Università degli studi del Molise
Dipartimento di Medicina e Scienze della Salute

Dottorato di Ricerca in
Medicina Traslazionale e Clinica

Ciclo XXXIV

Settore Scientifico Disciplinare:

MED/04 - PATOLOGIA GENERALE

Amphiphilic and ganglioside-binding properties of SARS CoV-2 Spike protein variants dissected by charge shift electrophoresis in detergent solution

Coordinatore del Dottorato:
Prof. Scapagnini Giovanni

Supervisore:
Prof. Gentile Fabrizio

DOTTORANDA:
Cimino Anna Rita
matricola 164352

Anno accademico 2020/2021

INDEX

1. INTRODUCTION

1.1. SARS-CoV-2 Disease (COVID-19)

Coronavirus-19 disease (also known as COVID-19) is a new disease caused by a coronavirus which emerged in 2019. Coronaviruses are a diverse group of viruses that infect many different animals and can cause mild to severe respiratory infections in humans (Ben *et al.*, 2021). Studies claim that coronaviruses belong to the *Orthocoronavirinae* subfamily of the *Coronaviridae* family of the order *Nidovirales*. They are further classified into four other CoV genera: Alphacoronavirus (α-CoV), Betacoronavirus (β-CoV), Deltacoronavirus (δ-CoV), and Gammacoronavirus (γ-CoV). The β-CoV genus is further classified into five subgenera or lineages. Genomic imaging has shown that perhaps bats and rodents are the gene sources of α-CoV and β-CoV (Shah *et al.*, 2020). In 2002 and 2012, respectively, two highly pathogenic coronaviruses with zoonotic origin, the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV), emerged in humans and caused respiratory diseases, making emerging coronaviruses a new public health concern. In late 2019, a new coronavirus designated SARS-CoV-2 emerged in the city of Wuhan, China, and caused unusual viral pneumonia. Being highly communicable, this new coronavirus disease, also known as coronavirus disease 2019 (COVID-19), has spread rapidly around the world, overwhelmingly outperforming SARS and MERS both in terms of number of infected people and extent of epidemic areas and representing an extraordinary threat to global public health (Ben *et al.*, 2021).

In late December 2019, several health facilities in Wuhan, China's Hubei province, reported groups of patients with pneumonia of unknown cause. Similar to patients with SARS and MERS, these

patients exhibited symptoms of viral pneumonia, including fever, cough and chest discomfort and, in severe cases, dyspnea and bilateral pulmonary infiltration. Among the first 27 documented hospitalized patients, most of the cases were epidemiologically linked to the Huanan Seafood Wholesale Market, located in downtown Wuhan, which sells not only seafood but also live animals, including poultry and wildlife. Independent teams of Chinese scientists have identified that the causative agent of this emerging disease was a β-coronavirus that had never been seen before. The first genome sequence of the coronavirus was published on the Virological website on January 10, and nearly complete genomic sequences determined by several research institutes were then published via the GISAID database on January 12, 2020. Subsequently, more patients with no history of exposure to the Huanan Seafood Wholesale Market were identified. Several family clusters of infection have been reported and nosocomial infection has also occurred in health care facilities. All of these cases have provided clear evidence of human-to-human transmission of the new virus. This new coronavirus pneumonia soon spread to other cities in Hubei Province and other parts of China. Within 1 month, it had spread massively to all 34 provinces of China. However, despite China's strict containment measures to the spread of the virus, the international spread of COVID-19 has accelerated since late February 2020. The high transmission efficiency of SARS-CoV-2 and the abundance of international travel have allowed for a rapid worldwide spread of COVID-19, so much that in March 2020 the WHO officially defined the global outbreak of COVID-19 a pandemic. According to the COVID-19 dashboard from the Johns Hopkins University Center for System Science and Engineering, as of August 11, 2020, 216 countries and regions on all six continents had reported more than 20 million cases of COVID-19 and more of 733,000 patients had died (Ben *et al.*, 2021).

1.2 The SARS-CoV-2 coronavirus

Coronaviruses belong to a group of enveloped viruses with a single-stranded positive-sense RNA belonging to the β genus of the *Coronaviridae* family and viral particles resembling the shape of a corona, hencefrom the name 'coronavirus'. They possess the largest identified RNA genome of up to 33.5 kilobases (kb), containing a 5 'cap structure alongside with a 3' poly-(A) tail, which allows them to act as mRNA for translation of replicase polyproteins. The viral genes coding for non-structural proteins (nsps) occupy two thirds of the genome (about 20 kb), as opposed to the genes coding for structural and accessory proteins, which make up only about 10 kb of the viral genome. The membrane of SARS-CoV-2 bears the transmembrane glycoprotein (M), the spike glycoprotein (S) and the envelope protein (E) and surrounds the flexible helical nucleocapsid (N). (Figure 1B). As a novel β-coronavirus, SARS-CoV-2 shares 79% genome sequence identity with SARS-CoV and 50% with MERS-CoV (Shang *et al.*, 2020). Six functional open reading frames (ORFs) are arranged in 5'-to-3' order. Furthermore, seven putative ORFs encoding accessory proteins are scattered among the structural genes. Most of the proteins encoded by SARS-CoV-2 are similar in length to the corresponding proteins in SARS-CoV. Of the four structural genes, SARS-CoV-2 shares more than 90% amino acid identity with SARS-CoV, with the exception of the S gene, which diverges (Figure 1A). From early genomic comparisons, it was clear that SARS-CoV-2 had a similar genomic organization to SARS-CoV. The spike proteins of both viruses have similar three-dimensional structures, suggesting that these viruses could use the same cell surface receptor: human angiotensin converting enzyme 2 (ACE2). This was soon confirmed *in vitro* and by structural biology data (Tian *et al.*, 2020). However, SARS-CoV-2 differs from SARS-CoV in two fundamental ways. First, there are six amino acid positions in the receptor binding domain (RBD) of the spike protein that mediate the attachment of SARS-CoV and SARS-CoV-2 spike proteins to the human ACE2 receptor. However, the amino acids in five of the six locations differ between SARS-CoV and SARS-CoV-2.

Notably, such differences result in SARS-CoV-2 having greater avidity for binding to the human ACE2 receptor and may have contributed to the higher transmissibility of SARS-CoV-2 compared to SARS-CoV. Second, there is an insertion of 12 nucleotides (nt) into the cleavage site of the SARS-CoV-2 Spike protein encoding four amino acids, PRRA, which can be recognized by the furin protease, which is widely expressed in different tissues and organs. This insertion may decrease the overall stability of the SARS-CoV-2 Spike protein, thus facilitating the adoption of the open conformation required for the binding of the Spike protein to human ACE2 (Hussain *et al.*, 2020).

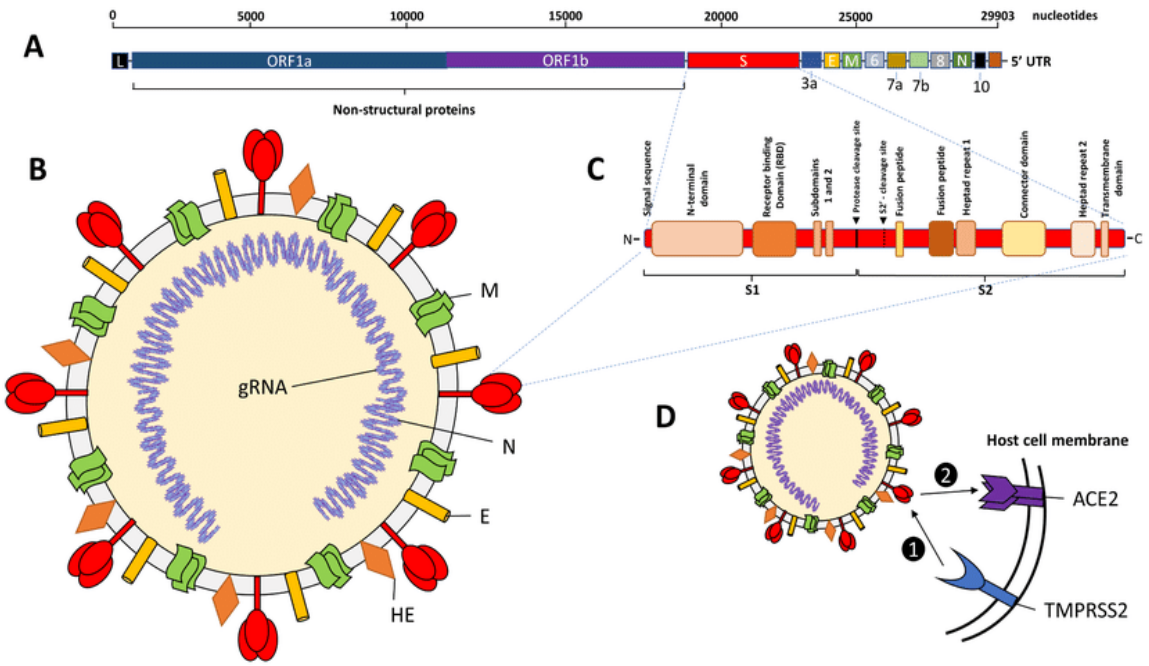

Figure 1: Genomic and molecular characteristics of the SARS-CoV-2 virus. *A.* The SAR-CoV-2 genome*: a* single-stranded, unsegmented RNA genome of 29,903 nucleotides*, w*ith 2 open read frames (ORF), ORF1a and ORF1b, which encode for non-structural proteins (nsps), viral spike protein (S), envelope protein (E), membrane protein (M), and nucleocapsid protein (N), as well as several presumed accessory proteins (3a, 6, 7a, 7b, 8 and 10). L: Leader 3 'sequence. UTR: 5 'untranslated region. B. Structure of the SARS-CoV-2 vir*i*on:

an enveloped virus containing major surface antigens, including the hemagglutinin esterase (HE) and the spike protein trimer (S), which surrounds the genomic RNA packaged in the nucleocapsid (N). C. Spike protein monomer protein structure (S) showing key molecular domains involved in pathogenesis. D. Primary cell host and co-receptor for SAR-CoV-2.

1.3 The Spike glycoprotein

The major determinant of coronavirus tropism is the Spike glycoprotein (S), with a full length of 1,273 amino acids, longer than the Spike of SARS-CoV-1 (1,255 amino acids) and bat SARSr-CoV (1,245–1,269 amino acids). It is distinct from the S proteins of most members of the Sarbecovirus subgenus and shares amino acid sequence similarities of 77.0% with human SARS-CoVs. Protein S is an important target for antiviral drug development, as it plays a vital role in binding to ACE2 receptors, viral attachment and entry into host cells (Qingxin *et al.*, 2022), and undergoes extensive structural rearrangements to promote membrane fusion. The protein is strongly glycosylated, with each protomer containing 22 N-linked glycosylation sites (Zhang *et al.*, 2021).

1.4. The SARS-CoV-2 variants

The SARS-CoV-2 genome continuously undergoes mutations, due to multiple factors that most frequently include errors during replication, damage to nucleic acids and modifications induced by the host's RNA editing systems as an antiviral defense. Recently, two deaminase families, adenosine deaminase and the catalytic polypeptide-like apolipoprotein B mRNA-modifying enzyme (ADAR and APOBEC, respectively), have been reported to be involved in mammalian antiviral defense. ADAR modifies double-stranded RNA while APOBEC can modify single stranded RNA or DNA. Editing of

ADAR and APOBEC type RNA was observed in the sequences derived from bronchoalveolar lavage fluids obtained from COVID-19 patients, along with the detection of APOBEC in the transcriptome from the same samples. APOBEC is otherwise undetectable in healthy tissues; hence, its detection supports host machinery-mediated viral RNA editing. Although these two enzymes act for the defense of host cells, their activity could contribute to mutations and evolution of the viral genome (Swapnil *et al.*, 2021). The mutations have led to the formation of new, genetically closely related variants with multiple variants documented in the United States and around the world during this pandemic. Common variants are alpha, first isolated in the UK associated with a 50% increase in transmission and increased mortality (likely to escape neutralization by most monoclonal antibodies targeting the N-terminal domain (NTD) of the spike protein); beta, first isolated in South Africa, with increased immune evasion, increased transmission by 50% (able to escape neutralization by most monoclonal antibodies targeting RBD and neutralizing antibodies from the plasma of convalescent patients from COVID) (Walls *et al.*, 2020); delta, isolated for the first time in India, almost twice as contagious as the previous variants, causes more serious forms of disease, is associated with increased mortality and able to evade people who are fully vaccinated; gamma, first isolated in Brazil and Japan, associated with probable increased transmissibility and disease severity; epsilon, first isolated in California, with a 20% increased risk of transmissibility; eta and iota, isolated for the first time in New York, with a probable increase in transmission speed. Other variants monitored are kappa, mu and zeta, and the two most recent variants of concern are the delta plus and omicron, currently found in more than 150 countries. Omicron was first isolated in South Africa and is currently the most common strain in both the UK and the US with over 50 spike protein mutations. To assess the genetic variation of different SARS-CoV-2 strains, the 2019 China National Center for Bioinformation Novel Coronavirus Resource aligned 77,801 genomic sequences of SARS-CoV-2 detected globally and identified a total of 15,018 mutations, 14,824 of which were single-nucleotide

polymorphisms. In protein S, four amino acid alterations, V483A, L455I, F456V, and G476S, are found near the binding interface in RBD, but their effects on host receptor binding are unknown. The D614G alteration in the S1 subunit was found much more frequently than other S variant sites and is the marker of a major subclade of SARS-CoV-2 (clade G). Since March 2020, SARS-CoV-2 variants with G614 in protein S have replaced the original D614 variants and have become the dominant form circulating globally. Compared to the D614 variant, higher viral loads were found in patients infected with the G614 variant, but clinical data did not suggest any significant link between the D614G alteration and disease severity. The D614G mutation in protein S appears to be associated increasingly with the spread of the virus and its meaning is functionally characterized. In the human lung epithelial cell line calu-3, this mutation increases the infectivity of the virus compared to the non-mutant type of the virus. The D614G mutation also does not appear to affect the S1/S2 cleavage, a characteristic which remains in favor of the virus. Therefore, there are still no concrete correlations between the reported mutations and the pathogenicity and/or spread of the virus. The study of these mutations will certainly be useful for predicting the future path of virus epidemics and also for the development of new antiviral drugs and the optimization of vaccines.

1.5. Fusogenic activity of viral spike glycoproteins

Fusogens are defined as proteins that catalyze membrane fusion. They pertain both to eukaryotic cells and to viruses. On one hand, in eukaryotes fusogens mediate such vital processes as gamete fusion, muscle differentiation, neurotransmitter release, ocular lens and placenta formation.

On the other hand, enveloped viruses gain access to eukaryotic cells by merging membranes with to release viral genetic material into the host cell cytoplasm. In all cases, membrane fusion cannot occur without the assistance of specialized proteins to overcome the repulsive forces between

membranes. Viruses have evolved four classes of viral fusion proteins (viral glycoproteins or fusogens) widely different as for their tertiary and oligomeric structures:

- class I viral fusion proteins (as found in influenza viruses, coronaviruses, HIV and Ebola virus) form coiled-coil trimers; class I viral fusion proteins are crucial for some of the deadliest viral infections of our era, including those caused by influenza viruses, coronaviruses, Ebola virus, Lassa virus and HIV; syncytin-1 and syncytin-2 are two fusion proteins encoded by human endogenous retroviruses (HEV), which are vital for placenta formation, which are grouped together with class I fusion proteins, on the base of structural and evolutionary connections;
- class II fusion proteins (as found in Dengue fever virus, West Nile virus, Zika virus and tick-borne encephalitis virus) transition from dimers to trimers during fusion, producing an elongated ectodomain composed of β sheets that settles into a hairpin trimer after fusion;
- class III proteins (as found in vesicular stomatitis virus, herpes simplex virus 1 and rabies virus) combine elements from the former two classes;
- class IV fusion-associated small transmembrane (FAST) proteins are cell-cell fusogens made by reoviruses to merge multiple host cells into a syncytium.

Class I, II and III fusogens) catalyze membrane fusion through a conserved mechanism (Figure _). Fusion proteins are presented on the surface of the virion in a metastable state with their hydrophobic 'fusion peptide' (class I) or 'fusion loop' (class II and III, and certain class I fusogens) buried in the fusogen. **An external trigger (e.g., receptor binding or a pH change) promotes a conformational change that releases the fusion peptide/loop, so that it can embed in the host cell membrane**. This short-lived, extended intermediate forms a physical tether between virus and host cell and readily collapses into a hemi-fusion state, where the two outer lipid layers fuse. The fusion pore expands as the inner lipid leaflets also fuse and the fusogen acquires a stable post-fusion conformation.

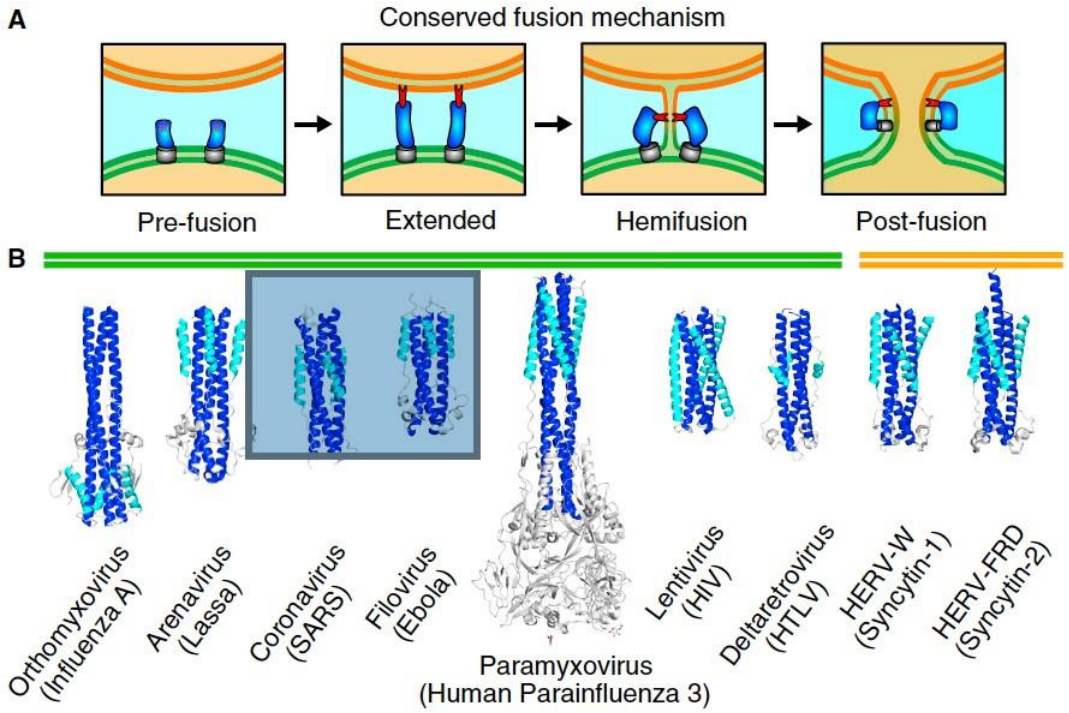

Figure 2 - Conserved mechanism for membrane fusion in the class I fusogen superfamily. A) Viral fusion proteins in *blue* are displayed in their pre-fusion state, with their carboxy-terminal transmembrane region (*gray*) attached to the virion membrane (*green*) and their hydrophobic fusion peptide or loop (*red*) buried. The protein then undergoes a series of conformational changes, extending to embed its fusion peptide/loop in the host cell membrane (*orange*), thus bringing the carboxy-terminal transmembrane region and fusion peptide/loop together in a post-fusion state. B) Structures of the major class I fusogens in their post-fusion state: influenza A (PDB: 1HTM), Lassa virus (PDB: 5OMI), SARS-CoV (PDB: 1WNC), Ebola (PDB: 2EBO), human parainfluenza virus-3 (PDB: 1ZTM), HIV-1 (PDB: 3WFV), human T-lymphocytic virus-1 (PDB: 1MG1), human syncytin-1 (PDB: 6RX1) and syncytin-2 (PDB: 6RX3). The amino-terminal helix (heptad repeat-1) is *blue*, the carboxy-terminal helix (heptad repeat-2) is *cyan*, and the connecting chain reversal region (if present) is *gray* (from Vance and Lee, 2020).

1.6. Fusogenic activity of Spike glycoproteins from SARS-CoV, SARS-CoV-2, and other coronaviruses

SARS-CoV-2 uses its trimeric Spike SARS-2-S glycoprotein (UniProtKB: P0DTC2) to bind to host cell ACE2 receptor and attain the fusion of its membrane envelope with the host cell membrane to gain entry into cells. This is a multi-step process involving proteolytic cleavages in different SARS-2-S glycoprotein domains, in a way that closely resembles the attachment and entry of SARS-CoV mediated by the SARS-S glycoprotein (UniProtKB: P59594). Both Spike glycoproteins are cleaved into S1 and S2 subunits. The S1 subunit serves the function of receptor-binding and contains a signal peptide (SP) at the N terminus, an N-terminal domain (NTD), and receptor-binding domain (RBD). The S2 subunit functions mainly in membrane fusion and contains the four elements that in all class I viral fusion proteins are required for membrane fusion: a fusion peptide (FP) or loop, an internal fusion peptide (IFP), two heptad-repeats (HR1 and HR2), and a transmembrane domain followed by a C-terminal domain (Figure 3) (Bosch *et al.*, 2003).

There appear to be two distinct ways of membrane fusion, associated with different pathways of viral entry into host cells. The first one, which permits the entry of an extracellular virus from the airway of human respiratory system, requires the fusion of the viral membrane and the target cell membrane. The second one, which involves the transfer of a virus from an infected cell to a neighboring cell, requires the formation of a syncytium through cell-cell fusion (Hoffmann *et al.*, 2020b, Belouzard *et al.*, 2009). The SARS-S and SARS-2-S glycoproteins undergo two crucial proteolytic cleavages, with the first cleavage at site S1/S2 (site 2) (residues 696-697 in SARS-2-S) splitting them into the S1 subunit (residues 14-685 in SARS-2-S) and the S2 subunit (residues 697-1273). The cleavage of the S protein into S1 and S2 is an essential step in viral entry into a host cell, and needs to occur before viral fusion with the host cell membrane (Lu *et al.*, 2015). Compared to

SARS-S, SARS-2-S contains an additional polybasic furin cleavage site (site 1) between residues 685-686, at the boundary between the S1 and S2 subunits. Cleavage at this site is essential for efficient viral entry into human lung cells, especially for cell-cell fusion into syncitia that facilitate viral spread (Hoffmann *et al.*, 2020b, Belouzard *et al.*, 2009). This additional furin site confers a broader cell tropism to SARS-CoV-2 compared to SARS-CoV, as furin is ubiquitously expressed in human tissues.

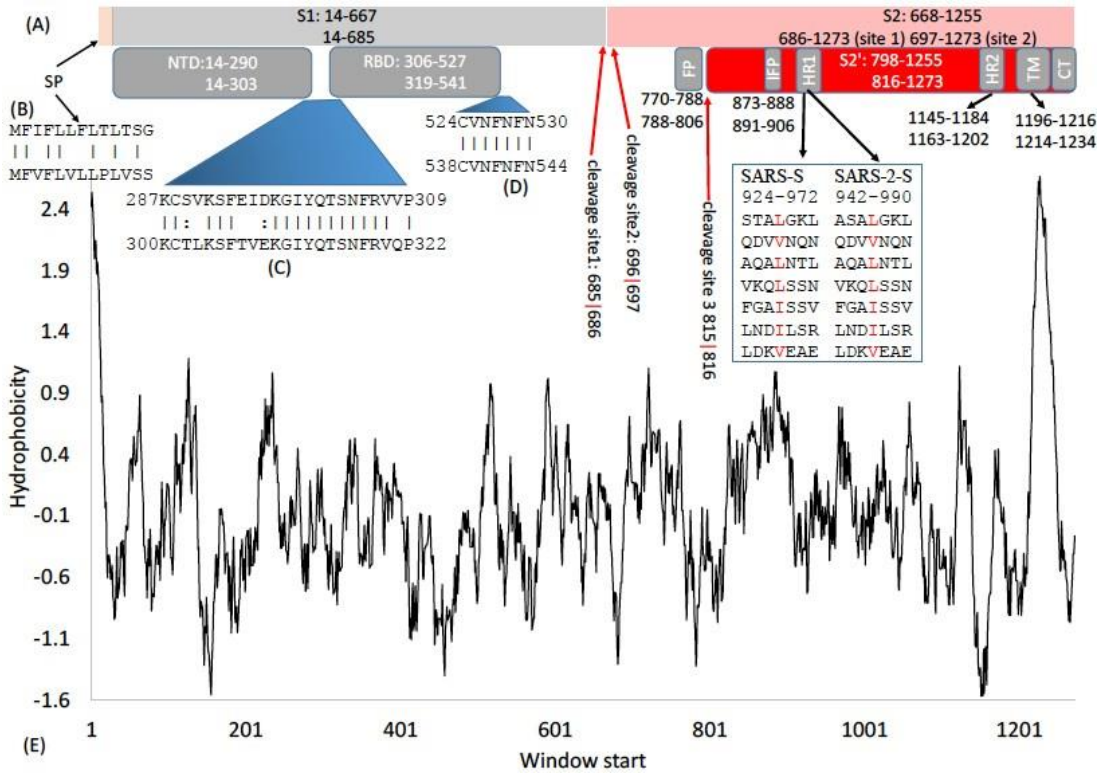

Figure 3 - Domain structure of SARS-S and SARS-2-S Spike proteins. A) Key domains. Legend: SP, signal peptide; NTD, *N*-terminal domain; *RBD*, receptor-binding domain; FP, fusion peptide; IFP, internal fusion peptide; HR, heptad repeats; TM, transmembrane domain; CT, cytoplasmic tail. Top and bottom numbers in each domain refer to SARS-S and SARS-2-S, respectively. *Red* arrows indicate cleavage sites (numbers refer to SARS-2-S); B, C and D) alignment of SP and two interdomain regions between SARS-S (*top*) and SARS-2-S (*bottom*), respectively; E) Hydrophobicity plot of SARS-2-S (from Xia, 2021).

Cleavage site 1 (Figure 3) is known to be cleaved during SARS-CoV-2 assembly, most likely by furin in the Golgi apparatus (Hoffmann *et al.*, 2020b; Walls *et al.*, 2020; Song *et al.*, 2004; Jaimes *et al.*, 2020). The S1 subunit remains non-covalently bound to the S2 subunit in the prefusion conformation after cleavage at site 1 (Kirchdoerfer *et al.*, 2016; Walls *et al.*, 2020; Belouzard *et al.*, 2009). In order to stabilize the prefusion conformation to facilitate vaccine design (Kirchdoerfer *et al.*, 2016; Hsieh *et al.*, 2020) or structural determination (Walls *et al.*, 2020; Wrapp *et al.*, 2020), the furin site is often mutated so that it is not cleaved.

In spite of the additional furin site 1, cleavage site S1/S2 (site 2) may still be important, being highly conserved in all sequenced SARS-CoV-2, as well as in all of its close relatives, such as SARS-CoV. This site appears to be cleaved by cathepsin L in endosomes in both SARS-S (Belouzard *et al.*, 2009; Simmons *et al.*, 2005; Bosch *et al.*, 2008; Burkard *et al.*, 2014) and SARS-2-S (Ou *et al.*, 2020).

A third cleavage at site S2' (residues 815-816 of SARS-2-S) splits S2 into the fusion peptide (FP) (residues 788-806) and the S2' subunit (residues 816-1273). This site is cleaved by trypsin-like membrane protease 2 (TMPRSS2) (Matsuyama *et al.*, 2010; Glowacka *et al.*, 2011; Kleine-Weber *et al.*, 2018). Site S2' is not cleaved during SARS-CoV assembly, as TMPRSS2 is active mainly in the membrane or the extracellular space (Song *et al.*, 2004; Matsuyama *et al.*, 2010).

Membrane fusion requires two anchors, one at the virion side and the other at the target cell side. With SARS-S and SARS-2-S, the transmembrane (TM) domain and cytoplasmic tail (CT) at the C terminus form an anchor inside the virion envelope, while the FP domain of the S2 subunit (or the IFP domain of the S2' subunit when FP is cleaved off) serves to penetrate the target cell membrane and anchor inside it. All known viral fusion peptides form trimers (White *et al.*, 2008; Bosch *et al.*, 2004; Broer *et al.*, 2006). The S2 (or S2') domain between the two anchors needs to undergo a conformational change which may bring the two membranes together for fusion. A trigger is

required for this conformational transition, which is typically a cleavage that occurs either at the cell surface at neutral pH or within an endosome at a reduced pH. These alternative cleavages correspond to two viral entry pathways in SARS-S and SARS-2-S, i.e., the membrane-TMPRSS2 pathway and the endosome-cathepsin L pathway (Glowacka *et al.*, 2011; Kleine-Weber *et al.*, 2018; Matsuyama *et al.* 2005; Matsuyama *et al.*, 2010). The cleavage of S2 at site S2' (i.e., the Arg797-Ser798 bond in SARS-S or the Arg815-Ser816 bond in SARS-2-S) by TMPRSS2, exposing IFP at the N-terminal of the S2' subunit, appears to be a reliable signal to the virus that a target cell is within reach. This is consistent with the finding that a target cell needs to express TMPRSS2 to be infected (Matsuyama *et al.*, 2010). Some residues located between FP and IFP in SARS-S, i.e., Cys822, Cys833 (with the two forming a disulfide bond with each other), Asp830 and Leu831 (corresponding to Cys840, Cys851, Asp848 and Leu849 in SARS-2-S) are critically important for membrane fusion (Madu *et al.*, 2009), which may explain why cleavage at site S2' markedly enhances membrane fusion and viral entry (Hoffmann *et al.*, 2020a; Bosch *et al.*, 2008; Matsuyama *et al.*, 2010). Thus, cleavage at site S2' by TMPRSS2 exposed at the surface of the target cell triggers membrane fusion, syncytium formation and SARS-CoV entry into target cells via the membrane-TMPRSS2 pathway (Hoffmann *et al.*, 2020a; Walls *et al.*, 2020; Belouzard *et al.*, 2009). In the absence of TMPRSS2, viral entry can occur via the endosome-cathepsin L pathway, through endocytosis and cleavage at site S2' to split S2 into FP and S2' by endosomal cathepsin L (Matsuyama *et al.*, 2010; Matsuyama *et al.* 2005). Protease cleavage at site S2' likely follows S1/S2 cleavage and may not occur until host-receptor engagement at the plasma membrane or viral endocytosis (Millet *et al.*, 2015).

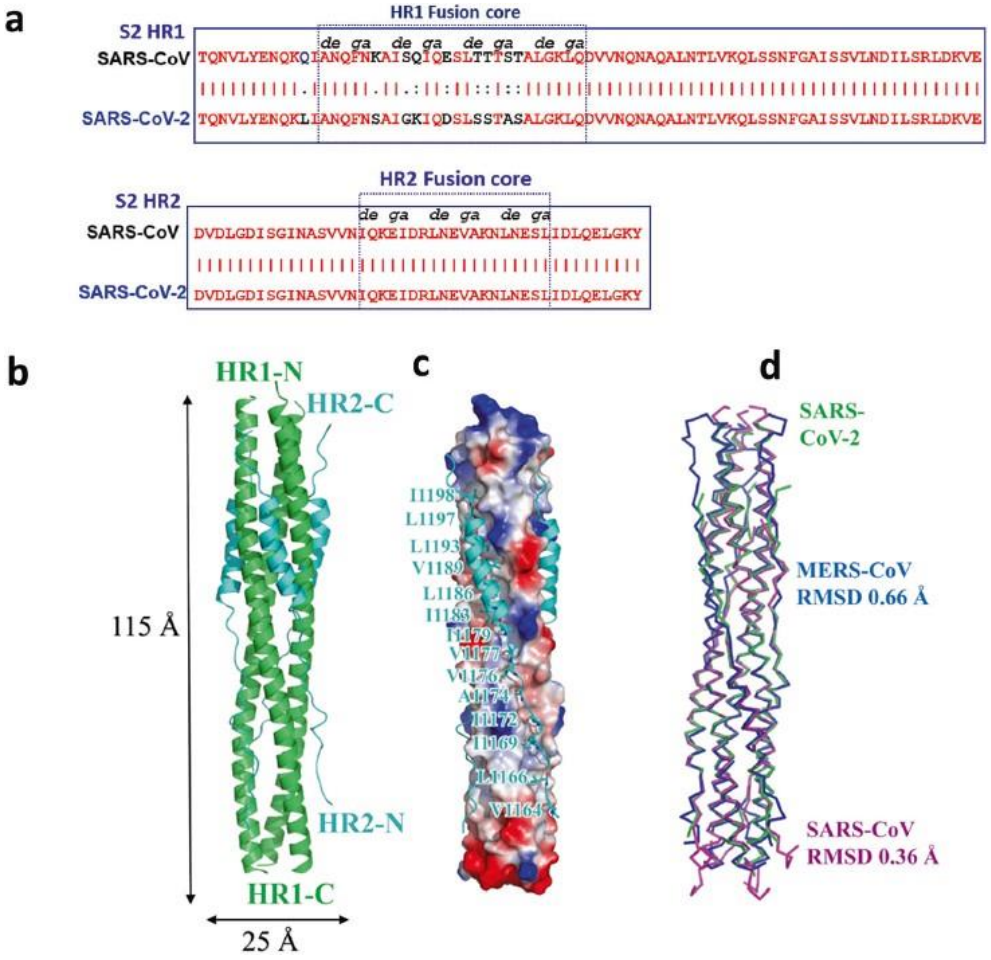

Figure 4 - Structure of post-fusion six-helical bundle (67-HB) of heptad repeats in SARS-CoV-2. a) Sequence alignment of HR1 and HR2 domains in SARS-CoV and SARS-CoV-2; b) structure of SARS-CoV 6-HB in cartoon representation, with HR1 colored in *green* and HR2 in *cyan*; c) HR1 trimer of SARS-CoV-2 6-HB shown in electrostatic surface, and HR2 domain in cartoon representation, with the critical residues shown in sticks and labeled; d) superposition of 6-HB structure of SARS-CoV (PDB: 1WYY), MERS (PDB: 4NJL) and SARS-CoV-2 (PDB: 6LXT) shown in ribbon, with root-mean-square deviation (RMSD) indicated (from Xia *et al.*, 2020b).

Heptad repeats HR1 and HR2 are highly conserved domains located downstream the FP and IFP domains of Spike glycoproteins of human coronaviruses (Figure 4) which also appear to play a role in the membrane fusion process mediated by the MERS-S (Lu *et al.*, 2014), SARS-S (Liu *et al.*, 2004),

and SARS-2-S glycoproteins (Xia *et al.*, 2020a). They consist of repeated, helix-forming $(abcdefg)_n$ 7mers which, in helical wheel representation, present hydrophobic residues on the same side of the helix and form a hydrophobic interface which permits the formation of a six-helix bundle (6-HB) between the three HR1 domains and the three HR2 domains of a Spike homotrimer (Bosch *et al.*, 2004; Lu *et al.*, 2015; Basak *et al.*, 2008). Xia *et al.* developed a cholesteryl derivative of a fusion-inhibiting peptide targeting the HR1 domain, which potently inhibited membrane fusion and pseudoviral infection in an *in vitro* assay of cell-cell fusion between Huh-7 or 293T/ACE2 target cells and 293T cells transfected with expression vectors for the Spike glycoproteins of SARS-CoV-2, MERS-CoV, SARS-CoV and other SARS-related coronaviruses (SARSr-CoVs) (Xia *et al.*, 2019; Xia *et al.*, 2020).

HR1 and HR2 are separated by a central helix. Structural studies on the Spike glycoproteins of the β-coronaviruses HCoV-HKU1 (Kirchdoerfer *et al.*, 2016) and MERS-CoV (Xia *et al.*, 2020b) have demonstrated that replacing two consecutive amino acids with proline near the transition from HR1 to the central helix contribute to the stabilization of the resulting Spike glycoproteins at the prefusion conformation, which is important for structure determination and vaccine development. These amino acids are located at sites 986 and 987 in SARS-2-S. A mutagenized SARS-2-S Spike protein with the two amino acid replacements Lys986Pro and Val987Pro is encoded in the mRNA vaccine from both Pfizer/BioNTech (BNT162b2) and Moderna (mRNA-1273) and is referred to as S-2P (Anderson *et al.*, 2020; Jackson *et al.*, 2020). Another Spike protein variant (HexaPro) which includes four additional amino acid replacements by proline (Phe817Pro, Ala892Pro, Ala899Pro, and Ala942Pro) is even more stable and expressed in higher yield than S-2P (Hsieh *et al.*, 2020).

The transmembrane (TM) domain of the Spike glycoprotein consists of a juxtamembrane part enriched in aromatic amino acid residues, a central hydrophobic region, and a cysteine-rich part. The TM domain is followed by a highly hydrophilic cytoplasmic tail (CT). The central hydrophobic region forms a helix, so that three transmembrane helices interact with each other in Spike

homotrimers. The TM and CT domains contribute to the stabilization of the trimeric structure, which is important for membrane fusion, as its destabilization is associated with reduced fusogenic activity and infectivity (Song *et al.*, 2004; Xiao *et al.*, 2004; Broer *et al.*, 2006). In fact, the efficiency of infection is reduced by the replacement of hydrophobic residues in the central part by lysine (Corver *et al.*, 2009), while substitution of the palmitoylated cysteine residues closest to the central hydrophobic part inhibits membrane fusion (Petit *et al.*, 2007).

Kirchdoerfer *et al.* have provided **a structural basis to support a model of membrane fusion mediated by progressive destabilization of the Spike protein through the concurrent effects of receptor binding and proteolytic cleavage** (Kirchdoerfer *et al.*, 2005). They presented the cryo-EM structure to 4.0 Å resolution of the human-endemic β-coronavirus HCoV-HKU1 in the prefusion conformation and confronted it with the structure of other viral fusion membrane proteins which are also cleaved into a receptor-binding subunit and a membrane fusion machinery, such as the external domain of influenza A hemagglutinin (HA) (Wilson *et al.*, 1981), the cleaved trimeric HIV-1 envelope (Env) glycoprotein in its prefusion state (Julien *et al.*, 2013; Lyumkis *et al.*, 2013) and SARS-CoV Spike protein in the post-fusion conformation (Duquerroy *et al.*, 2005). Similar to HA and HIV-1 Env protein, a region in the HCoV-HKU1 S1 CTD caps the S2 central helix, thereby preventing the fusion machinery from springing into action, which makes cleavage at the furin-cleavage site at the S1/S2 junction crucial for the entry process. Cleavage at site S2' likely follows S1/S2 cleavage and requires the action either of TMPRSS2 expressed at the surface of target cells or of cathepsin L in endosomes, following S protein binding to host plasma membrane receptors (Millet *et al.*, 2015).

As in all class I viral fusion proteins, also in HCoV-HKU1 the refolding of the HR1 domain into a long α-helix thrusts the fusion peptide (FP) into the host cell membrane, and as HR1 and HR2 domains of the three subunits of the Spike protein interact to form an anti-parallel six-helix bundle (6-HB), the host and viral membranes are brought together. The formation of such a coiled-coil structure in

the post-fusion conformation is a unifying feature of class I viral fusion proteins. The fusion peptide is located on the exterior of the HCoV-HKU1 S protein, with most of its hydrophobic amino acid residues buried in an interface with the S2 subunit. It is separated from the N terminus of HR1 by a stretch of 60 amino acids, encompassing four short α-helices and several non-structured segments, which is buried beneath subdomain SD-2 of the S1 subunit and the S2' cleavage site, suggesting that **cleavage at the S2' site may strongly affect the propensity of S2 to undergo the transition to the post-fusion conformation**. The post-fusion 6-HB structures of SARS-CoV and MERS-CoV heptad repeats indicate that the conversion to the post-fusion conformation requires that the structural elements composing HR1 undergo a transition into a single long α-helix (Duquerroy *et al.*, 2005; Lu *et al.*, 2005). In Spike proteins of coronaviruses such a transition entails five loop-to-helix transitions and is more complex (Figure 5), compared to the analogous transition of influenza HA, whose HR1 is located at a distance of only 14 amino acid residues from the fusion peptide and consists of two α-helices separated by a long loop (Bullough *et al.*, 1994). In addition, the C-terminal portions of the three HR1 segments, which form a tight three-helix bundle in the center of the 6-HB in the post-fusion state, are diverted outwards in the prefusion state. The formation of this three-helix bundle may be prevented by interactions between the C-terminal end of HR1 and the CTD of the S1 subunit, and that **disruption of these interactions through metastable conformational changes may be an additional mechanism whereby receptor binding by the S1 subunit may initiate membrane fusion mediated by the S2 subunit**. In summary, the prefusion Spike protein of HCoV-HKU1 appears to be progressively destabilized and matured by receptor binding and proteolytic cleavage. Following dissociation of the S trimer into S1 and S2 subunits, each HR1 would undergo a transition to a long α-helix and form a six-helix bundle, while the fusion peptides would be released from the sides of the S2 subunits and get inserted into the host cell membrane (Kirchdoerfer *et al.*, 2016).

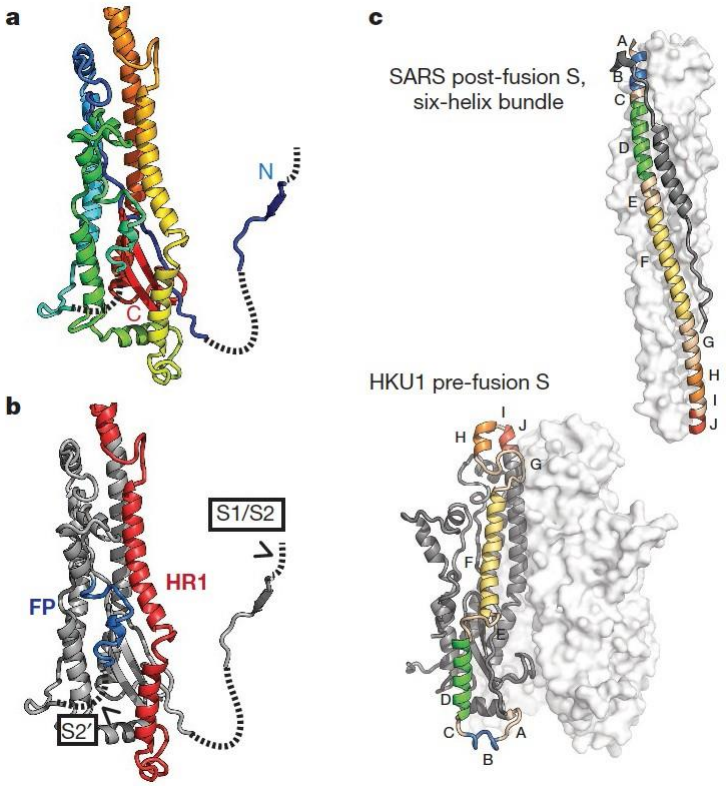

Figure 5 - Structure of post-fusion S2 subunit fusion machinery of the HCoV HKU1 β-coronavirus. a) The HKU1 S2 subunit is coloured like a rainbow from the N terminus (*blue*) to the C terminus (*red*) before the start of HR2; b) HKU1 S2 structure containing the fusion peptide (FP) and heptad repeat HR1. Proteolytic cleavage sites S1/S2 and S2' are indicated within disordered regions of the protein (*dashed lines*); c) a comparison of coronavirus S2 in the pre- and post-fusion conformations. Five HR1 a-helices are labeled and coloured like rainbow from *blue* (N terminus) to *red* (C terminus) (from Kirchdoerfer *et al.*, 2016).

1.7. Sialoside-binding domains and the conserved sialoside attachment strategy of Coronaviruses and Orthomyxoviruses

A broad range of viruses use sialic acid residues linked to glycoproteins and gangliosides as attachment sites for cell entry, including influenza viruses (Verma *et al.*, 2018) and coronaviruses

(Lu *et al.*, 2008; Tortorici *et al.*, 2019). Coronaviruses preferentially interact with 9-*O*-acetyl-N-acetylneuraminic acid (9-*O*-Ac-Sia, or 9-*O*-SIA) (Vlasak *et al.*, 1988; Peng *et al.*, 2012; Huang *et al.*, 2015; Matrosovich *et al.*, 2015). The depletion of sialic acids linked to glycoconjugates at the host cell surface by neuraminidase treatment was shown to inhibit the entry of MERS-CoV into human epithelial cells (Li *et al.*, 2017).

Coronaviruses of interest for human health include:

- two zoonotic β-coronaviruses, such as the severe acute respiratory syndrome coronavirus (SARS-CoV), which emerged in 2002 and was responsible for an epidemic in five continents with a fatality rate of 10%, and the Middle East respiratory syndrome coronavirus (MERS-CoV), a β-coronavirus which emerged in the Arabian Peninsula in 2012 and caused recurrent outbreaks in humans with a fatality rate of 35%; these were both zoonotic viruses that crossed the species barriers using bats/palm civets and dromedary camels, respectively;

- four coronaviruses of zoonotic origin, which are endemic in the human population, such as the α-coronaviruses HCoV-NL63 and HCoV-229E, and the β-coronaviruses HCoV-OC43 and HCoV-HKU1; these account for up to 30% of mild respiratory tract infections and cause severe complications or fatalities in young children, the elderly and immunocompromised individuals. HCoV-OC43 originated from zoonotic transmission of a bovine coronavirus (BCoV).

The attachment of viruses to sialic-acid-containing glycoproteins and gangliosides at the surface of the epithelial lining of the host respiratory tract is mediated by receptor-binding proteins that belong to the viral spike. In coronaviruses, this function is fulfilled by the Spike (S) glycoprotein. Some human coronaviruses depend for their binding both upon specific plasma membrane protein receptors, such as the angiotensin-converting enzyme-2 (ACE-2), which acts as a primary attachment site for both SARS-CoV (Li *et al.*, 2003) and SARS-CoV-2 (Yan *et al.*, 2020; Hoffmann *et al.*, 2020a), and upon sialic-acid-containing glycoconjugates.

Coronaviruses use homotrimers of the Spike (S) glycoprotein, a class I viral fusion protein, to attach to host cells and promote the fusion of viral and cellular membranes for entry. The S glycoprotein is the main antigenic target of neutralizing antibodies at the viral surface. As discussed in the previous section, in many coronaviruses the S glycoprotein is hydrolysed by host proteases into subunits S1 and S2, which remain non-covalently bound in the pre-fusion conformation. The N-terminal S1 subunit includes four β-rich domains, with domain A or B acting as the receptor-binding domain in different coronaviruses. The trans-membrane C-terminal S2 subunit is the metastable fusion machinery. During entry, S2 is further cleaved at site S2', immediately upstream the fusion peptide. This second cleavage step activates the S glycoprotein for membrane fusion, via irreversible conformational changes.

The Spike glycoproteins of HCoV-OC43, BCoV and porcine hemagglutinating encephalomyelitis virus (PHEV) bind to 9-*O*-acetyl-N-acetylneuraminic acid (9-*O*-Ac-Sia) terminally linked to oligosaccharides of glycoproteins and gangliosides at the host cell surface (Vlasak *et al.*, 1988; Peng *et al.*, 2012; Huang *et al.*, 2015; Matrosovich *et al.*, 2015). The 9-*O*-Ac-Sia binding site is conserved among these viruses and resides in domain A. MERS-CoV binds to α2,3-linked (and to a lesser extent to α2,6-linked) sialic acids, with sulfated sialyl-Lewis X being the preferred binder (Li *et al.*, 2017). On the other hand, the viral hemagglutinin-esterase (HE) protein exerts a receptor-destroying, sialate-*O*-acetyl-esterase activity, which facilitates the release of viral progeny from infected cells and escape from attachment to non-permissive host cells (De Groot, 2006; Desforges *et al.*, 2013; Bakkers *et al.*, 2017; Zeng *et al.*, 2008). HE proteins share these properties with hemagglutinin-fusion-esterase (HEF) glycoproteins of influenza C and D viruses (Rosenthal *et al.*, 1998; Song *et al.*, 2016). Other coronaviruses like infectious bronchitis virus (IBV, γ-coronavirus), porcine epidemic diarrhea virus (α-coronavirus) and transmissible gastroenteritis virus (α-coronavirus) also bind to sialoglycans

distinct from 9-*O*-Ac-sialosides, via their A domains in the course of host infection (Wickramasinghe *et al.*, 2011; Liu *et al.*, 2015; Schultze *et al.*, 1996).

HKU1-S S1 was shown to bind *O*-acetylated (other than 9-*O*-acetylated) sialic acids on human rhabdomyosarcoma (RD) host cells, whose pretreatment with neuraminidase and trypsin greatly reduced the binding. *O*-acetylated sialoconjugates were required for infection of primary human airway epithelial (HAE) cell cultures by HCoV-HKU1, which could be inhibited by pretreatment of HAE cells with HKU1 hemagglutinin esterase (HKU1-HE) endowed with receptor-destroying sialate-*O*-acetylesterase activity (Huang *et al.*, 2015). However, antibodies directed against the carboxy-terminal domain (CTD) of the S1 subunit, but not those directed against the N-terminal domain (NTD) of HCoV-HKU1 blocked infection of target cells by HCoV-HKU1 (Qian *et al.*, 2015), suggesting that S1 CTD harbors the primary HCoV-HKU1 protein-receptor-binding site, like in SARS-CoV (Li *et al.*, 2003) and MERS-CoV (Mou *et al.*, 2013), even though a protein receptor has not yet been identified for HCoV-HKU1.

Kirchdoerfer *et al.* presented the structure of the HCoV-HKU1 Spike protein ectodomain determined using cryo-electron microscopy (cryo-EM) (Kirchdoerfer *et al.*, 2016). The S1 NTD had strong structural and sequence homology to the BCoV S1 NTD, which, as mentioned above, recognizes 9-*O*-acetyl-sialic acid on glycosylated cell surface receptors (Peng *et al.*, 2012). The glycan-binding site found in the BCoV S1 NTD was conserved in the HCoV-HKU1 S1 NTD at the apex of the Spike homotrimer. Structural alignment of the SARS-CoV and MERS-CoV CTD-receptor complexes (Li *et al.*, 2005; Lu *et al.*, 2013) with the HCoV-HKU1 prefusion Spike protein revealed that the protein-receptor-binding surface of the S1 CTD is buried in the HCoV-HKU1 Spike protein homotrimer and is therefore unable to make equivalent interactions without some initial breathing and transient exposure of these domains. This is the result of the S1 CTDs being interdigitated and forming extensive quaternary interactions that occlude surfaces known in other coronavirus to bond protein

receptors. These observations suggest that **the initial attachment of HCoV-HKU1 to target cells via glycan binding mediated by the NTD may be responsible for the conformational transition needed to expose the receptor-binding site located in the S1 CTD** (Kirchdoerfer *et al.*, 2016). In principle, the metastable modifications initiated by sialoside binding to the S1 NTD may affect also the capping of the S2 central helix by the S1 CTD, the release of the fusion peptide from the side of the S2 subunit and the loop-to-helix transitions within HR1 that are required for the formation of the 6-HB.

More recently, the cryo-EM structures were determined of a stabilized HCoV-OC43 Spike glycoprotein trimer (in which the S1/S2 furin cleavage site was abrogated to prevent proteolytic processing during biogenesis), in isolation and in complex with 5-N-acetyl,9-*O*-acetyl-neuraminic acid α-methyl glucoside (9-*O*-Ac-Me-Sia) (Tortorici *et al.*, 2019). 9-*O*-Ac-Me-Sia bound with fast association/dissociation kinetics in a groove at the surface of domain A of HCoV-OC43 S protein and the residues involved in ligand binding proved to be essential for HCoV-OC43 entry into host cells.

The sialoside-interacting groove defined two hydrophobic pockets, designated P1 and P2, separated by the Trp90 indole side chain, and was delineated by two loops forming the rims of the binding site, termed L1 and L2. The 9-*O*-Ac-Me-Sia C1-carboxylate formed a salt bridge with the Lys81 side chain amine and a H-bond with the Ser83 side chain hydroxyl, while the 5-nitrogen atom of the ligand participated in a H-bond with the Lys81 backbone carbonyl; on the other hand, the ligand N-acetyl methyl inserted into the P2 hydrophobic pocket, defined by residues Leu80, Trp90 and Phe95, while the 9-*O*-acetyl methyl docked into the P1 hydrophobic pocket, comprising residues Leu85, Leu86 and Trp90, whereas the 9-*O*-acetyl carbonyl made a H-bond with the Asn27 side chain amide (Figure 7). A similar combination of hydrogen and hydrophobic bonding was observed in the 9-*O*-Ac-Sia binding sites of coronavirus Spike glycoproteins and of orthomyxovirus HEs/HEFs. In particular, all but one among the residues participating in the interaction with 9-*O*-Ac-Me-Sia in HCoV-OC43 Spike were conserved in BCoV Spike and PHEV Spike (with Ser83 being substituted by

Thr83), and many of the ligand-interacting residues or resides indirectly involved in the formation of the recognition site were also conserved in HCoV-HKU1 Spike (Tortorici *et al.*, 2019).

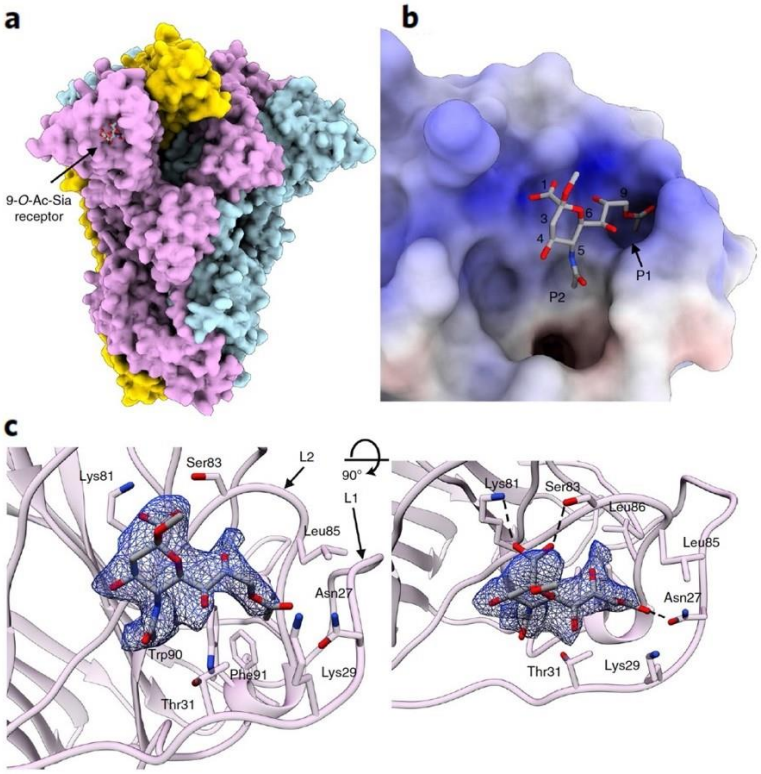

Figure 7. Sialoglycan-binding site in the holo-HCoV-OC43 S glycoprotein structure. a) Molecular surface representation of the holo-HCoV-OC43 S ectodomain trimer structure with the bound 9-*O*-Ac-Me-Sia ligand shown as sticks, with protomers individually colored; b) surface representation of the ligand-binding site colored by electrostatic potential; c) two orthogonal views of the binding site. The A domain is rendered as a ribbon diagram with the side chains of surrounding residues shown as sticks and the cryo-EM density shown as a *blue* mesh. In a) to c), the ligand is rendered as sticks with atoms colored by element (carbon, *gray*; nitrogen, *blue*; oxygen, *red*). Dashed lines show a salt bridge and hydrogen bonds (from Tortorici *et al.*, 2019).

Noticeably, **the binding of free 9-*O*-Ac-Me-Sia did not induce fusogenic conformational changes of the Spike glycoprotein, suggesting that multivalent interactions with sialoglycans (via mechanical destabilization of the pre-fusion trimers) and/or binding to some proteinaceous receptor are essential for promoting membrane fusion**. The receptor-interacting site is conserved in all coronavirus Spike proteins that are known to bind 9-*O*-Ac-sialoglycans and shares similarity with the ligand-binding pockets of coronavirus HE and influenza C/D viruses HEF glycoproteins (Langereis *et al.*, 2015; Bakkers *et al.*, 2016). Differences in the 3D structure of sialoside-binding sites account for the selectivity of different viruses for unmodified or modified sialic acids. The ligand-binding sites of a subset of coronavirus Spike glycoproteins, BCoV HE and influenza HEFs specifically recognize 9-*O*-Ac-Sia via hydrogen bonding with the 9-*O*-acetyl carbonyl moiety and formation of a hydrophobic pocket accommodating the 9-*O*-acetyl methyl (Bakkers *et al.*, 2017; Zeng *et al.*, 2008; Rosenthal *et al.*, 1998; Hulswit *et al.*, 2019; Langereis *et al.*, 2012). In contrast, influenza A/B hemagglutinin cannot bind 9-*O*-acetylated sialic acids, but a subset can bind to *N*-glycolyl neuraminic acids (Xiong *et al.*, 2013; Ito *et al.*, 1997). It is of note that, although the Spike glycoproteins of HCoV-OC43, HCoV-HKU1, BCoV and PHEV share the ligand specificity of influenza C/D HEFs, they function more like influenza A/B hemagglutinin, as they mediate receptor attachment and membrane fusion, whereas a dedicated hemagglutinin esterase (in coronaviruses) or neuraminidase (in influenza A/B viruses) is responsible for the receptor-destroying activity.

1.8. A putative ganglioside-binding domain (GBD) in SARS-CoV-2 Spike glycoprotein revealed by molecular modelling

Fantini *et al.* (2020) identified a ganglioside-binding site in the N-terminal domain (NTD) (amino acid residues 14-303) of the Spike (S) glycoprotein of SARS-CoV-2. Like all coronavirus Spike proteins, the

SARS-CoV-2 S glycoprotein consists of a trimer of S proteins, each possessing two distinct domains, namely the receptor-binding domain (RBD) and the N-terminal domain (NTD), which are distant from the viral envelope (Wrapp *et al.*, 2020). The NTD comprises 290 amino acid residues (14-303). A flat interface was found at its tip, which was proposed to target ganglioside-rich plasma membrane lipid rafts (Fantini *et al.*, 2020) (Figure 8).

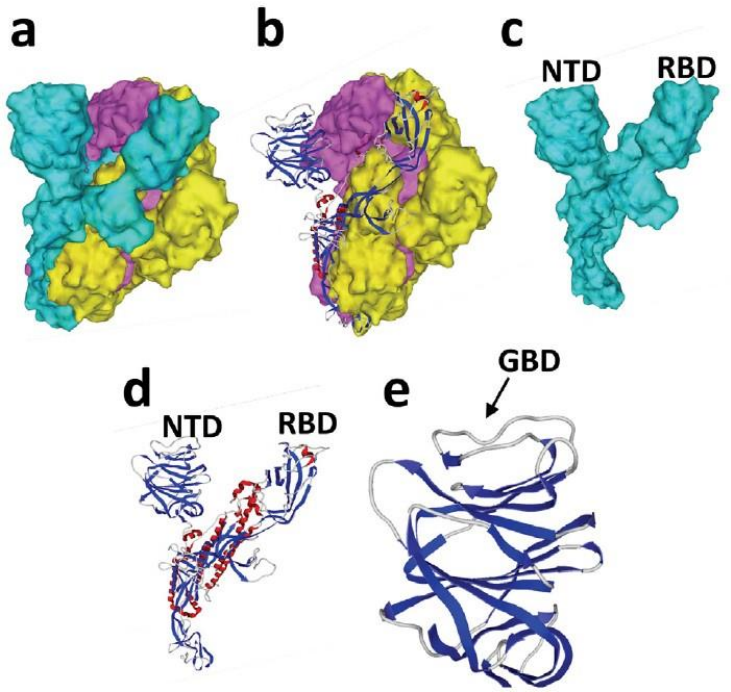

Figure 8 - Structural characteristics of the SARS-CoV-2 Spike (S) protein. a) Trimeric structure: each subunit has a distinct surface color, *blue*, *yellow* and *purple*; b) Ribbon representation of the *blue* S subunit, with alpha-Helices (*red*), beta-strands (*blue*) and random coil (*gray*); c) Surface structure of the isolated *blue* S subunit; d) Ribbon structure of the isolated *blue* S subunit; e) Zoom on the N-terminal domain (NTD) of the *blue* S subunit, in which the 'flat' interface and the putative ganglioside-binding domain are highlighted (from Fantini *et al.*, 2020).

Yahi *et al.* defined a sequence motif conferring ganglioside-binding properties to proteins, which was composed by a triad of conserved amino acid residues such as (K,R)-X_n-(F,Y,W)-X_n-(K,R), with two X_n intercalating segments of four-to-five non-conserved residues, with a preference for Gly, Pro and/or Ser (Yahi and Fantini, 2014). The same Authors investigated the presence of such a motif within the NTD of the SARS-CoV-2 S protein: even though they could not find strict conformity to the consensus, they noticed an over-representation of aromatic and basic residues in a 30-amino acid stretch, spanning amino acid residues 129-158 (Fantini *et al.*, 2020):

129-KVCEFQFCNDPFLGVYYHKNNKSWMESEFR-158

Molecular dynamic simulations of a structural subdomain encompassing amino acid residues 100-175 of the NTD (PDB: 6VSB) merged with ganglioside GM1 (whose coordinates were from CHARMM-GUI Glycolipid Modeler http://www.charmmgui.org/?doc=input/glycolipid) supported the concept that a large flat surface at the tip of the NTD could harbor a ganglioside-attachment interface, as it fitted precisely with the protruding oligosaccharide mojety of the GM1 ganglioside (Figure 9).

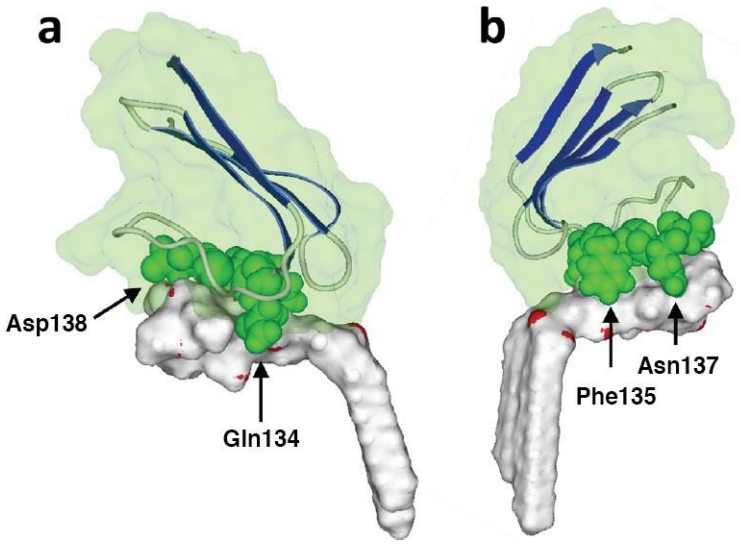

Figure 9 – Symmetric views of the molecular complex between the N-terminal domain (NTD) of SARS-CoV-2 Spike protein and a single GM1 molecule, with the saccharide moiety of GM1 providing a landing surface for the tip of the NTD. The NTD part involved in the interaction with GM1 is represented in ribbons superimposed onto a transparent *light green* surface rendering of the rest of the NTD. Amino acid residues !134 and D138 located in the center of the GM1-binding site are represented as *green* spheres (from Fantini *et al.*, 2020).

Several amino acid residues appeared to be critical for the interaction with GM1, particularly Phe135, Asn137 and Arg158 (Table _). The complex involved 10 amino acid residues for a total energy of interaction of -100 kJ•mol^{-1}. About 50% of the NTD flat surface was involved in the complex, leaving the remaining 50% available for the possible interaction with a second GM1 molecule, whose merging with the preformed GM1-NTD complex led to a trimeric complex consisting of two GM1 molecules in a symmetrical chalice-like structure, into which the putative GBD of the S protein NTD could be inserted (Figure 10).

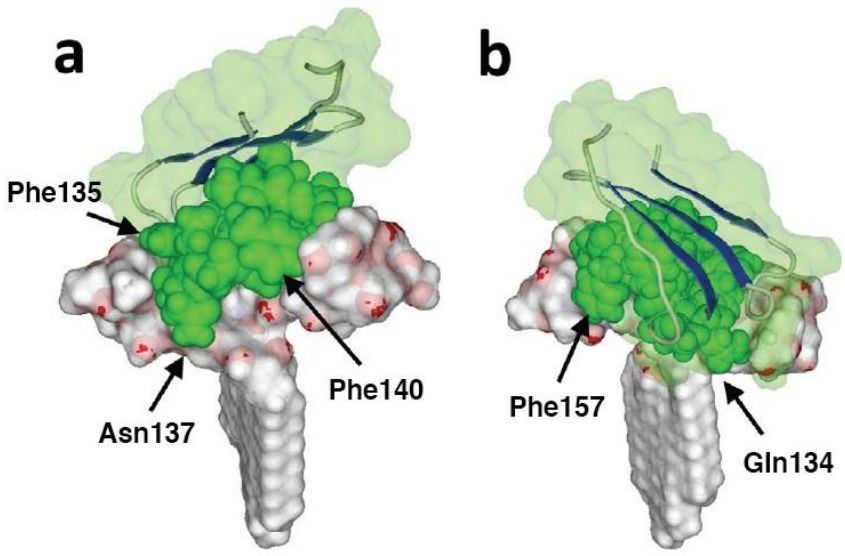

Figure 10 – Double view of the molecular trimeric complex between the N-terminal domain (NTD) of SARS-CoV-2 Spike protein and a GM1 dimer. The NTD is represented as in Figure _. Amino acids Gln134 to Ser162 belonging to the proposed ganglioside-binding domain (GBD) are visualized as *green* spheres. The GM1 dimer offers a larger interfacial surface to the NTD, compared to a single GM1 molecule (from Fantini *et al.*, 2020).

The binding of the second GM1 molecule entailed a conformational rearrangement of the first GM1-NTD complex. The energy of the interaction of the trimolecular complex reached the estimated value of -137 $kJ \bullet mol^{-1}$. Sixteen surface-accessible amino acid residues within the Asp111-Ser162 stretch were involved in the interaction, with Asp111, Gln134, Phe135, Arg158 and Ser161 being the most critical ones (Table _). In particular, the NTD binds to GM1 by a dyad of functional interactions: a) a network of hydrogen bonds between Asn137 and the GalNAc residue of GM1 and b) a geometrically perfect CH-π stacking interaction between the aromatic ring of Phe135 and the pyranose ring of the Glc residue of GM1.

Table 1 - Amino acid residues composing the proposed ganglioside-binding domain (GBD, residues 111-162) in the N-terminal domain (NTD) of SARS-CoV-2 Spike protein, with energies of interaction

A. Ganglioside-binding motif			
(K,R)- X n -(F,Y,W)-X n -(K,R)			
B. Amino acid residues engaged in GM1 binding predicted from ganglioside-binding motif			
111-DS**K**TQSLLIVNNATNVVI**K**VCE**F**Q**F**CNDP**F**LGV**YY**H**K**NN**K**S**W**MESE**FR**VYSS-162			
C. Amino acid residues involved in GM1 binding detected in modelling studies (*bold blue*)			
111-**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**VYYHKNNKSWMESE**FR**...V**YSS**-162			
D. Energies of interaction of amino acid residues in contact with GM1 gangliosides			
First step: one GM1 molecule	Energy of interaction (kj × mol^{-1})	Second step: two GM1 molecules	Energy of interaction (kj × mol^{-1})
Asp111	**-5.6**	**Asp111**	**-15.8**
		Ser112	-10.7
Lys113	**-8.2**	**Lys113**	**-9.2**
Gln134	**-8.6**	**Gln134**	**-11.2**
Phe135*	**-20.1**	**Phe135***	**-10.5**
Cys136	-7.0	Cys136	-6.2
Asn137**	**-15.2**	**Asn137****	**-4.7**
Asp138	-6.4		
		Phe140	-5.2
		Gly142	-5.6
		Glu156	-9.0
		Phe157	-13.8
Arg158	**-17.4**	**Arg158**	**-19.8**
		Tyr160	-3.2
Ser161	**-9.7**	**Ser161**	**-9.7**
Ser162	-2.0	Ser162	-2.0
Total	-100.2	Total	-136.6

Legend: in section A and B of this table, relevant residues are marked in *bold red*; in section C, relevant residues are marked in *bold blue*; in section D, residues indicated as most critical for the interaction of the SARS-2-S protein with ganglioside GM1 are also marked in *bold blue*.

* CH-π stacking with Glc pyrane ring of GM1

** H-bonded to GalNAc of GM1

Ref.: Fantini J, *et al.* Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents* 55 (2020) 105960

It is on these bases that Fantini *et al.* proposed that region 111-162 in the NTD of the SARS-CoV-2 Spike protein has the properties of a ganglioside-binding domain (GBD), containing an extended consensus with a basic residue at each end (Lys113 and Arg158) and a critical aromatic residue (Phe135) in the middle. In their dual receptor model of the attachment of SARS-CoV-2 to epithelial cells of the respiratory tract, the RBD binds to ACE-2 receptor, while the NTD interacts with a ganglioside-rich patch, such as a lipid raft, at the cell surface nearby (Figure 11).

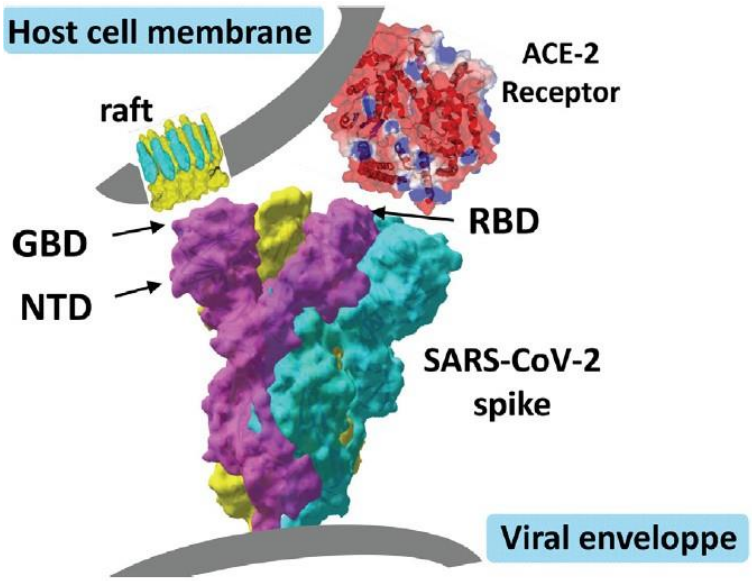

Figure 11 – Proposed dual recognition and attachment strategy of SARS-CoV-2 Spike protein to angiotensin-converting enzyme 2 (ACE-s) and gangliosides. The RBD binds to the ACE-2 receptor, while the NTD binds to ganglioside-rich lipid rafts (from Fantini *et al.*, 2020).

1.9. Probing the amphiphilic character of SARS-CoV-2 Spike glycoprotein by charge shift electrophoresis in detergent solutions

On the base of what was stated in the previous sections, it may be asked whether SARS-CoV-2 Spike glycoprotein exhibits amphiphilic properties. Amphiphilicity is a general property of any compound that contains two distinct, covalently bonded components, one of which has a high affinity for polar solvents, such as water, and another has a strong affinity for nonpolar solvents. Such a property is met in detergents and surfactants consisting of a polar water-soluble group attached to a water-insoluble hydrocarbon chain. Amphiphiles may have several hydrophilic and lipophilic parts. Glyco(sphingo)lipids are typical amphiphilic constituents of cell membranes, composed of a hydrophilic polar sugar headgroup and a hydrophobic apolar lipid mojety anchoring the molecule in the lipid bilayer. The sugar mojety may vary from small saccharide units to very large polysaccharide chains. Gangliosides are typical glycosphingolipids, with a distinct amphiphilic character.

Integral membrane proteins are normally amphiphilic, due to the hydrophobic domains that anchor them to the lipid bilayer, so that their extraction from membranes in a form in which they retain both solubility and a conformation as native as possible requires the use of detergents as substitutes for lipids normally interacting with transmembrane domains. Instead, soluble proteins and extrinsic membrane proteins are generally hydrophilic in their native conformations. Particular interest is deserved by amphipathic membrane active peptides, which exhibit the ability to reside at lipid-water interfaces and interact with membrane lipids: examples are antimicrobial peptides (AMP),

capable of disrupting cell membranes, causing cell lysis, and cell-penetrating peptides (CPP), which are able to translocate through membranes to deliver cargos into cells (Avci *et al.*, 2018).

Some degree of amphiphilicity of SARS-CoV-2 Spike glycoprotein can be expected on the base of the following two assumptions:

1. the need for the Spike glycoprotein to undergo the formation of transient intermediates, upon the transition from the prefusion to the post-fusion conformation, triggered by receptor binding and proteolytic cleavage, in which hydrophobic amino acid stretches and/or patches previously buried in the hydrophobic interior of the protein or hidden by interfacial interactions become at least transiently accessible to the solvent, as seen in the previous sections;

2. the presumed ability of the SARS-CoV-2 glycoprotein to bind sialosides, such as the GM1 ganglioside, which still awaits experimental confirmation. In regard, the structure of HCoV-OC43 Spike glycoprotein in complex with 5-N-acetyl,9-*O*-acetyl-neuraminic acid α-methyl glucoside has shown that binding requires a combination of salt bridges, hydrogen and hydrophobic bonds between charged, polar and apolar groups of the ligand and the Spike glycoprotein (Tortorici *et al.*, 2019). Also the molecular modelling of the interaction between GM1 and the SARS-2-S glycoprotein evidenced a mixed network of ionic, polar and hydrophobic interactions (Fantini *et al.*, 2020).

Helenius and Simons described a charge shift electrophoretic method for distinguishing between hydrophilic and amphiphilic proteins, based on their different interactions with mild detergents such as Triton X-100 (*p*-*t*-octylphenyl polyoxyethylene$_{9\text{-}10}$). Whereas soluble ordinary proteins and extrinsic membrane proteins bind little or no Triton X-100, amphiphilic membrane proteins bind large amounts of it (usually 80-100 moles of Triton X-100 per protein mole) when solubilized from membranes. The bound detergent forms micelle-like clusters around the hydrophobic domains of these proteins, which usually retain their native conformation. These Authors used mixtures of Triton X-100 and charged detergents, i.e., sodium deoxycholate (DOC) or cetyltrimethylammonium

bromide (CTAB) (Figure 12). When solubilized with such mixtures, the amphiphilic proteins form detergent-protein complexes containing both neutral and charged detergent molecules. The net charges of the complexes are thus dependent on the charge of the detergent used, resulting in a clear-cut difference in electrophoretic mobility of the amphiphilic proteins when electrophoresed in cationic or anionic detergent mixtures. Integral membrane proteins with an extensive hydrophobic domain display both an anodal shift in Triton X-100/DOC and a cathodal shift in Triton X-100/CTAB, as compared to their mobility in Triton X-100 alone. On the other hand, the electrophoretic mobility of the hydrophilic proteins, which do not interact with the detergents, remains unaffected by changes in the detergents used (Helenius and Simons, 1977).

Helenius and Simons tested 17 hydrophilic proteins and 5 amphiphilic membrane proteins. The former included acidic and basic proteins and glycoproteins, some of which were composed by various subunits. Of the amphiphilic membrane proteins tested, three could be obtained both as holoproteins and in a proteolytically cleaved form that contained only the polar moiety: the polar heme-containing domain of cytochrome b_5, the trypsin form of aminopeptidase, and the bacterial penicillinase. Rather than DOC and CTAB alone (which may have conferred a greater shift in mobility to amphiphilic proteins), combinations of charged detergents with excess Triton X-100 were used to keep the structure of protein-detergent complexes as constant as possible in all three detergent systems. Notice that, although CTAB is known to be a denaturant, when mixed with sufficient Triton X-100 it has a lower chemical potential than it is required for massive binding and denaturation.

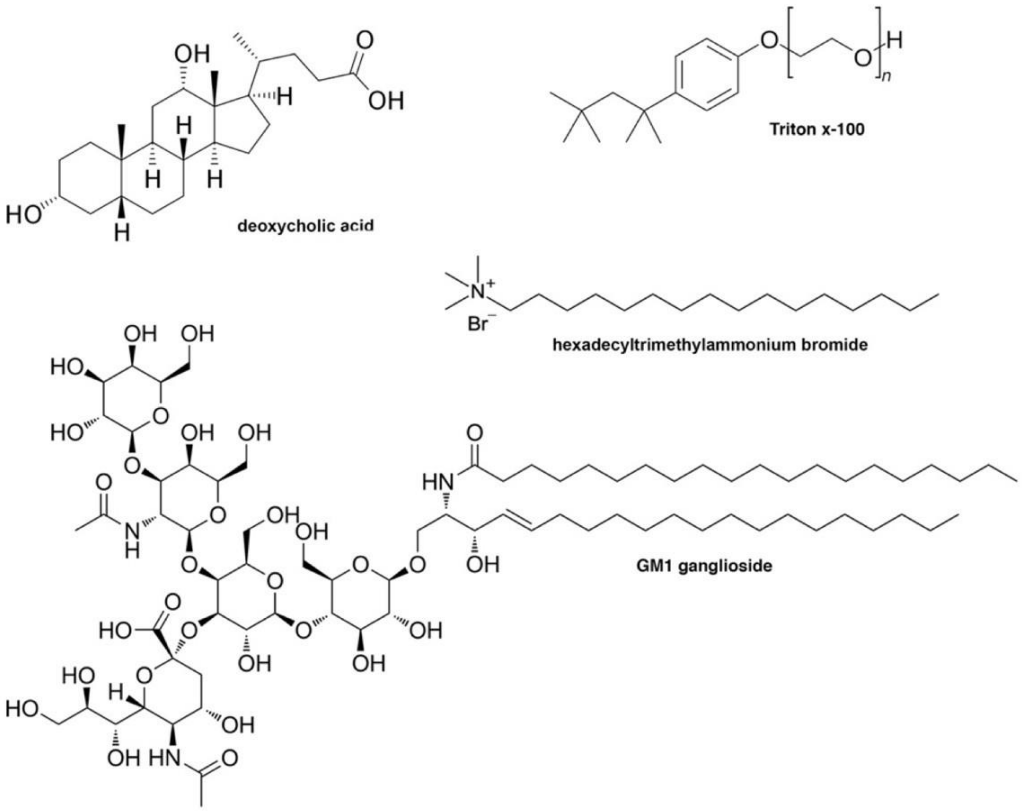

Figure 12 – Chemical structures of deoxycholic acid (DOC), Triton X-100 (*p*-*t*-octylphenyl polyoxyethylene$_{9-10}$), cetyltrimethylammonium bromide (CTAB) and ganglioside GM1.

Only minor differences were observed in the electrophoretic mobility of ordinary soluble proteins in the absence of detergent and in the three different detergent media (Fig. 13, panel a). The same occurred with exopenicillinase (Fig. 13, panel b), the trypsin form of aminopeptidase (Fig. 13, panel c) and the polar fragment of cytochrome b5 (not shown). In contrast, all amphiphilic integral membrane proteins tested, i.e., membrane penicillinase (Fig. 13, panel b), membrane aminopeptidase (Fig. 13, panel c), cytochrome b5 (not shown) and the Semliki Forest virus spike glycoproteins E1 (Fig. 13, panel d) and E2 (not shown) displayed a more anodal migration when electrophoresis was performed in Triton X-100 and deoxycholate than in the presence of Triton X-100 alone. On the other hand, the migration in the mixture of Triton X-100 and cetyltrimethylammonium bromide was more cathodal than in the presence of Triton X-100 alone.

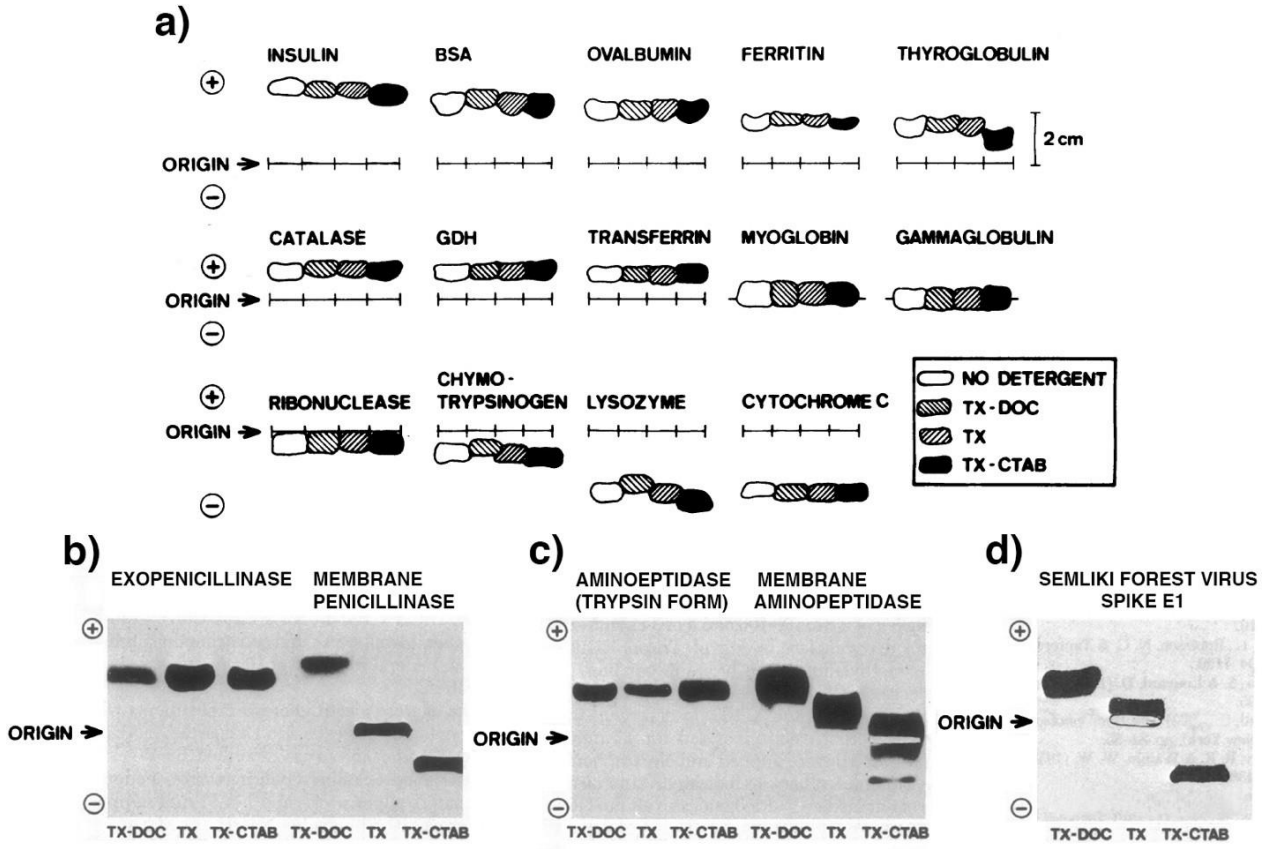

Figure 13 – Diagramatic representation of the combined results from agarose gel electrophoresis of: a) soluble proteins in the absence of detergent, in Triton X-100 (TX) and sodium deoxycholate (DOC), Triton X-100 alone, and Triton X-100 and cetyltrimethylammonium bromide (CTAB); b) exopenicillinase and membrane penicillinase; c) trypsin form of aminopeptidase and membrane aminopeptidase; d) spike glycoprotein E1 from Semliki Forest Virus, all in the same buffer systems. Protein stain used in b-d): Coomassie Brilliant Blue. *Legend*: BSA, bovine serum albumin; GDH, glutamate dehydrogenase (from Helenius and Simons, 1977).

2. AIM OF THE STUDY

The purpose of this study was twofold.

1. We aimed to investigate possible amphiphilic properties of the Spike glycoprotein of SARS-CoV-2 (formerly 2019-nCoV) (NCBI Reference Sequence: YP_009724390.1) and of some of its variants of concern, i.e., the B1.617.2 Spike (δ, delta) variant, the AY.2 (δ plus, delta plus) variant, and the B.1.1.529 (ο, omicron) variant. To this aim, we employed charge shift electrophoresis (CSE) in mixed detergent solutions (Helenius and Simons, 1977), to probe the interactions of the SARS-CoV-2 Spike protein and its variants with detergents, taken as models of membrane lipids. Because it would not be surprising that the S holoproteins, being integral membrane protein, exhibited amphiphilic properties in CSE, we studied their recombinantly expressed, monomeric extra-cellular domains (ECD) containing the S1 and S2 subunits, but devoid of the leader peptide and the transmembrane helix and cytoplasmic tail, encompassing amino acid residues Val16-Pro1213.

That the SARS-2-S glycoprotein from SARS-CoV-2 coronavirus, like the Spike glycoproteins of other coronaviruses, may exhibit amphiphilic properties is conceivable in the light of their activities as viral fusogens. These have been described as 'spring-loaded machineries' presented on the surfaceof the virions in a metastable state, with their hydrophobic fusion peptide buried in the hydrophobicinterior of the protein or hidden by interfacial interactions. Destabilization by receptor binding and proteolytic cleavage, separating the two subunits responsible for receptor binding (S1) and membrane fusion (S2) from each other, allows the exit of fusion peptides from their hydrophobic hideouts, followed by their insertion in the plasma membrane of target cells and membrane fusion (Kirchdoerfer *et al.*, 2016; Vance and Lee, 2016). Such a transition from a prefusion to a post-fusion

conformation implies the formation of intermediates in which hydrophobic amino acids previously buried in a hydrophobic environment become at least transiently exposed to the solvent.

A better understanding of how the Spike proteins of coronaviruses promote membrane fusion and how this process is affected by genetic variation may help elucidate the mechanism of viral entry into host cells and, most importantly, some of the determinants of the differences in infectivity, cell and tissue tropism and severity of disease associated with different SARS-CoV-2 variants.

2. We also aimed to investigate whether the Spike glycoproteins of the SARS-CoV-2 β-coronavirus and of some of its variants were able to bind GM1 ganglioside. Convincing proofs were provided for the binding of the Spike glycoproteins of HCoV-OC43, BCoV and porcine hemagglutinating encephalomyelitis virus (PHEV) to 9-*O*-acetyl-N-acetylneuraminic acid (9-*O*-Ac-Sia) terminally linked to oligosaccharides of glycoproteins and gangliosides (Vlasak *et al.*, 1988; Peng *et al.*, 2012; Matrosovich *et al.*, 2015), and of the Spike glycoprotein of HCoV-HKU1 to *O*-acetyl-N-acetylneuraminic acid (other than 9-*O*-acetylated) of similar sialoconjugates (Huang *et al.*, 2015) and for the binding of MERS-CoV to α2,3-linked and α2,6-linked) sialic acids at the host cell surface(Li *et al.*, 2017). The structures of the HCoV-HKU1 Spike glycoprotein trimer in complex with 5-N- acetyl,9-*O*-acetyl-neuraminic acid α-methyl glucoside (9-*O*-Ac-Me-Sia) (Kirchdoerfer *et al.*, 2016), and of the HCoV-OC43 Spike glycoprotein trimer in complex with the same glucoside (Tortorici *et al.*, 2019) have been worked out in detail. Glycan binding by the S1 NTD of HCoV-HKU1 was implicated in the conformational transition required to expose the receptor-binding site located in the S1 CTD (Kirchdoerfer *et al.*, 2016). Also, the existence of a ganglioside-binding site in the N- terminal domain (NTD) of the Spike glycoprotein of SARS-CoV-2 has been recently proposed by molecular modelling (Fantini *et al.*, 2020). However, to our knowledge no direct experimental prooffor the ability of SARS-2-S to bind gangliosides has been provided so far.

An understanding of the interactions of the Spike glycoproteins of SARS-CoV-2 with cell gangliosides, such as GM1, may expand our knowledge of the membrane fusion and infection processes in which it seems so deeply implicated. Elucidating the contribution of GM1 binding to the viral attachment and entry into host cells may lead to the development of anti-infectious and prophylactic drugs.

3. MATERIALS AND METHODS

3.1. Materials

- Recombinant SARS-CoV-2 Spike (formerly 2019-nCoV, NCBI Ref.: YP_009724390.1) S1+S2 ECD His-tagged (Sino Biological 40589-V08B1, MW 134.36 kDa) 100 μg in 0.02 M Tris/HCl, 0.3 M NaCl, pH 8.0, 10% glycerol, 5% trehalose, 5% mannitol, 0.01% Tween-80
- Recombinant SARS-CoV-2 B1.617.2 Spike (δ) S1+S2 ECD His-tagged (L452R T478K, S1/S2 furin site abrogated; K986P, V987P, pre-fusion stabilized) (R&D 10942-CV-100, MW 134 kDa) 100 μg in 0.02 M Tris/HCl, 0.3 M NaCl pH 8.0, 10% glycerol, 5% trehalose, 5% mannitol, 0.01% Tween-80
- Recombinant SARS-CoV-2 Spike AY.2 (δ plus) S1+S2 ECD His-tagged (Sino Biological 40589-V08B26, MW 134,200 Da) 100 μg in 0.02 M Tris/HCl, 0.3 M NaCl, pH 8.0, 10% glycerol, 5% trehalose,5% mannitol, 0.01% Tween-80
- Recombinant SARS-CoV-2 Spike B.1.1.529 (ο, omicron) S1+S2 ECD 6-His-tagged (L452R T478K, S1/S2 furin site abrogated; K986P, V987P, pre-fusion stabilized) (R&D Systems 11060-CV, MW 134 kDa) 100 μg in PBS, pH 7.4, 5% threalose
- GM1 ganglioside from bovine brain, lyophilized (Sigma G-9652, MW 1540), 2 mg/mL in dH_2O (GM1 solubility in water 3 mg/mL)
- GM1 pentasaccharide (oligo-GM1, GM1os) (Carbosynth 52659-37-7, FW 998,88), 2 mg/mL in dH_20
- fatty-acid-free BSA (MW 66400) 1 mg/mL in dH_2O. ^{280}O.D. (1/50 dilution) = 0.0216; ε = 43,824$M^{-1}cm^{-1}$ → [ffBSA] = 0.0227 x 50 / 43,824 x 66,400 = 1.63 mg/mL

- apoferritin (Sigma, MW 444,000) 54 mg/mL diluted to 1 mg/mL. ^{280}O.D. (1/50 dilution) = 0.0381; ε = 480,000 $M^{-1}cm^{-1}$ → [apoferritin] = 0.0381 x 50 / 480,000 x 444,000 = 1.76 mg/mL

- IgG (prepared in house) 33 mg/mL, diluted to 1 mg/mL in PBS

- cytochrome c (MW 12,000) 1 mg/mL in dH_2O. ^{280}O.D. = 1.713; ε = 28,500 $M^{-1}cm^{-1}$ → [cytochrome c] = 1.713 / 28,500 x 12,000 = 0.72 mg/mL

3.2. Methods

Charge-shift electrophoresis in 1% agarose:

- 1% agarose gel (14.25 mL, 10x7 cm) on GelBond® Film in CSE gel buffer:

  No detergent buffer: 0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl

  TX gel buffer (non-ionic): 0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl, 0.5 g/dL Triton X-100

  TX/DOC gel buffer (anionic): 0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl, 0.5 g/dL Triton X-100, 0.25 g/dL deoxycholic acid (DOC)

  TX/CTAB gel buffer (cationic): 0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl, 0.5 g/dL Triton X-100, 0.5 g/dL cetyltrimethylammonium bromide (CTAB)

- mini-Sub cell (10x7 cm)
- electrode buffer: CSE gel buffer
- run: 15 min at 15V, 1 min at 45V and 90 min at 60V, at 10 °C with electrode buffer recirculation with Bio-Rad 491 Buffer Recirculation Pump
- stain: Coomassie Brilliant Blue R-250 in 25% isopropanol, 10% acetic acid overnight
- destain: 25% methanol, 10% acetic acid

4. RESULTS

4.1. Charge shift electrophoresis of SARS CoV-2 wild-type, B1.617.2 (δ), AY.2 (δ plus) and B.1.1.529 (ο) Spike glycoproteins

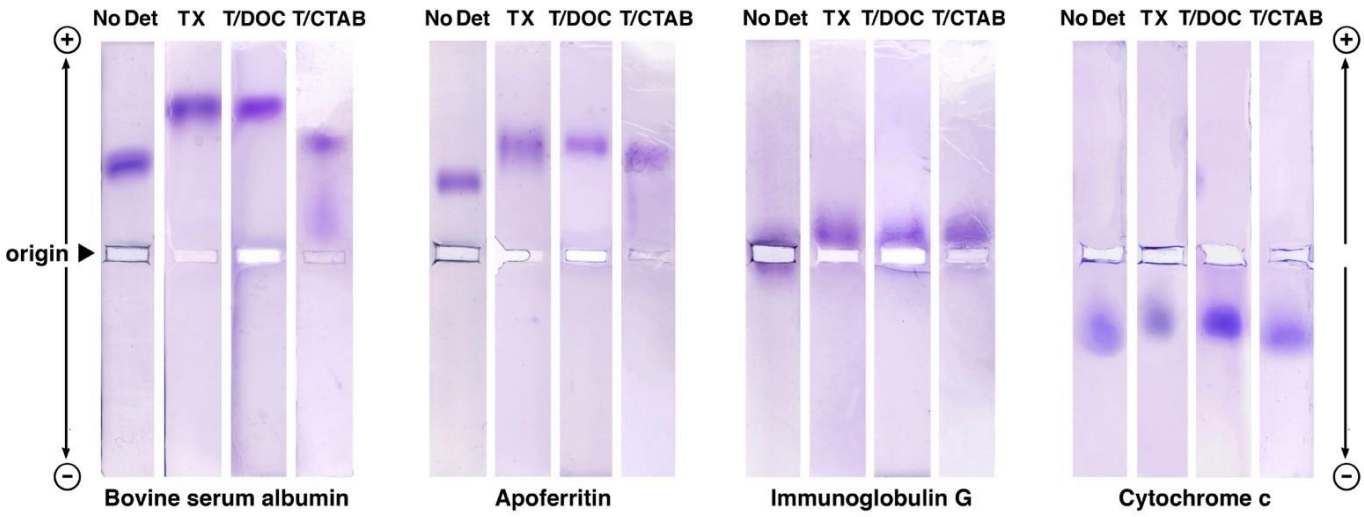

Figure 14 – Charge shift electrophoresis (CSE) of reference proteins (fatty-acid-free Bovine Serum Albumin, Apoferritin, Immunoglobin G, and Cytochrome c) in 1% agarose, in CSE buffer (0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl) with no detergent (No Det), 0.5 % Triton X-100 (TX), 0.5 % TX + 0.25 % deoxycholic acid (T/DOC) or 0.5 % TX + 0.5 % cetyltrimethylammonium bromide (T/CTAB). Protein load was 5 μg per lane. Stain: Coomassie Brilliant Blue R-250.

We first examined the electrophoretic behavior of four soluble reference proteins. These included: a) two acidic proteins, i.e., a special formulation of fatty-acid-free bovine serum albumin (BSA) and apoferritin; b) a slightly acidic pool of class G immunoglobulins (IgG's), and c) a basic protein, i.e.,

cytochrome c (Figure 14). BSA and apoferritin exhibited a markedly anodal migration both in the absence of detergents and in the three different detergent mixtures used (Triton X-100 alone or in combination with sodium deoxycholate or with cetyltrimethylammonium bromide). Only, the anodal migration of BSA was more marked in TX-100 and in TX-100/DOC than in the absence of detergents or in TX-100/CTAB. The anodal migration of apoferritin was also increased, although to a smaller extent in comparison with BSA, in all detergents tested. On the other hand, basic cytochrome c exhibited a marked cathodal migration in all conditions tested. As for IgGs, they remained at the origin in the absence of detergents, but showed a very limited anodal migration in all other conditions tested (Figure 14). Such behavior was in all cases as expected for soluble proteins devoid of amphiphilic properties.

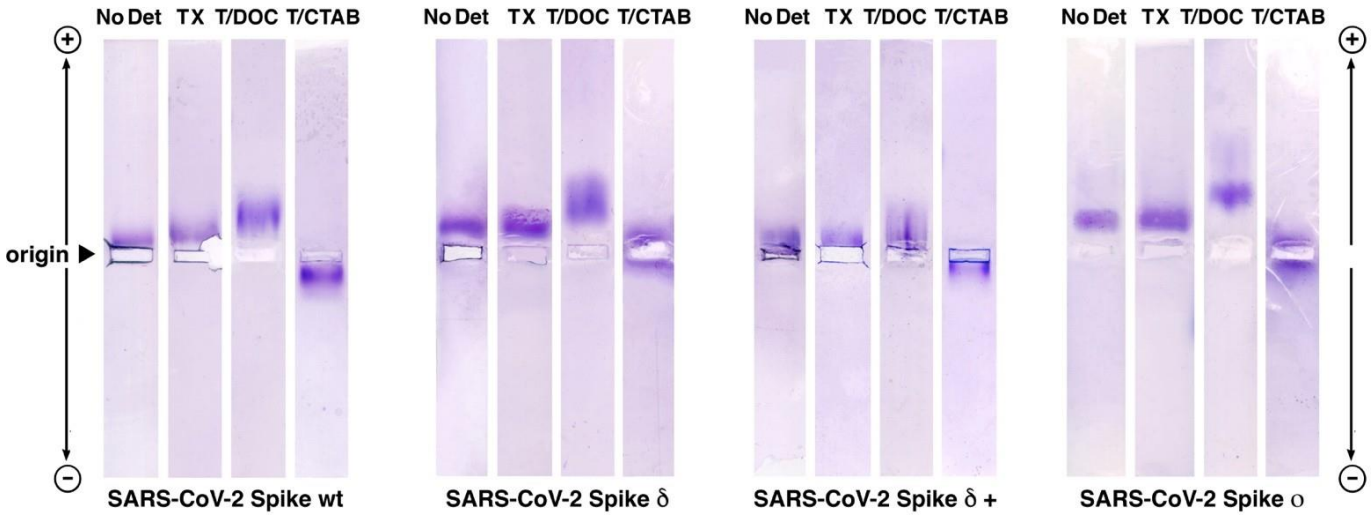

Figure 15 – Charge shift electrophoresis (CSE) of wild-type, δ, δ plus and ο SARS-2-S in 1% agarose, in CSE buffer (0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl) with no detergent (No Det), 0.5 % Triton X-100 (TX), 0.5 % TX + 0.25 % deoxycholic acid (T/DOC) or 0.5 % TX + 0.5 % cetyltrimethylammonium bromide (T/CTAB). Protein load was 5 μg per lane. Stain: Coomassie Brilliant Blue R-250.

Next, we examined the behavior in charge shift electrophoresis of wild-type SARS-2-S and its δ, δ plus and ο variants (Figure 15). Wild-type, δ and ο SARS-2-S exhibited a more anodal migration when electrophoresis was performed in the presence of TX-100 and DOC than in the presence of TX-100 alone. This reflected the ability of wild-type SARS-2-S and its δ and ο variants to participate in the formation of mixed TX-100/DOC micelles of homogeneous size, which allowed them to undergo a well-focused anodal shift by the virtue of the multiple negative charges provided by DOC. The anodal migration of wild-type SARS-2-S was negligible in the absence of detergent and a little more appreciable in TX-100, as was the anodal migration of the SARS-2-S δ and ο variants in both these conditions. On the other hand, the cathodal shift of wild-type SARS-2-S in the mixture of TX-100 and CTAB was quite distinct. The cathodal shift of the SARS-2-S δ and ο variants was not less so, although it may seem less apparent, as both migrated in a more anodal position in the absence of detergent. These cathodal shifts reflected the ability of wild-type SARS-2-S and its δ and ο variants to participate in the formation of mixed TX-100/CTAB micelles of homogeneous size, with multiple positive charges. It is of note that the SARS-2-S δ band in TX-100/DOC was less focused than those of wild-type SARS-2-S and the SARS-2-S ο variant. A case of its own was the SARS-2-S δ plus variant, which exhibited a negligible anodal migration both in the absence of detergent and in TX-100. Unlike the other SARS-2-S variants, the SARS-2-S δ plus variant did not exhibit a clear anodal shift in TX-100/DOC, but rather a narrow smear which started at the origin and was fading in the anodal direction. This could reflect a limited ability of the SARS-2-S δ plus variant to participate in the formation of large mixed TX-100/DOC micelles, and its preference for heterogeneous mixed TX-100/DOC micelles of smaller average size, containing a lower average number of negatively charged DOC molecules than those formed by the other SARS-2-S variants. This could be due to a significant number of charge-specific interactions between DOC molecules and positively charged residues of

SARS-2-S δ plus, which are not expected to entail changes of electrophoretic mobility, as they would abrogate both the electrostatic repulsion of positively charged residues by the anode and the contribution of negatively charged DOC carboxylates to anodal migration. Nevertheless, the SARS-2-S δ plus variant exhibited a clear, although limited cathodal shift in TX-100/CTAB (Figure 15).

4.2. Charge shift electrophoresis of SARS CoV-2 wild-type, B1.617.2 (δ), AY.2 (δ plus) and B.1.1.529 (ο) Spike glycoproteins pre-incubated with GM1 ganglioside

Next, we examined the effects of the GM1 ganglioside and the GM1 pentasaccharide, i.e., the isolated oligosaccharide part of the GM1-ganglioside, devoid of the ceramide moiety (oligo-GM1, GM1os) on the migration of wild-type SARS-2-S and its δ, δ plus and ο variants in CSE buffer (0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl) in the absence of detergents (Figure 16). Protein samples (4 μg/well) were pre-incubated with either ligand for 3 hours at room temperature at the SARS-2-S:ligand molar ratio of 1:100 in the case of wild-type SARS-2-S and the SARS-2-S ο variant, or at two different SARS-2-S:ligand molar ratios (1:100 or 1:350) in the case of SARS-2-S δ and δ plus variants. The migration of wild-type SARS-2-S pre-incubated with GM1 underwent a distinct anodal shift in comparison with the control, and the same occurred, to an even larger extent, with the SARS-2-S ο variant. Instead, pre-incubation with the GM1 pentasaccharide did not affect the migration of wild-type SARS-2-S. In the case of SARS-2-S δ, migration was only partially affected by pre-incubation with GM1 ganglioside at the SARS-2-S:GM1 ratio of 1:100, with part of the protein migrating like in the absence of GM1 and part of it being anodally shifted, with a markedly polydisperse appearance. Instead, at the SARS-2-S:GM1 ratio of 1:350, all of the SARS-2-S δ protein underwent such changes. A similar behavior was seen, although less distinctly, with the SARS-2-S δ plus variant. Pre-incubation with the GM1 pentasaccharide did not affect the migration of SARS-2-S δ and δ plus (Figure 16).

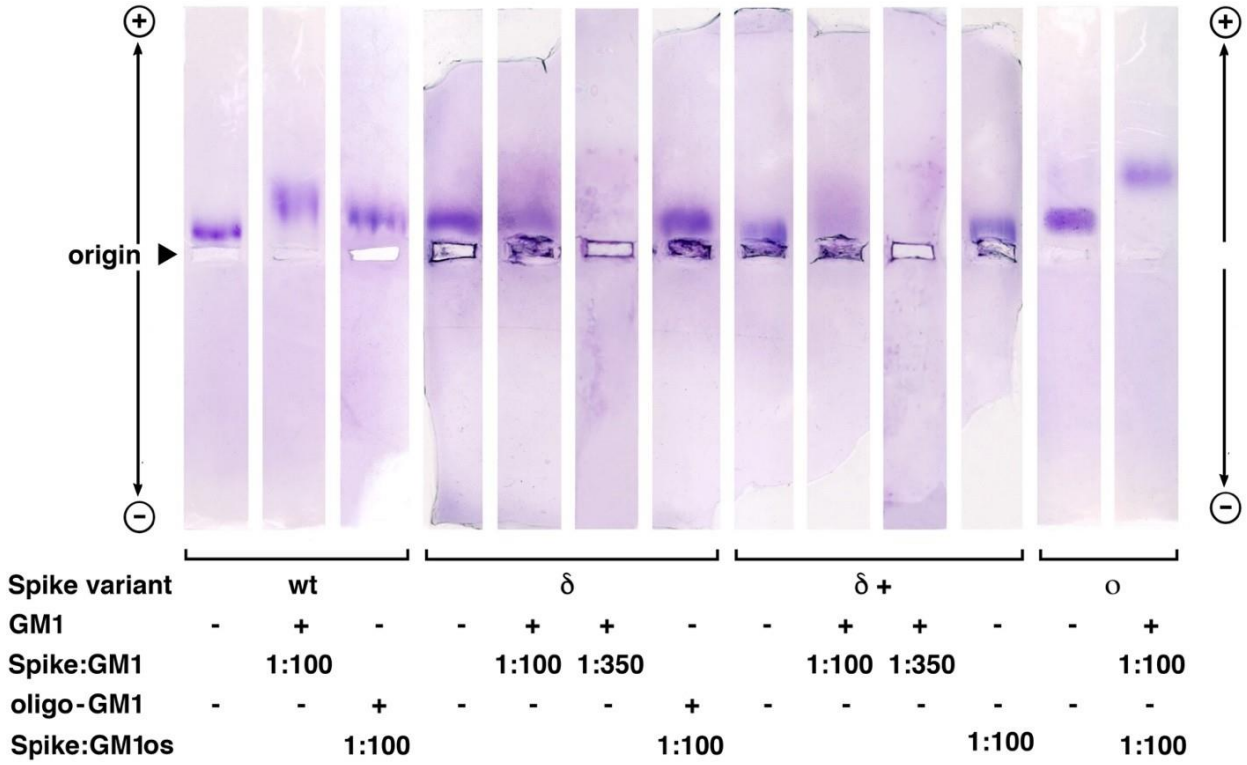

Figure 16 – Charge shift electrophoresis (CSE) of wild-type, δ, δ plus and o SARS-2-S and in 1% agarose, in CSE buffer with no detergent (0.05 M glycine/NaOH, pH 9.0, 0.1 M NaCl). SARS-2-S protein load was 4 μg per lane. Samples were pre-incubated for 3 h at room temperature at the indicated molar ratios of GM1 ganglioside or GM1 pentasaccharide (oligo-GM1, GM1os). The concentration of the GM1 ganglioside stock solution was 2 mg/mL and the final GM1 ganglioside concentration range in samples was 0.33-1.33 mg/mL.

The comparable extent of the anodal shift that was observed with wild-type SARS-2-S and the SARS-2-S o variant after pre-incubation with GM1, with respect to their migration in the presence of TX-100/DOC without GM1 addition, most probably reflected the participation of wild-type SARS-2-S and SARS-2-S o molecules in the formation of mixed micelles of homogeneous size with multiple GM1 molecules, in keeping with the amphiphilic properties that both glycoproteins demonstrated

in TX-100/DOC, rather than the charge-specific binding of individual GM1 ganglioside molecules to the SARS-2-S glycoprotein, mediated by their oligosaccharide moiety. The lack of anodal shift after incubation with the GM1 pentasaccharide supports this interpretation.

On the other hand, the dose-dependent anodal shift and polydisperse appearance that we observed with the SARS-2-S δ and δ plus variants after incubation with GM1 suggests that GM1 addition may have promoted the formation of a highly polydispersed, heterogeneous population of micelles and/or aggregates of variable size and charge, composed of a variable number of GM1 molecules bound to different SARS-2-S δ and δ plus conformers, most likely by heterogenous chemical bonds (H-bonds, CH-π stacks and hydrophobic interactions). Such mode of binding is expected to increase the anodal migration of SARS-2-S δ and δ plus molecules to a variable extent via the contribution of a variable number of GM1 carboxylates. The lack of effect of GM1 pentasaccharide on the mobility of SARS-2-S δ and δ plus suggests that, even though charge-specific interactions between GM1 carboxylates and positively charged amino acid residues of the SARS-2-S glycoproteins cannot be excluded, these were not the predominant interactions.

4.3. Correlation between genetic variation and the amphiphilic and GM1-binding properties of SARS-CoV-2 Spike glycoprotein variants

Table 2 summarizes the behavior exhibited by the four SARS-2-S variants studied, when subjected to: a) charge shift electrophoresis in 1% agarose in four different CSE buffers with/without various non-ionic and ionic detergents at pH 9.0 (described in section 4.1); b) electrophoresis in 1% agarose in CSE buffer without detergents after pre-incubation for 3 hours with GM1 ganglioside at a final concentration varying from 0.33 to 1.33 mg/mL (described in section 4.2). Table 2 also shows the

mutations and the changes in electric charge in the putative ganglioside-binding domain (Fantini *et al.*, 2020) and in the full-length ECD of the four SARS-2-S variants studied.

Concerning the shifts in electrophoretic mobility indicating the amphiphilic character of the SARS-2-S protein (Figure 15), the lack of a clear anodal shift of the SARS-2-S δ plus variant and its smeared migration in TX-100/DOC possibly reflected its limited ability to participate in the formation of mixed TX-100/DOC micelles of homogeneous size, possibly also due to the establishment of electrostatic interactions between the DOC carboxylate and positively charged residues of SARS-2-S δ plus molecules. In regard, it is of note that the mutations in the full-length ECD of the SARS-2-S δ plus variant determined an overall charge change of +5, to be compared to -1 in the SARS-2-S δ variant and +2 in the SARS-2-S ο variant (Table 2). It is also to be noted, however, that the positive charge increment produced by genetic variation in the δ and ο variants was limited by the purposeful introduction, in the commercial preparation that we used, of two mutations abrogating the S1/S2 furin site (Arg682Ser, Arg685Ser) and two more mutations in the transitional bend following the HR1 domain, which were intended to stabilize the prefusion conformation (Lys986Pro, Val987Pro).

Moreover, given the markedly polydisperse anodal migration of SARS-2-S δ and δ plus variants pre-incubated with GM1 ganglioside (Figure 16), these two variants showed, in comparison with wild-type SARS-2-S and the SARS-2-S ο variant, an enhanced tendency to form, upon interaction with GM1 ganglioside, a highly heterogeneous population of micelles and/or aggregates of variable size and charge, composed of a variable number of GM1 molecules bound to a number of different SARS-2-S δ and δ plus conformers. An exam of the mutations in the putative ganglioside-binding domain (GBD) which characterize the different SARS-2-S variants (Table 3) reveals that the δ and δ plus variants have in common the mutations of three amino acid residues which contributed strongly to the binding of GM1 gangliosides in a molecular model of the GBD of SARS-2-S (Fantini *et al.*, 2020) and were not mutated in the SARS-2-S ο variant and, obviously, in wild-type SARS-2-S: Glu156Gly,

Phe157del, and Arg158del. Of the residues involved, Glu156 and Phe157 were predicted to take part in the binding of a second GM1 molecule, while Arg158 contributed to the binding of both a first and a second GM1 molecule, with the third and second strongest energies of interaction among those measured for all sixteen residues involved in GM1 binding in the GBD (residues 111-162) (Table 1) (Fantini *et al.*, 2020). As already noted, of the two SARS-2-S δ and δ plus variants, the latter exhibited the scantiest amphiphilic properties (Figure 15). The SARS-2-S δ variant presented a Gly142Asp mutation in the GBD, which was missing in the δ plus variant (Table 3), which in turn presented a unique Val70Phe substitution outside the GBD (Table 2).

Table 2 - Impact of mutations on the amphiphilic and GM1-binding properties of SARS-2-S

SARS-2-S variant	mutations (GBD)	charge change (GBD)	additional mutations (full-length)	overall charge change	charge shift (DOC)	charge shift (CTAB)	charge shift (GM1)
wild type					anodic	cathodic (marked)	anodic
B.1.617.2 (δ)†	**G142D**, **E156G**, **F157del**, **R158del**	-1	T19R, A222V, K417N, L452R, T478K, D614G, P681R, D950N (R682S, R685S, K986P, V987P)	-1	anodic	cathodic (slight)	anodic polydisperse
AY.2 (δ +) °	~~G142D~~, **E156G**, **F157del**, **R158del**	0	T19R, **V70F**, A222V, K417N, L452R, T478K, D614G, P681R, D950N	+5	(minor)	cathodic (marked)	anodic polydisperse
B.1.1.529 (ο)	**G142D**, V143del, Y144del, Y145del	-1	~~T19R~~, A67V, H69del, **V70del**, T95I, N211del, L212I, ins214EPE, ~~A222V~~, G339D, S371L, S373P, S375F, K417N, N440K, G446S, ~~L452R~~, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, **P681H**, N764K, D796Y, N856K, ~~D950N~~, Q954H, N969K, L981F (R682S, R685S, K986P, V987P)	+2	anodic	cathodic (slight)	anodic

Legend: *blue bold* characters mark mutations of amino acid residues that were reported to be involved in interactions with GM1 ganglioside (Fantini *et al.*, 2020); *purple bold* characters mark novel mutations that emerged in successive variants along the timeline; *strikenthrough* characters indicate mutations which were not maintained in successive variants along the timeline.

† R&D # 10942-CV

° Sino Biological # 40589-V08B26

Table 3 - Mutations in the putative Ganglioside-Binding Domain of different SARS-2-S variants

SARS-2-S variant	mutations in SARS-2-S putative Ganglioside-Binding Domain (111-162)	charge change
Wild type	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**VYYHKNNKSWMES**EFR**...V**YSS**	
B.1.1.7 (α, UK)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**V-YHKNNKSWMES**EFR**...V**YSS**	0
B.1.351 (β, S. Africa)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**V-YHKNNKSWMES**EFR**...V**YSS**	0
P.1/P.1.1 (γ, Brazil)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCN**Y**P**F**L**G**V**YYHKNNKSWMES**EFR**...V**YSS**	0
B.1.617.2 (δ)†	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**G- -**...V**YSS**	**-1**
AY.1 (δ +) *	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**EFR**...V**YSS**	-1
AY.1 (δ +) **	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**G- -**...V**YSS**	-1
AY.2 (δ +) °	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**VYYHKNNKSWMES**G- -**...V**YSS**	**0**
AY.2 (δ +) °°	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**G- -**...V**YSS**	-1
AY.3 (δ +)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**VYYHKNNKSWMES**G- -**...V**YSS**	0
AY.4.2 (δ +)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VY**H**HKNNKSWMES**G- -**...V**YSS**	0
B.1.427/29 (ε, USA)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**G**VYYHKNNKS **C** MES**EFR**...V**YSS**	0
B.1.1.529 (ο)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D- - -**HKNNKSWMES**EFR**...V**YSS**	**-1**
BA.2 (ο 2)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**EFR**...V**YSS**	-1
AY.4/BA.1 (δ-cron)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**G- -**...V**YSS**	-1
BA.4 (ο 4)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**EFR**...V**YSS**	-1
BA.5 (ο 5)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**EFR**...V**YSS**	-1
BA.1/BA.2 (ο-XE)	**DSK**TQSLLIVNNATNVVIKVCEF**QFCND**P**F**L**D**VYYHKNNKSWMES**EFR**...V**YSS**	-1

Legend: in column 1, the SARS-2-S variants examined in study are indicated in *bold red*; in column 2, amino acid residues of critical importance in GM1 ganglioside binding in molecular modelling studies (Fantini *et al.*, 2020) are indicated in *bold blue*, while those that were mutated are indicated in *bold red* and evidenced in *yellow* (substitutions) or *green* (deletions).

† R&D # 10942-CV

* Sino Biological # 40589-V08B25

** R&D # 10922-CV

° Sino Biological # 40589-V08B26

°° R&D # 10923-CV

5. DISCUSSION

1. The data contained in this thesis provide a demonstration of the amphiphilic properties of SARS-2-S, the Spike glycoprotein of SARS-CoV-2, as well as of its δ (B.1.617.2, delta) and ο (B.1.1.529, omicron) variants. The property of amphiphilicity of these viral spike glycoproteins is in tune with their crucial role in the process of membrane fusion between the viral membrane envelope and the plasma membrane of target cells, or between the plasma membranes of an infected cell and of a second target cell in the process of cell-to-cell spreading of infection. Amphiphilicity is one of the modes by which membrane-active peptides can perturb biological membranes, the others being a polycationic or a hydrophobic nature (Avci *et al.*, 2018). Viral fusogens, such as spike proteins, can be described as metastable structures, or spring-loaded machineries endowed with hydrophobic fusion peptide or loops that are buried in the fusogen in the prefusion state, but can be released by conformational changes induced by receptor binding and/or proteolytic cleavage. Once released, they embed in the host cell membrane and promote membrane fusion, while the fusogen acquires its stable post-fusion conformation. The conformational transitions implicated in this process of progressive destabilization have been worked out in detail in the Spike glycoprotein of HCoV-HKU1 (Kirchdoerfer *et al.*, 2005). They include: 1) the disruption of interactions between the C-terminal end of HR1 and the CTD of the S1 subunit preventing the formation of a three-helix bundle in the center of the anti-parallel six-helix bundle (6-HB), triggered by receptor binding by the S1 subunit; 2) the proteolytic removal of the S1 CTD domain capping the S2 central helix, by cleavage at the S1/S2 site; 3) the release of the fusion peptide from its hydrophobic interaction with the S2 surface (Kirchdoerfer *et al.*, 2005); 4) the refolding of HR1 into a long α-helix (Duquerroy *et al.*, 2005; Lu *et al.*, 2005). In all of these transitions, hydrophobic amino acid stretches buried in the hydrophobic protein core or patches hidden by interfacial interactions acquire transient exposure to the solvent.

We employed a technique for probing the amphiphilic character of SARS-2-S which was both direct and simple, such as charge shift electrophoresis (Helenius and Simons, 1977). Wild-type, δ and ο SARS-2-S presented distinct amphiphilic properties in this system, exhibiting a more anodal migration when electrophoresis was performed in the presence of TX-100 and DOC, and a more cathodal migration when electrophoresis was performed in the mixture of TX-100 and CTAB than in the presence of TX-100 alone (Figure 15). This reflected the ability of wild-type SARS-2-S and its δ and ο variants to participate in the formation of mixed TX-100/DOC or TX-100/CTAB micelles of homogeneous size, which allowed them to undergo well-focused anodal or cathodal shifts by the virtue of the multiple charges provided by the negatively or positively charged detergent, respectively. Unlike the other SARS-2-S variants, the SARS-2-S δ plus variant did not exhibit a clear anodal shift in TX-100/DOC, but rather a narrow smear, which started at the origin and was fading in the anodal direction. This pattern of migration might reflect its limited ability to participate in the formation of large mixed TX-100/DOC micelles, and its preference for heterogeneous mixed TX-100/DOC micelles of smaller average size, containing a lower average number of negatively charged DOC molecules than those formed by the other SARS-2-S variants. As already mentioned, this could be due to a significant proportion of DOC molecules establishing charge-specific interactions with positively charged residues of SARS-2-S δ plus molecules. Such interactions are not expected to entail changes of electrophoretic mobility, as they would abrogate at one time both the electrostatic repulsion of positively charged residues by the anode and the contribution of negatively charged DOC carboxylates to anodal migration. This is in keeping with the overall charge change of +5 produced by genetic variation in SARS-2-S δ plus (Table 2). Such an interpretation is also consistent with the persisting cathodal shift of SARS-2-S δ plus in TX-100/CTAB. Further studies are needed to understand how the amphiphilic properties of SARS-2-S and its variants are determined and how they can affect: a) their binding to target cell receptors and membranes, 2) membrane fusion, c)

infection of target cells, and d) release of viral progeny from infected cells. The simple method for assessing amphiphilicity that we employed may be a useful screening method for emerging variants.

2. We also present data that provide a demonstration of the ability of different variants of the SARS-CoV-2 Spike glycoprotein to bind the GM1 ganglioside. This is in keeping with the amphiphilic character that was demonstrated for SARS-2-S, as gangliosides are also inherently amphiphilic molecules, with a pronounced surface and interfacial activity and a tendency to interact with molecules that are also amphiphilic. Even though the ability to bind sialoconjugates at the surface of target cells was already demonstrated for a number of animal and human coronaviruses (see Section 1.7), to our knowledge this aspect was not previously investigated with SARS-CoV-2. A ganglioside-binding site, spanning amino acid residues 111-162 in the NTD of SARS-CoV-2, was predicted by molecular modelling studies (Fontana *et al.*, 2020). However, ours is the first experimental demonstration of the GM1-ganglioside-binding ability of the SARS-2-S glycoprotein. We employed the same technique of charge shift electrophoresis in 0.05 M glycine, 0.1 M NaCl at pH 9.0 that we used to probe the amphiphilic character of SARS-2-S also to assess the GM1-binding ability of wild-type SARS-2-S and its δ, δ plus and ο variants, although in the absence of detergents. GM1 ganglioside from a 2 mg/mL stock in dH_2O was incubated with SARS-2-S at final concentrations ranging from 0.33 to 1.33 mg/mL, with all of these concentrations being well above the GM1 critical micellar concentration of the GM1 ganglioside and lower than its maximal solubility in water (Tomasi *et al.*, 1980). Both wild-type SARS-2-S and its ο variant exhibited a distinct GM1-binding ability, evidenced by a distinct anodal shift in comparison with the respective controls. Instead, pre-incubation with the GM1 pentasaccharide did not affect electrophoretic migration of wild-type SARS-2-S. As mentioned in Section 4.3 of the Results, the comparable extent of the anodal shift that was observed with wild-type SARS-2-S and the SARS-2-S ο variant after pre-incubation with GM1, with respect to their migration in the presence of TX-100/DOC without GM1 addition, most likely

reflected the participation of wild-type SARS-2-S and SARS-2-S ο molecules in the formation of mixed micelles of homogeneous size with multiple GM1 molecules, in keeping with the amphiphilic properties that both glycoproteins demonstrated in TX-100/DOC, rather than the charge-specific binding of individual GM1 ganglioside molecules to the SARS-2-S glycoprotein, mediated by their oligosaccharide moiety, as envisaged in the structural and modelling studies exploring the interaction of gangliosides with ganglioside-binding sites of SARS-S molecules (Huang *et al.*, 2015; Kirchdoerfer *et al.*, 2016; Tortorici *et al.*, 2019). The lack of effect of the pre-incubation with the GM1 pentasaccharide supports this interpretation. On the other hand, the dose-dependent anodal shift and polydisperse appearance that we observed with the SARS-2-S δ and δ plus variants after pre-incubation with GM1 suggests that GM1 addition may have promoted the formation of a highly polydispersed, heterogeneous population of micelles and/or aggregates of variable size and charge, composed of a variable number of GM1 molecules bound to different SARS-2-S δ and δ plus conformers, most likely by heterogenous chemical bonds (H-bonds, CH-π stacks and hydrophobic interactions). Such mode of binding is expected to increase the anodal migration of SARS-2-S δ and δ plus molecules to a variable extent via the contribution of a variable number of GM1 carboxylates. Thus, the interaction with the GM1 ganglioside seemed to affect the structure of the SARS-2-S δ and δ plus variants more pronouncedly than it did with wild-type SARS-2-S and the ο variant.

That the interaction with the GM1 ganglioside may trigger conformational transitions in the SARS-2-S glycoproteins is implicit in the metastable modifications of SARS-2-S that were implicated in the structural studies of the trimeric Spike glycoproteins of HCoV-HKU1 and of HCoV-OC43 in complex with 9-*O*-Ac-Me-Sia. It was suggested that glycan binding by the S1 NTD of HCoV-HKU1 might trigger a conformational transition which was required to expose the receptor-binding site located in the S1 CTD (Kirchdoerfer *et al.*, 2016). Also, the residues involved in the binding of 9-*O*-Ac-Me-Sia in a groove at the surface of domain A of the HCoV-OC43 Spike glycoprotein proved to be essential for

virus entry into host cells. The lack of fusogenic effects of free 9-*O*-Ac-Me-Sia on the Spike glycoprotein was taken as an indication that multivalent interactions with sialoglycans (via mechanical destabilization of the pre-fusion trimers) and/or some proteinaceous receptor were essential for membrane fusion (Tortorici *et al.*, 2019).

The interaction of GM1 gangliosides with bovine serum albumin was investigated previously. By employing UV and fluorescence spectroscopy, light scattering and gel sizing chromatography, Tomasi *et al.* observed that the amount of GM1 bound to BSA was dependent upon the concentration of GM1. A maximum of one ganglioside molecule bound per molecule of protein was observed at submicellar GM1 ganglioside concentrations (< 0.1 mM, or < 0.15 mg/mL), whereas at concentrations higher than the critical micellar concentration (CMC) mixed micelles of GM1 and BSA were formed, composed by one BSA molecule and approximately 310-345 molecules of GM1 ganglioside (Tomasi *et al.*, 1980). GM1 micelles interacted with BSA mostly by hydrophobic interactions and induced detectable alterations in UV absorption and fluorescence spectra of BSA, indicating differences in protein structure in the neighborhood of aromatic amino acids. These alterations enabled BSA to bind GM1 micelles, forming mixed GM1:BSA micelles. A further change induced by GM1 binding induced BSA dimerization (Tomasi *et al.*, 1980). Hirai *et al.* employed neutron and synchrotron x-ray small-angle scattering to study the complexes of GM1 and GD1a with various forms of BSA modified with galactose, glucose, fucose, maltitol and cellobiose, taken as models of membrane glycoproteins. They observed in most cases the formation of polydispersed large aggregates, except in the case of GM1 addition to maltitol-BSA, in which the initial GM1 micelles were destroyed and monodispersed small complexes formed (Hirai *et al.*, 1998). A better understanding of the effects of GM1 ganglioside addition on the electrophoretic mobility of SARS-2-S and its δ, δ plus and ο variants will require detailed studies of their modes of interaction by the use of UV and fluorescent spectroscopy, circular dichroism, scanning calorimetry and light

scattering, which should include other major gangliosides, such as GD1a, GD1b, GT1b, also in comparison with neutral glycolipids, such as asialo-GM1, glucosylceramide and lactosylceramide. We did our experiments at GM1 ganglioside concentrations well above its CMC. It would be interesting to investigate what differences in structure between wild-type SARS-2-S and its ο variant, on one hand, and the SARS-2-S δ and δ plus variants, on the other hand, may account for such diverse responses to the interaction with the GM1 ganglioside. As noted in Section 4.3 of the Results, the δ and δ plus SARS-2-S variants had in common the mutations of three amino acid residues which contributed to the binding of GM1 in a molecular model of the ganglioside-binding domain of SARS-2-S (Fantini *et al.*, 2020), these being Glu156Gly, Phe157del, and Arg158del (Tables 1 and 3). Intriguingly, one of the two SARS-2-S variants, i.e., δ plus, uniquely exhibited limited amphiphilic character in comparison with all other variants studied (Figure 15). One may wonder whether pre-incubation of SARS-2-S δ plus with GM1 might restore its amphiphilic character, manifested as an anodal shift upon electrophoresis in the presence of Triton X-100 and DOC.

It might be particularly interesting to establish what the impact may be of any differences in the modes of interaction of SARS-2-S and its variants with gangliosides on their fusogenic activities, and on the infectivity of the viral variants expressing them. This may require the use of cell-cell fusion assays *in vitro*, using 293T/ACE2 cells as target cells and effector cells prepared by transfecting 293T cells with Spike expression vectors (Xia *et al.*, 2019), in combination with neuraminidases. As already stated, understanding the mode of interactions of the Spike glycoprotein of SARS-CoV-2 with cell gangliosides, such as GM1, may broaden our knowledge of membrane fusion and viral infection.

3. Moreover, elucidating in further detail the contribution of GM1 binding to the processes of viral attachment and viral entry into target cells mediated by the SARS-2-S glycoprotein might lead to the development of innovative anti-infectious strategies, on the model of those recently developed to limit the intoxicating power of a typical ligand of the GM1 ganglioside, such as cholera toxin (CT).

Zuilhof investigated the properties of inhibitors that mimic membrane-bound GM1 molecules. In order to bind CT strongly and selectively, these should display multiple copies of oligosaccharides resembling as much as possible the naturally occurring cell-surface pentasaccharide to which CT normally binds (GM1os). Multivalency in CT binding was attained by the synthesis of dendrimers with up to eight GM1os, or by the construction of sugar-coated platforms in which the oligosaccharides were appended to a synthetic scaffold like corannulene or calix[5]arene, or else by the development of a neolectin CT mimic that itself carried five GM1os groups (Zuilhof, 2012). Another strategy for scavenging CT by the use of competing binding motifs was producing GM1a-functionalized and biocompatible microgels with a mesoporous and widely meshed network structure, using drop-based microfluidics. These exhibited strong and multivalent, high-capacity binding to the CT binding domain and the ability to scavenge and retain CT in direct binding competition to colorectal cells (Boesveld *et al.*, 2019). Similar strategies could be applied to the synthesis of Spike glycoprotein-binding molecules, platforms or glycan-functionalized microgels.

6. CONCLUSIONS

1. We provide a demonstration of the amphiphilic properties of wild-type SARS-2-S and its δ (B.1.617.2, delta) and ο (B.1.1.529, omicron) variants, obtained by a direct and simple technique, such as charge shift electrophoresis (CSE) in detergent solution. Wild-type, δ and ο SARS-2-S presented distinct amphiphilic properties in this system, likely reflecting their ability to participate in the formation of large mixed TX-100/DOC and TX-100/CTAB micelles of homogeneous size. The SARS-2-S δ plus variant exhibited a more limited amphiphilicity in TX/DOC, possibly reflecting its propensity for TX-100/DOC micelles with a lower DOC content, perhaps due to enhanced charge-specific interactions between DOC and positively charged amino acid residues.

Further studies are needed to understand how the amphiphilicity of SARS-2-S and its variants is determined and how it may affect binding, membrane fusion, infection, and release of viral progeny.

2. We also provide a demonstration of the ability of different variants of the SARS-CoV-2 Spike glycoprotein to bind the GM1 ganglioside, by the same CSE approach, although in the absence of detergents. Even though the sialoside-binding ability of a number of coronaviruses was known, ours is the first direct demonstration of the GM1-ganglioside-binding ability of the SARS-2-S glycoprotein. Wild-type SARS-2-S and the SARS-2-S ο variant exhibited, after pre-incubation with GM1, anodal shifts that likely reflected their participation in large mixed micelles of homogeneous size with multiple GM1 molecules, in keeping with the amphiphilic properties demonstrated in TX-100/DOC. Instead, the polydispersity of migration exhibited by the SARS-2-S δ and δ plus variants pre-incubated with GM1 seemed to reflect the formation of a highly heterogeneous population of micelles and/or aggregates of GM1 ganglioside and SARS-2-S δ and δ plus conformers, as though the interaction with GM1 had a disordering effect upon these two SARS-2-S variants. Investigation

of the determinants of such diverse responses to the interaction with GM1 may help decipher how genetic variation can affect the fusogenic activity and infectivity of emerging SARS-CoV-2 variants.

3. Elucidating the contribution of GM1 binding to SARS-2-S-mediated viral attachment and entry into target cells might pave the way to the development of inhibitors of SARS-CoV-2 infection, based on multimers of oligosaccharides, such as the GM1os pentasaccharide, that mimic membrane-bound GM1, in the form of dendrimers, GM1os-coated scaffolds, or GM1a-functionalized microgels.

7. REFERENCES

- Anderson EJ, Rouphael NG, Widge AT, Jackson LA, Roberts PC, Makhene M, Chappell JD, Denison MR, Stevens LJ, Pruijssers AJ, *et al.* Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *N Engl J Med* 2020;383:2427–2438

- Avci FG, Akbulut BS, Ozkirimili E. Membrane active peptides and their biophysical characterization. *Biomolecules* 2018;8:77

- Bakkers MJ, *et al.* Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. *Cell Host Microbe* 2017;21:356–366

- Bakkers MJ, *et al*. Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. *Proc Natl Acad Sci USA* 2016;113:E3111–E3119

- Basak S, Hao X, Chen A, Chrétien M, Basak A. Structural and biochemical investigation of heptad repeat derived peptides of human SARS corona virus (hSARS-CoV) spike protein. *Protein Pept Lett* 2008;15:874–886

- Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci USA* 2009;106:5871–5876

- Boesweld S, Jans A, Rommel D, Bartneck M, Möller M, Elling L, Trautwein C, Strnad P, Kuehne AJC. Microgels sopping up toxins – GM1-functionalized microgels as scavengers for cholera toxin. *ACS Appl Mat Interfaces* 2019;11:25017-25023

- Bosch BJ, Bartelink W, Rottier PJM. Cathepsin L Functionally Cleaves the Severe Acute Respiratory Syndrome Coronavirus Class I Fusion Protein Upstream of Rather than Adjacent to the Fusion Peptide. *J Virol* 2008;82:8887

- Bosch BJ, Martina BE, Van Der Zee R, Lepault J, Haijema BJ, Versluis C, Heck AJ, De Groot R, Osterhaus AD, Rottier PJ. Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc Natl Acad Sci USA* 2004;101: 8455–8460

- Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 2003;77:8801–8811

- Broer R, Boson B, Spaan W, Cosset FL, Corver J. Important role for the transmembrane domain of severe acute respiratory syndrome coronavirus spike protein during entry. *J Virol* 2006;80:1302–1310

- Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 1994;371:37–43

- Burkard C, Verheije MH, Wicht O, van Kasteren SI, van Kuppeveld FJ, Haagmans BL, Pelkmans L, Rottier PJ, Bosch BJ, de Haan CA. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog* 2014;10:e1004502

- Corver J, Broer R, van Kasteren P, Spaan W. Mutagenesis of the transmembrane domain of the SARS coronavirus spike glycoprotein: Refinement of the requirements for SARS coronavirus cell entry. *Virol J* 2009;6:230

- de Groot RJ. Structure, function and evolution of the hemagglutininesterase proteins of corona- and toroviruses. *Glycoconj J* 2006;23:59–72

- Desforges M, Desjardins J, Zhang C, Talbot PJ. The acetyl-esterase activity of the hemagglutinin-esterase protein of human coronavirus OC43 strongly enhances the production of infectious virus. *J Virol* 2013;87:3097–3107

- Duquerroy S, Vigouroux A, Rottier PJ, Rey FA, Bosch BJ. Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology* 2005;335:276–285

- Fantini J, Di Scala C, Chahinian H, Yahi N. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents* 55 (2020) 105960

- Glowacka I, Bertram S, Müller MA, Allen P, Soilleux E, Pfefferle S, Steffen I, Tsegaye TS, He Y, Gnirss K, *et al*. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* 2011;85:4122–4134

- Helenius and Simons. Charge shift electrophoresis: simple method for distinguishing between amphiphilic and hydrophilic proteins in detergent solution. *PNAS* 1977;74(2):529-532

- Hirai M, Iwase H, Arai S, Takizawa T, Hayashi K. Interaction of gangliosides with proteins depending on oligosaccharide chain and protein surface modification. *Biophys J* 1998;74:1380-1387

- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020a pii:S0092-8674(20)30229-4[Epub ahead of print]. doi:10.1016/j.cell.2020.02.052

- Hoffmann M, Kleine-Weber H, Pöhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol. Cell 2020b, 78, 779–784.e775.

- Hsieh C-L, Goldsmith JA, Schaub JM, DiVenere AM, Kuo H-C, Javanmardi K, Le KC, Wrapp D, Lee AG, Liu Y, et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* 2020;369:1501

- Huang X. Human coronavirus HKU1 spike protein uses *O*-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase proteins as a receptor-destroying enzyme. *J Virol* 2015;89:7202-7213

- Hulswit RJG, *et al.* Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc Natl Acad Sci USA* 2019;116:2681–2690

- Hussain A, Hasan A, Nejadi Babadaei MM, Bloukh SH, Chowdhury MEH, Sharifi M, Haghighat S, Falahati M. Targeting SARS-CoV2 spike protein receptor binding domain by therapeutic antibodies. *Biomed Pharmacother* 2020;130:110559

- Ito T, *et al.* Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. *Virology* 1997;227:493–499

- Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, McCullough MP, Chappell JD, Denison MR, Stevens LJ, *et al.* An mRNA Vaccine against SARS-CoV-2—Preliminary Report. *N Engl J Med* 2020;383:1920-1931

- Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR Phylogenetic Analysis and Structural Modeling of SARSCoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *J Mol Biol* 2020;432:3309–3325

- Julien JP, *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 2013;342:1477–1483

- Kadam SB, Sukhramani GS, Bishnoi P, Pable AA, Barkvar VT. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights. J Basic Microbiol 2021;61(3):180-202

- Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL, Corbett KS, Graham BS, McLellan JS, Ward AB. Pre-fusion structure of a human coronavirus spike protein. *Nature* 2016;531:118–121

- Kleine-Weber H, Elzayat MT, Hoffmann M, Pöhlmann S. Functional analysis of potential cleavage sites in the MERS coronavirus spike protein. *Sci Rep* 2018;8:16597

- Langereis MA, *et al.* Complexity and diversity of the mammalian sialome revealed by nidovirus virolectins. *Cell Rep* 2015;11:1966–1978

- Langereis MA, Zeng Q, Heesters BA, Huizinga EG, de Groot RJ. The murine coronavirus hemagglutinin-esterase receptor-binding site: a major shift in ligand specificity through modest changes in architecture. *PLoS Pathog* 2012;8:e1002492

- Li W, Hulswit RJG, Widjaja I, Raj VS, McBride R, Peng W, *et al.* Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci USA* 2017;114:E8508-8517

- Li F, *et al.* Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J Virol* 2006;80:6794–6800

- Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005;309:1864–1868

- Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;426:450-454

- Liu C, *et al.* Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. *J Virol* 2015;89:6121–6125

- Liu S, Xiao G, He Y, Niu J, Escalante CR, Xiong H, Farmar J, Debnath AK, Tien P, Jiang S. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus:

implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* 2004;363(9413):938–947

- Lu G,Wang Q, Gao GF. Bat-to-human: Spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol 2015;23:468–478

- Lu L, Liu Q, Zhu Y, Chan K-H, Qin L, Li Y, Wang Q, Chan JF-W, Du L, Yu F, Ma C, Ye S, Yuen K-Y, Zhang R, Jiang S. Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat Commun* 2014;5:3067

- Lu G, *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 2013;500:227–231

- Lu Y, Liu DX, Tam JP. Lipid rafts are involved in SARS-CoV entry into Vero E6 cells. *Biochem Biopys Res Commun* 2008;369:344-349

- Lyumkis D, *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 2013;342:1484–1490

- Madu IG, Belouzard S, Whittaker GR. SARS-coronavirus spike S2 domain flanked by cysteine residues C822 and C833 is important for activation of membrane fusion. *Virology* 2009;393:265-271

- Matrosovich M, Herriere G, Klenk HD. Sialic acid receptors of viruses. *Top Curr Chem* 2015; 367: 1-28

- Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. *J Virol* 2010;84:12658

- Matsuyama S, Ujike M, Morikawa S, Tashiro M, Taguchi F. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc Natl Acad Sci USA* 2005;102:12543

- Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 2015;202:120–134

- Mou H, *et al.* The receptor binding domain of the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *J Virol* 2013;87:9379–9383

- Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, *et al.* Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 2020;11:1620

- Peng G, Xu L, Lin Y-L, Chen L, Pasquarella JR, Holmes KV, Li F. Crystal structure of bovine coronavirus spike protein lectin domain. J Biol Chem 2012;287:41931–41938

- Petit CM, Chouljenko VN, Iyer A, Colgrove R, Farzan M, Knipe DM, Kousoulas KG. Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology* 2007;360:264-274

- Qian Z, *et al.* Identification of the receptor-binding domain of the spike glycoprotein of human betacoronavirus HKU1. *J Virol* 2015;89:8816–8827

- Qingxin L, Huang Q, Kang C. Secondary Structures of the Transmembrane Domain of SARS-CoV-2 Spike Protein in Detergent Micelles, *Int J Mol Sci* 2022;23(3):1040

- Rosenthal PB, *et al.* Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* 1998;396:92–96

- Schultze B, *et al.* Transmissible gastroenteritis coronavirus, but not the related porcine respiratory coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. *J Virol* 1996;70:5634–5637

- Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li FStructural basis of receptor recognition by SARS-CoV-2. *Nature* 2020;581(7807):221-224

- Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA* 2005;102:11876

- Song H, *et al.* An open receptor-binding cavity of hemagglutinin-esterasefusion glycoprotein from newly-identified influenza D virus: basis for its broad cell tropism. *PLoS Pathog* 2016;12:e1005411

- Song HC, Seo MY, Stadler K, Yoo BJ, Choo QL, Coates SR, Uematsu Y, Harada T, Greer CE, Polo JM, *et al.* Synthesis and characterization of a native, oligomeric form of recombinant severe acute respiratory syndrome coronavirus spike glycoprotein. *J Virol* 2004;78:10328–10335

- Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, Lu L, Jiang S, Yang Z, Wu, Y, *et al.* Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect* 2020;9:382–385

- Tomasi M, Roda LG, Ausiello C, D'Agnolo G, Venerando B, GHidoni R, Sonnino S, Tettamanti G. Interaction of GM1 ganglioside with bovine serum albumin. Formation and isolation of multiple complexes. *Eur J Biochem* 1980;111:315-324

- Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol* 2019;26:481-489

- Vance TDR, Lee JE. Virus and eukaryote fusogen superfamilies. *Curr Biol* 2020;30:R737-R758

- Verma DK, Gupta D, Lal SK. Host lipid rafts play a major role in binding and endocytosis of influenza A virus. *Viruses* 2018;10 pii:E650

- Vlasak R, Luytjes W, Spaan W, Palese P. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc Natl Acad Sci USA* 1988;85:4526–4529

- Wickramasinghe IN, de Vries RP, Grone A, de Haan CA, Verheije MH. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J Viro*l 2011;85:8903–8912

- Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020;181:281–292.e286

- White, J.M.; Delos, S.E.; Brecher, M.; Schornberg, K. Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. Crit. Rev. *Biochem Mol Biol* 2008;43: 189–219

- Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3Å resolution. *Nature* 1981;289:366–373

- Wrapp D, Wang N, Corbert KS, Goldsmith JA, Hsieh CL, Abiona O, *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367:1260-1263

- Xia X. Domains and functions of Spike protein in SARS-CoV-2 in the context of vaccine design. *Viruses* 2021;13:109. doi: 10.3390/v13010109

- Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, Ying T, Liu S, Shi Z, Jiang S, Lu L. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol* 2020a;17:765-767. Doi:10.1038/s41423-020-0374-2

- Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, Qi F, Bao L, Du L, Liu S, Qin C, Sun F, Shi Z, Zhu Y, Jiang S, Lu L 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* 2020b;30:343-355

- Xia S, Yan L, Xu W, Agrawal AS, Algaissi A, Tseng A-T, Wang Q, Du L, Tan W, Wilson IA, Jiang S, Yang B, Lu L. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci Adv* 2019;5(4):eaav4580

- Xiao X, Feng Y, Chakraborti S, Dimitrov DS. Oligomerization of the SARS-CoV S glycoprotein: Dimerization of the N-terminus and trimerization of the ectodomain. *Biochem. Biophys Res Commun* 2004;322:93–99

- Xiong X, *et al.* Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* 2013;497:392–396

- Yahi N, Fantini J. Deciphering the glycolipid code of Alzheimer's and Parkinson's amyloid protein allowed the creation of a universal ganglioside-binding peptide. *PLoS One* 2014;9:e104751

- Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science* 2020 Mar 4 pii:eabb2762[Epub ahead of print]. Doi: 10.1126/science.abb2762

- Zeng Q, Langereis MA, van Vliet AL, Huizinga EG, de Groot RJ. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc Natl Acad Sci USA* 2008;105:9065–9069

- Zhang J, Xiao T, Cai Y, Chen B. Structure of SARS-CoV-2 spike protein. *Curr Opin Virol* 2021;50:173–182

- Zuilhof H. Fighting cholera one-on-one: The development and efficacy of multivalent cholera toxin-binding molecules. *Accounts Chem Res* 2016;49:274-285