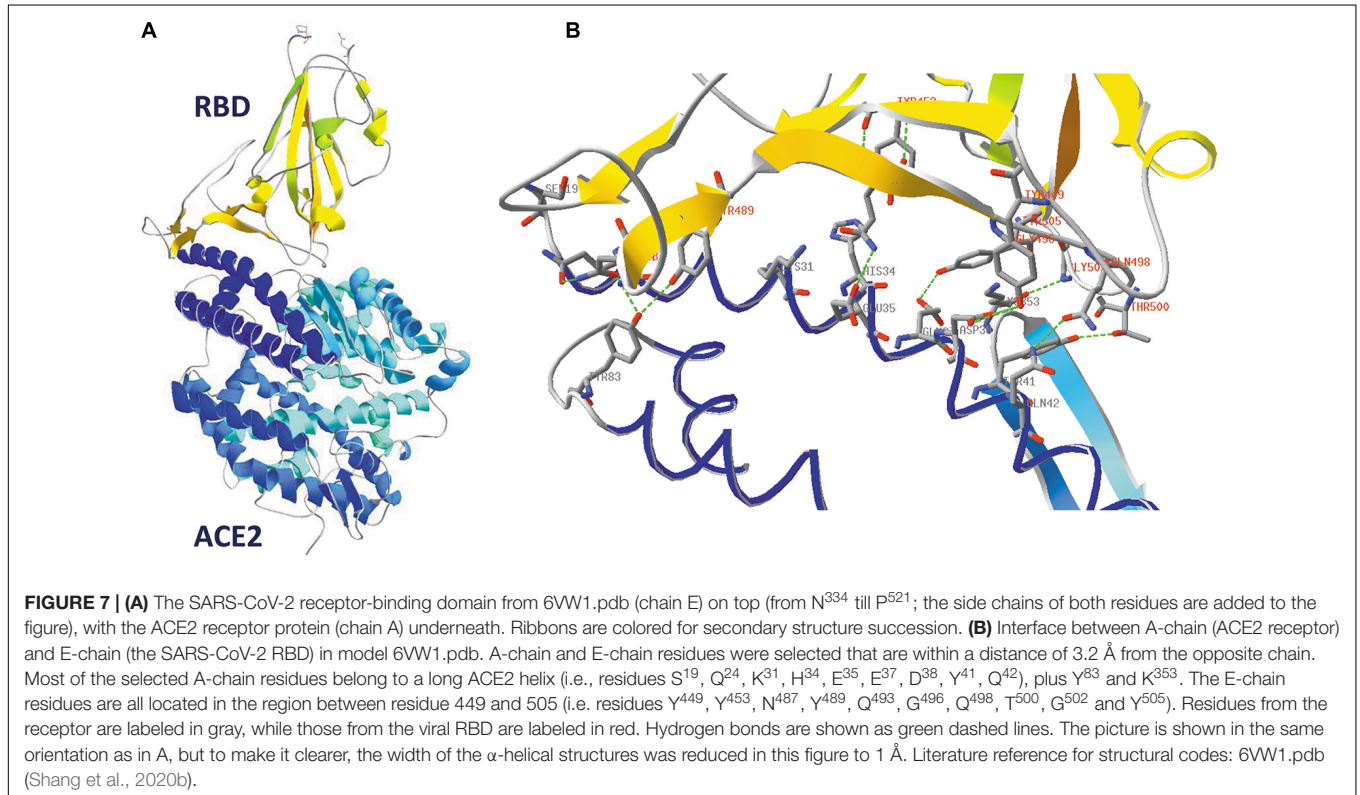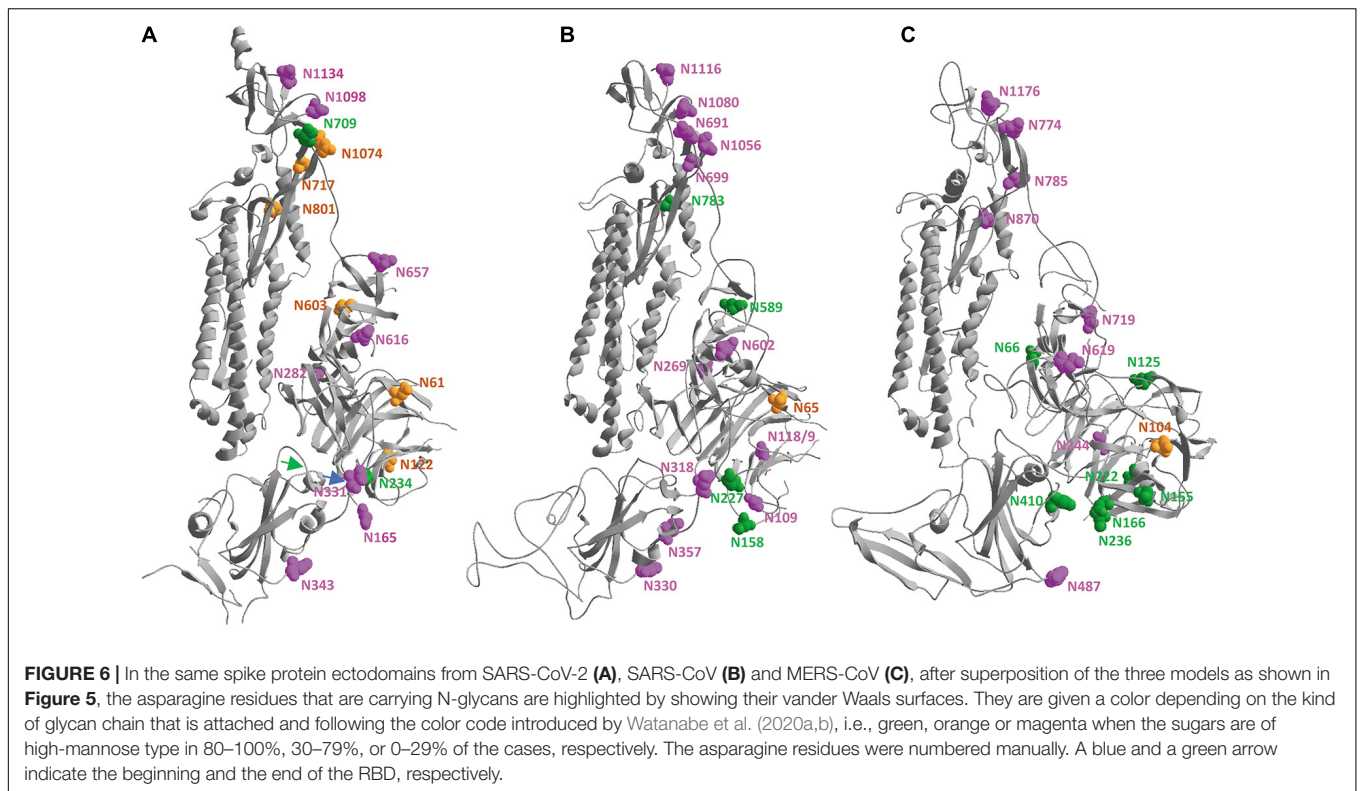
facilitates cellular uptake of the virus (Hofmann and Pöhlmann, 2004; Heurich et al., 2014). Co-expression of ACE2 and the protease TMPRSS2 occurs not only in lung cells but in many different other cell types as well (Sungnak et al., 2020; Ziegler et al., 2020). TMPRSS2 is a single-pass type-II membrane protein of 492 residues with a C-terminal serine protease domain and seems to be involved in various physiological and pathological processes (Thunders and Delahunt, 2020). Its expression is developmentally regulated and increases with aging, which may contribute to the enhanced susceptibility of the elderly to SARS-CoV-2.

For SARS-CoV-s, a correlation has been observed between the grade of infection and the virus load received (Magleby et al., 2020), but also with the affinity of the viral RBD for the host receptor (Ou et al., 2020). It was shown that for SARS-CoV-2 the affinity is very high (with K_D-values in the nM range), which contributes to the severity of the symptoms (Wrapp et al., 2020b). These affinities cannot be assessed from structural data but need to be measured by other experimental techniques, mostly based on ELISA or biosensor type of technologies (e.g., SPR or BLI) (for some examples: Huo et al., 2020; Ju et al., 2020; Pinto et al., 2020; Shang et al., 2020b; Shi et al., 2020; Walls et al., 2020; Wrapp et al., 2020a).

Binding of SARS-CoV-2 to Its ACE2 Receptor

A picture of SARS-CoV-2 RBD binding to its human receptor ACE2 can be made using the structure coordinates in model 6VW1.pdb. **Figure 7A** shows the SARS-CoV-2 RBD on top, with the ACE2 receptor molecule underneath. This figure shows both chains as ribbons colored for secondary structure succession, which enables to easily follow the progression in both chains from N- to C-terminus (alpha-helices and beta-strands are colored following rainbow colors and starting from blue to red, while loops are left gray).

When looking in detail to the interface between the two molecules (**Figure 7B**), it is clear that the RBD residues making contact with ACE2 are all located in the region from residue Y^{449} till Y^{505}. This is precisely that part of the RBD that is very different in MERS-CoV (**Figure 5**) and explains why MERS-CoV is not using ACE2 as receptor molecule. All the residues lining the contact surface between SARS-CoV-2 and its ACE2 receptor are shown in **Figure 7B** (and **Supplementary Figure 6**, see **Supplementary Material**), together with hydrogen bonds

FIGURE 6 | In the same spike protein ectodomains from SARS-CoV-2 **(A)**, SARS-CoV **(B)** and MERS-CoV **(C)**, after superposition of the three models as shown in **Figure 5**, the asparagine residues that are carrying N-glycans are highlighted by showing their vander Waals surfaces. They are given a color depending on the kind of glycan chain that is attached and following the color code introduced by Watanabe et al. (2020a,b), i.e., green, orange or magenta when the sugars are of high-mannose type in 80–100%, 30–79%, or 0–29% of the cases, respectively. The asparagine residues were numbered manually. A blue and a green arrow indicate the beginning and the end of the RBD, respectively.

FIGURE 7 | (A) The SARS-CoV-2 receptor-binding domain from 6VW1.pdb (chain E) on top (from N^{334} till P^{521}; the side chains of both residues are added to the figure), with the ACE2 receptor protein (chain A) underneath. Ribbons are colored for secondary structure succession. **(B)** Interface between A-chain (ACE2 receptor) and E-chain (the SARS-CoV-2 RBD) in model 6VW1.pdb. A-chain and E-chain residues were selected that are within a distance of 3.2 Å from the opposite chain. Most of the selected A-chain residues belong to a long ACE2 helix (i.e., residues S^{19}, Q^{24}, K^{31}, H^{34}, E^{35}, E^{37}, D^{38}, Y^{41}, Q^{42}), plus Y^{83} and K^{353}. The E-chain residues are all located in the region between residue 449 and 505 (i.e. residues Y^{449}, Y^{453}, N^{487}, Y^{489}, Q^{493}, G^{496}, Q^{498}, T^{500}, G^{502} and Y^{505}). Residues from the receptor are labeled in gray, while those from the viral RBD are labeled in red. Hydrogen bonds are shown as green dashed lines. The picture is shown in the same orientation as in A, but to make it clearer, the width of the α-helical structures was reduced in this figure to 1 Å. Literature reference for structural codes: 6VW1.pdb (Shang et al., 2020b).

that are formed between them. It is striking that it is precisely the RBD residue Y^{489}, which forms a hydrogen bond with H^{83} from the ACE2 receptor, that was tentatively identified as one of the most mobile residues in the viral spike protein (see section "Flexibility in the Spike Glycoprotein").

A Viral Spike Protein Cannot Bind to the ACE2 Receptor When All Its Subunits Are in Closed Conformation

In order to visualize that a spike protein subunit needs to be in open conformation before it can bind the ACE2 receptor, an overlay was made between the spike protein RBD in model 6VW1.pdb and the RBD of one of the subunits of the spike protein trimer in model 6VYB.pdb. **Figure 8A** shows the full spike protein trimer that has one of its subunits in open state (chain B, of which the ribbons are colored blue), with a ACE2 receptor molecule (colored orange) bound to it. No clashes occur between the receptor and any of the spike protein subunits in this conformation. However, when the same exercise was performed using model 6VXX.pdb, which has all its subunits in closed state, obviously many clashes occur between the receptor molecule and spike protein subunits, as evidenced in **Figure 8B**. Clashing residues are highlighted in more detail in **Figure 8C**, where they are made visible as pink dashed lines.

Flexibility in the Spike Glycoprotein

Another way of representing the spike protein is by coloring the model for "B-factor". As discussed by T.E. Creighton already in 1993, B-factors (alternatively called temperature factors, or atomic displacement parameters, or Debye-Waller factors) describe the displacement of an atomic position from its average or mean position (Sun et al., 2019). B-factors (expressed in $Å^2$) tell us, for each of the atoms in the model, how well determined and steady their position is. Several studies suggested that, in high quality models, B-factors might be used to identify flexibility and mobility in proteins, proposing that high B-factors indicate higher than average flexibility as opposed to low B-factors, which are believed to occur at more rigid positions.

Supplementary Figure 7 (see **Supplementary Material**) shows the SARS-CoV-2 spike protein trimer colored for B-factor. The region where we find residues having the highest B-factors corresponds to the receptor-binding domain, i.e., from residue P^{330} till P^{521} (from light green up to orange). The residue with the highest B-factor in each subunit is Y^{489}. This high flexibility in the RBD may be thought to greatly assist the spike protein in finding a receptor on a host cell.

Spectacular flexibility in the spike protein is indeed described in two papers. One study (Ke et al., 2020) is based on cryo-EM and tomography to investigate the distribution of spike protein trimers and their flexibility using virus-infected VeroE6 and Calu-3 cells. Roughly 24 spike trimers were seen per virion, of which 97% in pre-fusion (about 31% of them with all monomers in closed state) and only 3% in post-fusion state (this state is described in section "Events Causing Virus Entry Into Host Cells: The Spike Protein "Post-Fusion" State"). The study showed

that protruding spike proteins can extensively be tilted (up to 90°) towards the viral membrane. They seem to be rather sparsely but evenly distributed, without clustering, occurring at a density of about 1 trimer per 1,000 nm^2 of membrane surface. Based on these calculations, it was hypothesized that multiple binding to ACE2 receptors, leading to avidity, will be an exception rather than the rule. A second study (Turoňová et al., 2020) is based on cryo-electron tomography, combined with molecular dynamics simulation. It shows that, in the pre-fusion state, the spike protein is extremely mobile and its stalk contains three hinges that were coined hip, knee and ankle (with estimated flexibilities of 16.5° ± 8.8°, 23° ± 11.7 and 28° ± 10.2°, respectively). This is assumed to give the head of the spike protein a lot of freedom and helps it to accurately scan the host cell for ACE2 receptors. Contrarily, in the post-fusion state the structure is apparently inflexible. Linked to this publication, a video demonstrates the pronounced flexibility in the pre-fusion state[1]. The extreme structural adaptability of the SARS-CoV-2 spike protein is also visualized in a publication presenting at least ten structures and transition phases occurring over the course of ACE2 binding and priming of the protein for membrane fusion (Benton et al., 2020).

Endocytosis Is an Alternative Way to Enter Host Cells

SARS-CoV-s (but also other coronaviruses) may also invade a host cell by an alternative mechanism based on clathrin-mediated endocytosis. In this case, not only is the virus internalized but also the ACE2 receptor protein, which may lead to serious secondary effects due to reduced ACE2 activity (Delpino and Quarleri, 2020; Gheblawi et al., 2020; Lanza et al., 2020; Magalhaes et al., 2020; Ni et al., 2020; Samavati and Uhal, 2020; Saponaro et al., 2020). This occurs when there is no protease available nearby at the cellular surface to perform the necessary proteolytic cleavage into S1 and S2 (Heurich et al., 2014; Fung and Liu, 2019). Endocytosis is then followed by delivery of the virus inside an early endosome, which evolves towards a late endosome and finally towards a lysosome (Neefjes et al., 2017). The required proteolysis of their spike proteins then takes place in the context of either of these organelles, depending on which protease is able to perform the cleavage. Proteolysis leads to fusion of the virion (which is now inside the organelle) with the membrane of the respective organelle, followed by delivery of the viral RNA in the host cell's cytoplasm by the same mechanism, involving the regions HR1 and HR2, together with the FP, as is explained in section "Events Causing Virus Entry Into Host Cells: The Spike Protein "Post-Fusion" State" (Burkard et al., 2014). Host proteolytic enzymes of the cathepsin family, comprising aspartic as well as cysteine and serine proteases with a broad substrate specificity, are occasionally mentioned to help priming SARS-CoV-s for membrane fusion, though this is sometimes disputed (Turk et al., 2012; Patel et al., 2018). Also, other viruses have been

---

[1]https://www.pei.de/EN/newsroom/press-releases/year/2020/14-new-findings-spike-protein-sars-cov-2.html
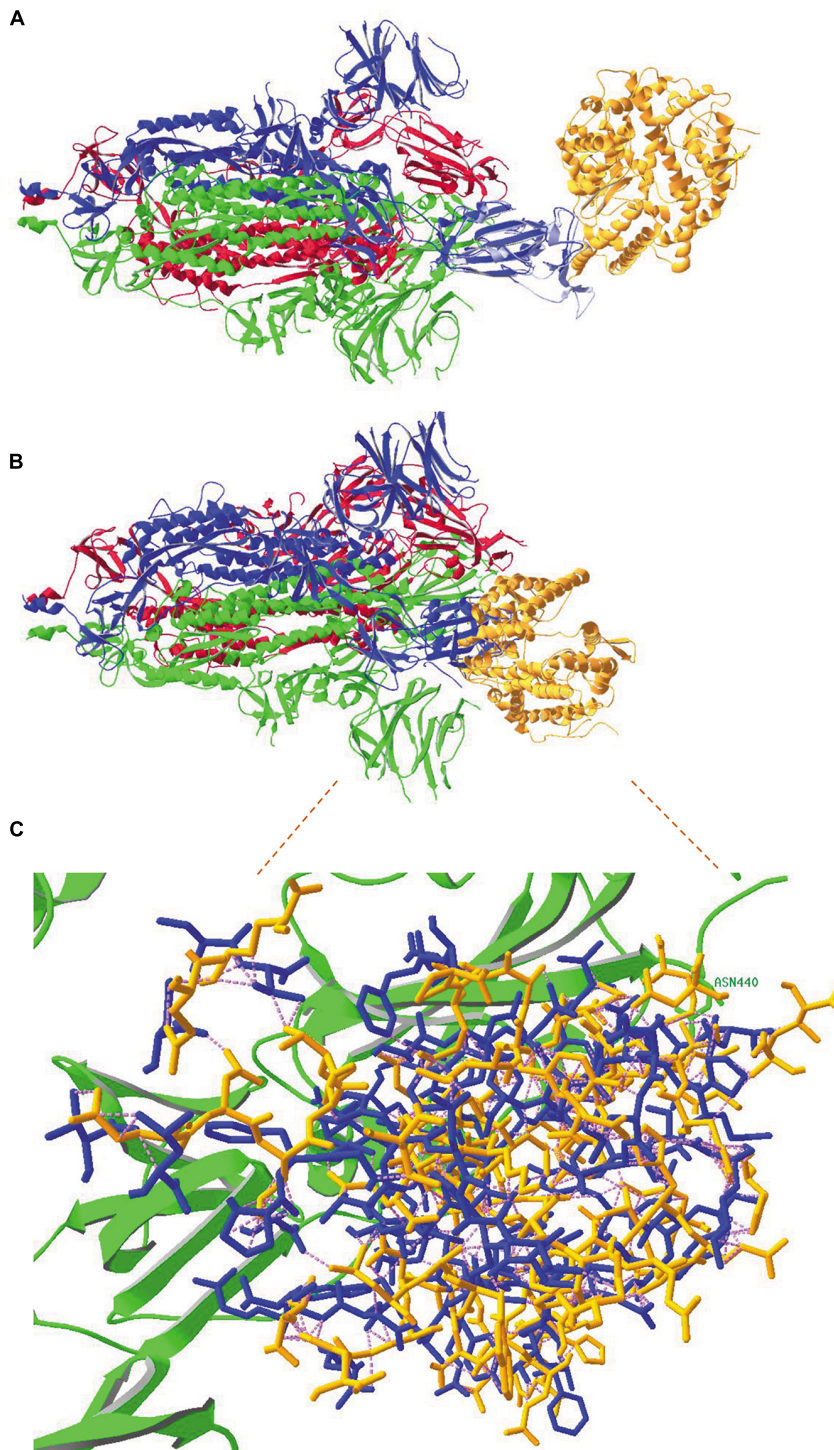
FIGURE 8 | (A) An overlay was made between the SARS-CoV-2 RBD in model 6VW1.pdb (which shows binding of the virus to its ACE2 receptor) and the viral RBD of subunit B in model 6VYB.pdb (showing the spike protein trimer, subunits A and C in closed conformation and subunit B in open conformation). This figure shows the complete spike protein trimer (chains A, B and C with ribbons colored red, blue and green, respectively). The ribbons of the RBD in model 6VW1 are colored light blue, to demonstrate the overlay with the RBD from chain B in 6VYB. The ACE2 receptor protein is colored orange. There are obviously no clashes here. **(B)** An overlay was made between the SARS-CoV-2 RBD in model 6VW1.pdb (which shows binding of the virus to its ACE2 receptor) and the viral RBD of subunit A in model 6VXX.pdb (showing the spike protein trimer, all subunits in closed conformation; ribbons of chains A, B and C colored red, blue and green, respectively). Extensive clashes occur between the ACE2 receptor (colored orange), especially with the RBD of chain B. **(C)** A more detailed view of the clashes between the ACE2 receptor and the spike protein. The residues that are clashing are shown with their backbones and side chains, colored blue when they belong to chain-B and green when belonging to the C-chain of the spike protein (only one residue: N[440], labeled), and orange when they belong to the ACE2 receptor. Clashes appear as pink dashed lines. How to make figures demonstrating the occurrence of clashes is explained in the **Supplementary Material** (**Supplementary Figure 13**). Literature references for structural codes: 6VW1.pdb (Shang et al., 2020b); 6VYB.pdb (Walls et al., 2020); 6VXX.pdb (Walls et al., 2020).

reported to rely on cathepsins at some stages of their life cycle (Brix, 2018).

The Role of Spike Protein Cleavage by Furin and of Neuropilin-1 for SARS-CoV-2 Cellular Uptake

Contrary to all other coronaviruses known so far, SARS-CoV-2 acquired a furin-cleavage sequence, right within the peptide where cleavage occurs in the spike protein to remove the S1 half of the subunits (**Figure 2**). Furin is a single-pass type-I membrane protein that is ubiquitously expressed in vertebrates and has serine endoprotease activity. It cleaves at doublets or clusters of basic amino acids (e.g., KR↓ and RR↓), Rx(K/R)R↓ being the canonical cleavage sequence (Thomas, 2002). It occurs in the trans-Golgi network (where it cycles between sorting compartments), but also at the cell surface and in early endosomes, i.e., at all locations where the virus might pass by at the onset of infection.

Proteolytic cleavage of SARS-CoV-2 spike by furin results in exposure of the $R^{682}RAR^{685}$ sequence at the C-terminus of S1, thereby converting S1 into a so-called C-end-Rule (CendR) peptide. CendR peptides (conform to a R/KxxR/K motif, where the spacing of the basic residues is important), but not their cryptic motifs, are known to bind to neuropilin-1 (NRP1) and then get internalized, together with molecular structures that are attached to them, by a mechanism similar to but different from endocytosis (Teesalu et al., 2009). NRP1 is an essential pleiotropic surface receptor, present on endothelial and epithelial cells, acting as co-receptor molecule (Parker et al., 2012; Wild et al., 2012; Kumanogoh and Kikutani, 2013; Guo and Vander Kooi, 2015). It is a single-pass type-I membrane protein with five ectodomains essential for ligand-binding (two CUB domains, followed by two coagulation factor-homology domains, the first of which has a binding pocket for peptides with C-terminal arginine, and one MAM domain) and its short cytoplasmic domain interacts with PDZ-domain proteins. It is believed that, after furin cleavage but before further priming the spike by the secondary S2' proteolytic step, S1 and S2 temporarily remain associated, giving time to the cleaved spike subunit to bind to NRP1. Two publications describe experiments showing that mAbs directed against NRP1, as well as a small molecule binding in the CendR pocket of NRP1, reduce SARS-CoV-2 infectivity, and also a mutant lacking the original furin cleavage site is less infective (Cantuti-Castelvetri et al., 2020; Daly et al., 2020). NRP1 can thus be considered as an important host factor facilitating cell entry of SARS-CoV-2 and explaining its enhanced infectivity when compared to SARS-CoV and MERS-CoV.

The Potential Implication of the Glycocalyx and/or Host Lectins at the Onset of a Coronavirus Infection

When a virus invades a host, the first structure it encounters is the glycocalyx, a 50–200 nm thick layer made up as an intricate network of glycoproteins (N- and/or O-glycosylated) and proteoglycans (containing glycosaminoglycans, or GAGs) that are covalently attached to the outer surface of the plasma membrane, either by means of transmembrane domains or through GPI-anchors (Koehler et al., 2020). The glycan chains of these glycoproteins are often decorated with terminal sialic acids, while the GAGs contain extended chains of heparan-, chondroitin- or keratan-sulfate, these building blocks being heavily negatively charged. Numerous viruses are known to interact with these charged glycans, though in general with low affinity (K_D-values in the mM range), leading to substantial binding strengths through multivalency. It is assumed that those initial interactions, mostly electrostatic in nature, bring the virions in close proximity and in elevated concentrations to the cell surface, increasing their chance to find their true receptors (Cagno et al., 2019; Koehler et al., 2020).

In some coronaviruses, the NTD, preceding the RBD, displays lectin activity and recognizes glycan ligands (Belouzard et al., 2012; Li, 2016). A finding that often goes unnoticed is that the MERS-CoV spike trimer as well binds to sialoglycans with the NTD proven to be responsible for that (Li et al., 2017). This binding is highly selective but of low affinity, and a multivalent sialoglycan presentation is required for interaction. It was argued that sialoglycans may guide MERS-CoV search for its true receptor, the DPP4, on the host cell surface. A similar mechanism of sialic acid recognition acting as an infection facilitator was proposed for other coronaviruses (Qing et al., 2020), including SARS-CoV and SARS-CoV-2 (Morniroli et al., 2020).

It was observed (Clausen et al., 2020) that the SARS-CoV-2 spike protein also binds heparan-sulfate (HS), which consists of linear chains of disaccharide building blocks comprising D-glucuronic acid (some of them modified to D-iduronic acid) and N-acetyl-D-glucosamine, extensively substituted with sulfate groups. Through modeling, the HS binding was pinpointed to the RBD, close to the ACE2 binding site, and both molecules bind independently from each other. Heparan sulfate (HS) binding is hypothesized to result from electrostatic interactions between the highly negative HS molecule and the overall positively charged RBD surface (see **Supplementary Material**, **Supplementary Figure 8**), the latter being able to accommodate a HS chain of up to 20 monosaccharides. Importantly, binding of HS promotes the RBD open conformation, thereby stimulating binding to ACE2 (Clausen et al., 2020). It was further discussed that the SARS-CoV-2 RBD surface is more electropositive than the one from SARS-CoV, mainly as the result of two mutations, i.e., $T^{431} \rightarrow K^{444}$ and $E^{341} \rightarrow N^{354}$.

Since not only do the SARS-CoV-s' receptor molecules have covalently attached N-glycans, but also the SARS-CoV-s' spike proteins are heavily glycosylated, it would not be unthinkable that the host's own cellular surface lectins might be involved in capturing virions, or at least act as binding facilitators. C-type lectins (CLRs) are important receptors on patrolling myeloid cells that recognize glycans at the surface of foreign invaders, leading to the induction of immune responses. However, certain viruses have "learnt" how to modulate the response of macrophages and dendritic cells, and how to (mis)use them for promoting infection instead. Although detailed knowledge of mechanisms is still missing, it was hypothesized that capture by host lectins on myeloid cells does not always lead to normal antigen processing in the lysosomes followed by peptide presentation at the cell

surface. Instead, the virions are temporarily contained, leading to their release at a later stage, followed by trans-infection of other susceptible target cells expressing the genuine ACE2 receptors (Geijtenbeek and van Kooyk, 2003). In case of SARS-CoV-2, preliminary reports using pseudovirus particles show that both DC/L-SIGN and MGL on antigen-presenting cells bind to spike protein glycan chains and promote virus transfer to permissive ACE2-containing cells (Thépaut et al., 2020). Moreover, the sialic acid-binding immunoglobulin-type lectins Siglec-3, -9 and -10 that are present on myeloid and/or B-cells were also found to bind to spike glycans (Chiodo et al., 2020).

Binding of the SARS-CoV-2 RBD to the ACE2 Receptor Does Not Extensively Affect the RBD Conformation

To compare the structure of the SARS-CoV-2 RBD in the absence (model 6VXX, chain B) and the presence of the ACE2 receptor molecule (model 6VW1, chain E), an overlay needs first to be made between the RBDs in both models. The result is shown in **Figure 9A**, in which the RBD of model 6VW1 is colored orange, while the ACE2 receptor is colored blue. Superposed is the RBD of model 6VXX, of which the ribbons are colored for RMS (RMS coloring, because the root-mean-square distances were calculated between corresponding backbone atoms to arrive at the color assignment for a group). RMS coloring for a model means that groups (backbones/side chains/atoms) in that model are colored according to how far they lie from corresponding groups in the other model that is considered the reference model (in this example, 6VW1 is taken as the reference). Regions that superimpose exactly are colored dark blue, with colors farther up the visible spectrum assigned for greater distances from corresponding atoms in the reference model. **Figure 9A** shows that the overall conformation in both models does not change dramatically.

In **Figure 9B**, we look in more detail to some amino acid side chains. In model 6VXX.pdb, the stretches N^{450}-L^{455} plus Y^{489}-Q^{506} from chain A are made visible (other residues are missing here). These peptides are very close to the ACE2 receptor (see above, **Figure 7**). In model 6VW1.pdb, the same stretches are added to the picture. The backbone plus side chains of these residues are shown in both models, with the aromatic residues and two glutamines, labeled. Finally, ribbons were added for the stretches on display. We may conclude that, locally, the orientation of some side chains is slightly modified in the model where the RBD is bound to the ACE2 receptor, which is not surprising, but the backbone is hardly affected.

Only ACE2 but Not ACE Can Act as a Receptor for SARS-CoV-2

An homolog of ACE2 exists in humans, i.e., angiotensin-I-converting enzyme, or ACE. This enzyme is a peptidyl-dipeptidase, cleaving a dipeptide at the C-terminus of angiotensin, and is ubiquitously expressed throughout the human body (Riordan, 2003). Just as ACE2, ACE is also a membrane-bound protein. It consists of two very similar domains that originated by duplication, and of which the

amino acid sequences can also easily be aligned with ACE2 (see **Supplementary Figure 9** in **Supplementary Material**). Therefore, one could speculate on the possibility that ACE might also act as a receptor for SARS-CoV and SARS-CoV-2.

In **Figure 10**, the overall structural similarities between both ACE domains and with ACE2 are visualized. The two domains of ACE are strikingly similar in structure, but also, the ACE2 structure is very similar to that of an ACE domain. However, though the three domains are structurally very similar, there are important differences to note in their primary structure (see **Supplementary Material**, **Supplementary Figure 9**). Particularly in the region of the contact surface of the ACE2 receptor protein with the SARS-CoV and SARS-CoV-2 RBDs, the sequences differ greatly from each other. Therefore, ACE is very unlikely to be able to act as receptor molecule for both viruses.
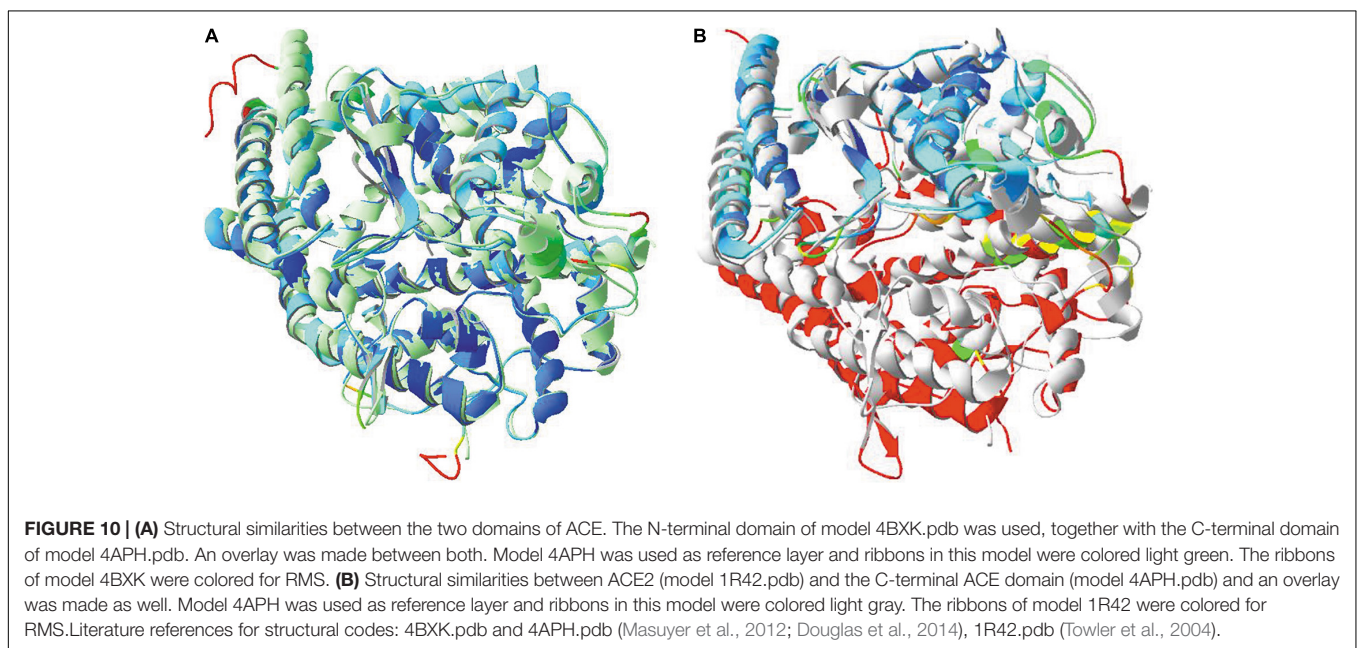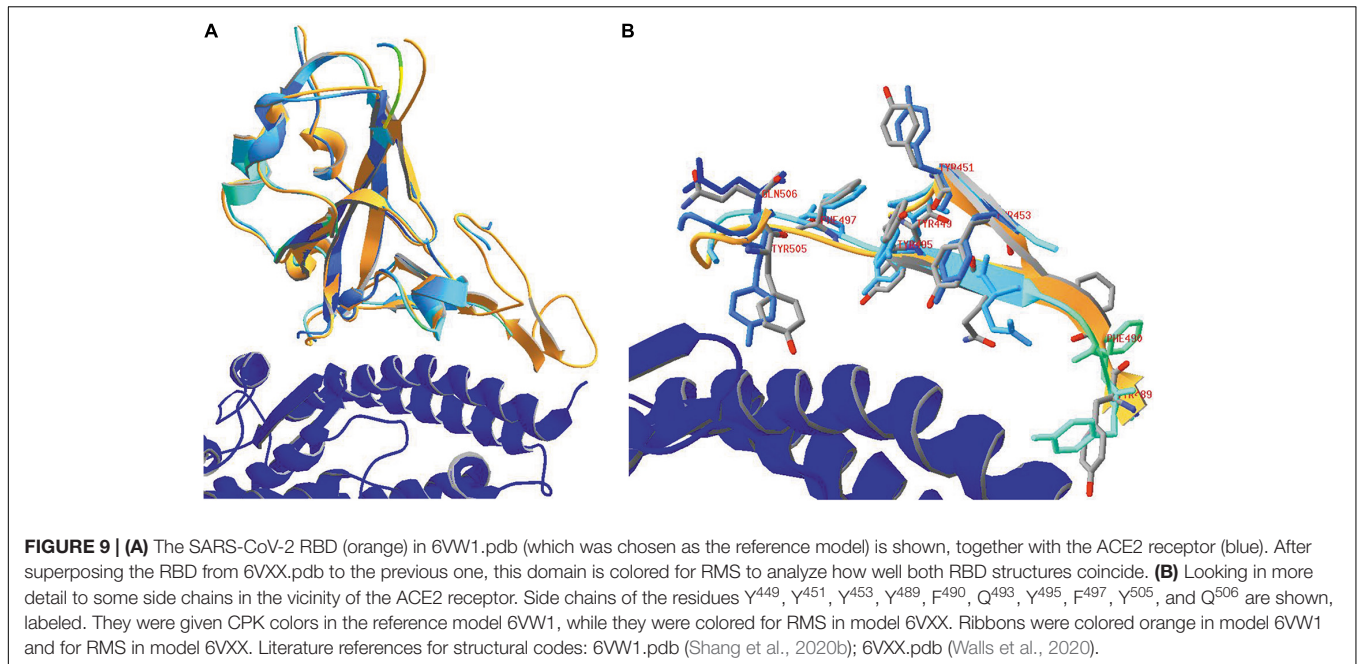
Proteolytic Events Occurring in the Spike Glycoprotein Upon Binding to Its Receptor

Soon after binding of the SARS-CoV-2 virus to its ACE2 receptor, a first proteolytic step occurs in the spike glycoprotein to split S1 (the N-terminal portion of the spike protein, containing the RBD) from S2 (the central and more rigid portion of the spike protein). The peptide of 12 residues in which cleavage takes place (i.e., T^{676} QTNSPRRARSVA S^{689}) is missing in the structure, but the flanking residues T^{676} and S^{689} are clearly located at the outer surface of the spike protein trimer (**Figures 3**, **11**). It can easily be imagined that this very hydrophilic peptide will be exposed and readily available for proteolysis.

In **Figure 11A**, we see the domains that are removed by the first cleavage as ribbons, colored for secondary structure succession, with the remainder of chain B in model 6VXX shown as backbone with side chains, and the fusion peptide (as far as its structure is available in the model) overlaid as red ribbons with the RBD domain at the top right (and colored yellow). After the first proteolytic cleavage, a second cleavage step (indicated as S2') occurs just before the fusion peptide. In this step, the peptide shown as green ribbons in **Figure 11B** will be removed. Finally, what is left from the spike protein subunit is shown in **Figure 11C**. The heptad repeat (HR1) is important for the next events, which will lead to fusion of viral and host membranes, to allow entry of the viral RNA into the host cell.

Events Causing Virus Entry Into Host Cells: The Spike Protein "Post-Fusion" State
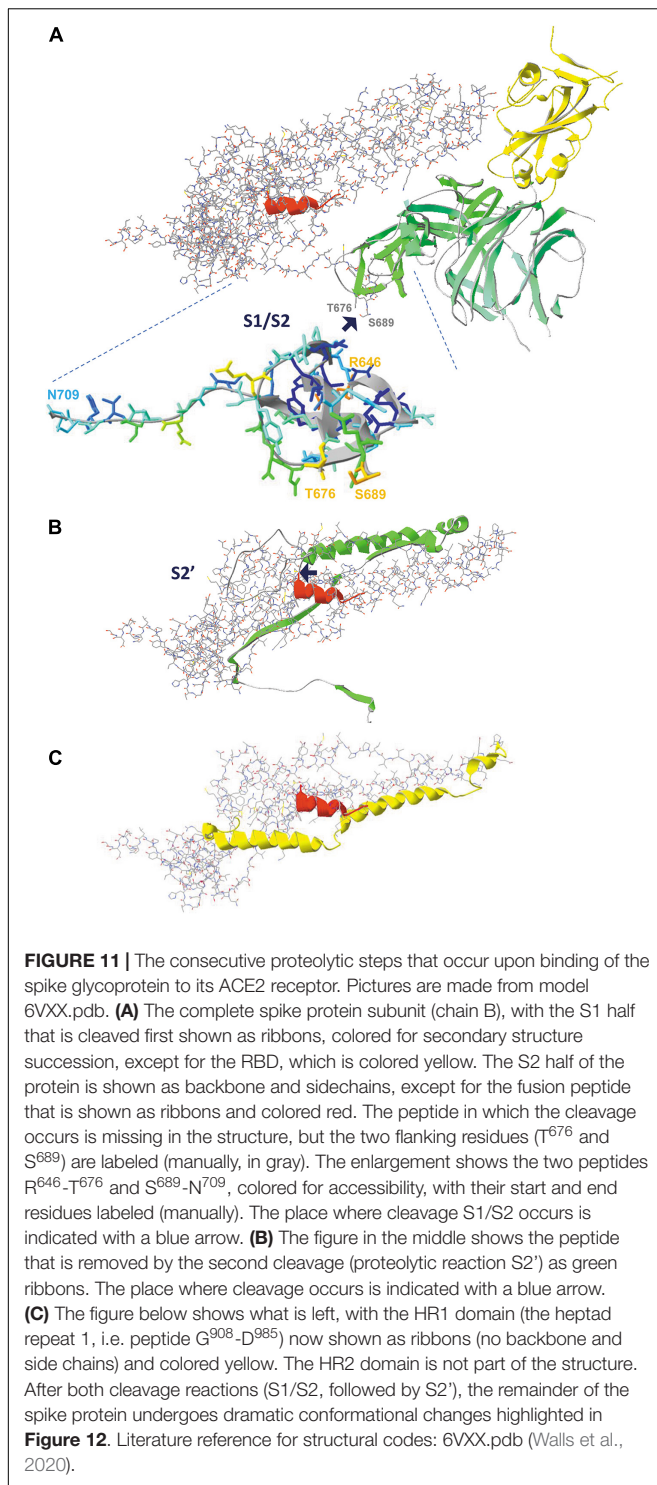
After both proteolytic cleavage events (S1/S2, followed by S2'), the remaining S2 domain undergoes an instantaneous and dramatic change in conformation to adopt the "post-fusion" state. In this state, the coiled coil-forming heptad repeats (HR1 and HR2) of each of the three subunits in the trimeric S protein form a strong and extended six-helix bundle, which prepares the virion for membrane fusion with the host cell plasma membrane (Cai et al., 2020). Towards one end of this bundle, the three fusion peptides, one in each subunit, are now brought juxtaposed to the host cell membrane and catapulted into it, after which HR2 domains

FIGURE 9 | (A) The SARS-CoV-2 RBD (orange) in 6VW1.pdb (which was chosen as the reference model) is shown, together with the ACE2 receptor (blue). After superposing the RBD from 6VXX.pdb to the previous one, this domain is colored for RMS to analyze how well both RBD structures coincide. **(B)** Looking in more detail to some side chains in the vicinity of the ACE2 receptor. Side chains of the residues Y^{449}, Y^{451}, Y^{453}, Y^{489}, F^{490}, Q^{493}, Y^{495}, F^{497}, Y^{505}, and Q^{506} are shown, labeled. They were given CPK colors in the reference model 6VW1, while they were colored for RMS in model 6VXX. Ribbons were colored orange in model 6VW1 and for RMS in model 6VXX. Literature references for structural codes: 6VW1.pdb (Shang et al., 2020b); 6VXX.pdb (Walls et al., 2020).

FIGURE 10 | (A) Structural similarities between the two domains of ACE. The N-terminal domain of model 4BXK.pdb was used, together with the C-terminal domain of model 4APH.pdb. An overlay was made between both. Model 4APH was used as reference layer and ribbons in this model were colored light gray. The ribbons of model 4BXK were colored for RMS. **(B)** Structural similarities between ACE2 (model 1R42.pdb) and the C-terminal ACE domain (model 4APH.pdb) and an overlay was made as well. Model 4APH was used as reference layer and ribbons in this model were colored light gray. The ribbons of model 1R42 were colored for RMS.Literature references for structural codes: 4BXK.pdb and 4APH.pdb (Masuyer et al., 2012; Douglas et al., 2014), 1R42.pdb (Towler et al., 2004).

fold back to bring FP and the TM domain segments together, leading to fusion of the viral and the host membranes. This results in release of the viral RNA, decorated with N proteins, into the host cell (Shulla and Gallagher, 2009; Li, 2016; Cai et al., 2020; Shang et al., 2020a; Tang et al., 2020). It was shown that, due to differences in the HR1 domain sequences, SARS-CoV-2 has a significantly higher capacity for membrane fusion than SARS-CoV, which might also contribute to its higher infectivity (Xia et al., 2020).

Figure 12A shows the formation of a 6-helix bundle structure, obtained for SARS-CoV-2 peptides, as a side and a top view.

Hydrophobic interactions are the major force driving the formation of this helix bundle. **Figure 12B** shows HR1 and HR2, with only the side chains of the hydrophobic residues. In the peptide T^{912}-E^{988}, 34 residues out of 77 are hydrophobic in nature and in the peptide V^{1164}-E^{1202}, 18 out of 39 are hydrophobic (44% and 46%, respectively). They form two lines of hydrophobicity on these peptides that slowly twist around the long helices, which results in wrapping both HR regions around each other through the formation of an antiparallel coiled coil in each of the spike protein subunits. These regions are further assembled to form the 6-helix bundle structure. **Figure 12C**
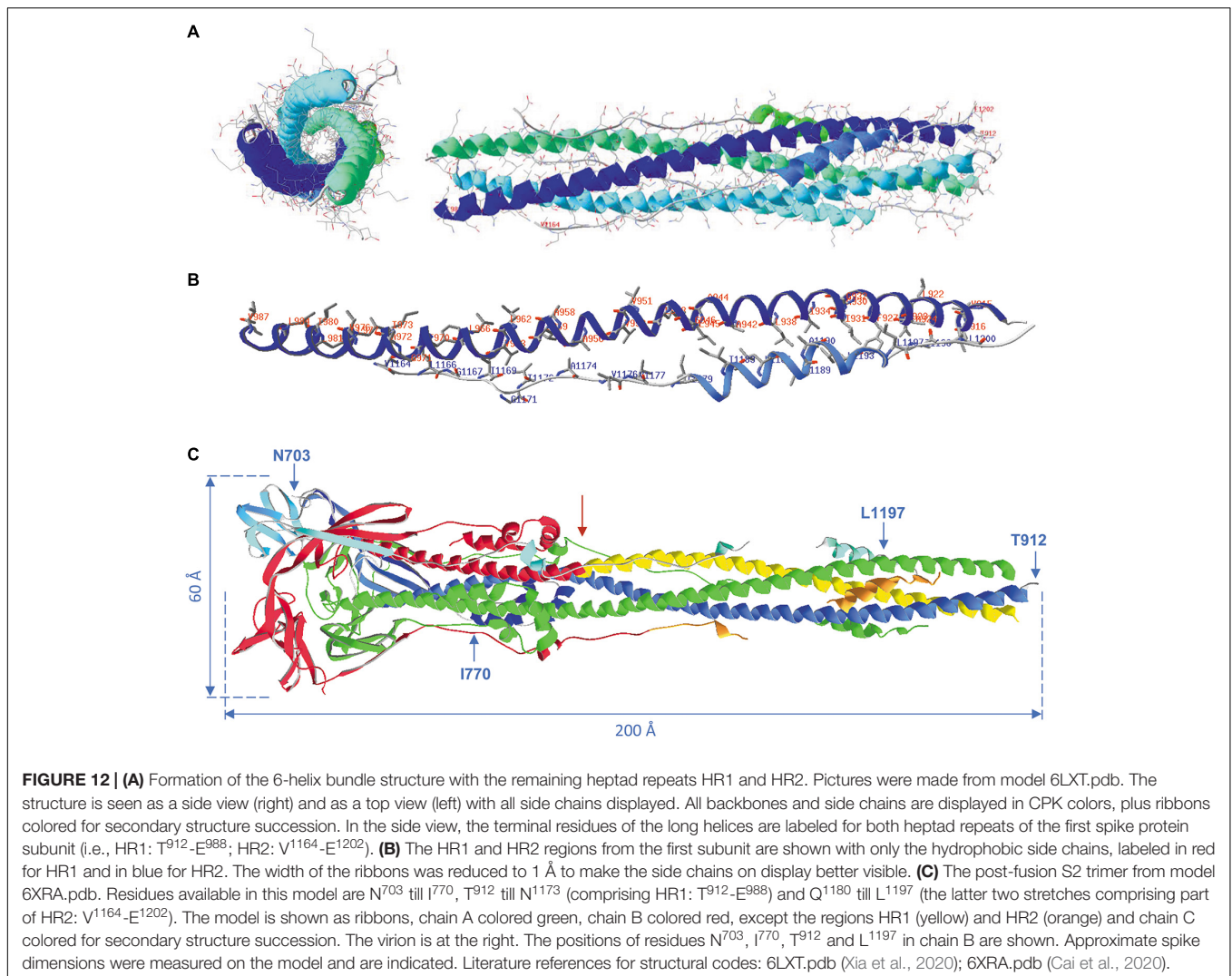
F^{833}, missing in this structure) that will integrate in the host membrane, resulting in fusion.

The Much-Debated Lucrative Spike Protein Mutant D^{614}G

From February 2020 onwards, a point mutation (D^{614}G) in the SARS-CoV-2 spike protein emerged (Korber et al., 2020). Within no time it supplanted the original protein worldwide and it was, and still is, wondered why this mutation spread at such an incredible speed. The apparently successful mutation was said to confer not only increased transmissibility to the virus, but also increased mortality. Studies using different kinds of pseudoviruses equipped with SARS-CoV-2 spike proteins indicated that spikes having the G^{614} mutation infect cells far more competently than the original D^{614} ones (summarized in Callaway, 2020b). Whether this will also be the case with the real virus in humans has yet to be confirmed, though it was further observed that genuine SARS-CoV-2 viruses were also more infectious in lab experiments on human lung cell lines and were present in increased concentrations in the upper airways of infected hamsters (Plante et al., 2020). These puzzling observations could neither to be attributed to a difference in numbers of virions produced, nor to an increased affinity of the variant to the ACE2 receptor (Daniloski et al., 2020). Certain studies ascribe the increased effectivity of the mutant to a decrease in premature S1/S2 cleavage of the G^{614} variant during assembly of new virions in the host (Daniloski et al., 2020; Zhang et al., 2020b), though an increased susceptibility to proteases was suggested as well from other experiments (Eaaswarkhanth et al., 2020; Hu et al., 2020b). The reason behind an alleged greater or lesser susceptibility to proteolysis remains unclear. Nevertheless, it needs to be kept in mind that mutations such as this one could influence the antigenic properties of the protein and might reduce the efficacy of vaccines that are currently under development using the original spike protein as it was isolated in Wuhan. Additionally, simple and seemingly harmless mutations may also have a pronounced effect on how the host immune system does recognize and react to the virus.

It was inferred from molecular modeling that the G^{614} mutation would destabilize the open conformation, thus promoting the closed state, which is unable to bind to the RBD (Becerra-Flores and Cardozo, 2020). Why then would the G^{614} mutant display a much higher fatality rate when compared to the original D^{614} protein? The authors suggested two possible hypotheses for these seemingly contradictory observations, i.e., the now more prevalent closed form might (i) be better shielded from attack by the host immune system, and/or (ii) elicit a harmful immune response, e.g., through the production of detrimental antibodies. Other studies, on the contrary, suggest that the D^{614}G mutant rather loosens the spike protein and brings its subunits more easily in the open state, which should facilitate the binding to its ACE2 receptor (Mansbach et al., 2020). Either way, many contradictory conclusions are still circulating and the final word on this mutation has clearly not yet been said. A paper expressed the stand of affairs (August 2020) in its title as follows: "*Making sense of mutation: what D^{614}G means*

FIGURE 11 | The consecutive proteolytic steps that occur upon binding of the spike glycoprotein to its ACE2 receptor. Pictures are made from model 6VXX.pdb. **(A)** The complete spike protein subunit (chain B), with the S1 half that is cleaved first shown as ribbons, colored for secondary structure succession, except for the RBD, which is colored yellow. The S2 half of the protein is shown as backbone and sidechains, except for the fusion peptide that is shown as ribbons and colored red. The peptide in which the cleavage occurs is missing in the structure, but the two flanking residues (T^{676} and S^{689}) are labeled (manually, in gray). The enlargement shows the two peptides R^{646}-T^{676} and S^{689}-N^{709}, colored for accessibility, with their start and end residues labeled (manually). The place where cleavage S1/S2 occurs is indicated with a blue arrow. **(B)** The figure in the middle shows the peptide that is removed by the second cleavage (proteolytic reaction S2') as green ribbons. The place where cleavage occurs is indicated with a blue arrow. **(C)** The figure below shows what is left, with the HR1 domain (the heptad repeat 1, i.e. peptide G^{908}-D^{985}) now shown as ribbons (no backbone and side chains) and colored yellow. The HR2 domain is not part of the structure. After both cleavage reactions (S1/S2, followed by S2'), the remainder of the spike protein undergoes dramatic conformational changes highlighted in **Figure 12**. Literature reference for structural codes: 6VXX.pdb (Walls et al., 2020).

shows the post-fusion state, after removal of the S1 half of the spike protein. The right half of this structure, from residue T^{912} till the red arrow, is very similar to the structure of **Figure 12A** (as a side view). At one end, this structure is still attached to the virion (which is to the right), and somewhere, in between I^{770} and T^{912}, are the fusion peptides (the peptides S^{816} till

FIGURE 12 | (A) Formation of the 6-helix bundle structure with the remaining heptad repeats HR1 and HR2. Pictures were made from model 6LXT.pdb. The structure is seen as a side view (right) and as a top view (left) with all side chains displayed. All backbones and side chains are displayed in CPK colors, plus ribbons colored for secondary structure succession. In the side view, the terminal residues of the long helices are labeled for both heptad repeats of the first spike protein subunit (i.e., HR1: T^{912}-E^{988}; HR2: V^{1164}-E^{1202}). **(B)** The HR1 and HR2 regions from the first subunit are shown with only the hydrophobic side chains, labeled in red for HR1 and in blue for HR2. The width of the ribbons was reduced to 1 Å to make the side chains on display better visible. **(C)** The post-fusion S2 trimer from model 6XRA.pdb. Residues available in this model are N^{703} till I^{770}, T^{912} till N^{1173} (comprising HR1: T^{912}-E^{988}) and Q^{1180} till L^{1197} (the latter two stretches comprising part of HR2: V^{1164}-E^{1202}). The model is shown as ribbons, chain A colored green, chain B colored red, except the regions HR1 (yellow) and HR2 (orange) and chain C colored for secondary structure succession. The virion is at the right. The positions of residues N^{703}, I^{770}, T^{912} and L^{1197} in chain B are shown. Approximate spike dimensions were measured on the model and are indicated. Literature references for structural codes: 6LXT.pdb (Xia et al., 2020); 6XRA.pdb (Cai et al., 2020).

for the Covid-19 pandemic remains unclear" (Grubaugh et al., 2020) and that statement is surely still true today. **Supplementary Figure 10** (see **Supplementary Material**) gives an impression of the surroundings of residue D^{614}.

It needs to be stressed that mutations in the spike protein continue to emerge and by early May 2020, 329 naturally occurring variants were already reported (Li et al., 2020b), some of which make the virus resistant to certain monoclonal antibodies. Moreover, certain glycosylation deletions were found to reduce viral infectivity.

The Problem of Antibody-Dependent Enhancement

There are clear indications that SARS-CoV-s also may infect certain cell types of the PBMC (collection of peripheral blood mononuclear cells) that do not express the ACE2 receptor, i.e., those belonging to the immune system (such as monocytes and macrophages, the former potentially also leading to productive virus replication). The immune response, which is specifically designed to clear infections, sometimes shows a dysregulated response leading to the opposite outcome (Taylor et al., 2015). This kind of response is due to the presence of anti-spike protein antibodies and known as antibody-dependent enhancement (ADE). This immunopathological situation is mediated by antibody Fc domains and occurs when virus-antibody immune complexes interact with cells carrying receptors for Fc. The ADE pathway is very complex with virus- as well as host-dependencies, and not all details are fully understood. Nevertheless, it is an important issue to be taken into consideration during development of vaccination strategies.

STRATEGIES TO PREVENT BINDING OF THE VIRUS TO ITS RECEPTOR

Soluble Mutated ACE2 Analogs as a Decoy Receptor

The possibility of fooling the SARS-CoV-2 virus by administering high-affinity soluble ACE2 analogs as decoy receptors, thereby preventing virus binding to and entry in host cells by

competition, was launched as an interesting idea to combat Covid-19 (Chan et al., 2020). The authors created an extensive library of 2,340 human sACE2 (soluble receptor) coding mutants that were expressed in human Expi293F cells (each cell expressing only one type of single mutant), which were then tested for SARS-CoV-2 RBD-binding using *in vitro* assays. Based on the results obtained, single mutants were combined and a series of sACE2 molecules with triple up to septuple mutations were generated for more detailed analysis. Several most interesting findings resulted from this study: (i) a mutation modifying residue T^{92}, resulting in a sACE2 mutant that is not glycosylated anymore in position N^{90}, favors RBD-binding, suggesting that the glycan at N^{90} hinders (but does not prevent) RBD-binding; (ii) a number of sACE2 mutants at the interface with RBD enhance binding, which opens perspectives for the aforementioned type of approach in fighting Covid-19; (iii) the variant called sACE2.v2.4 (carrying mutations $T^{27}Y$, $L^{79}T$ and $N^{330}Y$, thus still leaving the N-glycosylation site at N^{90} intact, and which is very well expressed and shows enzymatic activity on angiotensin II, albeit reduced), was purified and extensively analyzed: it was found to display a 65-fold higher affinity for immobilized SARS-CoV-2 RBD than the soluble wild type (using biosensor and ELISA technology) and also efficiently competes with antibodies from serum of Covid-19 patients for binding to the RBD. In **Supplementary Figure 11** (see **Supplementary Material**) we are looking to the result of a proposed triple mutation $T^{27}Y$, $L^{79}T$, $N^{330}Y$.

In another study, part of the sACE2 receptor (Q^{18}-A^{614}) was engineered after computational design and experimental affinity maturation, fused to the ACE2-collectrin domain (D^{615}-S^{740}) and dimerized by adding a human antibody Fc, resulting in avidity as well as long half-life times *in vivo*. A variant with seven amino acid changes ($Q^{18}R/K^{31}F/N^{33}D/H^{34}S/E^{35}Q/W^{69}R/Q^{76}R$), and of which ACE2 enzyme activity was destroyed by a $H^{345}L$ mutation, was found to bind the spike RBD 170-fold more tightly than the wild-type ACE2. This (or some alternative) construct was proposed to be potentially useful as "trap" to neutralize SARS-CoV-2 and prevent viral entry into host cells (Glasgow et al., 2020).

In small-scale clinical studies, a human recombinant sACE2 molecule has already been used as a potential drug candidate with promising results (Zoufaly et al., 2020). These studies were based on earlier research using non-mutated hrsACE2 (human recombinant soluble ACE2) (Monteil et al., 2020).

'Mini-Protein Inhibitors' as Prophylactic Molecules and/or for Use in Therapeutic Treatments

Another interesting avenue in the search for prophylactic and/or therapeutic treatments of Covid-19 was published (Cao et al., 2020). In this study, researchers intended to find high affinity and thermostable mini-binders to the SARS-CoV-2 spike RBD that would compete with ACE2 receptor binding. Such molecules were devised both by incorporating the ACE2 long α-helix that interacts with the RBD (see above, **Figure 7**) in small proteins that were further designed to make additional interactions with the spike protein to enhance the affinity, as well as modeled from

scratch. Such molecules would (i) not require obligatory storage at low temperatures, (ii) circumvent possible side effects inherent to using antibodies (e.g., ADE: see above), (iii) those peptides, being 20-fold smaller than antibodies, have a much higher binding site density per weight, (iv) potentially be applicable for internasal administration, e.g., as a gel or an aerosol, (v) make viral mutational escape very unlikely when being used in combinations. Promising peptides (56-64 amino acid residues long) were created that display excellent stability as well as high affinity for the spike protein (K_D-values ranging from 100 pM to 10 nM) and they were found to prevent infection of Vero cells with an IC_{50} between 24 pM and 35 nM (Cao et al., 2020).

Binding of Antibodies to the Spike Glycoprotein

Ever since the onset of the pandemic, numerous efforts have been made to track down neutralizing antibodies against SARS-CoV-2 that would help to combat the infection by using them in passive immunization. Initially, already available antibodies against SARS-CoV were tested for their potency against SARS-CoV-2 and later on, new antibodies were specifically generated and analyzed. A number of structural data have been made available in the PDB database. Antibodies in the pipeline are either of the conventional IgG type (or their Fab fragments), but also of the camelid type (heavy chain-only, or VHH, alternatively called single-domain antibodies). Because their concept is very different, both types of antibodies are conceived by Nature to recognize different types of epitopes: while classical antibodies are designed to grasp smaller groups or peptides sticking out from proteins' surfaces by using their two extended antigen-binding regions as two scoops, camelid antibodies (from which "nanobodies" are derived) form rather "finger-like" structures to penetrate in cavities of the antigen (Romão et al., 2016; Jovčevska and Muyldermans, 2020). Nanobodies have several advantages, one being that, because of their limited size (only 15 kDa, which is ten times smaller than classical H_2L_2 antibodies), they can be administered as inhalable drugs, which for Covid-19 is an indisputable asset.

Structures of SARS-CoV-2 RBD with various Fab fragments are available in the PDB database and they were used for making overlays. In **Figure 13**, six such Fab fragments and one nanobody are seen bound to the SARS-CoV-2 RBD. This figure was made after superposing all structures using the RBD available in model 6VW1.pdb as the reference chain.

Most of the antibodies analyzed compete for binding to the ACE2 receptor, as can be seen from **Supplementary Figures 12**, **Supplementary Figures 13** (see **Supplementary Material**). When an antibody binds to the RBD, several clashes are seen with the ACE2 receptor. This is true for Fab fragments in models 6XC2, 6XC4, 7BZ5, and 7C01, which all bind to the same region of the RBD. This is also seen for the Fab fragment in model 7BWJ where some, though less prominent, clashes are observed. These five antibodies were described in literature to be neutralizing (Ju et al., 2020; Shi et al., 2020; Wu et al., 2020; Yuan et al., 2020b). Of course, to really compete with the ACE2 receptor for binding, the affinity of such an antibody for the spike protein
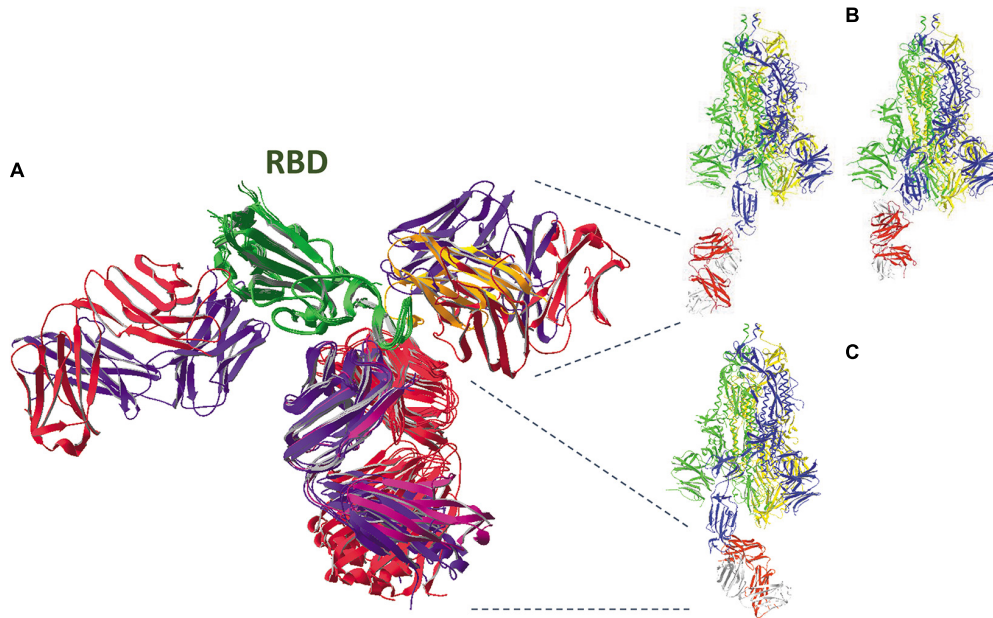
FIGURE 13 | (A) An overlay was made of the SARS-CoV-2 spike protein RBD (from residues N^{334} till P^{527}, colored dark green) with six antibody Fab fragments (all H- and L- chains are colored bluish and reddish, respectively) and one nanobody (colored orange). The Fab fragments in models 6XC2, 6XC4, 7BZ5, and 7C01 are pointing down, while Fab from model 6W41 (antibody CR3022) is pointing left, and the Fab from model 7BWJ is pointing right, which is overlapping with the nanobody from model 6Z2M.pdb. All these SARS-CoV-2-binding H$_2$L$_2$ antibodies, having neutralizing capacity, were cloned and expressed from memory-B cells present in PBMCs either isolated from Covid-19 recovered patients [mAb cc12.1 (CXC2.pdb) and mAb cc12.3 (6XC4.pdb): Rogers et al., 2020; mAb B38 (7BZ5.pdb): Wu et al., 2020; mAb CB6 (7C01.pdb): Shi et al., 2020); mAb P2B-2F6 (7BWJ.pdb): Ju et al., 2020] or from a convalescent SARS-CoV patient [mAb CR3022 (6W41.pdb): Yuan et al., 2020b]. The nanobody (H11-D4) was developed earlier against SARS-CoV and analyzed for its SARS-CoV-2 binding capacity (Huo et al., 2020). **(B)** Binding of antibody Fab fragment 7BWJ to the SARS-CoV-2 full spike trimer with chain B in open state (model 6VYB.pdb, left) and with all subunits in closed state (model 6VXX.pdb, right). An overlay was made between the structures (using the RBD of chain B) and no clashes were detected. **(C)** Binding of antibody Fab fragment 6XC4 to the SARS-CoV-2 full spike trimer with chain B in the open state (model 6VYB.pdb). An overlay was made between both structures (using the RBD of chain B) and no clashes were found. However, when chain B is also in the closed state, extensive clashes are seen with chain C residues (figure not shown; see **Supplementary Material** about how to detect clashes). In B and C, the ribbons of the spike protein subunits are colored yellow, blue and green for chains A, B and C, respectively, and red and gray for the Fab H- and L-chains, respectively.

is of utmost importance: when the affinity of the antibody is too low, the spike glycoprotein might nevertheless preferentially bind to the ACE2 receptor, leading to delivery of the viral RNA into the host cell's cytoplasm.

Some structures are also available of a complete SARS-CoV-2 spike protein trimer with antibody Fab fragments. The Fab fragment of antibody S309 was determined to potently neutralize both SARS-CoV and SARS-CoV-2 (Pinto et al., 2020). **Supplementary Figures 14A,B** (see **Supplementary Material**) shows binding of three Fab molecules to the RBDs of each of the subunits of the SARS-CoV-2 spike protein trimer.

Another interesting antibody is CR3022, which was previously isolated from a SARS-CoV patient. It is directed against the RBD, but the epitopes to which it binds are different compared to the other antibodies (**Figure 13**). Using an *in vitro* assay, CR3022 proved to be neutralizing for SARS-CoV but not for SARS-CoV-2, though it is able to bind to its RBD, albeit with 100-fold lower affinity (Yuan et al., 2020a). The neutralizing effect of this antibody for SARS-CoV was explained through structural modeling: it was envisaged that the epitope to which CR3022 binds can only be reached by the antibody molecules when at least two RBDs are in the open conformation and,

moreover, they need to be slightly rotated (Yuan et al., 2020a). Otherwise, there would be clashes with other parts (e.g., the NTDs) of the spike protein trimer (**Supplementary Figure 16D**, see **Supplementary Material**). It was further discussed in the same paper that, enigmatically, antibodies not having an *in vitro* neutralizing effect may nevertheless display *in vivo* protection for reasons that need to be further explored (Yuan et al., 2020a). **Figure 13** also shows how some Fabs are only able to bind to a 'one-up' (6XC4), while another binds a 'one-up' as well as a 'none-up' spike trimer (7BWJ), and another needs more than one subunit in open state (6W41).

Finally, when comparing binding of the Fab fragment to the SARS-CoV-2 RBD in model 7BWJ.pdb with that of a nanobody in model 6YZ5.pdb, the difference in the principle of antigen recognition between both antibodies catches the eye. As shown in **Figure 13**, both antibodies bind to the same region of the antigen. A more detailed picture of the binding is shown in **Figure 14**. Binding by the Fab fragment is due to residues belonging to two RBD loops, i.e., K^{444} till N^{450} and V^{483} till F^{490}, which are grasped by the binding sites formed by the antibody H- and L-chains, respectively (**Figure 14A**). On the other hand, binding of the nanobody occurs because essentially two VHH loops, i.e., R^{27} till

FIGURE 14 | (A) Contact surface between the SARS-CoV-2 RBD (chain E, colored green) and an antibody Fab fragment from model 7BWJ.pdb, with the H- and L-chains colored purple and yellow, respectively. Amino acid residues that are within a distance of 3.5 Å from the opposing protein are shown with their backbone and side chains and were manually labeled. Hydrogen bonds are shown as green dashed lines; one hydrogen bond is colored gray because the distance between hydrogen donor and acceptor (3.32 Å) is slightly above the default maximal value of 3.20. (B) Contact surface between the SARS-CoV-2 RBD (chain E, colored green) and a nanobody (chain F, colored orange) from model 6YZ5.pdb. Amino acid residues that are within a distance of 3.5 Å from the opposing protein are shown with their backbone and side chains and were manually labeled. Hydrogen bonds are shown as green dashed lines; one hydrogen bond is colored gray because the distance between hydrogen donor and acceptor (3.33 Å) is again slightly above the default maximal value of 3.20. Literature references for structural codes: 7BWJ.pdb (Ju et al., 2020); 6YZ5.pdb (Huo et al., 2020).

S^{30} and E^{100} till L^{106}, fit into a shallow depression that is formed on the RBD between residues K^{444} till F^{456} and E^{484} till Y^{495} (**Figure 14B**). In the example of the Fab binding, seven hydrogen bonds are formed between the RBD and the antibody (five with the H- and two with the L-chain), while in the example of the nanobody the interaction is stabilized by eleven hydrogen bonds.

Neutralizing Antibodies That Bind to the NTD Prevent Required Conformational Changes in the Spike Protein Trimer

Monoclonal antibodies with neutralizing activity were isolated from convalescent Covid-19 patients and characterized, some of which do not bind to the RBD, but rather to the NTD instead. From one of them (mAb 4A8), which binds with high (nM) affinity, the structure of the Fab in complex with the spike trimer was intensively analyzed (Chi et al., 2020). The potent neutralizing activity of this mAb was speculatively ascribed to its restraining effect on the conformational changes in the spike trimer, which are essential for activation of the spike leading to invasion of the host cell. In the **Supplementary Material**,

Supplementary Figures 14C,D,E show a structure of the SARS-CoV-2 spike trimer in complex with three Fab fragments, each of them obviously binding to a different NTD and **Supplementary Figure 15** visualizes the interface between the spike protein's NTD and the Fab fragment.

OTHER VIRAL MEMBRANE PROTEINS

The Abundant Membrane Protein M

Protein M, with a molecular mass of 24–28 kDa in various coronaviruses, is the most abundant protein in the viral membrane. It is known to be involved in the organization of viral assembly and binds to the nucleocapsid (Nal et al., 2005; Dhama et al., 2020). It is a multi-pass trans-membrane protein with three TM helices that are connected by short peptides, and with a long C-terminal endodomain. In the SARS-CoV and MERS-CoV, the N-terminal ectodomain is N-glycosylated at one single position (Fung and Liu, 2018). It has been suggested from electron microscopy and statistical analyses that protein M occurs in two conformations, which is supposed to regulate virus particle

shape and size: an elongated structure that makes the membrane more rigid, with less curvature and high spike density, and a more compact one that renders the membrane more flexible and with less spike density (Neuman et al., 2011). However, there are no structural data available as yet for protein M from any of the coronaviruses. Therefore, we have to rely exclusively on predictions (see **Supplementary Material**, **Supplementary Figures 17**, **18**). Properties of the protein are summarized in **Table 1**.

The Minor Membrane Component Protein E

Protein E is the smallest of the SARS-CoV-s' structural proteins, being 8.5–12 kDa in size, and its properties are summarized in **Table 1**. It has several functions, acting as an ion channel that is formed by homopentameric assembly of protein E subunits, but it is also involved in virus assembly and release, and interaction with the host (Yuan et al., 2006; Surya et al., 2018; Schoeman and Fielding, 2019; Dhama et al., 2020). Protein E is predicted to be a single-pass membrane protein with a short N-terminal peptide, followed by a TM helix and a longer C-terminal domain. However, the exact location of N- and C-termini is still a matter of debate (see **Supplementary Material**, **Supplementary Figure 19**). Some studies indicate that N- and C-termini might be located in the same compartment and it was also proposed that this protein might adopt different conformations in the viral membrane. Protein E from SARS-CoV was found to be S-palmitoylated at central cysteine residues (see **Supplementary Material**, **Supplementary Figure 20**). Very shortly after these cysteines are two potential N-glycosylation sites, N^{48} and N^{66}, which were shown to be partially occupied (Fung and Liu, 2018). This would mean that, at least during biosynthesis, this part of the protein must face the ER lumen. MERS-CoV protein E, on the other hand, does not have N-glycosylation sites.

The structure of part (E^8 till L^{65}) of protein E from SARS-CoV embedded in LMPG (lyso-myristoyl phosphatidylglycerol) micelles was unraveled by NMR technology. The cysteine residues C^{40}, C^{43} and C^{44} were replaced in the protein by alanines. All 16 models nicely coincide showing that there are no very flexible regions in the pentamer. The predicted transmembrane helices seem to form a central structure with the potential N-glycosylation sites at the outskirts (see **Supplementary Material**, **Supplementary Figure 20**). The side views suggest that, at least when protein E is taken up in micelles, the N- and C-terminal amino acid residues are located at the same side of the membrane, with the potential N-glycans and the palmitoyl chains at the opposite side.

WRAPPING UP THE VIRAL RNA: THE SOLUBLE NUCLEOCAPSID PROTEIN N

Protein N (45–50 kDa) is the only soluble structural protein in the SARS-CoV-s. It is used by the virions to wrap up their RNA molecules (Chang et al., 2014; Dhama et al., 2020). It consists of two major domains that each contribute to RNA-binding: an N- and a C-terminal domain, the latter of which is used by

the protein for dimerization. Both domains are linked to each other by a serine-arginine-rich peptide. A third domain at the C-terminus is important for interacting with protein M.

SARS-CoV and MERS-CoV protein N molecules are phosphorylated on serine and threonine residues at multiple sites, especially within the SR-rich peptide, by host kinases (Fung and Liu, 2018). Moreover, SARS-CoV protein N was proven to be modified by sumoylation (on residue K^{62}) but the effect of this reaction needs further investigation (Fung and Liu, 2018). Finally, ADP-ribosylation also seems to occur in both SARS-CoV and MERS-CoV (Fung and Liu, 2018).

Supplementary Figure 21 (see **Supplementary Material**) shows that, despite the rather modest sequence similarity, both domains of protein N are structurally very similar in all three SARS-CoV-s. The dimeric C-terminal domains are attached with their N-terminal residues to the C-termini of two non-interacting N-terminal domains, making an extended overall structure. The C-terminal tails of the N-protein point in opposite directions. Both domains as well as the linker region have many basic residues, explaining the elevated theoretical pI value of 10, and the number of hydrophobic residues is very limited (the aliphatic index of protein N is very small: see **Table 1**). The excess positive charges on this protein are considered essential for wrapping up the polyanionic viral RNA. Protein-RNA interactions are proposed to be guided further by base stacking interactions using the protein's aromatic residues that are amply present in the two RNA-binding domains: the YWF content amounts to 10.4% and 11% in the N- and C-terminal domains, respectively (see **Supplementary Material**, **Supplementary Figure 22**). Curiously, protein N is also predicted to have extended disordered regions (see **Supplementary Material**, **Supplementary Figure 23**), despite the fact that well-ordered structures were determined by X-ray crystallography. Only three regions (roughly residues G^{99}-P^{142}, A^{217}-L^{230} and W^{301}-Y^{360} in SARS-CoV-2) are predicted by the program IUPred to be ordered, i.e., some parts of the N- and C-terminal domains and part of the SR-rich peptide. It has been discussed (Chang et al., 2014) that inclusion of disordered regions (IDRs) within the structured regions of protein N not only increases the binding affinity for nucleotides, but also its binding cooperativity (making a next domain binding better and stronger). This may be explained in the light of the following known IDR's properties: enhanced binding/speed of interaction; promiscuity in binding partners; enabling larger interaction surfaces with partners upon complex formation (the IDR is wrapping itself tightly around its binding partner); facilitating introduction/removal of post-translational modifications (Tompa, 2012; Habchi et al., 2014).

DISCUSSION

This paper summarizes and discusses the current knowledge of the structural proteins that make up the coronaviruses in general, and the beta-coronaviruses SARS-CoV-2, SARS-CoV and MERS-CoV in particular. We demonstrate how these proteins are well-designed by Nature for their function, how they cooperate with each other to make very successful virions, and how these

viruses mislead and hijack the host for their own benefit. Certain aspects are well-known since they are explained and illustrated in other papers, but some others remain often unnoticed or their importance underestimated, such as, for instance, the observation that coronaviruses might transiently interact with sialoglycans/heparan-sulfate prior to binding to their true receptors, thereby facilitating and speeding up invasion of a host cell. Another point is that these beta-coronaviruses developed very different ways of entering a host cell, i.e., either by directly releasing their RNA after membrane fusion, or after invading the host cell making use of the endocytotic pathway (occasionally with the help of NRP1), and in all routes they rely on the action of a plethora of host proteases that are ubiquitously available. These invasion routes exist side-by-side, and some virions may take one route, while others, at the same time and in the same host, may take the other. The way by which new virions leave a host cell through de-acidified lysozomes is also peculiar. Furthermore, it is of the utmost importance to keep an eye on new mutants that may develop in the future, which might turn these SARS-CoV-s into even smarter particles than they already are today, possibly rendering our developed defense strategies ineffective. Finally, although already a lot is known about these SARS-CoV-s, at several points knowledge and essential details are still missing. It is hoped that the near future will see these gaps being filled in, and that smart solutions, maybe still not been considered today, will emerge that will put an end to the pandemic that is currently straining the health systems globally. Very promising strategies to combat Covid-19 are in the pipeline, amongst others, the development of decoy receptors and mini-protein inhibitors, monoclonal antibodies and nanobodies that might find applications in nasal sprays, new and repurposed antivirals, and, of course, vaccines. Today, and after only one single year of development and clinical trials, four vaccines have already received approval in EU/United Kingdom and United States and are now being successfully applied. Two of them are viral vector vaccines in which recombinant DNA is packaged in a harmless adenovirus, either from chimpanzee (Oxford-AstraZeneca) or of human origin (Johnson and Johnson). Two others apply newer technology, based on a synthetic piece of mRNA packaged in lipid nanoparticles (Pfizer/BioNTech and Moderna) (Kaur and Gupta, 2020; Silveira et al., 2021). All four use coding sequences for the spike protein. The latest developments will certainly also contribute to our fight, not only against other types of viral infections, but also against cancer (Pardi et al., 2018; Zhang et al., 2019; Xu et al., 2020b; Miao et al., 2021).

The problematic spreading of human coronaviruses early in this century, with SARS-CoV, MERS-CoV and the most recently developed SARS-CoV-2 as known culprits, unmistakably ushered a huge variety of structural studies dealing with all aspects of these viruses. This will be even more so if new pandemics emerge in the near or more distant future, a situation that is predicted by many researchers and healthcare workers to occur. Consequently, it is of the utmost importance to understand these structures and to be able to look at them in detail, using a combination of a series of bioinformatic tools, most of which are freely available these days through the internet. It is hoped that this publication will stimulate more researchers and students to visualize the available structures on their computers and to use the bioinformatic tools that become available, which will help advance science in this and related fields. With this goal in mind, we present in the **Supplementary Material** a set of guidelines, using the interactive program DeepView (Guex and Peitsch, 1997; Guex et al., 2009), that allows non-specialists in structural biology to upload protein structures and scrutinize them. This program was also used to make the figures in this paper.

EPILOG: RECENT DEVELOPMENTS

Circulating SARS-CoV-2 lineages were estimated to accumulate nucleotide mutations, mostly synonymous, at a rate of about 2.7 per month (Duchene et al., 2020). However, at least three examples of much faster mutation rates have recently emerged, one in United Kingdom (Rambaut et al., 2020), another one in South Africa (Tegally et al., 2020) and a third one in Brazil (Faria et al., 2021). The United Kingdom variant has nine amino acid mutations in the spike protein when compared to the original Wuhan strain, one of which in the RBD ($N^{501}Y$), while the SA and Brazilian variants display each ten non-synonymous mutations in the spike protein, three of which in the RBD ($K^{417}N$, $E^{484}K$, $N^{501}Y$ in the South African variant and $K^{417}T$, $E^{484}K$, $N^{501}Y$ in the Brazilian one). Although the reason for this rapid development remains enigmatic so far, intra-host evolution in an immune-deficient or immune-suppressed individual suffering from long-term infection was suggested, possibly leading to an accumulation of "immune-escape" mutants. A major problem is that such mutations might affect the efficacy of vaccines being developed to combat Covid-19. The variant strains are analyzed in **Supplementary Figures 24–26** (see **Supplementary Material**).

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

REFERENCES

Abd Ellah, N. H., Gad, S. F., Muhammad, K., Batiha, G. E., and Hetta, H. F. (2020). Nanomedicine as a promising approach for diagnosis, treatment and prophylaxis against COVID-19. Nanomed. (Lond) 15, 2085–2102. doi: 10.2217/nnm-2020-0247

Aebi, M. (2013). N-linked protein glycosylation in the ER. Biochim. Biophys. Acta 1833, 2430–2437. doi: 10.1016/j.bbamcr.2013.04.001

Aebi, M., Bernasconi, R., Clerc, S., and Molinari, M. (2009). N-glycan structures: recognition and processing in the ER. Trends Biochem. Sci. 35, 74–82. doi: 10.1016/j.tibs.2009.10.001

Amanat, F., and Krammer, F. (2020). SARS-CoV-2 Vaccines: Status Report. Immunity 52, 583–589. doi: 10.1016/j.immuni.2020.03.007

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. Nat. Med. 26, 450–452. doi: 10.1038/s41591-020-0820-9

Astuti, I., and Ysrafil, Y. (2020). Severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes Metab. Syndr. Clin. Res. Revs. 14, 407–412. doi: 10.1016/j.dsx.2020.04.020

Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., et al. (2020). SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. Cell 183, 1325–1339.e21. doi: 10.1016/j.cell.2020.10.004

Bar-On, Y.-M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19) by the Numbers. eLife. Available online at: https://bit.ly/2WOeN64. (accessed March 30, 2020)

Baum, A., Fulton, B. O., Wloga, E., Copin, R., Pascal, K. E., Russo, V., et al. (2020). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science 369, 1014–1018. doi: 10.1126/science.abd0831

Becerra-Flores, M., and Cardozo, T. (2020). SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int. J. Clin. Pract. 74:e13525. doi: 10.1111/ijcp.13525

Belouzard, S., Millet, J. K., Licitra, B. N., and Whittaker, G. R. (2012). Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4, 1011–1033. doi: 10.3390/v4061011

Benton, D. J., Wrobel, A. G., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., et al. (2020). Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature 588, 327–330. doi: 10.1038/s41586-020-2772-0

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. Nucl. Acids Res. 28, 235–242. doi: 10.1093/nar/28.1.235

Blaskovic, S., Blanc, M., and van der Goot, F. G. (2013). What does S-palmitoylation do to membrane proteins? Febs J. 280, 2766–2774. doi: 10.1111/febs.12263

Brix, K. (2018). "Host Cell Proteases: Cathepsins," in Activation of Viruses by Host Proteases, eds E. Böttcher-Friebertshäuser, et al. (Cham: Springer), 249–276. doi: 10.1007/978-3-319-75474-1_10

Burkard, C., Verheije, M. H., Wicht, O., van Kasteren, S. I., van Kuppeveld, F. J., Haagmans, B. L., et al. (2014). Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. PLoS Pathog. 10:e1004502. doi: 10.1371/journal.ppat.1004502

Cagno, V., Tseligka, E. D., Jones, S. T., and Tapparel, C. (2019). Heparan sulfate proteoglycans and viral attachment: true receptors or adaptation bias? Viruses 11:596. doi: 10.3390/v11070596

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M. Jr., et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. Science 369, 1586–1592. doi: 10.1126/science.abd4251

Callaway, E. (2020a). The race for corona virus vaccines. Nature 580, 576–577. doi: 10.1038/d41586-020-01221-y

Callaway, E. (2020b). Making sense of coronavirus mutations. Nature 585, 74–77.

Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., et al. (2020). Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. Science 370, 856–860. doi: 10.1126/science.abd2985

Cao, L., Goreshnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., et al. (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science. 370, 426–431. doi: 10.1126/science.abd9909

Capell, T., Twyman, R. M., Armario-Najera, V., Ma, J. K.-C., Schillberg, S., and Christou, P. (2020). Potential applications of plant biotechnology against SARS-CoV-2. Trends Plant Sci. 25, 635–643. doi: 10.1016/j.tplants.2020.04.009

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond shielding: the roles of glycans in SARS-CoV-2 spike protein. ACS Cent. Sci. 6, 1722–1734. doi: 10.1021/acscentsci.0c01056

Chan, J. F.-W., To, K. K.-W., Tse, H., Jin, D.-Y., and Yuen, K.-Y. (2013). Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends Microbiol. 21, 544–555. doi: 10.1016/j.tim.2013.05.005

Chan, K. K., Dorosky, D., Sharma, P., Abbasi, S. A., Dye, J. M., Kranz, D. M., et al. (2020). Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science 369, 1261–1265. doi: 10.1126/science.abc0870

Chang, C.-K., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T.-H. (2014). The SARS coronavirus nucleocapsid protein - Forms and functions. Antiviral Res. 103, 39–50. doi: 10.1016/j.antiviral.2013.12.009

Charollais, J., and van der Goot, F. G. (2009). Palmitoylation of membrane proteins. Mol. Membrane Biol. 26, 55–66. doi: 10.1080/09687680802620369

Chen, C.-Y., Chang, C.-K., Chang, Y.-W., Sue, S.-C., Bai, H.-I., Riang, L., et al. (2007). Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. J. Mol. Biol. 368, 1075–1086. doi: 10.1016/j.jmb.2007.02.069

Chen, Y., Liu, Q., and Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. J. Med. Virol. 92, 418–423. doi: 10.1002/jmv.25681

Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., et al. (2020). A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science 369, 650–655. doi: 10.1126/science.abc6952

Chiodo, F., Bruijns, S. C. M., Rodriguez, E., Li, R. J. E., Molinaro, A., Silipo, A., et al. (2020). Novel ACE2-independent carbohydrate-binding of SARS-CoV-2 spike protein to host lectins and lung microbiota. bioRxiv [Preprint] doi: 10.1101/2020.05.13.092478

Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. Cell 183, 1043–1057. doi: 10.1016/j.cell.2020.09.033

Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., et al. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368, 1331–1335. doi: 10.1126/science.abb4489

Daly, J. L., Simonetti, B., Klein, K., Chen, K.-E., Williamson, M. K., Antón-Plágaro, C., et al. (2020). Neuropilin-1 is a host factor for SARS-CoV-2 infection. Science. 370, 861–865. doi: 10.1126/science.abd3072

Daniloski, Z., Guo, X., and Sanjana, N. E. (2020). The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. bioRxiv [Preprint] doi: 10.1101/2020.06.14.151357

de Alwis, R., Chen, S., Gan, E. S., and Ooi, E. E. (2020). Impact of immune enhancement on Covid-19 polyclonal hyperimmune globulin therapy and vaccine development. EBioMedicine 55, 102768. doi: 10.1016/j.ebiom.2020.102768

Delpino, M. V., and Quarleri, J. (2020). SARS-CoV-2 pathogenesis: imbalance in the renin-angiotensin system favors lung fibrosis. Front. Cell. Infect. Microbiol. 10:340. doi: 10.3389/fcimb.2020.00340

Dhama, K., Khan, S., Tiwari, R., Sircar, S., Bhat, S., Malik, Y. S., et al. (2020). Coronavirus disease 2019–COVID-19. Clin. Microbiol. Revs. 33:e00028-20. doi: 10.1128/CMR.00028-20

Dong, Y., Dai, T., Wei, Y., Zhang, L., Zheng, M., and Zhou, F. (2020). A systematic review of SARS-CoV-2 vaccine candidates. Signal Transduct. Target. Ther. 5:237. doi: 10.1038/s41392-020-00352-y

Douglas, R. G., Sharma, R. K., Masuyer, G., Lubbe, L., Zamora, I., Acharya, K. R., et al. (2014). Fragment-based design for the development of N-domain-selective angiotensin-1-converting enzyme inhibitors. Clin. Sci. 126, 305–313. doi: 10.1042/cs20130403

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015).). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43, W389–W394. doi: 10.1093/nar/gkv332

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., and Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 6:veaa061. doi: 10.1093/ve/veaa061

Eaaswarkhanth, M., Al Madhoun, A., and Al-Mulla, F. (2020). Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? *Int. J. Infect. Dis.* 96, 459–460. doi: 10.1016/j.ijid.2020.05.071

Faria, N. R., Morales Claro, I., Candido, D., Moyses Franco, L. A., Andrade, P. S., Coletti, T. M., et al. (2021). *Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in Manaus: Preliminary Findings*. Available online at: https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586. (accessed April 10, 2021)

Fehr, A. R., and Perlman, S. (2015). "Coronaviruses: an overview of their replication and pathogenesis," in *Coronaviruses: Methods and Protocols, Methods in Molecular Biology*, eds H. J. Maier, E. Bickerton, and P. Britton (New York, NY: Humana Press), doi: 10.1007/978-1-4939-2438-7_1

Ferron, F., Subissi, L., Silveira De Morais, A. T., Le, N. T. T., Sevajol, M., Gluais, L., et al. (2017). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U.S.A.* 115, E162–E171. doi: 10.1073/pnas.1718806115

Fung, T. S., and Liu, D. X. (2018). Post-translational modifications of coronavirus proteins: roles and function. *Future Virol.* 13, 405–430. doi: 10.2217/fvl-2018-0008

Fung, T. S., and Liu, D. X. (2019). Human coronavirus: host-pathogen interaction. *Annu. Rev. Microbiol.* 73, 529–557. doi: 10.1146/annurev-micro-020518-115759

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). Protein identification and analysis tools on the ExPASy server. In: "Walker J.M. (ed) The *Proteomics Protocols Handbook*, New York, NY: Humana Press, 571-607.

Geijtenbeek, T. B. H., and van Kooyk, Y. (2003). Pathogens target DC-SIGN to influence their fate: DC-SIGN functions as a pathogen receptor with broad specificity. *APMIS* 111, 698–714. doi: 10.1034/j.1600-0463.2003.11107803.x

Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J.-C., Turner, A. J., et al. (2020). Angiotensin-Converting Enzyme 2: SARS-CoV-2 receptor and regulator of the Renin-Angiotensin System. *Circ. Res.* 126, 1456–1474. doi: 10.1161/circresaha.120.317015

Ghosh, S., Dellibovi-Ragheb, T. A., Kerviel, A., Pak, E., Qiu, Q., Fisher, M., et al. (2020). β-Coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell* 183, 1520-1535.e14. doi: 10.1016/j.cell.2020.10.039

Glasgow, A., Glasgow, J., Limonta, D., Solomon, P., Lui, I., Zhang, Y., et al. (2020). Engineered ACE2 receptor traps potently neutralize SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* 117, 28046–28055. doi: 10.1073/pnas.2016093117

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A., et al. (2020). The species Severe acute respiratory syndrome related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z

Graham, B. S. (2020). Rapid COVID-19 vaccine development. *Science* 368, 945–946. doi: 10.1126/science.abb8923

Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020). 3D Models of glycosylated SARS-CoV-2 spike protein suggest challenges and opportunities for vaccine development. *bioRxiv* [Preprint] doi: 10.1101/2020.04.07.030445

Grubaugh, N. D., Hanage, W. P., and Rasmussen, A. L. (2020). Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell* 182, 794–795. doi: 10.1016/j.cell.2020.06.040

Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. doi: 10.1002/elps.1150181505

Guex, N., Peitsch, M. C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30, S162–S173. doi: 10.1002/elps.200900140

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., et al. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* 27, 119–129. doi: 10.1038/cr.2016.152

Guo, H.-F., and Vander Kooi, C. W. (2015). Neuropilin functions as an essential cell surface receptor. *J. Biol. Chem.* 290, 29120–29126.

Habchi, J., Tompa, P., Longhi, S., and Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chem. Rev.* 114, 6561–6588. doi: 10.1021/cr400514h

Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pöhlmann, S. (2014). TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J. Virol.* 88, 1293–1307. doi: 10.1128/JVI.02202-13

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14, 378–379. doi: 10.1093/bioinformatics/14.4.378

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. doi: 10.1016/j.cell.2020.02.052

Hofmann, H., and Pöhlmann, S. (2004). Cellular entry of the SARS coronavirus. *Trends Microbiol.* 12, 466–472. doi: 10.1016/j.tim.2004.08.008

Hotez, P. J., Corry, D. B., and Bottazzi, M. E. (2020). COVID-19 vaccine design: the Janus face of immune enhancement. *Nature Rev. Immunol.* 20, 347–348. doi: 10.1038/s41577-020-0323-4

Hu, B., Guo, H., Zhou, P., and Shi, Z.-L. (2020a). Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev. Microbiol.* 19, 141–154. doi: 10.1038/s41579-020-00459-7

Hu, J., He, C.-L., Gao, Q.-Z., Zhang, G.-J., Cao, X.-X., Long, Q.-X., et al. (2020b). The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity and decreases neutralization sensitivity to individual convalescent sera. *bioRxiv* [Preprint] doi: 10.1101/2020.06.20.161323

Huo, J., Le Bas, A., Ruza, R. R., Duyvesteyn, H. M. E., Mikolajek, H., Malinauskas, T., et al. (2020). Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nature Struct. Mol. Biol.* 27, 846–854. doi: 10.1038/s41594-020-0469-6

Ikai, A. (1980). Thermostabililty and aliphatic index of globular proteins. *J. Biochem.* 88, 1895–1898. doi: 10.1093/oxfordjournals.jbchem.a133168

Iwasaki, A., and Yang, Y. (2020). The potential danger of suboptimal antibody responses in COVID-19. *Nat. Rev. Immunol.* 20, 339–341. doi: 10.1038/s41577-020-0321-6

Jiang, Y., Yin, W., and Xu, H. E. (2020). RNA-dependent RNA polymerase: Structure, mechanism, and drug discovery for COVID-19. *Biochem. Biophys. Res. Commun.* 538, 47–53. doi: 10.1016/j.bbrc.2020.08.116

Jovčevska, I., and Muyldermans, S. (2020). The therapeutic potential of nanobodies. *BioDrugs* 34, 11–26. doi: 10.1007/s40259-019-00392-z

Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., et al. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* 584, 115–119. doi: 10.1038/s41586-020-2380-z

Junier, T., and Pagni, M. (2000). Dotlet: diagonal plots in e Web browser. *Bioinformatics* 16, 178–179. doi: 10.1093/bioinformatics/16.2.178

Käll, L., Krogh, A., and Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027–1036. doi: 10.1016/j.jmb.2004.03.016

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sinica B.* 10, 1228–1238. doi: 10.1016/j.apsb.2020.04.009

Kaslow, D. C. (2020). Certainty of success: three critical parameters in coronavirus vaccine development. *NPJ Vaccines* 5:42. doi: 10.1038/s41541-020-0193-6

Kaur, S. P., and Gupta, V. (2020). COVID-19 Vaccine: a comprehensive status report. *Virus Res.* 288, 198114. doi: 10.1016/j.virusres.2020.198114

Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., et al. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature.* 588, 498–502. doi: 10.1038/s41586-020-2665-2

Kikkert, M. (2020). Innate immune evasion by human respiratory RNA viruses. *J. Innate Immun.* 12, 4–20. doi: 10.1159/000503030

Klasse, P. J. (2014). Neutralization of virus infectivity by antibodies: old problems in new perspectives. *Adv. Biology* 2014:157895. doi: 10.1155/2014/157895

Koehler, M., Delguste, M., Sieben, C., Gillet, L., and Alsteens, D. (2020). Initial step of virus entry: virion binding to cell-surface glycans. *Annu. Rev. Virol.* 7, 143–165. doi: 10.1146/annurev-virology-122019-070025

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G

increases infectivity of the COVID-19 virus. *Cell* 182, 812–827.e19. doi: 10. 1016/j.cell.2020.06.043

Krammer, F. (2020). SARS-CoV-2 vaccines in development. *Nature* 586, 516–527. doi: 10.1038/s41586-020-2798-3

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Kumanogoh, A., and Kikutani, H. (2013). Immunological functions of the neuropilins and plexins as receptors for semaphorins. *Nat. Rev. Immunol.* 13, 802–814.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5

Lanza, K., Perez, L. G., Costa, L. B., Cordeiro, T. M., Palmeira, V. A., Ribeiro, V. T., et al. (2020). Covid-19: the renin-angiotensin system imbalance hypothesis. *Clin. Sci.* 134, 1259–1264. doi: 10.1042/CS20200492

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* 5, 562–575. doi: 10.1038/s41564-020-0688-y

Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3, 237–261. doi: 10.1146/annurev-virology-110615-042301

Li, G., Fan, Y., Lai, Y., Han, T., Li, Z., Zhou, P., et al. (2020a). Coronavirus infections and immune responses. *J. Med. Virol.* 92, 418–423. doi: 10.1002/jmv.25685

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., et al. (2020b). The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182, 1284–1294. doi: 10.1016/j.cell.2020.07.012

Li, W., Hulswit, R. J. G., Widjaja, I., Raj, V. S., McBride, R., Peng, W., et al. (2017). Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 114, E8508–E8517. doi: 10.1073/pnas.1712592114

Li, W., Shi, Z., Yu, M., Ren, W., Craig Smith, C., Epstein, J. H., et al. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310, 676–679. doi: 10.1126/science.1118391

Li, Y.-D., Chi, W.-Y., Su, J.-H., Ferrall, L., Hung, C.-F., and Wu, T.-C. (2020c). Coronavirus vaccine development: from SARS and MERS to COVID-19. *J. Biomed. Sci.* 27, 104. doi: 10.1186/s12929-020-00695-2

Liu, Y., Wang, K., Massoud, T. F., and Paulmurugan, R. (2020). SARS-CoV-2 vaccine development: an overview and perspectives. *ACS Pharmacol. Transl. Sci.* 3, 844–858. doi: 10.1021/acsptsci.0c00109

Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., et al. (2013). Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500, 227–231. doi: 10.1038/nature12328

Luini, A. (2011). A brief history of the cisternal progression-maturation model. *Cell. Logistics* 1, 6–11. doi: 10.4161/cl.1.1.14693

Magalhaes, G. S., Rodrigues-Machado, M. G., Motta-Santos, D., Campagnole-Santos, M. J., and Santos, R. A. S. (2020). Activation of Ang-(1-7)/Mas receptor is a possible strategy to treat coronavirus (SARS-CoV-2) infection. *Front. Physiol.* 11:730. doi: 10.3389/fphys.2020.00730

Magleby, R., Westblade, L. F., Trzebucki, A., Simon, M. S., Rajan, M., Park, J., et al. (2020). Impact of SARS-CoV-2 viral load on risk of intubation and mortality among hospitalized patients with coronavirus disease 2019. *Clin. Infect. Dis.* ciaa851. doi: 10.1093/cid/ciaa851

Makwana, K. M., and Mahalakshmi, R. (2015). Implications of aromatic-aromatic interactions: from protein structures to peptide models. *Prot. Sci.* 24, 1920–1933. doi: 10.1002/pro.2814

Mansbach, R. A., Chakraborty, S., Nguyen, K., Montefiori, D. C., Korber, B., and Gnanakaran, S. (2020). The SARS-CoV-2 spike variant D614G favors an open conformational state. *bioRxiv* [Preprint] doi: 10.1101/2020.07.26.219741

Marth, J. D., and Grewal, P. K. (2008). Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* 8, 874–887. doi: 10.1038/nri2417

Masuyer, G., Schwager, S. L., Sturrock, E. D., Isaac, R. E., and Acharya, K. R. (2012). Molecular recognition and regulation of human angiotensin-I converting enzyme (ACE) activity by natural inhibitory peptides. *Sci. Rep.* 2:717. doi: 10.1038/srep00717

Matheson, N. J., and Lehner, P. J. (2020). How does SARS-CoV-2 cause COVID-19? The viral receptor on human cells plays a critical role in disease progression. *Science* 369, 510–511.

McBride, C. E., and Machamer, C. E. (2010). Palmitoylation of SARS-CoV S protein is necessary for partitioning into detergent-resistant membranes and cell-cell fusion but not interaction with M protein. *Virology* 405, 139–148. doi: 10.1016/j.virol.2010.05.031

Mészáros, B., Erdős, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucl. Acids Res.* 46, W329–W337. doi: 10.1093/nar/gky384

Miao, L., Zhang, Y., and Huang, L. (2021). mRNA vaccine for cancer immunotherapy. *Mol. Cancer* 20, 41. doi: 10.1186/s12943-021-01335-5

Monteil, V., Kwon, H., Prado, P., Hagelkrüys, A., Wimmer, R. A., Stahl, M., et al. (2020). Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* 181, 905–913. doi: 10.1016/j.cell.2020.04.004

Moore, J. B., and June, C. H. (2020). Cytokine release syndrome in severe COVID-19. *Science* 368, 473–474. doi: 10.1126/science.abb8925

Morniroli, D., Giannì, M. L., Consales, A., Pietrasanta, C., and Mosca, F. (2020). Human sialome and coronavirus disease-2019 (COVID-19) pandemic: an understated correlation? *Front. Immunol.* 11:1480. doi: 10.3389/fimmu.2020.01480

Nal, B., Chan, C., Kien, F., Siu, L., Tse, J., Chu, K., et al. (2005). Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S. *M and E. J. Gen. Virol.* 86, 1423–1434. doi: 10.1099/vir.0.80671-0

Naveca, F., Nascimento, V., Souza, V., Corado, A., Nascimento, F., Silva, G., et al. (2021). *Phylogenetic Relationship of SARS-CoV-2 Sequences from Amazonas with Emerging Brazilian Variants Harboring Mutations E484K and N501Y in the Spike Protein.* Available online at: https://virological.org/t/phylogenetic-relationship-of-SARS-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585/1. (accessed April 10, 2021)

Neefjes, J., Jongsma, M. M. L., and Berlin, I. (2017). Stop or Go? Endosome positioning in the establishment of compartment architecture, dynamics, and function. *Trends Cell. Biol.* 27, 580–594. doi: 10.1016/j.tcb.2017.03.002

Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., et al. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* 174, 11–22. doi: 10.1016/j.jsb.2010.11.021

Nguyen, T. H. V., Lichière, J., Canard, B., Papageorgiou, N., Attoumani, S., Ferrona, F., et al. (2018). Structure and oligomerization state of the C-terminal region of the Middle East respiratory syndrome coronavirus nucleoprotein. *Acta Crystallogr.* 75(Pt 1), 8–15. doi: 10.1107/S2059798318014948

Ni, W., Yang, X., Yang, D., Bao, J., Li, R., Xiao, Y., et al. (2020). Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Critical Care* 24:422. doi: 10.1186/s13054-020-03120-0

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6. doi: 10.1093/protein/10.1.1

Nwanochie, E., and Uversky, V. N. (2019). Structure determination by single-particle cryo-electron microscopy: only the sky (and intrinsic disorder) is the limit. *Int. J. Mol. Sci.* 20:4186. doi: 10.3390/ijms20174186

Ott, C. M., and Lingappa, V. R. (2002). Integral membrane protein biosynthesis: why topology is hard to predict. *J. Cell Sci.* 115, 2003–2009.

Ou, J., Zhou, Z., Dai, R., Zhang, J., Lan, W., Zhao, S., et al. (2020). Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural 2 stability and human ACE2 receptor affinity of the spike protein. *bioRxiv* [Preprint] doi: 10.1101/2020.03.15.991844

Papageorgiou, N., Lichière, J., Baklouti, A., Ferron, F., Sévajol, M., Canard, B., et al. (2016). Structural characterization of the N-terminal part of the MERS-CoV nucleocapsid by X-ray diffraction and small-angle X-ray scattering. *Acta Crystallogr.* 72, 192–202. doi: 10.1107/S2059798315024328

Pardi, N., Hogan, M. J., Porter, F. W., and Weissman, D. (2018). mRNA vaccines—a new era in vaccinology. *Nature Rev. Drug Discov.* 17, 261–279. doi: 10.1038/nrd.2017.243

Parker, M. W., Guo, H.-F., Li, X., Linkugel, A. D., and Vander Kooi, C. W. (2012). Function of members of the neuropilin family as essential pleiotropic cell surface receptors. *Biochemistry* 51, 9437–9446.

Patel, S., Homaei, A., El-Seedi, H. R., and Akhtar, N. (2018). Cathepsins: proteases that are vital for survival but can also be fatal. *Biomed. Pharmacother.* 105, 526–532. doi: 10.1016/j.biopha.2018.05.148

Pedersen, S. F., and Ho, Y.-C. (2020). SARS-CoV-2: a storm is raging. *J. Clin. Invest.* 130, 2202–2205. doi: 10.1172/JCI137647

Petit, C. M., Chouljenko, V. N., Iyer, A., Colgrove, R., Farzan, M., Knipe, D. M., et al. (2007). Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology* 360, 264–274. doi: 10.1016/j.virol.2006.10.034

Pinto, D., Park, Y. J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., et al. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* 583, 290–295. doi: 10.1038/s41586-020-2349-y

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., et al. (2020). Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *bioRxiv* [Preprint] doi: 10.1101/2020.09.01.278689

Poland, G. A., Ovsyannikova, I. G., Richard, B., and Kennedy, R. B. (2020). SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *Lancet* 396, 1595–1606. doi: 10.1016/S0140-6736(20)32137-1

Qing, E., Hantak, M., Perlman, S., and Gallagher, T. (2020). Distinct roles for sialoside and protein receptors in coronavirus infection. *mBio* 11, e2764–19. doi: 10.1128/mBio.02764-19

Rabi, F. A., Al Zoubi, M. S., Kasasbeh, G. A., Salameh, D. M., and Al-Nasser, A. D. (2020). SARS-CoV-2 and coronavirus disease 2019: what we know so far. *Pathogens* 9:231. doi: 10.3390/pathogens9030231

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., et al. (2020). *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel set of Spike Mutations.* Available online at: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-SARS-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563. (accessed January 15, 2021)

Riordan, J. F. (2003). Angiotensin-I-converting enzyme and its relatives. *Genome Biol.* 4:225. doi: 10.1186/gb-2003-4-8-225

Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., et al. (2020). Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* 586, 113–119. doi: 10.1038/s41586-020-2577-1

Rogers, T. F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W.-T., et al. (2020). Rapid isolation of potent SARS-CoV-2 neutralizing antibodies and protection in a small animal model. *bioRxiv preprint* doi: 10.1101/2020.05.11.088674

Romano, M., Ruggiero, A., Squeglia, F., Maga, G., and Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* 9:1267. doi: 10.3390/cells9051267

Romão, E., Morales-Yánez, F., Hu, Y., Crauwels, M., De Pauw, P., Hassanzadeh, G. G., et al. (2016). Identification of useful nanobodies by phage display of immune single domain libraries derived from camelid heavy chain antibodies. *Curr. Pharm. Design* 22, 6500–6518. doi: 10.2174/1381612822666160923114417

Ruddock, L. W., and Molinari, M. (2006). N-glycan processing in ER quality control. *J. Cell Sci.* 119, 4373–4380. doi: 10.1242/jcs.03225

Sabino, E. C., Buss, L. F., Carvalho, M. P. S., Prete, C. A. Jr., Crispim, M. A. E., Fraiji, N. A., et al. (2021). Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet.* 397, 452–455. doi: 10.1016/S0140-6736(21)00183-5

Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., et al. (2007). Ribonucleocapsid formation of Severe Acute Respiratory Syndrome Coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol.* 81, 3913–3921. doi: 10.1128/JVI.02236-06

Samavati, L., and Uhal, B. D. (2020). ACE2, much more than just a receptor for SARS-COV-2. *Front. Cell. Infect. Microbiol.* 10:317. doi: 10.3389/fcimb.2020.00317

Saponaro, F., Rutigliano, G., Sestito, S., Bandini, L., Storti, B., Bizzarri, R., et al. (2020). ACE2 in the era of SARS-CoV-2: controversies and novel perspectives. *Front. Mol. Biosci.* 7:588618. doi: 10.3389/fmolb.2020.588618

Schoeman, D., and Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virology J.* 16, 69. doi: 10.1186/s12985-019-1182-0

Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* 30, 981–988. doi: 10.1093/glycob/cwaa042

Shang, J., Wan, Y., Liu, C., Yount, B., Gully, K., Yang, Y., et al. (2020a). Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog.* 16:e1008392. doi: 10.1371/journal.ppat.1008392

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., et al. (2020b). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224. doi: 10.1038/s41586-020-2179-y

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., and Siddique, R. (2020). COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* 24, 91–98. doi: 10.1016/j.jare.2020.03.005

Shi, R., Shan, C., Duan, X., Chen, Z., Liu, P., Song, J., et al. (2020). A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* 584, 120–124. doi: 10.1038/s41586-020-2381-y

Shulla, H., and Gallagher, T. (2009). Role of spike protein endodomains in regulating coronavirus entry. *J. Biol. Chem.* 284, 32725–32734. doi: 10.1074/jbc.M109.043547

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Silveira, M. M., Moreira, G. M. S. G., and Mendonça, M. (2021). DNA vaccines against COVID-19: perspectives and challenges. *Life Scie.* 267:118919. doi: 10.1016/j.lfs.2020.118919

Skariyachan, S., Challapilli, S. B., Packirisamy, S., Kumargowda, S. T., and Sridhar, V. S. (2019). Recent aspects on the pathogenesis mechanism, animal models and novel therapeutic interventions for Middle East Respiratory Syndrome coronavirus infections. *Front. Microbiol.* 10:569. doi: 10.3389/fmicb.2019.00569

Słomińska-Wojewódzka, M., and Sandvig, K. (2015). The role of lectin-carbohydrate interactions in the regulation of ER-associated protein degradation. *Molecules* 20, 9816–9846. doi: 10.3390/molecules20069816

Snijder, E. J., Limpens, R. W. A. L., de Wilde, A. H., de Jong, A. W. M., Zevenhoven-Dobbe, J. C., Maier, H. J., et al. (2020). A unifying structural and functional model of the coronavirus replication organelle: tracking down RNA synthesis. *PLoS Biol* 18:e3000715. doi: 10.1371/journal.pbio.3000715

Stanley, P., Taniguchi, N., and Aebi, M. (2017). "N-Glycans," in *Essentials of Glycobiology,* 3rd Edn, eds A. Varki, R. D. Cummings, J. D. Esko, et al. (Cold Spring Harbor NY: Cold Spring Harbor Laboratory Press).

Strasser, R., Altmann, F., and Steinkellner, H. (2014). Controlled glycosylation of plant-produced recombinant proteins. *Curr. Opin. Biotechnol.* 30, 95–100. doi: 10.1016/j.copbio.2014.06.008

Sun, Z., Liu, Q., Qu, G., Feng, Y., and Reetz, M. T. (2019). Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem. Rev.* 119, 1626–1665. doi: 10.1021/acs.chemrev.8b00290

Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., et al. (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Med.* 26, 681–692. doi: 10.1038/s41591-020-0868-6

Surya, W., Li, Y., and Torres, J. (2018). Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta (Biomembranes)* 1860, 1309–1317. doi: 10.1016/j.bbamem.2018.02.017

Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., and Daniel, S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res.* 178:104792. doi: 10.1016/j.antiviral.2020.104792

Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A., and Ng, L. F. P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* 20, 363–374. doi: 10.1038/s41577-020-0311-8

Taylor, A., Foo, S.-S., Bruzzone, R., Dinh, L. V., King, N. J. C., and Mahalingam, S. (2015). Fc receptors in antibody-dependent enhancement of viral infections. *Immunol. Rev.* 268, 340–364. doi: 10.1111/imr.12367

Teesalu, T., Sugahara, K. N., Kotamraju, V. R., and Ruoslahti, E. (2009). C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16157–16162.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et al. (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* [Preprint] doi: 10.1101/2020.12.21. 20248640

Thépaut, M., Luczkowiak, J., Vivès, C., Labiod, N., Bally, I., Lasala, F., et al. (2020). DC/L-SIGN recognition of spike glycoprotein promotes SARS-CoV-2 trans-infection and can be inhibited by a glycomimetic antagonist. *bioRxiv* [Preprint] doi: 10.1101/2020.08.09.242917

Thomas, G. (2002). Furin at the cutting edge: From protein traffic to embryogenesis and disease. *Nature Rev. Mol. Cell Biol.* 3, 753–766. doi: 10.1038/nrm934

Thunders, M., and Delahunt, B. (2020). Gene of the month: TMPRSS2 (transmembrane serine protease 2). *J. Clin. Pathol.* 73, 773–776. doi: 10.1136/jclinpath-2020-206987

Tisoncik, J. R., Korth, M. J., Simmons, C. P., Farrar, J., Martin, T. R., and Katze, M. G. (2012). Into the eye of the cytokine storm. *Microbiol. Mol. Biol. Reviews* 76, 16–32. doi: 10.1128/MMBR.05015-11

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 37, 509–516. doi: 10.1016/j.tibs.2012.08.004

Towler, P., Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., et al. (2004). ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* 279, 17996–18007. doi: 10.1074/jbc.M311191200

Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., et al. (2012). Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys.Acta (Proteins and Proteomics)* 1824, 68–88. doi: 10.1016/j.bbapap.2011.10.002

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J. H., Welsch, S., Blanc, F. E. C., et al. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* 370, 203–208. doi: 10.1126/science.abd5223

Tusnády, G. E., and Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* 283, 489–506. doi: 10.1006/jmbi.1998.2107

Valitutto, M. T., Aung, O., Tun, K. Y. N., Vodzak, M. E., Zimmerman, D., Yu, J. H., et al. (2020). Detection of novel coronaviruses in bats in Myanmar. *PLoS One* 15:e0230802. doi: 10.1371/journal.pone.0230802

Vankadari, N., and Wilce, J. A. (2020). Emerging COVID-19 coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerg. Microb. Infec.* 9, 601–604. doi: 10.1080/22221751.2020. 1739565

Veit, M. (2012). Palmitoylation of virus proteins. *Biol. Cell* 104, 493–515.

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARSCoV2 spike glycoprotein. *Cell* 181, 281.e–292.e. doi: 10.1016/j.cell.2020.02.058281-292.e6,

Walls, A. C., Xiong, X., Park, Y.-J., Tortorici, M. A., Snijder, J., Quispe, J., et al. (2019). Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* 176, 1026–1039. doi: 10.1016/j.cell.2018.12.028

Wan, Y., Jian Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virology* 94, e00127–20. doi: 10.1128/JVI.00127-20

Wang, L.-F., and Anderson, D. E. (2019). Viruses in bats and potential spillover to animals and humans. *Curr. Opinion Virol.* 34, 79–89. doi: 10.1016/j.coviro. 2018.12.007

Wang, P., Wang, H., Gai, J., Tian, X., Zhang, X., Lv, Y., et al. (2017). Evolution of protein N-glycosylation process in Golgi apparatus which shapes diversity of protein N-glycan structures in plants, animals and fungi. *Sci. Rep.* 7:40301. doi: 10.1038/srep40301

Warner, F. J., Smith, A. I., Hooper, N. M., and Turner, A. J. (2004). Angiotensin-converting enzyme-2: a molecular and cellular perspective. *Cell. Mol. Life Sci.* 61, 2704–2713. doi: 10.1007/s00018-004-4240-7

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020a). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 369, 330–333. doi: 10.1126/science.abb9983

Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., et al. (2020b). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* 11:2688. doi: 10.1038/s41467-020-16567-0

Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta (Gen. Subjects)* 1863, 1480–1497. doi: 10.1016/j.bbagen.2019.05.012

Wild, J. R. L., Staton, C. A., Chapple, K., and Corfe, B. M. (2012). Neuropilins: expression and roles in the epithelium. *Int. J. Exp. Path.* 93, 81–103. doi: 10. 1111/j.1365-2613.2012.00810.x

Wolff, G., Limpens, R. W. A. L., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W. M., et al. (2020). A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* 369, 1395–1398. doi: 10.1126/science.abd3629

Wrapp, D., De Vlieger, D., Corbett, K. S., Torres, G. M., Nianshuang Wang, N., Van Breedam, W., et al. (2020a). Structural basis for potent neutralization of betacoronaviruses by single-domain camelid antibodies. *Cell* 181, 1–12. doi: 10.1016/j.cell.2020.04.031

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020b). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263.

Wu, Y., Wang, F., Shen, C., Peng, W., Li, D., Zhao, C., et al. (2020). A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* 368, 1274–1278.

wwPDB consortium (2019). rotein Data Bank: the single global archive for 3D macromolecular structure data. *Nucl. Acid Res.* 47, D520–D528. doi: 10.1093/nar/gky949

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., et al. (2020). 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.* 30, 343–355. doi: 10.1038/s41422-020-0305-x

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., et al. (2020a). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.* 12:8. doi: 10.1038/s41368-020-0074-x

Xu, S., Yang, K., Li, R., and Zhang, L. (2020b). mRNA vaccine era—mechanisms, drug platform and clinical prospection. *Int. J. Mol. Sci.* 21:6582. doi: 10.3390/ijms21186582

Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., et al. (2020). Molecular architecture of the SARS-CoV-2 virus. *Cell* 183, 730-738.e13. doi: 10.1016/j.cell. 2020.09.018

Yuan, M., Liu, H., Wu, N. C., Lee, C.-C. D., Zhu, X., Zhao, F., et al. (2020a). Structural basis of a shared antibody response to SARS-CoV-2. *Science* 369, 1119–1123. doi: 10.1126/science.abd2321

Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., et al. (2020b). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* 368, 630–633. doi: 10.1126/science.abb 7269

Yuan, Q., Liao, Y., Torres, J., Tam, J. P., and Liu, D. X. (2006). Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells. *FEBS Lett.* 580, 3192–3200. doi: 10.1016/j.febslet.2006.04.076

Zhang, C., Maruggi, G., Shan, H., and Li, J. (2019). Advances in mRNA vaccines for infectious diseases. *Front. Immunol.* 10:594. doi: 10.3389/fimmu.2019.00594

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Rangarajan, E. S., Izard, T., et al. (2020). The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* [preprint] doi: 10.1101/2020.06.12. 148726

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020a). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* 368, 409–412. doi: 10.1126/science. abb3405

Zhao, X., Guo, F., Comunale, M. A., Mehta, A., Sehgal, M., Jain, P., et al. (2015). Inhibition of endoplasmic reticulum-resident glucosidases impairs severe acute respiratory syndrome coronavirus and human coronavirus NL63 spike protein-mediated entry by altering the glycan processing of angiotensin I-converting enzyme 2. *Antimicrob. Agents Chemother.* 59, 206–216. doi: 10.1128/AAC. 03999-14

Zhou, D., Tian, X., Qi, R., Peng, C., and Zhang, W. (2020). Identification of 22 N-glycosites on spike glycoprotein of SARS-CoV-2 and accessible surface

glycopeptide motifs: implications for vaccination and antibody therapeutics. *Glycobiology* 31, 69–80. doi: 10.1093/glycob/cwaa052

Ziegler, C. G. K., Allons, S. J., Nyquist, S. K., Mbano, I. M., Miao, V. N., Tzouanas, C. N., et al. (2020). 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* 181, 1016–1035. doi: 10.1016/j.cell.2020.04.035

Zoufaly, A., Poglitsch, M., Aberle, J. H., Hoepler, W., Seitz, T., Traugott, M., et al. (2020). Human recombinant soluble ACE2 in severe COVID-19. *Lancet Respir. Med.* 8, 1154–1158. doi: 10.1016/S2213-2600(20)30418-5