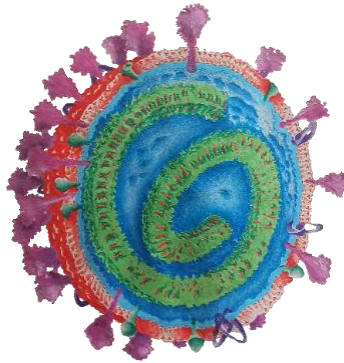PhD COURSE IN BIOCHEMISTRY

XXXV CYCLE

Structural bioinformatics analysis of the SARS-COV-2 proteome evolution to characterize the emerging variants of the virus and to suggest possible therapeutic strategies

PhD Student

MARTINA BIANCHI

Tutor   *Coordinator*

Prof. Stefano Pascarella   **Prof. Stefano Gianni**

December 2022

To my Grandmother Maria,

yesterday, today and tomorrow

ABSTRACT

SARS-CoV-2 is a new coronavirus responsible for the global COVID-19 pandemic, detected in China in December 2019 and that has spread rapidly across the world. Our unit, with its specific expertise in structural bioinformatics and molecular modelling, has been involved in collaboration with epidemiology and molecular genetics groups to study SARS-CoV-2 proteome and to suggest possible molecular strategies able to inhibit virus infection. All coronaviruses, including SARS-CoV-2, evolve and adapt to the host through accumulation of mutations generated by characteristics of the virus RNA-polymerase. This work can be divided into two parts: the first part is focused onto the predictions of the potential effects of the mutations on the functions of the SARS-CoV-2 Spike glycoprotein, whereas the second part is focused at suggesting possible therapeutic strategies. In particular, I performed docking analyses to study the possible mode ad sites of interaction of inorganic polyphosphates with ACE2 and SARS-CoV-2 RNA dependent RNA polymerase (RdRp) because the molecular genetics group with whom we collaborate suggested that polyphosphates can enhance ACE2 proteasomal degradation and impair synthesis of viral RNA. In addition, I developed a pipeline to predict the most frequent sites of interaction between Spike glycoprotein and neutralizing monoclonal antibodies in order to propose therapeutic alternatives more specific and selective.

TABLE OF CONTENTS:

1. INTRODUCTION

1.1. Characteristics of SARS-CoV-2

The SARS-CoV-2 is a new coronavirus (CoV), that causes a severe acute respiratory syndrome (SARS) called COVID-19, a contagious disease that was first identified in Wuhan, Hubei Province, China, in December 2019 and then has rapidly spread globally. The World Health Organization (WHO) declared the COVID-19 a public health emergency of international concern on 30 January 2020 and pandemic on 11 March 2020 [1,2]. Coronaviruses belong to the order Nidovirales, family Coronaviridae, and the subfamily Coronavirinae. They can infect domestic and wild animals causing mild to severe respiratory tract infection in birds and mammals including humans [3]. They are taxonomically divided into four coronavirus genera: α, β, γ and δ. The α and the β coronaviruses mainly infect mammals while the γ and the δ coronaviruses tend to infect birds [1]. Among the coronaviruses that cause human infections, there are HCoV 229E, HCoV and NL63 from the α-coronavirus genus and HCoV HKU1 and HCoV OC43 from the β-coronavirus genus that typically cause only the common cold symptoms. Infections caused by other β coronaviruses such as SARS-CoV-2, SARS-CoV and MERS-CoV may be asymptomatic or show mild respiratory symptoms but also severe acute respiratory disease and death [4]. A full genome sequence analysis revealed that SARS-CoV-2 shares 79% genome sequence identity with SARS-CoV and 50% with MERS- CoV [5]. Based on these data and on the results of phylogenetic analysis it can be deduced that SARS-CoV-2 and SARS-CoV form a distinct lineage within the subgenus *Sarbecovirus* and are relatively distant to MERS-CoV (belonging to the subgenus *Merbecovirus*) in the genus *Betacoronavirus* [6]. SARS-CoV and MERS-CoV caused the 2002–2004

SARS outbreak and the 2012 Middle East respiratory syndrome (MERS), respectively. The SARS-CoV outbreak originated in Southern China in November 2002 and spread very rapidly to other parts of the world mainly by international air travel. At the end of the epidemic in June 2003, 8422 cases with 916 deaths (case fatality rate of 11%) in 37 countries [7] were reported. In comparison, the MERS-CoV emerged in June 2012 in Saudi Arabia, has caused 2494 reported cases and 858 deaths in 38 countries [8]. Until a few months ago the origin of SARS-CoV-2 remained unclear, most scientists affirmed that the virus was likely of zoonotic origin but the possibility that the virus had a laboratory origin could not be excluded. The last August, Worobey et al have demonstrated that the SARS-CoV-2 is certainly of zoonotic origin, in fact their analyses indicate that the SARS-CoV-2 pandemic was caused by wildlife trade in China that began at the Huanan market food [9]. Similar evidences have been reported by a group of Chinese scientists that published a work where the full-length genome sequence of a bat coronavirus detected in *Rhinolophus affinis*, from China, named 'RaTG13', showed 96.2% identity to that of SARS-CoV-2, suggesting that bats can also be considered the most likely natural reservoir of this new coronavirus [10,11]. A group of scientists, in a study published on *Nature* the last February [12], affirm that they have found three viruses in bats in Laos that are more similar to SARS-CoV-2 that any known viruses. This fact confirms that the bats are a reservoir for this new coronavirus but also that it circulates in the Indochinese peninsula. Differences between the receptor binding domain (RBD) of the Spike glycoprotein (S) sites in bat coronavirus and SARS-CoV-2 suggest that this bat coronavirus did not directly infect humans, but it was transmitted through an intermediate host as SARS-CoV and MERS-CoV (e.g., palm civets for SARS-CoV, dromedary

camels for MERS-CoV) [6,13]. A study published in July 2020 on *Nature* [14] showed that a pangolin-CoV has 100%, 98.6%, 97.8% and 90.7% amino acid identity with SARS-CoV-2 in the Envelope (E), Membrane (M), Nucleocapsid (N) and Spike (S) proteins, respectively suggesting that the new coronavirus may have originated from the recombination of a virus similar to pangolin-CoV with one similar to RaTG13. However, other studies demonstrated the poor affinity of the pangolin virus for the human ACE2 receptor, excluding this animal as intermediate host. Therefore, it appears possible that the transmission occurs directly [15] (**Figure 1**).
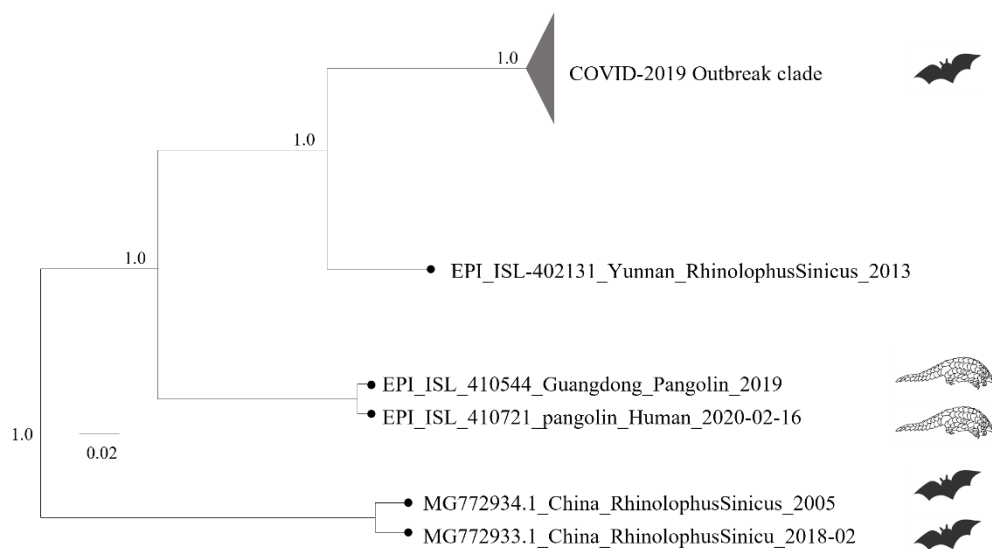
Figure 1: Maximum likelihood phylogeny was estimated with n = 143 complete genomes sequences from the current (2019–2020) SARS-CoV-2 epidemic plus n = 3 closely related bat strains plus n = 2 pangolin strains retrieved from GISAID. Number along branch represent bootstrap score. Scale bar represents expected substitutions per nucleotide site

1.2. Genomic and proteomic organization of SARS-CoV-2

The genome of SARS-CoV-2 is a single-stranded positive-sense ribonucleic acid (ssRNA+) that measures, on average, 30 kilobases. Two thirds of the genome consist of the replicase genes encoding for two large polyproteins, pp1a and pp1ab, which undergo a series of proteolytic cleavages to form 16 non structural proteins. The remaining one third of the genome contains ORFs that encode structural proteins like spike, envelope, membrane, and nucleocapsid proteins. In addition, the SARS-CoV-2 genome contains eight ORFs which code for accessory proteins named 3a, 3b, 6, 7a, 7b, 8, 9b, 10 and 14 (also named 9c). ORF3a and 3b are located between S and E genes, ORFs 3a and 3b. In particular, the ORF3b overlaps the 3' half of ORF3a and the 5′ end of the envelope protein gene. Four ORFs are found between M and N genes (ORF6, ORF7a, ORF7b, ORF8). The ORFs 9b and 14 called also 9c are encoded by an alternative open reading frame within the Nucleocapsid ORF (N). [16,17,18] (**Table 1 and Figure 2**).

Table 1: Summary of SARS-CoV-2 genes and corresponding proteins

GENES	PROTEINS
ORF1ab	nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp12, nsp13, nsp14, nsp15, nsp16
S	Spike
ORF3a	ORF3a
ORF3b	ORF3b
E	Envelope

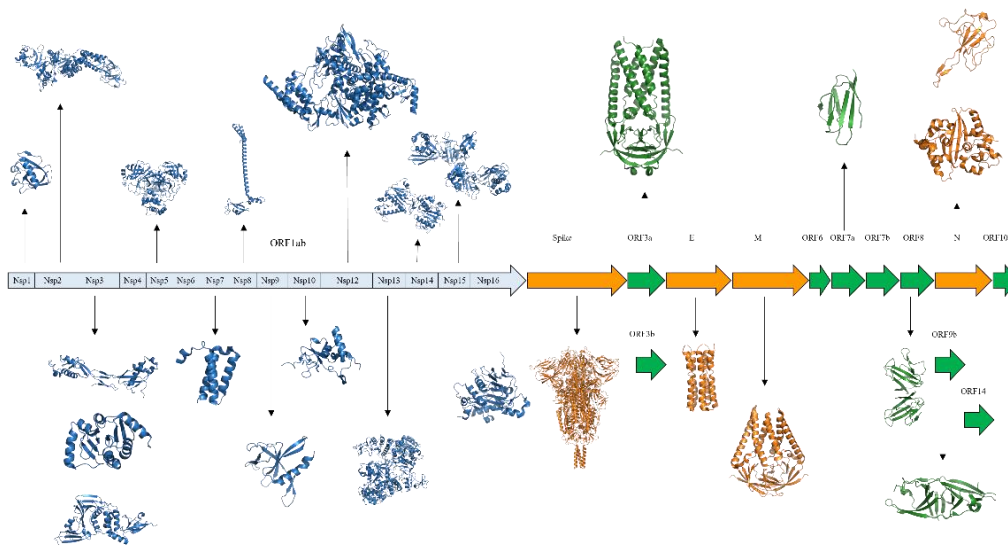| M	Membrane
ORF6	ORF6
ORF7a	ORF7a
ORF7b	ORF7b
ORF8	ORF8
N	Nucleocapsid
ORF9b	ORF9b
ORF14	ORF14
ORF10	ORF10

Figure 2: Genomic and proteomic organization of SARS-CoV-2. The ORF1ab gene is indicated as light blue rectangles, this region is proteolytically cleaved to form 16 nonstructural proteins, shown in sky-blue cartoon models. The structural proteins genes are indicated as orange rectangles and the proteins encoded are shown in orange cartoon models. The accessory

proteins genes are indicated as green rectangles and the protein encoded by it are shown in green cartoon models.

1.2.1. Non structural proteins (nsps)

The pp1ab and pp1a proteins located at the 5′-terminus of the genome comprise 16 non structural proteins including nsp1 to nsp10 and nsp12 to nsp16 [19]. The nonstructural protein 1 (nsp1) is a small 180 residue protein, a virulence factor that binds the host 40S subunit in ribosomal complexes. The N-terminal domain consists of seven antiparallel β-strands folded as a closed β-barrel. There is also one α-helix which is located at the opening of the barrel and other two helices which are positioned on one side of the barrel. [20]. The C-terminal domain of the protein blocks the mRNA entry channel of the ribosome to promote endonucleolytic cleavage of host mRNA facilitating the suppression of the host innate immune response [21]. The non structural protein 2 (nsp2) comprises 638 amino acids, it can interact with host proteins PHB1, PHB2 and with the actin-nucleation-promoting WASH protein [22,23,24]. The non structural protein 2 participates in biological processes such as the maintenance of the functional integrity of the mitochondria, host immune regulation and endosomal transport [25,26]. The N-terminal domain of nsp2 is composed of ten α-helices, fourteen β-strands, and three classic zinc-finger (ZnF) motives, while the C-terminal domain consists of only 14 β-strands, and one helix. The middle region contains three β-strands and nine α-helices [27]. The non structural protein 3 (nsp3) also called papain-like protease, PLpro, is the largest protease produced by coronaviruses. It contains 16 domains, namely Ubiquitin-like domain 1, (Ubl1), Hypervariable region (HVR) or acidic

domain, Macrodomain I/II/III (MacI/II/III), Domain Preceding Ubl2 and PL2pro (DPUP), Ubiquitin-like domain 2 (Ubl2), Papain-like protease two domain (PL2pro), Nucleic acid binding domain (NAB), betacoronavirus-specific marker domain (βSM), Transmembrane domain 1 (TM1), nsp3 ectodomain (3Ecto), Transmembrane domain 2 (TM2), Amphipathic helix region (AH1), domains specific to *Nidovirales* and *Coronaviridae* (Y1 & CoV-Y). These domains function synergistically to regulate viral infection. In particular, the nsp3 plays an essential role in viral suppression of the host immune response [28]. The full-length protein structure of nsp3 is currently not available, although there are domains of the protein which have been solved individually such as the Ubiquitin-like domain 1 (Ubl1), the Macrodomain 1 (Mac1) and the Papain-like protease domain (PLpro). The structure of Ubl1 domain folds like a human ubiquitin (Ub), the MacI domain contains seven-stranded β-sheet and six α-helices which are located in the central part and in the external part of the structure, respectively [29]. The structure of PLpro domain is similar to a right-hand with thumb, palm, and finger subdomains [30]. The non structural protein 4 (nsp4) is about 500 amino acids long. It participates to viral replication-transcription complex and it is also able to modify endoplasmic reticulum (ER) membranes [31]. Concerning the structure of nsp4, only the crystal structure of its C-terminal domain is currently available. The non structural protein 5 (nsp5) is the main-protease (Mpro) of SARS-CoV-2 and it comprises 306 amino acids. It cleaves the polyprotein 1ab at 11 sites with stringent substrate specificity to produce nsp14 and nsp16. For this reason, it plays an essential role in viral replication and maturation of non structural proteins and therefore is considered as a potential target for antiviral drugs [32]. Nsp5 is a homodimer in which each monomer

is composed of three domains: the domains 1 and 2 have an antiparallel β-barrel whereas the domain 3 contains five α-helices that are involved in protein dimerization [33]. The non structural protein 6 (nsp6) is 290 amino acid long. It plays a role in the initial induction of autophagosome and double membrane vesicle formation [34]. Currently, structural data are not available for this protein. The non structural proteins 7, 8 and 12 (nsp7, nsp8 and nsp12) are discussed together since they interact forming the coronavirus replication machinery, in which nsp12 is the RNA-dependent RNA polymerase (RdRp) and nsp7 and nsp8 function as cofactors of nsp12 [35,36]. Nsp7 is 83 amino acids long and its structure comprises only four α-helices (α1-α4), while nsp8 is 198 amino acids in length and its structure presents a long helical N-terminal domain and five α-helices and one four-stranded antiparallel β-sheet in the C-terminal domain. These two proteins form a hetero tetramer that binds nsp12 [37]. The non structural protein 12 comprises 932 amino acids and its structure is composed of three domains: starting from the N-terminal to the C-terminal side, the transferase domain, the interface domain, and the RNA-dependent RNA polymerase (RdRp) domain are found [38]. The non structural protein 9 (nsp9) is 113 amino acid long and plays an important role in viral replication. This protein is a homodimer, its structure presents a seven β-strand barrel and one flexible α-helix [39]. SARS-CoV-2 non structural protein 10 (nsp10) comprises 139 amino acids. It is able to bind and activate the exoribonuclease and methyltransferase activities of nsp14 and nsp16, respectively. Consequently, nsp10 plays an important role in the viral mRNAs capping apparatus. In the N-terminal region, the nsp10 is composed of two antiparallel α-helices connected to small β-sheet the C-terminal region [40]. The nonstructural protein 13 (nsp13) consists of 601 amino acids and it represents

a multi-functional protein which has helicase, RNA 5′-triphosphatase and NTPase activities. The structure of this protein contains five domains. Starting from the N-terminus to the C-terminus, they are: the zinc-binding domain (ZBD), the stalk domain (SD), the inserted domain 1B, and two helicase domains: RecA1 and RecA2 [41,42,43]. The non structural protein 14 (nsp14) is 527 amino acid long and it is an enzyme which possesses two different activities: 3′-to-5′ exoribonuclease (ExoN) and N7-guanine methyltransferase. These two activities are responsible for nascent RNA proofreading and mRNA capping during viral RNA replication, respectively [44,45,46]. The N-terminal ExoN domain forms a complex with nsp10. This region of nsp14 presents two zinc-finger motives which are associated with the stability and enzymatic activity of ExoN. The activity of C-terminal N7-MTase domain of nsp14 is independent of nsp14-nsp10 complex formation and its structure includes one three-stranded β-sheet and one zinc finger domain (ZnF3) [47,48]. The non structural protein 15 (nsp15) comprises 346 amino acids and it is an uridylate-specific endoribonuclease (NendoU), able to cleave the 5'-polyuridines from negative-sense viral RNA and to prevent the activation of host sensor system. The structure of this protein revealed a hexameric assembly, where each protomer is composed of three domains: the N-terminal domain, the variable middle domain, and the C-terminal domain [49]. The non structural protein 16 (nsp16) is 298 amino acid long and it is a methyltransferase which is able to mediate mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs. Nsp16 in complex with nsp10 can efficiently perform its function as methyltransferase. The structure of nsp16 consists of a β-sheet of eight strands flanked by two α helices on one side and three helices on the other [50]. Currently, the full-length protein structures of the nsp1, nsp2, nsp5, nsp7, nsp8,

nsp9, nsp10, nsp12, nsp13, nsp14, nsp15 and nsp16 are available in the Protein Data Bank (PDB).

1.2.2. Structural proteins

The SARS-CoV-2 proteome contains four important structural proteins: envelope (E) protein, membrane (M) protein, nucleocapsid (N) protein and Spike (S) glycoprotein. These proteins play key roles in the viral life cycle (**Figure 3).**
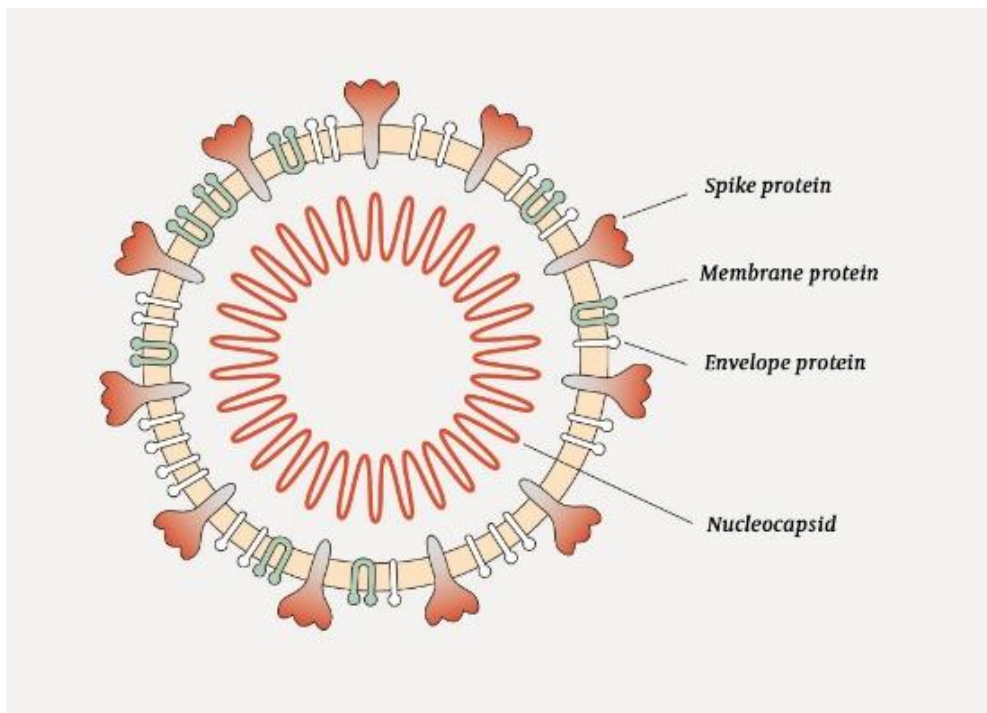

Figure 3: Schematic illustration of SARS-CoV-2 showing the four structural proteins: spike protein (S), membrane protein (M), envelope protein (E) and nucleocapsid (N). The S, M and

E proteins are anchored in the lipid bilayer that makes up the viral membrane. The N protein is associated with viral RNA on the inside of the membrane.

The Envelope protein is a tiny integral membrane protein composed of 75 amino acids, it plays a central role in virus morphogenesis, assembly and other functions which are not yet clear, and for this reason the E protein can be considered a critical multifunctional structural protein [51]. The full-length structure is currently available which presents three different domains: the N-terminal ectodomain (residues 1-7), transmembrane domain (TMD) (residues 8–38), and C-terminal domain (residues 39–75). It forms a homopentameric cation channel that is important for virus pathogenicity. The structure of pentameric TMD is formed by 5 α-helices, the stability of which is determined by several interactions in particular hydrophobic forces. Currently, only the structure of transmembrane domain of the protein is available in the Protein data Bank (PDB code: 7K3G) [52]. The Membrane protein is 222 amino acids in length, it is a component of the viral envelope involved in different processes such as modification and trafficking of multiple viral proteins, and virus assembly via its interactions with other proteins [53]. The cryo-EM structure of SARS-CoV-2 Membrane protein was recently deposited in the PDB (PDB ID: 8CTK). It is a homodimer structurally related to the SARS-CoV-2 ORF3a viroporin, suggesting a shared ancestral origin [54]. The nucleocapsid protein is composed of 419 amino acids and it is able to pack the positive strand viral genome RNA into the ribonucleoprotein particle (RNP) and to interfere with the interferon pathway of the host. The structure comprises two main domains: N-terminal domain and the C-terminal domain, joined by the flexible linker,

rich in serine and arginine. The full-length structure of this protein has not yet been resolved. Indeed, the only regions solved are the N- and C-terminal domains, while the coordinates of the linker portion are still missing. The N-terminal domain of the protein consists of a five-stranded antiparallel β-sheet interposed between two short α-helices whereas the C-terminal domain consists of six α-helices and two β-strands [55]. This protein is able to induce a protective immune response, and, for this reason, it is interesting as an antibody target. Kang, S. *et al.* isolated a human mAb from convalescent COVID-19 patients. This mAb, called nCoV396, effectively interacts with the N-terminal domain of the N protein [56]. The Spike glycoprotein (S) is 1273 amino acid long, it is a trimeric, transmembrane protein which protrudes from the viral surface [57]. This protein is able to recognize and bind the host receptor angiotensin-converting enzyme 2 (ACE2) mediating viral cell entry [58]. It is involved in indispensable functions for the virus such as receptor recognition, viral attachment, and entry into host cells. For this reason, it represents the main target of neutralizing monoclonal antibodies (mAbs), vaccines and anti-coronavirus drugs. Each monomer of the Spike glycoprotein is composed of two functional subunits: the subunit S1 and the subunit S2. In the S1 subunit, there is an N-terminal domain (residues 14–305) and a receptor-binding domain (residues RBD, 319–541), the main function of which is to bind the receptor on host cell. In the subunit S2, there is the fusion peptide (FP) (residues 788–806), the heptapeptide repeat sequence 1 (HR1) (residues 912–984), the HR2 (residues 1163–1213), the TM domain (residues 1213–1237), and the cytoplasm tail (residues 1237–1273). The function of S2 subunit is to mediate fusion of the virion and cellular membranes. The Spike glycoprotein has two different conformations: the closed state (inaccessible) and the open

state (accessible) characterized by the position, down or up, of its receptor-binding domains (RBDs) as shown in **Figure 4**.
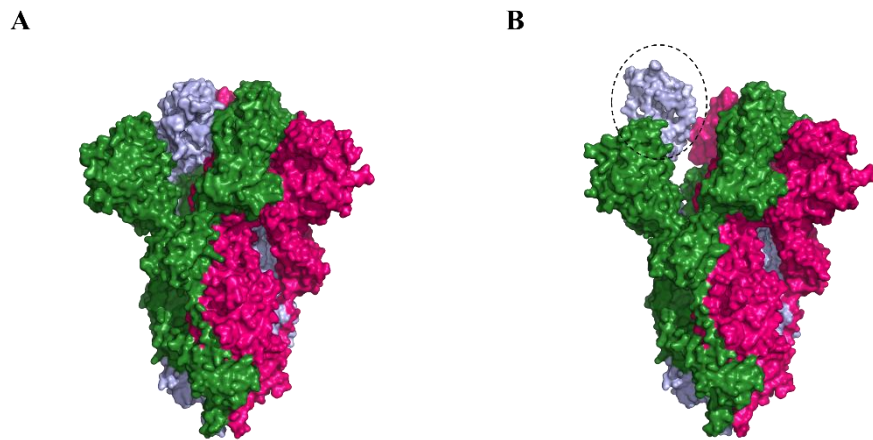
Figure 4: Schematic of SARS-CoV-2 spike- protein primary structure. Different chains are shown by different colors. The closed state (PDB: 6VXX) and the open state (PDB: 6VYB) of the SARS-CoV-2 S glycoprotein are shown in the panel **A** and **B**, respectively. The accessible RBD is highlighted in the dashed circle.

In the closed state, the subunits S1 and S2 remain non-covalently bound after a cleavage at the boundary between the S1 and S2 subunits (S1/S2) contributing to stabilize the prefusion conformation. The prefusion conformation indicates the Spike protein's state before the fusion between the viral membrane with the host cell membrane. To activate the protein for membrane fusion via extensive irreversible conformational changes the Spike

glycoprotein is further cleaved by host proteases at the so-called S2′ site located immediately upstream of the fusion peptide [59].

1.2.3. Accessory proteins

The accessory proteins of SARS-CoV-2 play important roles in pathogenesis. SARS-CoV-2 encodes nine accessory proteins including ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF10 and ORF14 (also named ORF9c) [60]. ORF3a is 275 amino acids in length, it represents an integral membrane protein able to function as an ion channel that may modulate virus release [61]. This protein is an oligomer that can exists as dimer or a tetramer; its monomer contains a transmembrane domain composed of three α-helices and a C-terminal domain composed of eight antiparallel β-strands [62]. ORF3b is an accessory protein which contains 22 amino acids. Despite its small size, it has been demonstrated that this protein is a potent interferon (IFN) antagonist [63]. ORF6 is a small protein which consists of 61 amino acids. Recently, it has been shown that ORF6 acts as a virulence factor to accelerate viral replication, resulting in disease progression [64]. The crystallographic structures of ORF3b and ORF6 are currently not available. ORF7a is a transmembrane protein which contains 121 amino acids, it represents another SARS-CoV-2 protein with the ability to hamper the IFN-I response [65]. Structurally, this protein is composed of a N-terminal signal peptide, an Ig-like ectodomain, a transmembrane region and one ER retention motif in its cytoplasm-exposed tail folded as seven β-strands divided into two tightly packed β-sheets [66]. ORF7b is 43 residues long, its function remains to be

investigated. ORF8 is a 121 amino acid protein, and it plays a role in host-virus interaction [67]. The structure of this protein contains a N-terminal transmembrane region and a central Ig-like domain. This protein is a homodimer, and each monomer consists of two antiparallel β-sheets [68]. ORF9b is an ORF located within the nucleocapsid (N) gene coding for a 97 amino acid long protein. It appears to be involved in the suppression of anti-viral interferon responses [69]. This protein is a homodimer where its monomers consist of β-strands only. ORF14 is a 73 amino acid protein, its expression impairs interferon signaling, antigen presentation, and complement signaling, while it induces IL-6 signaling. The crystal structure of ORF14 is not currently available. However, bioinformatics analyses suggest that this protein contains a putative transmembrane domain [70]. A putative 38 amino acids long protein has been described as ORF10. Pancer et al. have been shown that this protein is not essential in human SARS-CoV-2 infection [71]. Currently, the complete structures of ORF3a and ORF8 and the partial structures of ORF7a and ORF9b are available at the Protein Data Bank (PDB),

1.3. Mechanisms of SARS-CoV-2 entry into cell

The Spike glycoprotein binds the receptor angiotensin converting enzyme 2 (ACE2) through its receptor-binding domain (RBD) to enter into cell [58]. ACE2 is an 805 amino acid carboxypeptidase that removes a single amino acid from the C-terminus of its substrates including angiotensin II. It is highly expressed in the heart, kidneys and lungs and it was found to be the main host cellular receptor recognized by the Spike glycoprotein [72,73]. The RBD domain consists of two subdomains: a core domain folded as five-stranded antiparallel β-sheet and an external flexible loop named the receptor-binding motif (RBM) [74]. RBD recognizes and binds ACE2 through its RBM. Interaction involves 20 residues of ACE2 and 17 residues from RBD. [72]. ACE2 is not the only receptor of SARS-CoV-2; in fact, also a variety of other receptors, such as neuropilin-1 and AXL [75,76], are deemed to be involved in virus interaction. Upon engagement of ACE2 by the receptor binding domain (RBD), the Spike glycoprotein requires two proteolytic cleavages by host cell: the first is localized in the S1-S2 junction and it is cleaved by furin, whereas the second is localized to the site S2' and it is cleaved by the TMPRSS2, a protease presents on the cell surface or by the cathepsin, a protease localized in the endosomes [77,78,79]. The virus can enter in the cell through two distinct pathways depending by different proteases: the activation by TMPRSS2 or by the cathepsin. The latter occurs when there is an insufficient expression of TMPRSS2 and, in this way, the virus is internalized via clathrin-mediated endocytosis in endolysosomes [80,81]. After that, it follows the fusion between viral and cellular membranes forming a pore which enables the

penetration of the viral RNA into the host cell cytoplasm for uncoating and replication [82] (**Figure 5).**
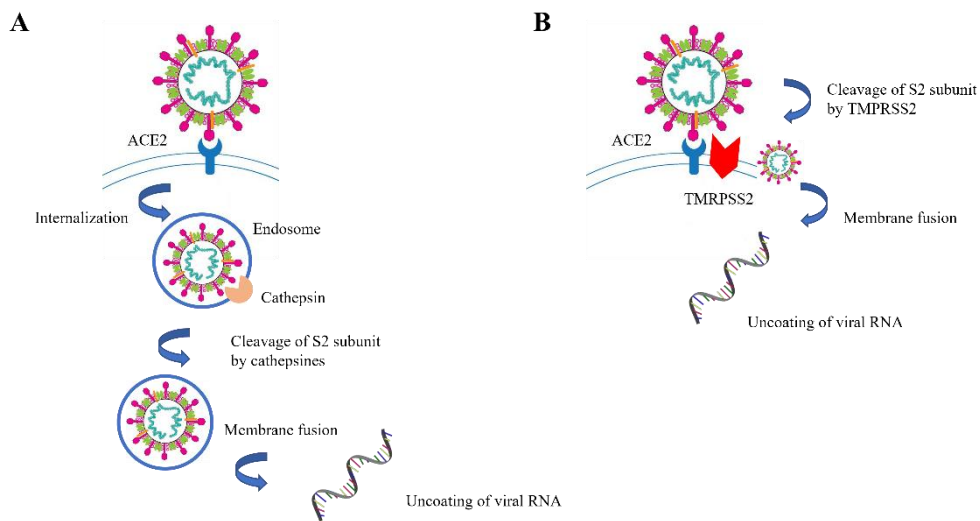
Figure 5: Description of two different entry pathways of the SARS-CoV-2. **A)** The endosomal entry pathway which uses cathepsins to cleave the S2' subunit of Spike protein. **B)** The cell surface pathway which uses TMPRSS2 to cleave the S2' subunit. Both processes enable membrane fusion and cell invasion.

1.4. Variants classification and definitions of SARS-CoV-2

All viruses, including SARS-CoV-2, when replicate, can evolve as changes (called mutations) are inserted in the nucleotide sequence due the slightly imprecise activity of the proofreading exoribonuclease. This causes incorporation of uncorrected errors during the activity of RNA-dependent RNA polymerase (RdRp) [83]. In general, most mutations have not strong

18

impact on the virus's properties, and they are considered relatively neutral. However, some of them may influence the virus's transmission or the severity of virus-associated disease. During the early phase of the pandemic, the genetic evolution of the SARS-CoV-2 was minimal, while since then SARS-CoV-2 generated several new variants. The World Health Organization classifies the variants as variants of interest (VOIs) and variants of concern (VOCs) and named them with letters of the Greek alphabet [84]. The researchers and the public health agencies worldwide used the Pango nomenclature to classify genetic lineages for SARS-CoV-2 [85].

1.4.1. Variants of interest (VOIs)

The variants of interest (VOIs) are variants with changes that may influence virus characteristics such as transmissibility, virulence, disease severity or immune escape and, for this reason, these variants should be monitored closely. Since the beginning of the pandemic, the World Health Organization has described eight variants of interest: Epsilon, Zeta, Eta, Theta, Iota, Kappa, Lambda and Mu. Among the variants mentioned above, the Epsilon (B.1.427 and B.1.429) variants are the only designated by the Centers for Disease Control and Prevention (CDC) as VOC in USA, whereas all the other variants are considered as VOIs [86,87]. Currently, there are no SARS-CoV-2 variants

designated as VOI(s). In the table below are shown the Pango lineage, the first detection and all mutations for each variant.

Table 2: List of variants of interest

Pango Lineage	First detection	Mutation list
Epsilon: B.1.427 and B.1.429	California (USA) July 2020	B.1.427 → nsp2: T85I nsp4: S395T nsp12: P323L nsp13: P53L, D260Y Spike: S13I, W152C, L452R, D614G ORF3a: Q57H N: T205I B.1.429→ nsp2: T85I nsp9: I65V nsp12: P323L nsp13: D260Y Spike: S13I, W152C, L452R, D614G ORF3a: Q57H N: T205I
Zeta: P.2	Brazil, April 2020	nsp5: L205V nsp7: L71F nsp12: P323L Spike: E84K, D614G, V1176F N: A119S, R203K, G204R, M234I
Eta: B.1.525	New York (USA), November 2020	nsp3: T1189I nsp6: Δ106-106 nsp6 nsp12: P323F Spike: Q52R, A67V, Δ69/70, Δ144, E484K, D614G, Q677H, F888L E: L21F M: I82T ORF6: Δ2/3 N: S2Y, Δ3, A12G, T205I
Iota B.1.526	New York (USA), November 2020	nsp2: T85I nsp4: L438P nsp6: Δ106-108 nsp12: P323L nsp13: Q88H Spike: L5F, T95I, D253G, E484K, D614G, A701V ORF3a: P42L, Q57H ORF8: T11I N: P199L, M234I
Theta: P3	Philippines and Japan, February 2021	nsp3: D736G, S1807F nsp4: D217N, L438P nsp6: D112E nsp7: L71F

		nsp12: P323L nsp13: L280F, A368V Spike: Δ141-143, Δ243/244, E484K, N501Y, D614G, P681H, E1092K, H1101Y, V1176F ORF8: K2Q N: R203K, G204R
Kappa: B.1.617.1	India, December 2020	nsp3: T749I nsp6: T77A nsp12: P323L nsp13: G206C, M429I nsp15: K259R, S261A Spike; G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H ORF3a: S26L M: I82S ORF7a: V82A N: R203M, D377Y
Lambda: C.37	Perù, June 2021	nsp3: T428I, P1469S, F1569V nsp4: L438P, T492I nsp5: G15S nsp6: Δ106-108 nsp12: P323L Spike: G75V, T76I, R246N, Δ247-253, L452Q, F490S, D614G, T859N N: P13L, R203K, G204R, G214C
Mu: B.1.621	Colombia, August 2021	nsp3: T237A, T720I nsp4: T492I nsp6: Q160R nsp12: P323L nsp13: P419S Spike: T95I, Y144S, Y145N, R346K, E484K, N501Y, D614G, P681H, D950N ORF3a: Q57H, Δ256/257 ORF8: T11K, P38S, S67F N: T205I

1.4.2. Variants of concern (VOCs)

The variants of concern are variants with changes that increase the transmissibility, the virulence or in clinical disease presentation and decrease the susceptibility to neutralizing antibodies or to antiviral drugs. Since the beginning of the pandemic, five SARS-CoV-2 variants of concern (VOCs)

have been identified: Alpha, Beta, Gamma, Delta, Omicron. The Alpha variant includes eight mutations in the Spike protein, among these N501Y within the RBD increases the binding affinity to ACE2 receptor, the Δ69/70 mutation is potentially associated with immune evasion and the P681H mutation facilitates epithelial-cell entry [88]. The Beta variant includes nine mutations in the Spike protein, among these the combination of K417N, E484K, and N501Y in the RBD enhances the binding affinity to ACE2 [89]. The Gamma variant shares the mutations E484K and N501Y with the Beta variant while the position 417 is mutated into threonine instead of asparagine as observed in the Beta variant. The presence of these mutations has been linked to increased risk of transmission and reduction of immune response [90]. The Delta variant, in its RBD, includes a mutation in position 452 where the leucine is replaced by arginine. This substitution may cause structural changes that enhance the binding affinity to ACE2 receptor [91,92]. In comparison to Alpha, Beta, Gamma and Delta variant, the Omicron variant (BA.1) presents a high number of mutations especially in the Spike protein increasing transmissibility and resistance to neutralizing antibodies [93]. Following the original Omicron variant, several subvariants of Omicron have emerged: BA.2, BA.3, BA.4, and BA.5 [94,95]. In the table below are shown the Pango lineage, the first detection and all mutations for each variant.

Table 3: List of variants of concern

Pango Lineage	First detection	Mutation list
Alpha: B.1.1.7	United Kingdom, December 2020	nsp3: T183I, A890D, I1412T nsp6: Δ106-108 nsp12: P323L Spike: Δ69/70, Δ144, N501Y, A570D, D614G, P681H, T716I, S982A D1118H ORF8: Q27stop, R52I, Y73C

		N: D3L, R203K, G204R, S235F
Beta: B.1.351	South Africa, October 2020	nsp2: T85I nsp3: K837N nsp5: K90R nsp6: Δ106-108 nsp12: P323L Spike: D80A, D215G, Δ241-243, K417N, E484K, N501Y, D614G A701V ORF3a: Q57H, S171L E: P71L N: T205I
Gamma: P.1	Brazil, December 2020	nsp3: S370L, K977Q nsp6: Δ106-108 nsp12: P323L nsp13: E341D Spike: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F ORF3a: S253P ORF8: E92K N: P80R, R203K, G204R
Delta: B.1.617.2	India, December 2020	nsp3: A488S, P1228L, P1469S nsp4: V167L, T492I nsp6: T77A nsp12: P323L, G671S nsp13: P77L nsp14: A394V Spike: T19R, G142D, E156G, Δ157/158, L452R, T478K, D614G, P681R, D950N ORF3a: S26L M: I82T ORF7a: V82A, T120I ORF7b: T40I ORF8: Δ119/120 N: D63G, R203M, G215C, D377Y
Omicron: BA.1	South Africa, November 2021	nsp3: K38R, S1265I, Δ1266, A1892T nsp4: T492I nsp5: P132H nsp6: Δ105-107, I189V nsp12: P323L nsp14: I42V Spike: A67V, Δ69/70, T95I, G142D, Δ143-145, N211I, Δ212, R214ins, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H,

23

		N764K, D796Y, N856K, Q954H, N969K, L981F
		E: T9I
		M: D3G, Q19E, A63T
		N: P13L, Δ31-33, R203K, G204R

1.5. Therapeutic strategies for COVID-19

Nowadays, the therapeutic strategies to treat the COVID-19 mainly include antiviral, anti-inflammatory drugs but also the antibody therapies seem to have an important role in the SARS-CoV-2 infection treatment. In addition, several scientists are testing different therapeutic options including monoclonal antibodies [96].

1.5.1. Monoclonal antibodies for COVID-19

The immunotherapy is an effective method to treat infectious diseases including SARS-CoV-2. Generally, the monoclonal antibodies (mAbs) can bind a specific part of the virus and "neutralize" it providing an effective therapeutic intervention [97]. The neutralizing mAbs for treatment of the COVID-19 derive from the B cells of convalescent exposed to SARS-CoV-2 antigens [98]. Most neutralizing mAbs to combat COVID-19 were generated against Spike glycoprotein of the SARS-CoV-2 [99]. Among these there are many human neutralizing mAbs isolated from COVID-19-convalescent donors which target the Spike receptor binding domain (RBD). Barnes et al [100] have categorized eight distinct COVID-19 human neutralizing antibodies in complex with RBD into four classes:

1. Neutralizing mAbs block ACE2 and bind only to 'up' conformation RBDs;

24

2. Neutralizing mAbs block ACE2 and bind both 'up' and 'down' RBDs and can contact adjacent RBDs;
3. Neutralizing mAbs bind outside the ACE2 site and recognize both 'up'and 'down' RBDs;
4. Neutralizing mAbs do not block ACE2 site and bind only to 'up' RBDs.

The neutralizing mAbs of the class 1 and 2 inhibit viral entry by directly competing with ACE2, on the other hand, the neutralizing mAbs of the class 3 and 4 do not overlap with ACE2 binding site (**Figure 6**).
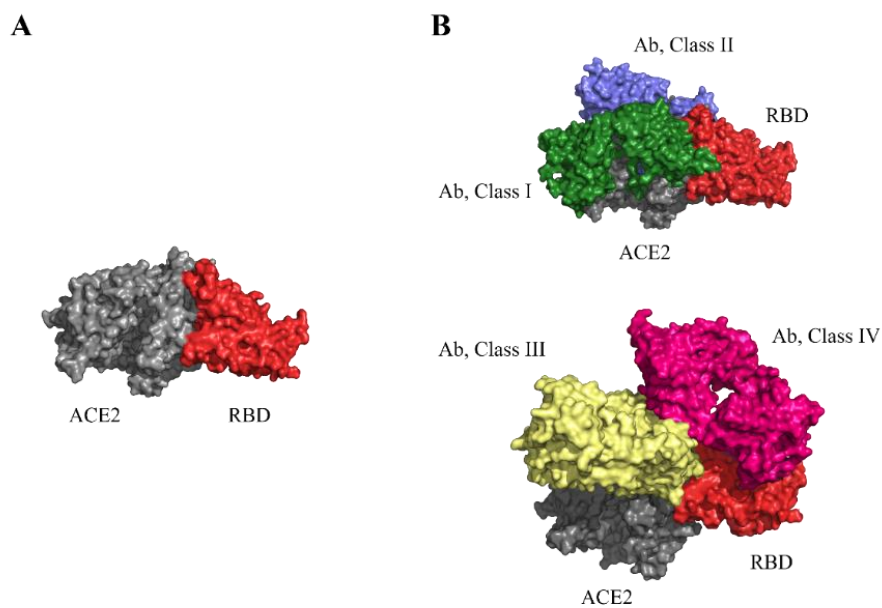

Figure 6: A) Binding region between ACE2 and RBD. **B)** On the top, the figure displays the complexes between the antibodies of the class 1 (blue) and the class 2 (green) and RBD superposed to the complex with ACE2 (grey); on the bottom, the figure shows the complexes

between the antibodies of the class 3 (yellow) and the class 4 (magenta) and RBD overlapped to the complex with ACE2 (grey).

Among these neutralizing mAbs, there are considered therapeutic for COVID-19 treatment:

- COV2-2130 (AZD1061/Cilgavimab) and COV2-2196 (AZD8895/Tixagevimab) are two human mAbs isolated from patient convalescing from COVID-19. These mAbs belong to the class 1 because structural studies revealed that they can bind the RBD in 'up' conformation blocking the access to the human receptor ACE2. From the combination of these two mAbs, the AstraZeneca biotechnology company has developed another antibody named AZD7442 which can neutralize SARS-CoV-2 and all tested SARS-CoV-2 variants of concern [101,102]. Concerning the Omicron variant, it was recently observed that AZD7442 treatments retained inhibitory activity against several Omicron subvariants [103]

- LyCoV555 (Bamlanivimab) is a neutralizing monoclonal antibody developed by AbCellera Biologics and Eli Lilly to treat the COVID-19 disease. Structural studies revealed that LyCoV555 can bind the RBD in both 'up' and 'down' conformations, suggesting that it belongs to the class 2 [104]. This mAbs was the first SARS-CoV-2 neutralinsg mAb authorized for clinical use. The Bamlanivimab activity is retained against the Alpha variant, it is slightly lost against the Beta, Gamma, and Delta variants and it is totally lost against the Omicron variant [105]

- Sotrovimab is a derivative of the S309 mAb, it belongs to the class 3 because structural studies revealed that it can bind the RBD in both 'up' and 'down' conformations, outside the ACE2 site [106]. The Sotrovimab activity is retained against the Alpha, Beta, Gamma, Delta, and Omicron variants [107,103].

Nowadays, there is little data on neutralizing mAbs binding to epitopes outside the RBD and on the mechanism with which they protect us against COVID-19 disease. Among these, there are some which recognize the N-terminal domain (NTD) of the SARS-CoV-2. Chi *et al* demonstrated that one mAb, named 4A8, binds the NTD of the Spike protein and has a potent neutralizing activity. In the **Figure 7** are shows that two loops (N3 and N5) of the NTD mediate the interaction with 48A, (**Figure 7**)
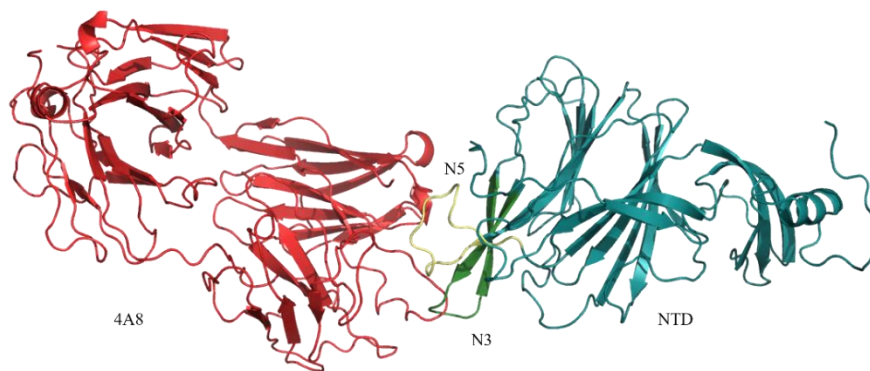
Figure 7: The figure shows the complex NTD-4A8 (PDB code: 7C2L). The two loops: N3 and N5 which mediate the interaction between the monoclonal antibody 4A8 and the N-terminal domain of the Spike protein are shown in green and yellow, respectively.

This mAb can recognize and bind an epitope of the NTD regardless of RBD inhibition by other mAbs and drugs, suggesting that the development of additional therapeutics strategies against SARS-CoV-2 could exploit Spike epitopes outside the RBD [108]. Concerning the SARS-CoV-2 variants, at the moment, there are not data which demonstrate if 4A8is effective.

2. AIM OF THE WORK

The genome of all viruses, including the SARS-CoV-2 genome, accumulate mutations rapidly. The analyses of the evolution of the virus are essential for tracing the virus transmission and the potential variants to manage the pandemic and to suggest possible therapeutic strategies.

The objective of this PhD project was, therefore, twofold.

First, it aimed at tracking the SARS-CoV-2 evolution from molecular point of view using a structural bioinformatic approach on several proteins of the SARS-CoV-2 proteome: nsp2, nsp3, nsp6, Envelope, Membrane, ORF3a and Spike glycoprotein

The second objective focused on the study of the interaction between SARS-CoV-2 proteins and polyphosphates (PolyPs) or monoclonal antibodies to understand the molecular basis of potential therapeutic strategies. Simulation and study of the modes of the interaction between polyphosphates and ACE2 or RdRp have been carried out using docking procedures in the context of a collaboration with the molecular genetics group of Prof. Zollo of the University of Naples Federico II. This group has suggested that polyphosphates can exert antiviral activities against SARS-CoV-2 as PolyPs can enhance ACE2 proteasomal degradation and impair synthesis of viral RNA.

In addition to this study, a software pipeline able to scrutinize the molecular interactions between Spike domains and monoclonal antibodies in the complexes available in the PDB, has been developed. The analysis should clarify the most important patterns of interactions between Spike and the mAbs. This information can be useful to predict the immunoevasion ability of

the emerging variants and, in perspective, to produce more effective and specific mAbs.

3. MATERIALS AND METHODS

3.1. Data retrival

All SARS-CoV-2 sequences have been downloaded from GISAID repository at www.gisaid.org/ [109], the GenBank platform (https://www.ncbi.nlm.nih.gov/genbank/) [110] and the Reference Sequence (RefSeq) database (https://www.ncbi.nlm.nih.gov/refseq/) [111]. All protein structures and corresponding functional informations were retrieved from Protein Data Bank (PDB) (https://www.rcsb.org/) and from UniProt (https://www.uniprot.org/) databases [112,113], respectively.

3.2. Data Processing

3.2.1. *In silico* analysis of early SARS-CoV-2 mutant

BLAST tool has been used for databank searches [114]. Jalview [115] and MAFFT [116] have been used for multiple sequence display and alignment, respectively. Transmembrane helix prediction has been obtained by TMHMM [117], MEMSAT [118] and Protter [119]. Cd-hit program [120] has been used for sequence clustering. Homology modelling has been carried out with Swiss-Model [121] and HHpred [122] servers. I-Tasser has salso been used as an alternative source of SARS-CoV-2 protein structure models [123]. Three-dimensional structures have been analyzed and displayed using PyMOL [124] and UCSF Chimera [125].

3.2.2. ORF3a

The Wuhan ORF3a protein (RefSeq accession number YP_009724391.1) has been taken as the reference (wild type) sequence with which its variants have been compared. The collection and selection of the ORF3a protein sequences coded by different SARS-CoV-2 genome isolates has been carried out using this workflow:

1. SARS-CoV-2 genome sequences have been downloaded as FASTA format from GISAID repository at www.gisaid.org [110]. Since the quality of the deposited sequences is not uniform, only complete sequences deposited with a high degree of coverage have been downloaded using the filters provided by the GISAID server.

2. The file containing the genomic sequences has been converted into a BLAST-formatted database with the "makeblastdb" tool [114].

3. The "tblastn" tool searches a protein sequence within a translated nucleotide sequence database. The reference ORF3a sequence has been used as a query to retrieve the other ORF3a coding sequences from the SARS-CoV-2 genomes. Incomplete sequences or sequences containing ambiguous codons (resulting in undetermined residues) have been eliminated. This step relied on the tools available in the EMBOSS suite [126] and on Linux bash shell scripts.

4. The clustering software "cd-hit" [120] has been applied to remove redundancies. Identical ORF3a sequences have been clustered and one representative sequence has been designated by the software. Each cluster

contains all the sequences of one ORF3a variant. As a matter of fact, the ORF3a sequences belonging to different clusters differ for at least one residue.

5. The representative ORF3a variants have been multiply aligned to the reference protein with the program MAFFT [116].

6. A R script has been written within the Rstudio environment to scrutinize the multiple sequence alignments and collect mutation statistics and for graphical output. The R script utilized input and output functions from the bio3d package [127]. Multiple sequence alignments display and editing relied on Jalview [115]. PyMOL [124] and UCSF Chimera [125] have been used for structure display and analysis. PyMOL plugin Caver 3.0.3 [128] has been utilized to study tunnels inside the protein structure.

3.2.3. Spike glycoprotein: variant characterizations

Mutations of the spike variants have been taken from the site cov-lineages.org [85]. *In silico* mutagenesis to generate the variant structural models was carried out with the ad hoc tools available within UCSF Chimera and PyMOL programs [124,125]. Energy minimization protocol embedded in the molecular graphics program Swiss-PdbViewer [129] has been applied to the model RBDs to remove residue steric overlaps at the interface. The protocol used the GROMOS96 43B1 forcefield, cut-off 10 Å, and 100 steps of steepest descent minimization followed by 1000 steps of conjugate gradients in vacuo. The minimization was stopped if the energy difference between two consecutive steps was lower than 0.05 kJ/mol. Only residues at the interface have been

minimized. This forcefield does not include the parameters to describe glycans which were therefore ignored during minimization. However, the glycans present in the RBD are apparently far from the complex interface and have not the potential to interfere with it. PyMOL suite was also utilized for its Adaptive Poisson Boltzmann Solver (APBS) plugin [130] which has been used to calculate the electrostatic potential of the wild-type and variants SARS-CoV-2 Spike RBDs. PROPKA [131] was used to predict the pka values of ionizable groups of this protein taking into account the influence of the local structural environment, inclusive of residue solvent exposure, presence of hydrogen bonds or interactions with other charged groups. The pKa values were used to predict the overall protein net charge at pH=7.0.

3.2.4. Spike glycoprotein: Lambda variant

The Lambda Spike structure was modelled using the method available in the web server Swiss-Model [121] using as a starting structure the coordinate set identified by the PDB code 7KRS. This structure was selected by Swiss-Model as the best trimeric template. This coordinate set contains the Cryo-EM structure of the Spike mutant D614G. The model was built using the sequence mode: the target sequence was given as an input and the program searches for the best templates, from which the user selects the one to be used for model building. At the end of the calculations, the stereochemistry and residue contacts in the model are automatically optimized. Energy minimization of the model complexes between the Spike N-terminal domain (NTD) and the antibody was applied to remove possible residue steric overlaps at the

interface. The same energy minimization protocol embedded in the molecular graphics program Swiss-PdbViewer [129] utilized for the characterization of the other variants, was applied. GlycoPred [132] and NetNGlyc [133] were used to predict potential glycosylation sites. GlycoPred utilizes a Random Forest predictor which is reported to reach 92.8% of correct predictions and a Matthews Correlation Cofficient equal to 0.85 for the N-glycosilation sites. NetNGlyc is based on artificial neural networks. Predictions were made using the recommended threshold score of 0.5. Protein-protein interaction energy was predicted with the method implemented in PRODIGY [134] using the default parameters (temperature 25 °C). This method predicts the binding energy and affinity of a protein complex on the basis of the number and type of interfacial residue-residue interactions. Structural analysis and visualization were carried out with PyMOL [124] or UCSF Chimera [125]. Discontinuous B-cell epitopes were predicted with the methods implemented in the software DiscoTope [135] and BePro [136]. DiscoTope predicts potential B-cell epitopes by attribution of an epitope propensity score to each residue and by analysis of the corresponding spatial neighborhood along with surface exposure. BePro utilizes a similar strategy. DiscoTope predictions were carried out with the default threshold for the combined score of −3.7, which corresponds to an expected sensitivity and specificity equal to 0.47 and 0.75, respectively. BePro epitope assignment was carried out using a score threshold equal to 0.95. For reference, it has been reported that a threshold equal to 1.3 corresponds to a sensitivity >0.3 and a specificity >0.9. Computational alanine scanning of the residues at the interfaces between the Spike NTD and the antibody was obtained through the webserver DrugScore^{PPI} [137]. The method available in the server provides a fast and accurate system to predict the

binding free energy changes upon alanine mutations at protein-protein interfaces using a knowledge-based scoring function. The method does not require any parameter input from the user.

3.2.5. SARS-CoV-2 therapeutic strategies: PolyP

The PolyP molecule containing 20 phosphate groups (PolyP20) was built with the tools available within the ZINC data bank web server [138] and then converted into PDB coordinates. The 3D structures of the receptor molecules were downloaded from the PDB. The atom types and charges were assigned to the ligands and receptors with the AutoDock Tools resources. All of the docking experiments were performed with AutoDock Vina [139]. Blind docking has been applied: the entire receptor molecule (ACE2 ectodomain or RdRp) has been enclosed in one grid; at least 100 docking experiments were carried out with the PolyP ligand and the most sampled sites on the receptor were collected. Subsequently, more specific docking experiments were applied at the most frequently sampled sites and the ligand binding energy estimated.

3.2.6. SARS-CoV-2 therapeutic strategies: development of pipeline to predict the critical residues of interaction between Spike glycoprotein and neutralizing monoclonal antibodies

The pipeline was structured in this way:

1.The structures of the complexes between RBD or NTD and neutralizing monoclonal antibodies have been downloaded from the Protein Data Bank and selected according to the following criteria. The X-ray had priority over the Cryo-electron microscopy, which in turn, was preferred over NMR spectroscopy. In cases of structures solved with the same method, we selected the one with the best resolution and/or the highest coverage of the chain of interest.

2.The neutralizing monoclonal antibodies collected were grouped on the base of the Spike domain recognized.

3.RING software [140] has been applied to find the type of the interactions between the RBD or NTD of the Spike glycoprotein and neutralizing monoclonal antibodies. DrugScore^{PPI} software [137] has been used to predict the binding free energy changes upon alanine mutations at protein-protein interfaces using a knowledge-based scoring function.

4.The data obtained were used to identify the most frequent residues and interactions at the RBD or NTD interface between the Spike domains and mAbs.

The frequency of the residues at the interface RBD or NTD-mAb is calculated in the following way:

$$f_i = \frac{n_i}{n_{tot}}$$

where n_i is the number of residue types (for example Ala, Val, Tyr etc.), n_{tot} is the total number of the RBD or NTD at the interface with the mAbs.

The margin of error (MOE) at 95% confidence interval was calculated with the equation:

$$MOE = z \cdot \sqrt{\frac{f_i \cdot (1 - f_i)}{n_{tot}}}$$

where z is 1.96 that corresponds to a confidence level of 95%.

5.The results obtained from the wild-type virus were compared to Omicron variants 1, 2 and 5 (BA.1, BA.2, BA.5).

To predict the effects of mutations on protein structural stability MutateX tool [141] has been used. This software is based on the FoldX energy function [142] to evaluate the change in Gibbs free energy upon *in silico* saturation mutagenesis. ($\Delta\Delta G$). It consists of calculating the free energy change associated with the systematic substitution of each protein residue to any of the other 19 natural amino acids. The $\Delta\Delta G$ is is a metric for predicting how a single point mutation will affect protein stability which is equal to: $\Delta\Delta G = \Delta G_{wild-type} - \Delta G_{mutant}$, (kcal/mol^{-1}) [142], where the ΔG is the difference between the free energies of the folded and unfolded states.

- $\Delta\Delta G > 0$ suggest that a mutation would be destabilizing.

- $\Delta\Delta G$ near 0 suggest that a mutation would be considered neutral.
- $\Delta\Delta G < 0$ suggest that a mutation would lead to a more stable protein

MutateX also allows performing partial mutational scans, in which only few positions at a time are considered. The positions were selected by the PyMOL script "InterfaceResidues.py" [143] that has been modified for this specific task. This script retrieves the interface residues between two proteins or chains. The classification of the variants in stabilizing, neutral and destabilizing according to the predicted change in ΔG has not yet been defined because the method was recently developed. However, the thresholds currently recommended by the authors of the program are: stabilizing ($\Delta\Delta G < -1.0$ kcal/mol), neutral ($-1.0 \leq \Delta\Delta G < 1.0$ kcal/mol), and destabilizing ($\Delta\Delta G \geq 1.0$ kcal/mol).

Additionally, we used the self-scan option of MutateX to mutate all the residues to themselves. We run the self scans with and without the repair step (where each model is subject to a Repair run in which residues having bad torsion angles, Van der Waals clashes or bad total energy are modified to more reasonable conformations) of FoldX as a control experiment to evaluate the quality of the initial structures before and after the repair. Since we are replacing each residue's side chain with itself, the $\Delta\Delta G$ values associated with self mutations are expected to be close to 0 kcal/mol. Any deviations from this behavior in the repaired model (repair function active) or in the initial structure (repair function inactive) should indicate that the sidechain and/or the surrounding residues are entrapped in an unfavorable conformation.

4. RESULTS AND DISCUSSION

4.1. *In silico* analysis of early SARS-CoV-2 mutants

At the beginning of the SARS-CoV-2 pandemic our unit with its specific expertise in structural bioinformatics and molecular modelling has been involved in collaboration with an epidemiology group. The analyses conducted by this group indicated two sites under selective pressure in nsp2 (Q321 and S543) and one site in nsp3 (P192) of early isolated Italian SARS-CoV-2 strain possibly related to higher human contagiousness (**Figure 8**).
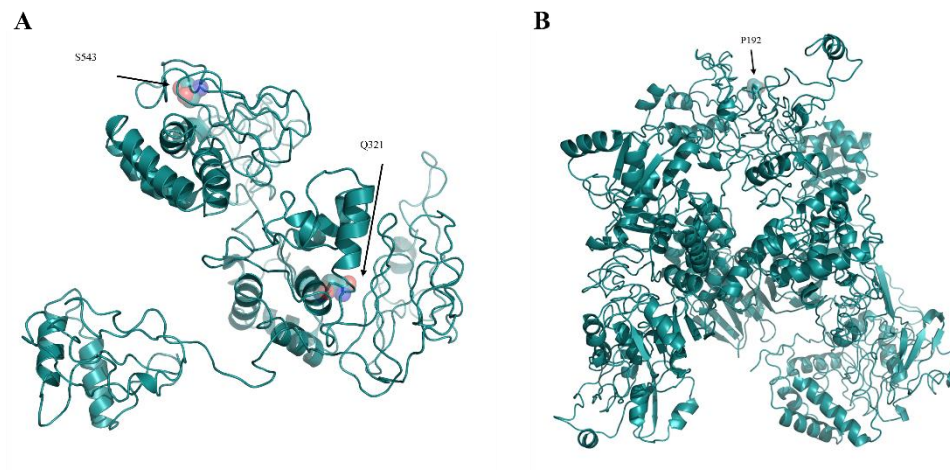

Figure 8: I-Tasser model of SARS-CoV-2 nsp2 on the left and nsp3 on the right. The residues under selective pressure with a $p<0.05$ are shown as sticks and transparent spheres and marked by corresponding labels.

In an attempt to understand the impact of these mutations on the structure and function of nsp2 and nsp3, the structural models available on the I-Tasser web

site [123] have been analyzed because experimental structures were not yet available at the time of this study. The position 321 of nsp2 appears to be exposed to the solvent while the position 543 is predicted to have a low solvent accessibility. The nsp3 position 192 is predicted to be partially accessible to the solvent. In an attempt to structurally characterize nsp2 and nsp3 TMHMM, MEMSAT and MEMPACK [117,118,119] analyses have been utilized showing the presence of several potential transmembrane helices. In particular, in nsp2 were predicted four transmembrane helices while in nsp3 were predicted six helices. The same epidemiology group spotted an interesting mutation in nsp6 in sequence position 37 that replaces a leucine with a phenylalanine. The interpretation of the mutation utilized the structural model available in the I-Tasser site [123]. This protein is located in the endoplasmic reticulum (ER) and it is also able to generate autophagosomes. The structural analysis performed using TMHMM and Protter servers [117,119] have shown that nsp6 protein has seven putative transmembrane helices like in other coronaviruses Analysis of the nsp6 structural model, suggested that the mutated residue is positioned within a constellation of aromatic residues. This can have several consequences: the phenyl ring of phenylalanine may stiffen the local secondary structure by means of aromatic-aromatic, hydropohobic or stacking interactions with the surrounding side chains as shown in the **Figure 9**.
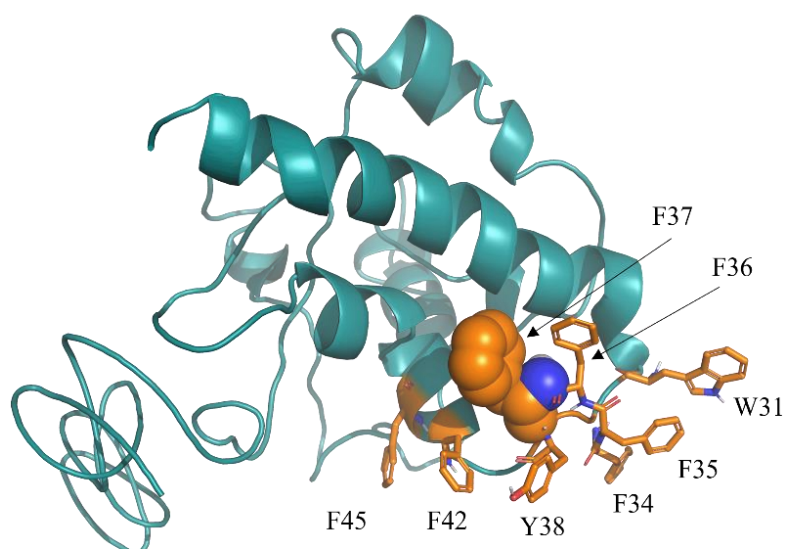
Figure 9: I-TASSER model of SARS-CoV-2 nsp6. Residue under positive selective pressure with a $p < 0.05$ is shown as a sphere. Residues found in the structure proximity are shown in sticks. All residues are marked by the corresponding labels.

Subsequently, from the GenBank repository 797 complete genomes of SARS-CoV-2 have been collected and the sequences of the Envelope and of the Membrane have been extracted from each genome using TblastN program. The sequences of these two proteins have been compared to the homologous counterparts from Bat and Pangolin SARS-like viruses to understand how the amino acid mutations can influence the virus properties. These two proteins appear rather conserved. However, the comparison with the other homologous sequences highlighted structural differences specific of SARS-CoV-2 proteins which may be correlated to cross-species transmission. About the Envelope protein the differences are located mainly at the C-terminal region of the

45

protein, in particular one mutation is at position 69 where arginine replaces glutamate or glutamine in the homologous CoV protein and the other mutation in at position 70. In this position, there is a deletion the SARS-CoV-2 sequence corresponding to glycine or cysteine in the other proteins. These two mutations may have a significant impact on conformational properties and possibly on protein-protein interactions (**Figure 10**).
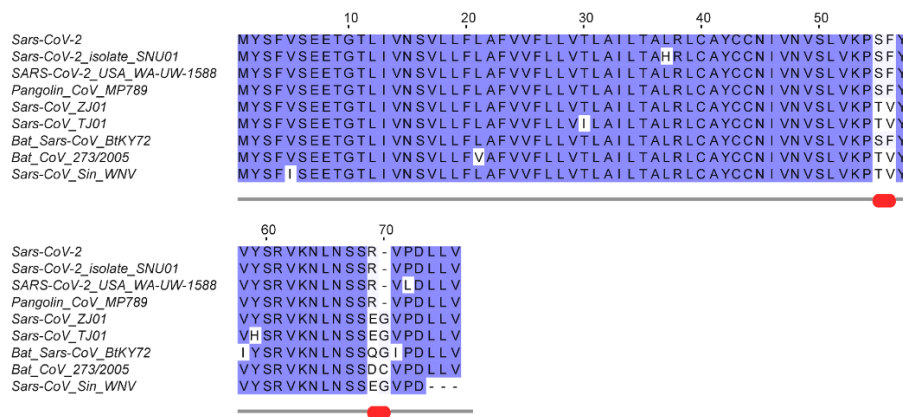
Figure 10: Multiple sequence alignment among SARS-CoV-2 Envelope protein variants and a set of the most similar homologous proteins. The first sequence refered to SARS-CoV-2 E. Red lines below the alignment indicate the changed sites discussed in the text. Blue background denotes conserved alignment positions.

A homology model of the E protein has been built with Modeller using as a template the pentameric ion channel structure of SARS-CoV protein identified by the PDB code 5X29.This sequence shares 91% identity to SARS-CoV-2 E protein and covers the segment encompassed by positions 8-65. **Figure 11** displays the structure of the homology model of the SARS-CoV-2 E protein

46

assembled as a pentameric viroporin-like protein. **Figure 13** displays also the position of the variant sites onto the three-dimensional model. Prediction of the transmembrane helices is difficult in a short protein. Therefore, transmembrane topology cannot be assigned reliably. Likewise, experiments have not clarified definitively which portions of the E protein are exposed to the external or internal side of the virus membrane [51].
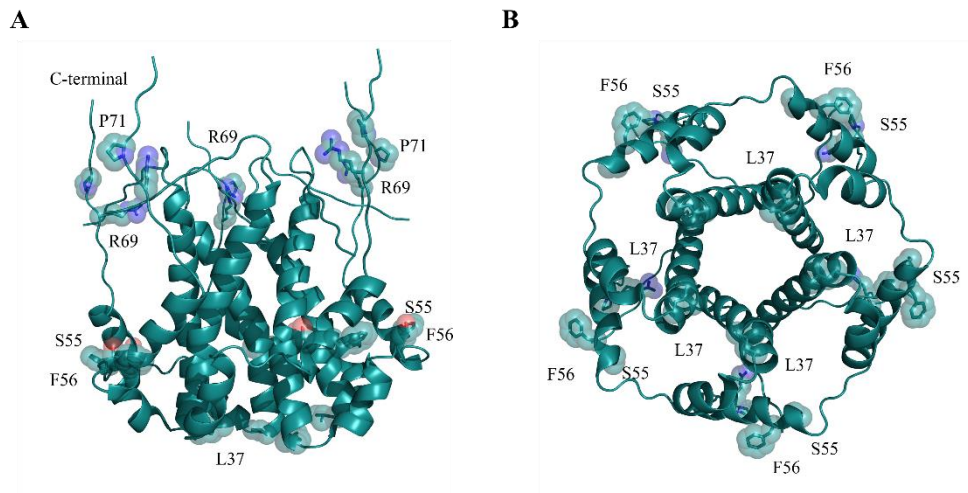
Figure 11: Three-dimensional model of the viroporin-like tetrameric assembly of the E protein from SARS-CoV-2 represented as a cartoon model. Residues corresponding to the mutated sites indicated in **Figure 10** are displayed as transparent space-filling spheres and labelled with the amino acid one-letter code. The C-terminal segments of the model are reported for completeness. However, they convey no structural information due to lack of a corresponding segment in the structural template used in homology modelling. Structure in panel **(B)** is rotated by approximately 180° along the x axis with respect to the orientation shown in panel **(A).**

About the Membrane protein the differences are located mainly at N-terminal region of the protein, among these the insertion of a serine residue at position 4 of the human SARS-CoV-2 seems to be a unique feature of this protein. This region is exposed to the virus surface, could play a key role in the host cell interaction. The mutation occurs within a predicted transmembrane helix and, if confirmed, may have a significant impact on the protein properties. (**Figure 12**) The three-dimensional model of the M protein has been taken from the I-Tasser server as other methods failed to find any suitable template. However, it should be mentioned that HHpred found a weak local affinity, albeit below the statistical significance level, to 4N31, a peptidase-like protein from *Streptococcus pyogenes* essential for pilus polymerization. **Figure 13** displays the positions of the variant sites onto the model structure. This model has been predicted by ab initio techniques. Therefore, it should be considered with great caution and should be treated as a low-resolution approximation of the real structure. According to the prediction of the transmembrane helix topology, the N- and C-terminal of the M protein are exposed outside and inside the virus particle, respectively.
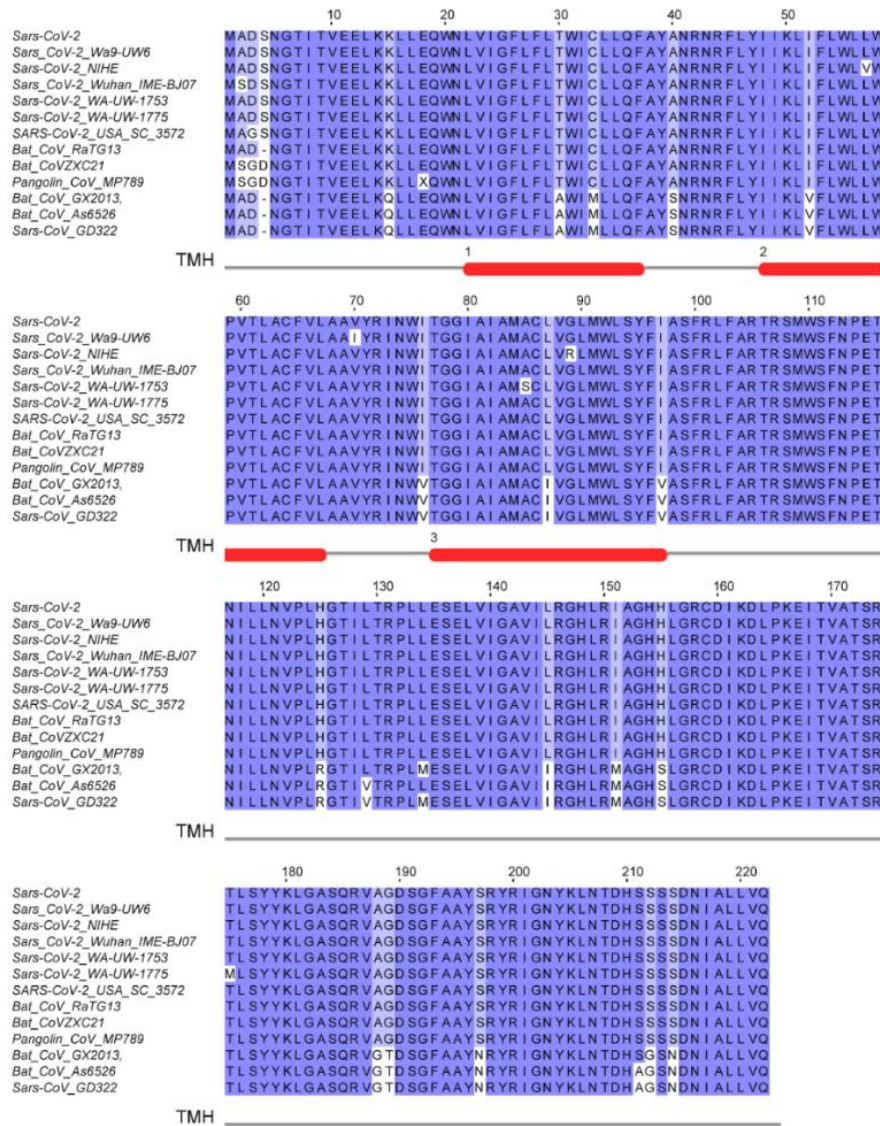
Figure 12: Multiple sequence alignment among SARS-CoV-2 M protein variants and a set of most similar homologous proteins. The first sequence refers to SARS-CoV-2 M. Red box indicates the variant sites at the N-terminal discussed in the text. Numbered red bars under the

49

multiple alignment mark the prediction of transmembrane helices. The location of the connect loop with respect to the virion surface is indicated as "in" or "out". Blue background denotes conserved alignment positions

Figure 13: I-Tasser model of the Membrane protein represented as cartoon model. Variant positions are displayed as transparent space-filling spheres and labelled with the amino acid one-letter code.

4.2. ORF3a

The SARS-CoV-2 ORF3a is an integral membrane protein able to function as an ion channel that may promote virus release. ORF3a may also play a role in virus pathogenesis. For this reason, we performed a specific analysis which was aimed at identifying its mutations over time. For this analysis, 70.752 SARS-CoV-2 genomes available in GISAID databank at the end of August 2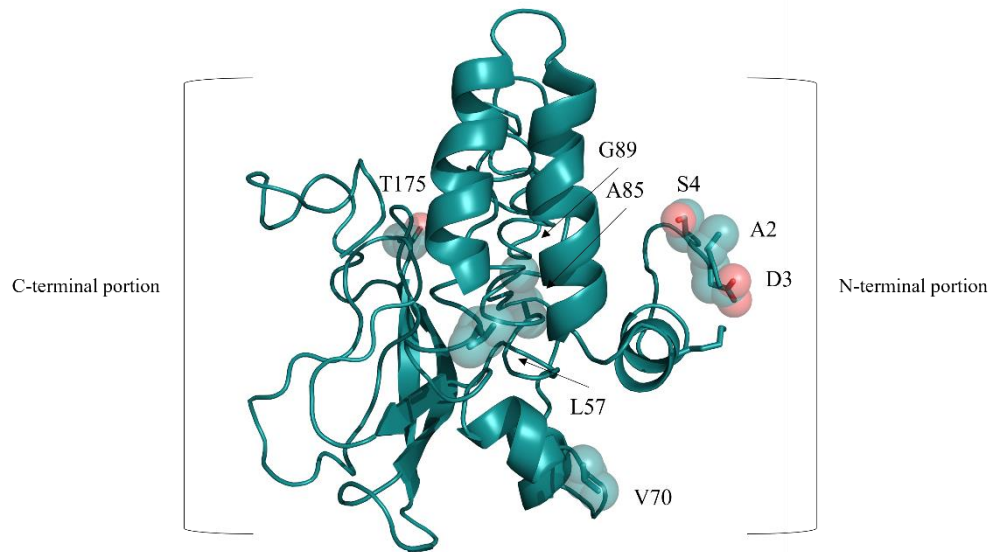020 have been scanned. All ORF3a mutations were grouped according to the collection date interval and over the entire data set. Intervals considered are:February, March, April, May, June, July and August. In each time period, the number of all the different ORF3a variants has been reported. Frequency is defined as the number of replicas of a single variant found in the data set considered. For example, if the ORF3a variant 1 is found 100 times in 1000 genomes collected, its frequency is 0.1. In total, seventeen mutations have been isolated. We focused our attention onto the five most frequent mutations within each month to avoid inclusion of possible statistical noise. Among these, the most frequent substitution was Q57H, that was detected from February to August. The other most frequent mutations, within the entire sample, were: V13L, A99V, G196V and G251V as shown in the **Figure 14**. To test whether mutations at these sites may influence channel shape, the mutant sites of ORF3a were modelled by Chimera 1.14. The impact of these mutations on the tunnel geometry delineated by Caver 3.0.3 [128] was visually assessed. None of these mutations influences significantly the central pore topography.
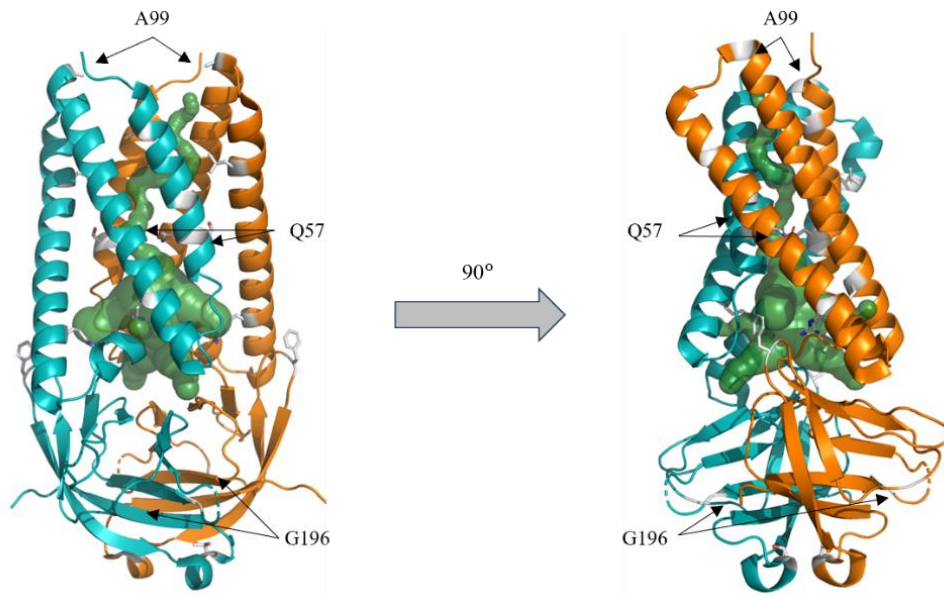
Figure 14: The five most frequent mutations of ORF3a are shown in both subunits colored in deepteal and orange respectively. All mutations are shown in this figure except V13L and G251V that belong to a ORF3a region with missing spatial coordinates.

The same analysis also identified twenty-eight sites conserved in all isolates. Among these, twelve sites are in the transmembrane domain, twelve in the β-sandwich cytosolic domain and four sites are in one region not visible by the cryo-EM analysis, and, for this reason, their spatial coordinates are not available, (**Figure 15**)
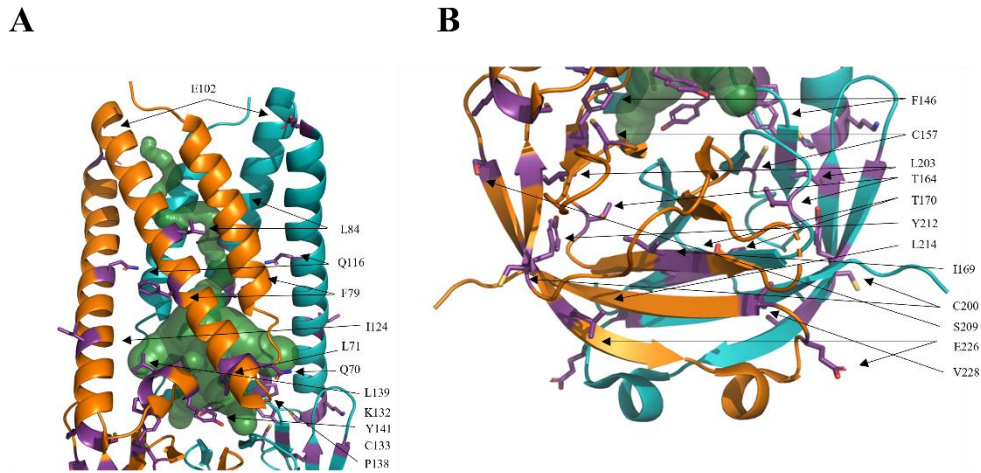
Figure 15: ORF3a dimer represented as cartoon model. The two subunits are coloured in orange and deepteal. Conseved sites are labelled and the corresponding side chains are reported as violet sticks. Transpartent internal spheres indicate the tunnels connecting to the extracellular environment (green). **(A)** transmembrane domain **(B)** extracellular domain.

Interestingly, one of the conserved sites, K132, is close to the putative dimerization surface suggesting that also this residue may contribute to the tetramerization interface, as suggested by other authors [144]. Noteworthy is the conservation of C133 and C157. Residue C133 is the most important for homodimerization and is conserved between different species [144]. The conservation of the two residues strongly supports this observation and suggests that C157 also is essential for ORF3a structural stability and function. Distance between the sulphur atoms of the two cysteine is 3.9 Å that is not compatible with the presence of a disulfide bridge. However, flexibility of the loop bearing C157 may allow, in certain circumstances, the formation of a

bond. Other conserved sites are exposed at different locations. E102 is exposed at the extracellular side. I124 is exposed to the bilayer interface. C200 is exposed to the intracellular compartment and may have, for its reactivity, a role in the interaction with other cellular components. S209 and E226 are also exposed to the intracellular compartment. Interestingly, the conserved Y141 and F146 belong to the Domain IV described by Issa et al. [144] that is deemed to be involved in interaction with caveolin. A multiple sequence alignment of 66 homologous ORF3a from other coronaviruses has been calculated to assess whether the SARS-CoV-2 unmutated positions are conserved in other species.

4.3. Spike glycoprotein

SARS-CoV-2 variants are characterized by specific mutations in the Spike glycoprotein; these mutations can influence the infectivity, the trasmissibility, the host immune response, or the replication capacity of the virus. Our studies aimed at structurally interpreting the effect of these mutations on Spike properties. The variants analyzed were: Kappa, Delta, Lambda, Mu and Omicron with its subvariants. In the table below are shown the mutations in the Spike glycoprotein of each variant analyzed.

Table 4: List of Spike glycoprotein variants

Variants	Mutations in the Spike glycoprotein
Beta	D80A, D215G, Δ241-243, K417N, E484K, N501Y, D614G A701V
Gamma	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F
Kappa	G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H
Delta	T19R, G142D, E156G, Δ157/158, L452R, T478K, D614G, P681R, D950N
Lambda	G75V, T76I, R246N, Δ247-253, L452Q, F490S, D614G, T859N N: P13L, R203K, G204R, G214C
Mu	T95I, Y144S, Y145N, R346K
Omicron BA.1	A67V, Δ69/70, T95I, G142D, Δ143-145, N211I, L212I, R214ins, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F
Omicron BA.2	T19I, L24S, Δ25-27, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R,

	N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K.
Omicron BA.5	T19I, L24S, Δ25-27, Δ69-70, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, L452R, S477N, T478K, E484A, F486V, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K

4.3.1. Kappa, Delta and Omicron variants

The Kappa and Delta variants were first detected in India in late 2020. The Spike proteins of the two variants are characterized by several mutations. Among these, two mutations occur in the RBD segment in both variants: E484Q and L452R, and L452R and T484K in Kappa and Delta variants, respectively. We examined the changes of the RBD biophysical properties caused by the mutations of these two variants and compared the results obtained with the properties of the wild-type and the Beta and Gamma variants. We calculated the surface electrostatic potential using Adaptive Poisson Boltzmann Solver (APBS) plugin function in PyMOL software [130]. The results shown a major shift toward increased positive electrostatic potential in the Delta variant compared with the other variants due the changes of the neutral leucine and threonine to positively charged residues Arg and Lys (**Figure 16**).
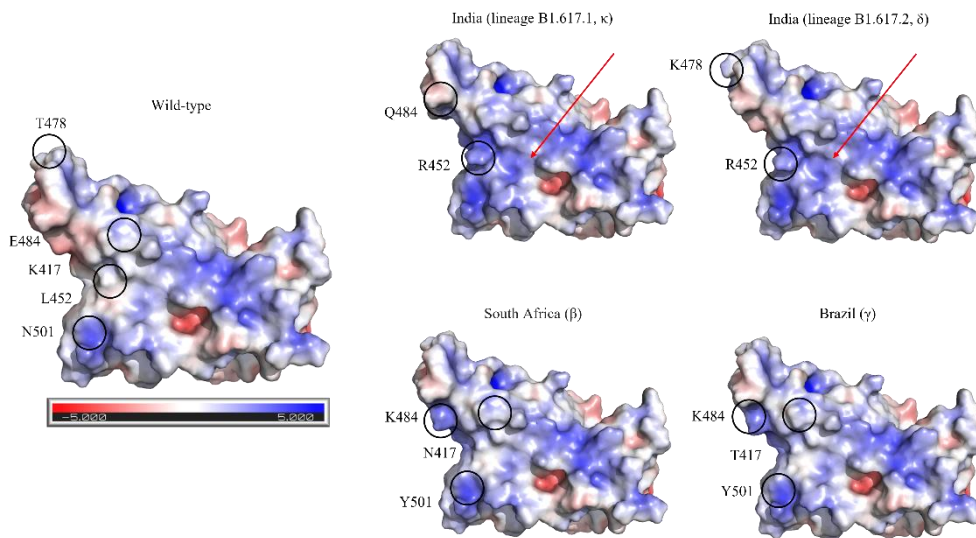
Figure 16: Comparison between the wild-type and variant spike receptor-binding domains (RBDs). Protein surface is colored according to the electrostatic potential. Color scale ranges from −5 kT/e (red) to +5 kT/e (blue) as reported by the bar under the wild-type RBD. Position of the mutant sites is indicated by a circle and an attached label. Red arrows mark the area of increased positive potential in the RBD Indian (delta) variants

This change can strengthen the interaction between the Delta RBD and the negatively charged ACE2, thus conferring a potential increase in the virus transmission or it could induce SARS-CoV-2 antibody evasion [145]. The same analysis was performed on the Omicron variant BA.1 which was first detected in South Africa in November 2021. This variant presents a high number of mutations on its Spike glycoprotein, much higher than Kappa or Delta variants **(Table 4).** The comparison between the Spike Wuhan, Delta and Omicron RBDs reveals that in Omicron RBD there is a marked increase of the

positivity of surface electrostatic potential at the interface with ACE2 with respect to the Delta RBD, previously observed [146] (**Figure 17**).
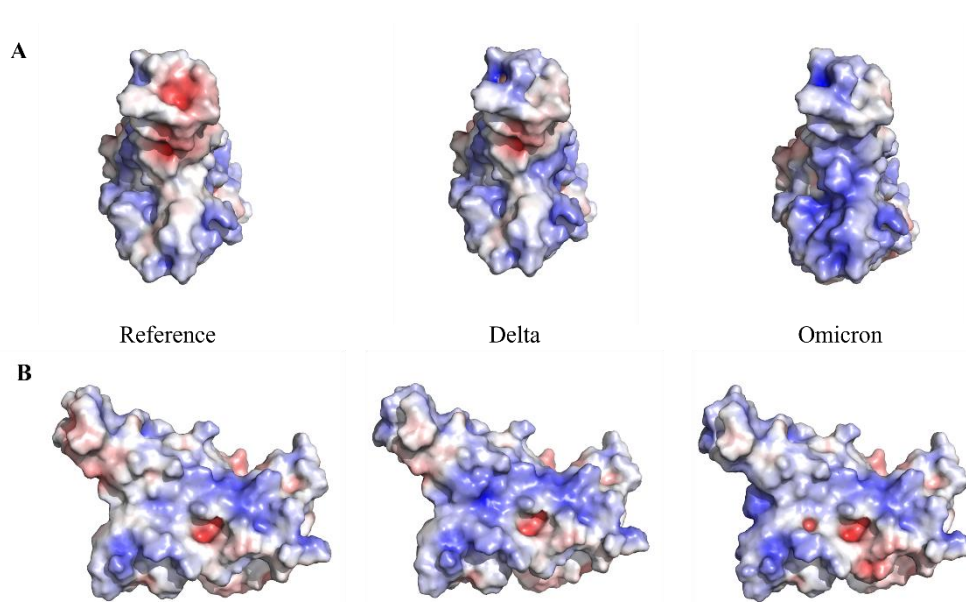
Figure 17: Electrostatic potential surface representation of the WT, Delta and Omicron variants of the Spike RBD. Red and blue colors indicate negative and positive potential, respectively. The color scale ranges from −5.0 to +5.0 kT/e. The RBD is oriented with the ACE2 interface in the front (A) or rotated by 90° to the left along the y-axis (B)

The change of the electrostatic potential towards positive charge values in the Omicron RBD could suggest an increase of the binding affinity between the positively charged Omicron RBD and the negatively charged ACE2. This may partly explain the observed high transmissibility of this variant [147,148]. The predicted net charge of the Spike RBD wt, Delta and Omicron at pH 7 are

shown in the table below and compared with the ranges of the estimated effective reproduction number (ERN) [149]. The RBD electrostatic potential is positive and linearly increased from the Wuhan strain (2.15) to successive variants, reaching its highest value (5.22) in Omicron VOC. Noteworthy is the comparison between the RBD electrostatic potential and the ERN. As the increase of positive charge at the surface also increase the estimated ERN.The ERN is the key epidemic parameter used to assess whether an epidemic is growing or not (high values means that infections are increasing) [149].

Table 5: Comparison between net charge of Spike RBD and ERN

SARS-CoV-2	Pango lineage	Net charge at pH 7	ERN
WILD-TYPE	B.1	2.15	1.4–2.5
DELTA	B.1.617.2	4.15	3.4–8.6
OMICRON	B.A.1	5.22	9.7--13.0

4.3.3. Lambda variant

The Lambda variant emerged in Perù in June 2021. This variant is characterized by several mutations which do not occur at the RBD interface with ACE2. However, among these mutations there is a deletion of seven residues at the N-terminal domain (NTD) of the Spike protein between the sequence positions 246-252. To predict the potential impact of the deletion on the NTD affinity to a human mAbs, the complex between the SARS-CoV-2 Spike and the 4A8 Ab deposited as 7C2L in PDB was used as a case study. The mutant NTD was modelled via Swiss-Model and superposed to the wild-type domain of the complex. Within the Spike trimer, the NTD model was built

using the chain A of 7C2L as the template. NTD-4A8 interface interactions were predicted by PRODIGY [134]. Overall, the partial deletion of the loop is predicted to weaken interaction with the 4A8 antibody with a consequent decrease in binding affinity owing to the loss of several interactions. In particular, the deletion in Lambda NTD removes interactions that in the wild-type complex take place between L249 and F60, Y54 of the 4A8 light chain. Moreover, a salt bridge and a π-cation interaction between R246 and 4A8 E31 and Y27, respectively, disappear in the Lambda variant **(Figure 18)**
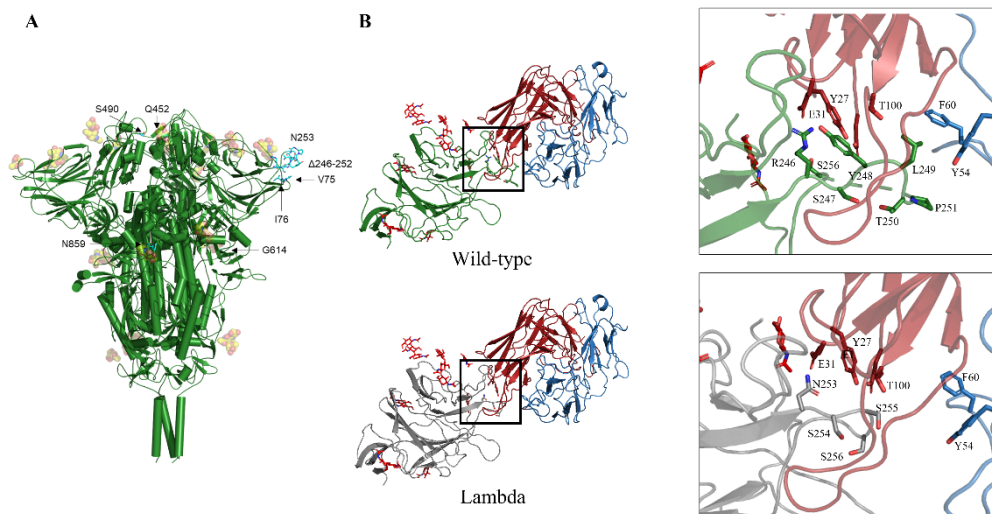
Figure 18: A) Map of the Lambda mutations of the trimeric Spike structure. Only one mutation per monomer is indicated. Glycans are indicated as orange sphere models. **B)** Interface chain NTD-4A8. The mutant NTD was modelled via Swiss-Model and superposed to the wild-type domain of the complex. Interface interactions and energies calculated by PRODIGY were compared. Comparison of the interfaces between the wild-type (green cartoon, top) and the Lambda NTD (grey cartoon, bottom) and the mAb 48A at the deletion region. Red and blue cartoons indicate the heavy and light chain, respectively. Side chains in the deleted loop and

the interacting residues are indicated with stick models and labelled. Sequence numbering refers to the wild-type Spike. Glycans are reported as red sticks

To support these observations, we also used DrugScore[PPI] *in silico* alanine scanning [137] which revealed that the residues R246 and Y248 are predicted to contribute significantly to the binding energy (**Table 6**)

Table 6: Variation of binding energy of R246 and Y248 at the interface between NTD and 4A8 calculated by DrugScore[PPI] alanine scanning

Residues	Wild type (ΔΔG, kcal/mol)
R246	1.20
Y248	2.96

To test the possible impact of the NTD deletion on the S protein antigenicity, the potential B-cell epitopes predicted for the reference and Lambda Spike were compared using two methods: DiscoTope [135] and BePro [136]. Both methods predict the presence of a B-cell neutralizer epitope in the sequence of the reference Spike protein at the positions corresponding to the deleted loop in the Lambda NTD. Finally, we used GlycoPred [132] and NetNGlyc [133] to predict N-glycosylation sites of the Lambda Spike glycoprotein. Both methods predicted D253N as a potential N-glycosylation site [150].

4.4. SARS-CoV-2 therapeutic strategies: PolyP

The inorganic polyphosphates are found in several cells in the blood. In 1997, Lorenz et al [151] demonstrated that at high concentrations linear polyphosphates have cytoprotective and antiviral activity particularly against HIV-1 infection. Based on these studies, the Professor Zollo's group decided to study the potential activities of inorganic polyphosphates (PolyPs) against SARS-CoV-2 infection. Lorenz et al [151] also demonstrated that the anti-HIV effects of PolyPs are due to binding to both the host cell surface and the virus. Starting from these results the molecular genetic's group decided to study the potential activities of inorganic polyphosphates (PolyPs) against SARS-CoV-2 infection. In particular, they tested whether the antiviral activity of PolyPs against SARS-CoV-2 could depend on their binding to ACE2 on the host cells and/or to the viral RdRp. To verify these hypotheses, our unit performed docking experiments using AutoDock Vina software [139] in order to study possible mode and sites of interaction between PolyPs and ACE2 and between PolyPs and RdRp. In the experimental analysis the Professor Zollo's group used PolyPs with 120 phosphate groups- For the *in silico* analysis, only PolyPs with 20 phosphate groups (20P) could be used because longer-chain PolyPs cannot be reliably treated with the current docking algorithms. Our analysis showed that four amino acid residues of ACE2: H378, R393, H401, and R514 are mainly involved in the binding with the 20 PolyPs (**Figure 19, A**) These residues are conserved across different vertebrates. Docking analyses between RdRp with Poly20 revealed that there are seven RdRp amino acids mainly involved in the binding with PolyPs, which are: K4892, K4937, R4945, D5152, R5228, D237, and K5241 (**Figure 19, B**) These amino acids are

conserved in the RdRp proteins of otherviruses in the family Coronoviridae (e.g., SARS-CoV and MERS-CoV). The experimental analyses confirmed that the sites predicted by docking experiments are essential for the interaction between ACE2 and PolyPs and between RdRp and PolyP [152].
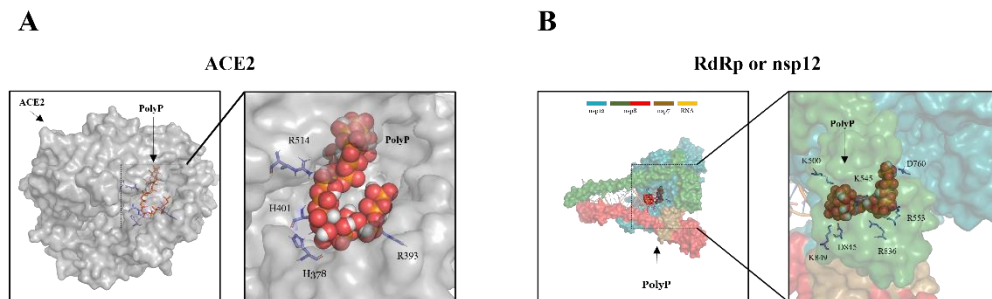
Figure 19: **A)** Molecular docking of PolyP20 on the SARS-CoV-2 ACE2 domain (PDB structure: 6M0J, chain A). Left, ACE2 is represented as grey transparent surface and the orange sticks represent PolyP20. Right: Magnified view of the ACE2 receptor to indicate the binding interface. The amino acid residues mainly responsible for the interactions between ACE2 and PolyP20 are shown as blue sticks and labelled. **B)** Molecular docking of PolyP20 on the SARS-CoV-2 RdRp domain (PDB structure: 6M71, chain A). Left panel RdRp is represented as cyan transparent surface and the orange spheres represent PolyP20 Right: Magnified view of RdRp as a cyan transparent surface to indicate the binding interface. The amino acid residues mainly responsible for the interactions between RdRp and PolyP20 are shown as blue sticks and labelled

4.5. SARS-CoV-2 therapeutic strategies: development of pipeline to predict the critical residues of interaction between Spike glycoprotein and neutralizing monoclonal antibodies

4.5.1. RBD-mAbs

Based on the information from literature about the mode of the interactions between the RBD and neutralizing monoclonal antibodies [100] we retrieved in total 68 complexes starting from June 2021 until May 2022, among which: 29 belonging to class 1, 24 belonging to class 2, 6 belonging to class 3 and 9 belonging to class 4. We used RING software [140] to identify covalent and non-covalent bonds in these complexes, including π-π stacking and π-cation interactions and DrugScore[PPI] [137] to predict the contribution of the RBD residues to the interface stabilization energy. Concerning the complexes belonging to class 1, RING software revealed that most residues of RBD located at the interface are hydrophobic (frequency equal to 0,56 ± 0,04). Considering all residues that are in the interface, the most frequent residues are tyrosines with frequency equal to 0,22 ± 0,03. DrugScore[PPI] identified as hot spots, i.e:residues proving the highest contribution to the interface stabilization energy, Y505 and Y489. Regarding the complexes belonging to class 2, RING software revealed that most of the RDB residues are hydrophobic with the frequency of 0.60 ± 0.05. Considering all residues that are in the interface, the most frequent residues are tyrosines with frequency equal to 0,16 ± 0,04. DrugScore[PPI] identified as hot spots: Y505, Y489 and Y449. Only 6 complexes for the class 3 and 9 complexes for the class 4 were available. Nevertheless, regarding the complexes of the class 3, RING software revealed that among

the RBD residues located at the interface there are hydrophobic residues with the frequency equal to 0,50 ± 0,11 and hydrophilic residues with frequency of 0,50 ± 0,11 and that the most frequent residues are asparagines with frequency of 0,14 ± 0,08. DrugScore^{PPI} identified as hot spots: T449, V503, L441, Y380. Concerning the complexes belonging to class 4, RING software revealed that among the RBD residues located at the interface there are hydrophobic residues with the frequency of 0.51 ± 0.09 and hydrophilic residues with frequency equal to 0.48 ± 0.09 and that the most frequent residues are phenylalanines with frequency of 0.12 ± 0.06. DrugScore^{PPI} identified as hot spots: Y369, R357, D427, Y380. As regards the latter (complexes belonging to class 3 and 4) the analyses revealed that there is not a correlation between the frequency and the contribution of a residue at the interface between RBD and mAbs. This aspect could depend on the scant number of samples available for class 3 and 4 which makes the observed frequencies affected by a large statistical error and therefore they should be considered with caution. In the table below the frequencies of residues for each class are shown. Among the complexes retrieved, we observed that some antibodies of the collected complexes are considered potential therapeutic antibodies. We focused our analysis on two complexes which contain the antibodies Regdanvimab (PDB code 7CM4) and Sotrovimab (PDB code 7BEP) belonging to class 2 and class 3, respectively, with the application of MutateX procedure. We chose Regdanvimab and Sotrovimab to study the molecular basis of their interactions with RBD since experimental data (available at https://covdb.stanford.edu/susceptibility-data/table-mab-susc/) [153] indicate that Regadanavimab is inactive against some SARS-CoV-2 variants whereas Sotrovimab activity is minimally affected against each variant. MutateX [141]

65

has been applied to interpret at a molecular level the different behavior of the two antibodies against the Omicron variants 1, 2 and 5. MutateX analyses predicted that in the complex between Regdanvimab and Omicron 5 F486V and L452R mutations are destabilizing ($\Delta\Delta G$ = 1.90 kcal/mol and $\Delta\Delta G$ = 1.94 kcal/mol, respectively). The RING software revealed that in this complex, the Omicron 5 mutation F486V implies the loss of two π-π stacking with the Y113 of the heavy chain and Y50 of the light chain and that L452R loses two van der Waals interactions with Y106 and R107 of the heavy chain of the Regdanvimab antibody. The lack of the two π-π stacking interactions and the two van der Waals interactions in the Omicron variant 2 and 5 in the RBD, could indicate a slightly destabilizing effect of the complex.A π-π stacking interaction and a van der Waals interaction has an energy of 2.25 kcal/mol and 1.43 kcal/mol, respectively, as mentioned in RING [140]. This observation could be consistent with the MutateX results. It has been observed that the predicted interactions between Omicron 1 and 2 RBDs and Regdanvimab are similar to those observed in the case of the wild type. For this reason, the results were not reported.

Regarding the Sotrovimab antibody, the RING software revealed that there were no differences in the interactions between the wild-type and the Omicron 1, 2 and 5 variants. These results could be confirmed by MutateX which predicted that the two mutated residues G339D and N440K in the Omicron variant 1, 2 and 5 do not influence the stability of the complex ($\Delta\Delta G$ = -0.769 kcal/mol and 0.005 kcal/mol, respectively).

Table 7: Frequencies of residues for each class.

Complex	Most Frequent class	Most frequent residues
Class 1	Hydrophobic (0,56 ± 0,04)	Tyrosine (0,22 ± 0,03)
Class 2	Hydrophobic (0.6 ± 0.05)	Tyrosine (0,16 ± 0,04)
Class 3	Hydrophilic (0,50 ± 0,11)	Asparagine (0,14 ± 0,08)
Class 4	Hydrophobic (0.51 ± 0.09)	Phenylalanine (0.12 ± 0.06)

4.5.2. NTD-mAbs

We retrieved in total 16 complexes, starting from June 2021 until April 2022: 14 complexes bind the NTD in the same position whereas 2 complexes bind the NTD in a different region as shown in **Figure 20**.
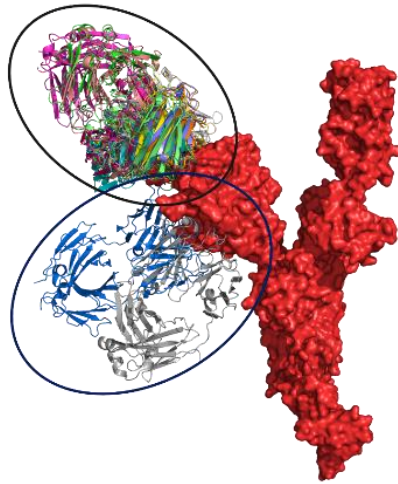

Figure 20: The Spike monomeric glycoprotein is displayed as a red molecular surface. Antibodies are reported as cartoon models. The fourteen Abs which bind the same NTD region are enclosed in the black circle while the other two are in the blue circle.

In consideration of the scant amount of data, the 16 complexes were considered in the same group. RING analysis revealed that most frequent interface residues of RBD are hydrophilic (frequency equal to 0.58 ± 0.07). Considering all residues that are in the interface, the most frequent residue is tyrosine with frequency of the $0,11 \pm 0.04$. In addition, it was observed that the positive charged residues (lysine, arginine, and histidine) have a frequency of 0.29 ± 0.06. DrugScore^{PPI} identified as hot spots Y144 and Y248. Among the complexes retrieved, we observed that in two of them were present two therapeutic antibodies: DH1052 (PDB code 7LAB) and 4A8 (PDB cod 7C2L). Regarding the DH102 antibody, RING analysis indicates that the Omicron 1, 2 and 5 NTD and DH1052 are similar to those observed in the case of the wild type. Concerning the 4A8 antibody, RING analyses suggest that the interface interactions observed in the complex between 4A8 and wild-type NTD are identical to those expected in the case of Omicron 1, 2 and 5 variants. Also in this case, MutateX has been applied to study the stability of the interface between the two mAbs and the variant NTD. Concerning the DH102-antibody, MutateX analysis predicts that none of the mutations in the Omicron 1, 2 and 5 variants have a significant impact on the complex stability. Regarding the 4A8-antibody the mutations characterizing these variants do not occur at the interface to the antibody, for this reason the MutateX procedure does not predict considerable results.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

5.1. *In silico* analysis of early SARS-CoV-2 mutants

The analysis of the early mutant of SARS-CoV-2 proteins suggested a possible molecular basis for explaining the virus evolution and change in pathogenicity. Concerning nsp2, the positions 321 and 543 fall within the region homologous to the endosome-associated protein similar to that of the avian infectious bronchitis virus (PDB ID: 3LD1). It is known that this region plays an essential role in the viral pathogenicity [154]. The nsp3 position 192 falls near the domain containing the protein similar to a phosphatase present also in the SARS-CoV (PDB ID: 2ACF) which plays a key-role in the replication process of the virus in infected cell [155].The structural similarity of the region in which fall the sites under positive selective pressure, in the case of nsp2, could explain why this virus is more contagious than SARS-CoV, while, in the case of nsp3, could suggest a potential mechanism differentiating the disease caused by a SARS-CoV and SARS-CoV-2. The results of this study could fill some gaps about COVID-2019 knowledge especially in the moment when the epidemic was starting, and the scientific community was trying to enrich knowledge about this new viral pathogen. Concerning the nsp6 the presence of multiple phenylalanine residues in the outer membrane region of nsp6 should favor the affinity between this region and the ER membrane inducing a more stable binding of the protein to ER. It has been shown that this binding may favor coronavirus infection by compromising the ability of autophagosomes to deliver viral components to lysosomes for degradation [156]. Thus, the effect of the L37F mutation could be to limit autophagosome expansion, directly or indirectly by starvation or chemical inhibition of mTOR signaling. [157]. We have speculated that the mutation, in the nsp6, could bring to some appreciable

change in the expression of SARS-CoV-2 relationship with its host, particularly concerning a critical host anti-viral defense, such as the autophagic lysosomal machinery. About the E and M proteins, the analyses show the structural similarity of E and M proteins to the counterparts from Pangolin and Bat coronavirus isolates. At the same time, comparisons have highlighted structural differences specific of SARS-CoV-2 proteins which may be correlated to the cross-species transmission and/or to the properties of the virus. Although further studies are needed, it is clear that these amino acid variations have been important for the virus evolutionary history, and the results may hint at how similar mutations within the coronavirus family can lead in the next years to other epizootic epidemic events similar to the one that we have been experiencing these days.

5.2. ORF3a

In general, the most frequent mutations found do not influence significantly the central pore topography. Interesting are the ORF3a positions conserved in the SARS-CoV-2 isolates that are variable in the other coronaviruses. For example, F28 seems to be unique to SARS-CoV-2. Likewise, C200 and Ser209 exposed to the cytosolic side are conserved only in a few other SARS-CoV-2 from pangolin or bats. This pattern points to functions specific to SARS-CoV-2 possibly connected to its peculiar pathogenicity, contagiousness, and ability to cross-species transmission.

5.3. Spike glycoprotein: variant characterizations

The analyses conducted over the different variants strongly suggested that the SARS-CoV-2 evolves toward an increase of the positivity of the electrostatic potential of the RBD from the original virus strain to the most recent Omicron variant. Since ACE2 possesses surface patches of negative electrostatic surface potential, it is reasonable to speculate that the higher positive potential of RBD may increase tropism of virus Spike for the ACE2 receptor and overall interaction affinity. We have previously reported on the likely relationship between the increase of positive electrostatic potential and affinity in the Delta VOC, in particular the Delta B.1.617.2 [146]. If a direct relationship between the electrostatic potential and receptor affinity exists, then the omicron VOC is expected to be more transmissible as some initial data do indeed suggest [147,148]. Additional experiments both *in vitro* and *in vivo* are needed to establish the biological significance of SARS-CoV-2 mutations and how the interactions between mutations and local cellular microenvironment influence the clinical outcome and the transmission dynamics of the virus. However, the methods that have been applied to study these variants can be easily applied to the characterization of the continuously emerging new variants that will be produced by the virus evolution.

It should be noted that all the observations reported here were obtained during the onset of the COVID-19 pandemics. The fast evolution of the SARS-CoV-2 has been monitored and new mutations emerged and were fixed in the new variants. Therefore, conclusions here reported may change in the light of the accumulating new molecular data.

5.4. Spike glycoprotein: Lambda variant

This variant is an example of an evolutionary path that exploits the remodeling of NTD peptide backbone to fine tune interaction with host. Overall, the most evident and likely functionally impact of the changes of the Lambda variant is represented by the 246–252 deletion. This amino acid loss could confer to the virus an enhancing capacity to escape from the host immune response by two theoretical, though likely and already reported, strategies: (i) shortening or fully deleting neutralizer epitopes located in the loops; (ii) exploiting increased glycosylation. These considerations would represent important points for the selection of vaccine candidates [158].

5.5. SARS-CoV-2 therapeutic strategies: PolyP

Our analysis combined with experimental analysis indicated the structural sites of the interaction between ACE2 and PolyPs and between RdRp and PolyP and corroborated the potential therapeutic use of these macromolecules. Because PolyPs are already known not to be toxic, the therapeutic use of PolyPs to prevent SARS-CoV-2 infections should be explored [159].

5.6 SARS-CoV-2 therapeutic strategies: development of pipeline to predict the critical residues of interaction between Spike glycoprotein and neutralizing monoclonal antibodies

The development of this pipeline to predict the critical residues of interaction between the Spike glycoprotein and the neutralizing monoclonal antibodies suggested that there are residues at the interface between RBD or NTD which are more important than other for the stabilization of the interaction with nmAbs. In particular, the interactions between RBD-mAbs, of the complexes of class 1 and 2 involve mainly hydrophobic residues and especially tyrosines. The same pattern has been found at the interface between RBD and ACE2 [160]. In the RBD Omicron variants 1, 2 and 5 there are some hydrophilic residues such as S371, S373, S375 and E484 which are replaced by hydrophobic residues. To be more specific, in the Omicron 1 variant the mutations are: S371L, S373P, S375F and E484A whereas in the Omicron 2 and variants the mutations are: S371F, S373P, S375F and E484A. This change at the RBD interface towards hydrophobicity could influence the interaction between the RBD and the mAbs, suggesting that mAbs with hydrophobic interfaces could bind more effectively the RBD. Concerning the interactions between NTD-mAbs, although there is a scant amount of data, it is noteworthy that 29% of the total residues at the interface are the positively charged as also seen in the Omicron 1, 2 and 5 variants. This observation could suggest that new mAbs negatively charged on their surface could bind more effectively the NTD. In addition, MutateX a new software recently developed [141] was used to rationalize the consequences of known mutations. The results obtained by

74

the application to the analysis on the protein structure stability of the complexes Regdanvimab-RBB and Sotrovimab-RBD are consistent with the results of RING and DrugScore^{PPI} software and with the data published in literature. Our predictions suggested that the Regdanvimab antibody seem not to be efficient against the Omicron variants, specifically against the Omicron variant 5 whereas the Sotrovimab antibody seem to maintain its activity. These observations are consistent with the work conducted by VanBlargan et al [161] which demonstrated that Regdanvimab completely lost neutralizing activity against Omicron variants and with the work conducted by Touret et al [162] which demonstrated *in vitro* that Sotrovimab is less active against Omicron 2 than against Omicron 1 and even less active against Omicron 5. Concerning the DH1052-NTD and 4A8-NTD, MutateX does not indicate any significant result. These observations can provide a molecular framework which can contribute to the development of strategies to design more specific and effective neutralizing monoclonal Ab directed to the present and the future emerging variants.

REFERENCES

1) Guo, YR., Cao, QD., Hong, ZS. et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. Military Med Res 7, 11 (2020). https://doi.org/10.1186/s40779-020-00240-0

2) Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020 Mar 19;91(1):157-160. doi: 10.23750/abm.v91i1.9397. PMID: 32191675; PMCID: PMC7569573

3) Hasöksüz M, Kiliç S, Saraç F. Coronaviruses and SARS-COV-2. Turk J Med Sci. 2020 Apr 21;50(SI-1):549-556. doi: 10.3906/sag-2004-127. PMID: 32293832; PMCID: PMC7195990

4) Chen B, Tian EK, He B, Tian L, Han R, Wang S, Xiang Q, Zhang S, El Arnaout T, Cheng W. Overview of lethal human coronaviruses. Signal Transduct Target Ther. 2020 Jun 10;5(1):89. doi: 10.1038/s41392-020-0190-2. PMID: 32533062; PMCID: PMC7289715

5) Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020 Feb 22;395(10224):565-574. doi: 10.1016/S0140-6736(20)30251-8. Epub 2020 Jan 30. PMID: 32007145; PMCID: PMC7159086.6.

6) Chen Z, Boon SS, Wang MH, Chan RWY, Chan PKS. Genomic and evolutionary comparison between SARS-CoV-2 and other human

coronaviruses. J Virol Methods. 2021 Mar;289:114032. doi: 10.1016/j.jviromet.2020.114032. Epub 2020 Dec 5. PMID: 33290786; PMCID: PMC7718587.

7) Chan-Yeung M, Xu RH. SARS: epidemiology. Respirology. 2003 Nov;8 Suppl(Suppl 1):S9-14. doi: 10.1046/j.1440-1843.2003.00518.x. PMID: 15018127; PMCID: PMC7169193

8) Lee J, Chowell G, Jung E. A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: A retrospective analysis on control interventions and superspreading events. J Theor Biol. 2016 Nov 7;408:118-126. doi: 10.1016/j.jtbi.2016.08.009. Epub 2016 Aug 10. PMID: 27521523; PMCID: PMC7094115

9) Worobey M, Levy JI, Malpica Serrano L, Crits-Christoph A, Pekar JE, Goldstein SA, Rasmussen AL, Kraemer MUG, Newman C, Koopmans MPG, Suchard MA, Wertheim JO, Lemey P, Robertson DL, Garry RF, Holmes EC, Rambaut A, Andersen KG. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science. 2022 Aug 26;377(6609):951-959. doi: 10.1126/science.abp8715. Epub 2022 Jul 26. PMID: 35881010; PMCID: PMC9348750

10) Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005 Oct 28;310(5748):676-9. doi: 10.1126/science.1118391. Epub 2005 Sep 29. PMID: 16195424

11) Zhou, P., Yang, XL., Wang, XG. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273 (2020). https://doi.org/10.1038/s41586-020-2012-7

12) Temmam, S., Vongphayloth, K., Baquero, E. et al. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 604, 330–336 (2022). https://doi.org/10.1038/s41586-022-04532

13) Cyranoski D. Mystery deepens over animal source of coronavirus. Nature. 2020 Mar;579(7797):18-19. doi: 10.1038/d41586-020-00548-w. PMID: 32127703

14) Xiao, K., Zhai, J., Feng, Y. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 583, 286–289 (2020). https://doi.org/10.1038/s41586-020-2313-x

15) Giovanetti M, Benedetti F, Campisi G, Ciccozzi A, Fabris S, Ceccarelli G, Tambone V, Caruso A, Angeletti S, Zella D, Ciccozzi M. Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. Biochem Biophys Res Commun. 2021 Jan 29;538:88-91. doi: 10.1016/j.bbrc.2020.10.102. Epub 2020 Nov 6. PMID: 33199021; PMCID: PMC7836704

16) Arya R, Kumari S, Pandey B, Mistry H, Bihani SC, Das A, Prashar V, Gupta GD, Panicker L, Kumar M. Structural insights into SARS-CoV-2 proteins. J Mol Biol. 2021 Jan 22;433(2):166725. doi: 10.1016/j.jmb.2020.11.024. Epub 2020 Nov 24. PMID: 33245961; PMCID: PMC7685130

17) Irwin Jungreis, Chase W. Nelson, Zachary Ardern, Yaara Finkel, Nevan J. Krogan, Kei Sato, John Ziebuhr, Noam Stern-Ginossar, Angelo Pavesi, Andrew E. Firth, Alexander E. Gorbalenya, Manolis Kellis, Conflicting and ambiguous names of overlapping ORFs in the SARS-CoV-2 genome: A homology-based resolution, Virology, Volume 558, 2021,Pages 145-151, ISSN 0042-6822, https://doi.org/10.1016/j.virol.2021.02.013.

18) Zella D, Giovanetti M, Cella E, Borsetti A, Ciotti M, Ceccarelli G, D'Ettorre G, Pezzuto A, Tambone V, Campanozzi L, Magheri M, Unali F, Bianchi M, Benedetti F, Pascarella S, Angeletti S, Ciccozzi M. The importance of genomic analysis in cracking the coronavirus pandemic. Expert Rev Mol Diagn. 2021 Jun;21(6):547-562. doi: 10.1080/14737159.2021.1917998. Epub 2021 Apr 28. PMID: 33849359; PMCID: PMC8095159

19) Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe. 2020 Mar 11;27(3):325-328. doi: 10.1016/j.chom.2020.02.001. Epub 2020 Feb 7. PMID: 32035028; PMCID: PMC7154514

20) Clark LK, Green TJ, Petit CM. Structure of Nonstructural Protein 1 from SARS-CoV-2. J Virol. 2021 Jan 28;95(4):e02019-20. doi: 10.1128/JVI.02019-20. PMID: 33234675; PMCID: PMC7851544

21) Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Mühlemann O, Ban N. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat Struct Mol Biol. 2020 Oct;27(10):959-966. doi: 10.1038/s41594-020-0511-8. Epub 2020 Sep 9. Erratum in: Nat Struct Mol Biol. 2020 Nov;27(11):1094. PMID: 32908316

22) Zheng YX, Wang L, Kong WS, Chen H, Wang XN, Meng Q, Zhang HN, Guo SJ, Jiang HW, Tao SC. Nsp2 has the potential to be a drug target revealed by global identification of SARS-CoV-2 Nsp2-interacting

proteins. Acta Biochim Biophys Sin (Shanghai). 2021 Aug 31;53(9):1134-1141. doi: 10.1093/abbs/gmab088. PMID: 34159380

23) Verba K, Gupta M, Azumaya C, Moritz M, Pourmal S, Diallo A, Merz G, Jang G, Bouhaddou M, Fossati A, Brilot A, Diwanji D, Hernandez E, Herrera N, Kratochvil H, Lam V, Li F, Li Y, Nguyen H, Nowotny C, Owens T, Peters J, Rizo A, Schulze-Gahmen U, Smith A, Young I, Yu Z, Asarnow D, Billesbølle C, Campbell M, Chen J, Chen KH, Chio US, Dickinson M, Doan L, Jin M, Kim K, Li J, Li YL, Linossi E, Liu Y, Lo M, Lopez J, Lopez K, Mancino A, Iii FM, Paul M, Pawar K, Pelin A, Pospiech T, Puchades C, Remesh S, Safari M, Schaefer K, Sun M, Tabios M, Thwin A, Titus E, Trenker R, Tse E, Tsui TKM, Feng F, Zhang K, Zhang Y, Zhao J, Zhou F, Zhou Y, Zuliani-Alvarez L, Agard D, Cheng Y, Fraser J, Jura N, Kortemme T, Manglik A, Southworth D, Stroud R, Swaney D, Krogan N, Frost A, Rosenberg O. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. Res Sq [Preprint]. 2021 May 19:rs.3.rs-515215. doi: 10.21203/rs.3.rs-515215/v1. PMID: 34031651; PMCID: PMC8142659

24) Tollenaere MAX, Tiedje C, Rasmussen S, Nielsen JC, Vind AC, Blasius M, Batth TS, Mailand N, Olsen JV, Gaestel M, Bekker-Jensen S. GIGYF1/2-Driven Cooperation between ZNF598 and TTP in Posttranscriptional Regulation of Inflammatory Signaling. Cell Rep. 2019 Mar 26;26(13):3511-3521.e4. doi: 10.1016/j.celrep.2019.03.006. PMID: 30917308

25) Cornillez-Ty CT, Liao L, Yates JR 3rd, Kuhn P, Buchmeier MJ. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and

intracellular signaling. J Virol. 2009 Oct;83(19):10314-8. doi: 10.1128/JVI.00842-09. Epub 2009 Jul 29. PMID: 19640993; PMCID: PMC2748024

26) von Brunn A, Teepe C, Simpson JC, Pepperkok R, Friedel CC, Zimmer R, Roberts R, Baric R, Haas J. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. PLoS One. 2007 May 23;2(5):e459. doi: 10.1371/journal.pone.0000459. PMID: 17520018; PMCID: PMC1868897

27) Ma J, Chen Y, Wu W, Chen Z. Structure and Function of N-Terminal Zinc Finger Domain of SARS-CoV-2 NSP2. Virol Sin. 2021 Oct;36(5):1104-1112. doi: 10.1007/s12250-021-00431-6. Epub 2021 Aug 16. PMID: 34398430; PMCID: PMC8365134

28) Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res. 2018;149:58-74. doi:10.1016/j.antiviral.2017.11.0011

29) Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res. 2018 Jan;149:58-74. doi: 10.1016/j.antiviral.2017.11.001. Epub 2017 Nov 8. PMID: 29128390; PMCID: PMC7113668

30) Zhao Y, Du X, Duan Y, Pan X, Sun Y, You T, Han L, Jin Z, Shang W, Yu J, Guo H, Liu Q, Wu Y, Peng C, Wang J, Zhu C, Yang X, Yang K, Lei Y, Guddat LW, Xu W, Xiao G, Sun L, Zhang L, Rao Z, Yang H. High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. Protein Cell. 2021 Nov;12(11):877-888. doi: 10.1007/s13238-021-00836-9. Epub 2021 Apr 17. PMID: 33864621; PMCID: PMC8052528

31) Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology. 2017 Oct;510:165-174. doi: 10.1016/j.virol.2017.07.019. Epub 2017 Jul 21. PMID: 28738245; PMCID: PMC7111695

32) Chen CC, Yu X, Kuo CJ, Min J, Chen S, Ma L, Liu K, Guo RT. Overview of antiviral drug candidates targeting coronaviral 3C-like main proteases. FEBS J. 2021 Sep;288(17):5089-5121. doi: 10.1111/febs.15696. Epub 2021 Feb 1. PMID: 33400393

33) Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature. 2020 Jun;582(7811):289-293. doi: 10.1038/s41586-020-2223-y. Epub 2020 Apr 9. PMID: 32272481

34) Qiu, Y., & Xu, K. (2020). Functional studies of the coronavirus nonstructural proteins. STEMedicine. https://doi.org/10.37175/stemedicine.v1i2.39

35) te Velthuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res. 2012 Feb;40(4):1737-47. doi: 10.1093/nar/gkr893. Epub 2011 Oct 29. PMID: 22039154; PMCID: PMC3287201

36) Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, Ferron F, Gorbalenya AE, Canard B. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J. 2006 Oct

18;25(20):4933-42. doi: 10.1038/sj.emboj.7601368. Epub 2006 Oct 5. PMID: 17024178; PMCID: PMC1618104

37) Biswal M, Diggs S, Xu D, Khudaverdyan N, Lu J, Fang J, Blaha G, Hai R, Song J. Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. Nucleic Acids Res. 2021 Jun 4;49(10):5956-5966. doi: 10.1093/nar/gkab370. PMID: 33999154; PMCID: PMC8191759

38) Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science. 2020 May 15;368(6492):779-782. doi: 10.1126/science.abb7498. Epub 2020 Apr 10. PMID: 32277040; PMCID: PMC7164392

39) Zhang C, Chen Y, Li L, Yang Y, He J, Chen C, Su D. Structural basis for the multimerization of nonstructural protein nsp9 from SARS-CoV-2. Mol Biomed. 2020;1(1):5. doi: 10.1186/s43556-020-00005-0. Epub 2020 Aug 20. PMID: 34765992; PMCID: PMC7438161

40) Joseph JS, Saikatendu KS, Subramanian V, Neuman BW, Brooun A, Griffith M, Moy K, Yadav MK, Velasquez J, Buchmeier MJ, Stevens RC, Kuhn P. Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. J Virol. 2006 Aug;80(16):7894-901. doi: 10.1128/JVI.00467-06. PMID: 16873246; PMCID: PMC1563791

41) Shu T, Huang M, Wu D, Ren Y, Zhang X, Han Y, Mu J, Wang R, Qiu Y, Zhang DY, Zhou X. SARS-Coronavirus-2 Nsp13 Possesses NTPase and

RNA Helicase Activities That Can Be Inhibited by Bismuth Salts. Virol Sin. 2020 Jun;35(3):321-329. doi: 10.1007/s12250-020-00242-1. Epub 2020 Jun 4. PMID: 32500504; PMCID: PMC7271831

42) Hao W, Wojdyla JA, Zhao R, Han R, Das R, Zlatev I, Manoharan M, Wang M, Cui S. Crystal structure of Middle East respiratory syndrome coronavirus helicase. PLoS Pathog. 2017 Jun 26;13(6):e1006474. doi: 10.1371/journal.ppat.1006474. PMID: 28651017; PMCID: PMC5501694

43) Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007;76:23-50. doi: 10.1146/annurev.biochem.76.052305.115300. PMID: 17506634

44) Chen Y, Cai H, Pan J, Xiang N, Tien P, Ahola T, Guo D. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. Proc Natl Acad Sci U S A. 2009 Mar 3;106(9):3484-9. doi: 10.1073/pnas.0808790106. Epub 2009 Feb 10. PMID: 19208801; PMCID: PMC2651275

45) Shannon, A., Selisko, B., Le, NTT. et al. Rapid incorporation of Favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. Nat Commun 11, 4682 (2020). https://doi.org/10.1038/s41467-020-18463-z

46) Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci U S A. 2015 Jul 28;112(30):9436-41. doi: 10.1073/pnas.1508686112. Epub 2015 Jul 9. PMID: 26159422; PMCID: PMC4522806

47) Chen Y, Tao J, Sun Y, Wu A, Su C, Gao G, Cai H, Qiu S, Wu Y, Ahola T, Guo D. Structure-function analysis of severe acute respiratory syndrome

coronavirus RNA cap guanine-N7-methyltransferase. J Virol. 2013 Jun;87(11):6296-305. doi: 10.1128/JVI.00061-13. Epub 2013 Mar 27. PMID: 23536667; PMCID: PMC3648086

48) Hackbart M, Deng X, Baker SC. Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. Proc Natl Acad Sci U S A. 2020 Apr 7;117(14):8094-8103. doi: 10.1073/pnas.1921485117. Epub 2020 Mar 20. PMID: 32198201; PMCID: PMC7149396

49) Kim Y, Jedrzejczak R, Maltseva NI, Wilamowski M, Endres M, Godzik A, Michalska K, Joachimiak A. Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. Protein Sci. 2020 Jul;29(7):1596-1605. doi: 10.1002/pro.3873. Epub 2020 May 2. PMID: 32304108; PMCID: PMC7264519

50) Viswanathan T, Arya S, Chan SH, Qi S, Dai N, Hromas RA, Park JG, Oladunni F, Martinez-Sobrido L, Gupta YK. Structural Basis of RNA Cap Modification by SARS-CoV-2 Coronavirus. bioRxiv [Preprint]. 2020 Apr 26:2020.04.26.061705. doi: 10.1101/2020.04.26.061705. Update in: Nat Commun. 2020 Jul 24;11(1):3718. PMID: 32511383; PMCID: PMC7263512

51) Schoeman, D., Fielding, B.C. Coronavirus envelope protein: current knowledge. Virol J 16, 69 (2019). https://doi.org/10.1186/s12985-019-1182-0

52) Mandala VS, McKay MJ, Shcherbakov AA, Dregni AJ, Kolocouris A, Hong M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nat Struct Mol Biol. 2020

Dec;27(12):1202-1208. doi: 10.1038/s41594-020-00536-8. Epub 2020 Nov 11. PMID: 33177698; PMCID: PMC7718435

53) Li J, Guo M, Tian X, Wang X, Yang X, Wu P, Liu C, Xiao Z, Qu Y, Yin Y, Wang C, Zhang Y, Zhu Z, Liu Z, Peng C, Zhu T, Liang Q. Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis. Med (N Y). 2021 Jan 15;2(1):99-112.e7. doi: 10.1016/j.medj.2020.07.002. Epub 2020 Jul 21. PMID: 32838362; PMCID: PMC7373048

54) Zhang, Z., Nomura, N., Muramoto, Y. et al. Structure of SARS-CoV-2 membrane protein essential for virus assembly. Nat Commun 13, 4399 (2022). https://doi.org/10.1038/s41467-022-32019-3

55) Bai Z, Cao Y, Liu W, Li J. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. Viruses. 2021 Jun 10;13(6):1115. doi: 10.3390/v13061115. PMID: 34200602; PMCID: PMC8227405

56) Kang, S., Yang, M., He, S. et al. A SARS-CoV-2 antibody curbs viral nucleocapsid protein-induced complement hyperactivation. Nat Commun 12, 2697 (2021). https://doi.org/10.1038/s41467-021-23036-9

57) Tortorici MA, Veesler D. Structural insights into coronavirus entry. Adv Virus Res. 2019;105:93-116. doi: 10.1016/bs.aivir.2019.08.002. Epub 2019 Aug 22. PMID: 31522710; PMCID: PMC7112261

58) Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol 5, 562–569 (2020). https://doi.org/10.1038/s41564-020-0688-y

59) 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 Dec 10;183(6):1735. doi: 10.1016/j.cell.2020.11.032. Erratum for: Cell. 2020 Apr 16;181(2):281-292.e6. PMID: 33306958; PMCID: PMC7833104

60) Redondo N, Zaldívar-López S, Garrido JJ, Montoya M. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. Front Immunol. 2021 Jul 7;12:708264. doi: 10.3389/fimmu.2021.708264. PMID: 34305949; PMCID: PMC8293742

61) Azad GK, Khan PK. Variations in Orf3a protein of SARS-CoV-2 alter its structure and function. Biochem Biophys Rep. 2021 Jan 27;26:100933. doi: 10.1016/j.bbrep.2021.100933. PMID: 33527091; PMCID: PMC7839395

62) Kern, D.M., Sorum, B., Mali, S.S. et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. Nat Struct Mol Biol 28, 573–582 (2021). https://doi.org/10.1038/s41594-021-00619-0

63) Konno Y, Kimura I, Uriu K, Fukushi M, Irie T, Koyanagi Y, Sauter D, Gifford RJ; USFQ-COVID19 Consortium, Nakagawa S, Sato K. SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. Cell Rep. 2020 Sep 22;32(12):108185. doi: 10.1016/j.celrep.2020.108185. Epub 2020 Sep 4. PMID: 32941788; PMCID: PMC7473339

64) Miyamoto, Y., Itoh, Y., Suzuki, T. et al. SARS-CoV-2 ORF6 disrupts nucleocytoplasmic trafficking to advance viral replication. Commun Biol 5, 483 (2022). https://doi.org/10.1038/s42003-022-03427-4

65) Cao, Z., Xia, H., Rajsbaum, R. et al. Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. Cell Mol Immunol 18, 746–748 (2021). https://doi.org/10.1038/s41423-020-00603-6

66) Zhou Z, Huang C, Zhou Z, Huang Z, Su L, Kang S, Chen X, Chen Q, He S, Rong X, Xiao F, Chen J, Chen S. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. iScience. 2021 Mar 19;24(3):102187. doi: 10.1016/j.isci.2021.102187. Epub 2021 Feb 12. PMID: 33615195; PMCID: PMC7879101

67) Zhang Y, Chen Y, Li Y, Huang F, Luo B, Yuan Y, Xia B, Ma X, Yang T, Yu F, Liu J, Liu B, Song Z, Chen J, Yan S, Wu L, Pan T, Zhang X, Li R, Huang W, He X, Xiao F, Zhang J, Zhang H. The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. Proc Natl Acad Sci U S A. 2021 Jun 8;118(23):e2024202118. doi: 10.1073/pnas.2024202118. PMID: 34021074; PMCID: PMC8201919

68) Flower TG, Buffalo CZ, Hooy RM, Allaire M, Ren X, Hurley JH. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. Proc Natl Acad Sci U S A. 2021 Jan 12;118(2):e2021785118. doi: 10.1073/pnas.2021785118. PMID: 33361333; PMCID: PMC7812859

69) Meier C, Aricescu AR, Assenberg R, Aplin RT, Gilbert RJ, Grimes JM, Stuart DI. The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus. Structure. 2006 Jul;14(7):1157-65. doi: 10.1016/j.str.2006.05.012. PMID: 16843897; PMCID: PMC7126280

70) Dominguez Andres A, Feng Y, Campos AR, Yin J, Yang CC, James B, Murad R, Kim H, Deshpande AJ, Gordon DE, Krogan N, Pippa R, Ronai ZA. SARS-CoV-2 ORF9c Is a Membrane-Associated Protein that Suppresses Antiviral Responses in Cells. bioRxiv [Preprint]. 2020 Aug

19:2020.08.18.256776. doi: 10.1101/2020.08.18.256776. PMID: 32839770; PMCID: PMC7444281

71) Pancer K, Milewska A, Owczarek K, Dabrowska A, Kowalski M, Łabaj PP, Branicki W, Sanak M, Pyrc K. The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. PLoS Pathog. 2020 Dec 10;16(12):e1008959. doi: 10.1371/journal.ppat.1008959. PMID: 33301543; PMCID: PMC7755277

72) Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q, Zhou H, Yan J, Qi J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell. 2020 May 14;181(4):894-904.e9. doi: 10.1016/j.cell.2020.03.045. Epub 2020 Apr 9. PMID: 32275855; PMCID: PMC7144619

73) Hamming I, Cooper ME, Haagmans BL, Hooper NM, Korstanje R, Osterhaus AD, Timens W, Turner AJ, Navis G, van Goor H. The emerging role of ACE2 in physiology and disease. J Pathol. 2007 May;212(1):1-11. doi: 10.1002/path.2162. PMID: 17464936; PMCID: PMC7167724

74) Han X, Qi J, Song H, Wang Q, Zhang Y, Wu Y, Lu G, Yuen KY, Shi Y, Gao GF. Structure of the S1 subunit C-terminal domain from bat-derived coronavirus HKU5 spike protein. Virology. 2017 Jul;507:101-109. doi: 10.1016/j.virol.2017.04.016. Epub 2017 Apr 19. PMID: 28432925; PMCID: PMC7111649

75) Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kuivanen S, van der Meer F, Kallio K, Kaya T, Anastasina M, Smura T, Levanov L, Szirovicza L, Tobi A, Kallio-Kokko H, Österlund P, Joensuu M, Meunier FA, Butcher SJ, Winkler MS, Mollenhauer B, Helenius A, Gokce O, Teesalu T, Hepojoki J, Vapalahti O, Stadelmann C, Balistreri G, Simons

M. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. Science. 2020 Nov 13;370(6518):856-860. doi: 10.1126/science.abd2985. Epub 2020 Oct 20. PMID: 33082293; PMCID: PMC7857391

76) Wang S, Qiu Z, Hou Y, Deng X, Xu W, Zheng T, Wu P, Xie S, Bian W, Zhang C, Sun Z, Liu K, Shan C, Lin A, Jiang S, Xie Y, Zhou Q, Lu L, Huang J, Li X. AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells. Cell Res. 2021 Feb;31(2):126-140. doi: 10.1038/s41422-020-00460-y. Epub 2021 Jan 8. PMID: 33420426; PMCID: PMC7791157

77) Bestle D, Heindl MR, Limburg H, Van Lam van T, Pilgram O, Moulton H, Stein DA, Hardes K, Eickmann M, Dolnik O, Rohde C, Klenk HD, Garten W, Steinmetzer T, Böttcher-Friebertshäuser E. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci Alliance. 2020 Jul 23;3(9):e202000786. doi: 10.26508/lsa.202000786. PMID: 32703818; PMCID: PMC7383062

78) Ou T, Mou H, Zhang L, Ojha A, Choe H, Farzan M. Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2. PLoS Pathog. 2021 Jan 19;17(1):e1009212. doi: 10.1371/journal.ppat.1009212. PMID: 33465165; PMCID: PMC7845965

79) Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr 16;181(2):271-280.e8. doi: 10.1016/j.cell.2020.02.052. Epub 2020 Mar 5. PMID: 32142651; PMCID: PMC7102627

80) Bayati A, Kumar R, Francis V, McPherson PS. SARS-CoV-2 infects cells after viral entry via clathrin-mediated endocytosis. J Biol Chem. 2021 Jan-Jun;296:100306. doi: 10.1016/j.jbc.2021.100306. Epub 2021 Jan 19. PMID: 33476648; PMCID: PMC7816624

81) Inoue Y, Tanaka N, Tanaka Y, Inoue S, Morita K, Zhuang M, Hattori T, Sugamura K. Clathrin-dependent entry of severe acute respiratory syndrome coronavirus into target cells expressing ACE2 with the cytoplasmic tail deleted. J Virol. 2007 Aug;81(16):8722-9. doi: 10.1128/JVI.00253-07. Epub 2007 May 23. PMID: 17522231; PMCID: PMC1951348

82) V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol. 2021 Mar;19(3):155-170. doi: 10.1038/s41579-020-00468-6. Epub 2020 Oct 28. PMID: 33116300; PMCID: PMC7592455

83) Smith EC, Blanc H, Surdel MC, Vignuzzi M, Denison MR. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. PLoS Pathog. 2013 Aug;9(8):e1003565. doi: 10.1371/journal.ppat.1003565. Epub 2013 Aug 15. Erratum in: PLoS Pathog. 2014 Jul;10(7):e1004342. Surdel, Matthew C [added]. PMID: 23966862; PMCID: PMC3744431

84) Scudellari M. How to coronavirus infects our cells? Nature. (2021) 595:640–5. doi: 10.1038/d41586-021-02039-y

85) O'Toole, Á., Pybus, O.G., Abram, M.E. et al. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. BMC Genomics 23, 121 (2022). https://doi.org/10.1186/s12864-022-08358-2

86) Rambaut, A., Holmes, E.C., O'Toole, Á. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5, 1403–1407 (2020). https://doi.org/10.1038/s41564-020-0770-5

87) Aleem A, Akbar Samad AB, Slenker AK. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). 2022 May 12. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 34033342

88) Walker AS, Vihta KD, Gethings O, Pritchard E, Jones J, House T, Bell I, Bell JI, Newton JN, Farrar J, Diamond I, Studley R, Rourke E, Hay J, Hopkins S, Crook D, Peto T, Matthews PC, Eyre DW, Stoesser N, Pouwels KB; Covid-19 Infection Survey Team. Tracking the Emergence of SARS-CoV-2 Alpha Variant in the United Kingdom. N Engl J Med. 2021 Dec 30;385(27):2582-2585. doi: 10.1056/NEJMc2103227. Epub 2021 Dec 8. PMID: 34879193; PMCID: PMC8693687

89) Khan A, Zia T, Suleman M, Khan T, Ali SS, Abbasi AA, Mohammad A, Wei DQ. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. J Cell Physiol. 2021 Oct;236(10):7045-7057. doi: 10.1002/jcp.30367. Epub 2021 Mar 23. PMID: 33755190; PMCID: PMC8251074

90) Ramanathan M, Ferguson ID, Miao W, Khavari PA. SARS-CoV-2 B.1.1.7 and B.1.351 Spike variants bind human ACE2 with increased affinity. bioRxiv [Preprint]. 2021 Feb 22:2021.02.22.432359. doi: 10.1101/2021.02.22.432359. Update in: Lancet Infect Dis. 2021 May 19;: PMID: 33655251; PMCID: PMC7924271

91) Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020 Sep 3;182(5):1295-1310.e20. doi: 10.1016/j.cell.2020.08.012. Epub 2020 Aug 11. PMID: 32841599; PMCID: PMC7418704

92) Ozono S, Zhang Y, Ode H, Sano K, Tan TS, Imai K, Miyoshi K, Kishigami S, Ueno T, Iwatani Y, Suzuki T, Tokunaga K. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. Nat Commun. 2021 Feb 8;12(1):848. doi: 10.1038/s41467-021-21118-2. PMID: 33558493; PMCID: PMC7870668

93) Tchesnokova V, Kulakesara H, Larson L, Bowers V, Rechkina E, Kisiela D, Sledneva Y, Choudhury D, Maslova I, Deng K, Kutumbaka K, Geng H, Fowler C, Greene D, Ralston J, Samadpour M, Sokurenko E. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. bioRxiv [Preprint]. 2021 Mar 11:2021.02.22.432189. doi: 10.1101/2021.02.22.432189. Update in: J Clin Microbiol. 2021 Aug 11;:JCM0092121. PMID: 33758861; PMCID: PMC7987020

94) Khandia R, Singhal S, Alqahtani T, Kamal MA, El-Shall NA, Nainu F, Desingu PA, Dhama K. Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. Environ Res. 2022 Jun;209:112816. doi: 10.1016/j.envres.2022.112816. Epub 2022 Jan 29. PMID: 35093310; PMCID: PMC8798788

95) Dhawan M, Saied AA, Mitra S, Alhumaydhi FA, Emran TB, Wilairatana P. Omicron variant (B.1.1.529) and its sublineages: What do we know so far amid the emergence of recombinant variants of SARS-CoV-2? Biomed Pharmacother. 2022 Oct;154:113522. doi: 10.1016/j.biopha.2022.113522. Epub 2022 Aug 15. PMID: 36030585; PMCID: PMC9376347

96) Niknam Z, Jafari A, Golchin A, Danesh Pouya F, Nemati M, Rezaei-Tavirani M, Rasmi Y. Potential therapeutic options for COVID-19: an update on current evidence. Eur J Med Res. 2022 Jan 13;27(1):6. doi: 10.1186/s40001-021-00626-3. PMID: 35027080; PMCID: PMC8755901

97) Lu, RM., Hwang, YC., Liu, IJ. et al. Development of therapeutic antibodies for the treatment of diseases. J Biomed Sci 27, 1 (2020). https://doi.org/10.1186/s12929-019-0592-z

98) Taylor, P.C., Adams, A.C., Hufford, M.M. et al. Neutralizing monoclonal antibodies for treatment of COVID-19. Nat Rev Immunol 21, 382–393 (2021). https://doi.org/10.1038/s41577-021-00542-x

99) Gavor E, Choong YK, Er SY, Sivaraman H, Sivaraman J. Structural Basis of SARS-CoV-2 and SARS-CoV Antibody Interactions. Trends Immunol. 2020 Nov;41(11):1006-1022. doi: 10.1016/j.it.2020.09.004. Epub 2020 Sep 17. PMID: 33041212; PMCID: PMC7498231

100) Barnes, C.O., Jette, C.A., Abernathy, M.E. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. Nature 588, 682–687 (2020). https://doi.org/10.1038/s41586-020-2852-1

101) Dong J, Zost SJ, Greaney AJ, Starr TN, Dingens AS, Chen EC, Chen RE, Case JB, Sutton RE, Gilchuk P, Rodriguez J, Armstrong E, Gainza C, Nargi RS, Binshtein E, Xie X, Zhang X, Shi P-Y, Logue J, Weston S, McGrath ME, Frieman MB, Brady T, Tuffy KM, Bright H, Loo Y-M,

McTamney PM, Esser MT, Carnahan RH, Diamond MS, Bloom JD, Crowe JE. Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. Nat Microbiol. 2021;6(10):1233–44

102) Loo YM, McTamney PM, Arends RH, Abram ME, Aksyuk AA, Diallo S, Flores DJ, Kelly EJ, Ren K, Roque R, Rosenthal K, Streicher K, Tuffy KM, Bond NJ, Cornwell O, Bouquet J, Cheng LI, Dunyak J, Huang Y, Rosenbaum AI, Pilla Reddy V, Andersen H, Carnahan RH, Crowe JE Jr, Kuehne AI, Herbert AS, Dye JM, Bright H, Kallewaard NL, Pangalos MN, Esser MT. The SARS-CoV-2 monoclonal antibody combination, AZD7442, is protective in nonhuman primates and has an extended half-life in humans. Sci Transl Med. 2022 Mar 9;14(635):eabl8124. doi: 10.1126/scitranslmed.abl8124. Epub 2022 Mar 9. PMID: 35076282; PMCID: PMC8939769

103) Case, J.B., Mackin, S., Errico, J.M. et al. Resilience of S309 and AZD7442 monoclonal antibody treatments against infection by SARS-CoV-2 Omicron lineage strains. Nat Commun 13, 3824 (2022). https://doi.org/10.1038/s41467-022-31615-7

104) Jones BE, Brown-Augsburger PL, Corbett KS, Westendorf K, Davies J, Cujec TP, Wiethoff CM, Blackbourne JL, Heinz BA, Foster D, Higgs RE, Balasubramaniam D, Wang L, Zhang Y, Yang ES, Bidshahri R, Kraft L, Hwang Y, Žentelis S, Jepson KR, Goya R, Smith MA, Collins DW, Hinshaw SJ, Tycho SA, Pellacani D, Xiang P, Muthuraman K, Sobhanifar S, Piper MH, Triana FJ, Hendle J, Pustilnik A, Adams AC, Berens SJ, Baric RS, Martinez DR, Cross RW, Geisbert TW, Borisevich V, Abiona O, Belli HM, de Vries M, Mohamed A, Dittmann M, Samanovic MI,

Mulligan MJ, Goldsmith JA, Hsieh CL, Johnson NV, Wrapp D, McLellan JS, Barnhart BC, Graham BS, Mascola JR, Hansen CL, Falconer E. The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates. Sci Transl Med. 2021 May 12;13(593):eabf1906. doi: 10.1126/scitranslmed.abf1906. Epub 2021 Apr 5. PMID: 33820835; PMCID: PMC8284311

105) Tada T, Zhou H, Dcosta BM, Samanovic MI, Chivukula V, Herati RS, Hubbard SR, Mulligan MJ, Landau NR. Increased resistance of SARS-CoV-2 Omicron variant to neutralization by vaccine-elicited and therapeutic antibodies. EBioMedicine. 2022 Apr;78:103944. doi: 10.1016/j.ebiom.2022.103944. PMID: 35465948; PMCID: PMC9021600

106) Dejnirattisai W, Zhou D, Ginn HM, Duyvesteyn HME, Supasa P, Case JB, Zhao Y, Walter TS, Mentzer AJ, Liu C, Wang B, Paesen GC, Slon-Campos J, López-Camacho C, Kafai NM, Bailey AL, Chen RE, Ying B, Thompson C, Bolton J, Fyfe A, Gupta S, Tan TK, Gilbert-Jaramillo J, James W, Knight M, Carroll MW, Skelly D, Dold C, Peng Y, Levin R, Dong T, Pollard AJ, Knight JC, Klenerman P, Temperton N, Hall DR, Williams MA, Paterson NG, Bertram FKR, Siebert CA, Clare DK, Howe A, Radecke J, Song Y, Townsend AR, Huang KA, Fry EE, Mongkolsapaya J, Diamond MS, Ren J, Stuart DI, Screaton GR. The antigenic anatomy of SARS-CoV-2 receptor binding domain. Cell. 2021 Apr 15;184(8):2183-2200.e22. doi: 10.1016/j.cell.2021.02.032. Epub 2021 Feb 18. PMID: 33756110; PMCID: PMC7891125

107) Hwang YC, Lu RM, Su SC, Chiang PY, Ko SH, Ke FY, Liang KH, Hsieh TY, Wu HC. Monoclonal antibodies for COVID-19 therapy and

SARS-CoV-2 detection. J Biomed Sci. 2022 Jan 4;29(1):1. doi: 10.1186/s12929-021-00784-w. PMID: 34983527; PMCID: PMC8724751

108) Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, Zhang Z, Fan P, Dong Y, Yang Y, Chen Z, Guo Y, Zhang J, Li Y, Song X, Chen Y, Xia L, Fu L, Hou L, Xu J, Yu C, Li J, Zhou Q, Chen W. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science. 2020 Aug 7;369(6504):650-655. doi: 10.1126/science.abc6952. Epub 2020 Jun 22. PMID: 32571838; PMCID: PMC7319273

109) Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017 Mar 30;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494. PMID: 28382917; PMCID: PMC5388101

110) A. Benson, M. Cavanaugh, K. Clark et al., "GenBank,"Nucleic Acids Research, vol. 46, no. D1, pp. D41–D47, 2018

111) O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. Reference sequence (RefSeq) database at NCBI: current status, taxonomic

expansion, and functional annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733-45

112) The RCSB PDB is a member of the worldwide PDB (wwPDB). The wwPDB should be cited with the URL www.wwpdb.org and the following citation: H.M. Berman, K. Henrick, H. Nakamura (2003) Announcing the worldwide Protein Data Bank Nature Structural Biology 10 (12): 980

113) UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015 Jan;43(Database issue):D204-12. doi: 10.1093/nar/gku989. Epub 2014 Oct 27. PMID: 25348405; PMCID: PMC4384041

114) S. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," Nucleic Acids Research, vol. 25, no. 17, pp. 3389–3402, 1997

115) A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, "Jalview version 2–a multiple sequence alignment editor and analysis workbench," Bioinformatics, vol. 25,no. 9, pp. 1189–1191, 2009

116) K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," Molecular Biology and Evolution, vol. 30, no. 4, pp. 772–780, 2013

117) Y. Chen, P. Yu, J. Luo, and Y. Jiang, "Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT,"Mammalian Genome, vol. 14, no. 12, pp. 859–865, 2003

118) L. J. McGuffin, K. Bryson, and D. T. Jones, "The PSIPRED protein structure prediction server," Bioinformatics, vol. 16, no. 4, pp. 404-405, 2000

119) U. Omasits, C. H. Ahrens, S. Müller, and B. Wollscheid, "Protter:interactive protein feature visualization and integration with experimental proteomic data," Bioinformatics, vol. 30, no. 6, pp. 884–886, 2014

120) L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "CD-HIT: accelerated for clustering the next-generation sequencing data," Bioinformatics,vol. 28, no. 23, pp. 3150–3152, 2012

121) A. Waterhouse, M. Bertoni, S. Bienert et al., "SWISS-MODEL:homology Modelling of protein structures and complexes,"Nucleic Acids Research, vol. 46, no. W1, pp. W296–W303,2018

122) L. Zimmermann, A. Stephens, S.-Z. Nam et al., "A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core," Journal of Molecular Biology, vol. 430, no. 15, pp. 2237–2243, 2018

123) J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER suite: protein structure and function prediction,"Nature Methods, vol. 12, no. 1, pp. 7-8, 2015

124) L. L. C. Schrodinger, The AxPyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC., 2015

125) Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." J. Comput. Chem. 25:1605-1612 (2004)

126) P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite, Trends Genet. 16 (2000) 276–277, https://doi.org/10.1016/s0168-9525 (00)02024-2

127) .B.J. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3d: an R package for the comparative analysis of protein structures, Bioinformatics 22(2006) 2695–2696, https://doi.org/10.1093/bioinformatics/btl461

128) E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V.Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, J. Damborsky, CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures, PLoS Comput. Biol. 8 (2012), e1002708, https://doi.org/10.1371/journal.pcbi.1002708

129) Guex, N.; Peitsch, M.C.; Schwede, T. Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: A Historical Perspective. Electrophoresis 2009, 30, S162–S173

130) Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA. 2001;98(18):10037-10041. https://doi.org/10.1073/pnas.181342398

131) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions Mats H. M. Olsson, Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen Journal of Chemical Theory and Computation 2011 7 (2), 525-537 DOI: 10.1021/ct100578z

132) Hamby, S.E.; Hirst, J.D. Prediction of Glycosylation Sites Using Random Forests. BMC Bioinform. 2008, 9, 500

133) Gupta, R.; Brunak, S. Prediction of Glycosylation across the Human Proteome and the Correlation to Protein Function. Pac. Symp.Biocomput. 2002, 310–322

134) Xue, L.C.; Rodrigues, J.P.; Kastritis, P.L.; Bonvin, A.M.; Vangone, A. PRODIGY: A Web Server for Predicting the Binding Affinity of Protein–Protein Complexes. Bioinformatics 2016, 32, 3676–3678

135) Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. PLoS Comput. Biol. 2012, 8, e1002829

136) Sweredoski, M.J.; Baldi, P. PEPITO: Improved Discontinuous B-Cell Epitope Prediction Using Multiple Distance Thresholds and Half Sphere Exposure. Bioinformatics 2008, 24, 1459–1460

137) Krüger, D.M.; Gohlke, H. DrugScorePPIWebserver: Fast and Accurate in Silico Alanine Scanning for Scoring Protein-Protein Interactions. Nucleic Acids Res. 2010, 38, W480–W486

138) Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005 Jan-Feb;45(1):177-82. doi: 10.1021/ci049714+. PMID: 15667143; PMCID: PMC1360656

139) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30(16), 2785–2791

140) Piovesan D, Minervini G, Tosatto SC. The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res. 2016 Jul

8;44(W1):W367-74. doi: 10.1093/nar/gkw315. Epub 2016 May 19. PMID: 27198219; PMCID: PMC4987896

141) Tiberti M, Terkelsen T, Degn K, Beltrame L, Cremers TC, da Piedade I, Di Marco M, Maiani E, Papaleo E. MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles. Brief Bioinform. 2022 May 13;23(3):bbac074. doi: 10.1093/bib/bbac074. PMID: 35323860

142) Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. Bioinformatics. 2019 Oct 15;35(20):4168-4169. doi: 10.1093/bioinformatics/btz184. PMID: 30874800; PMCID: PMC6792092

143) .https://pymolwiki.org/index.php/InterfaceResidues

144) E. Issa, G. Merhi, B. Panossian, T. Salloum, S. Tokajian, SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis, MSystems. 5 (2020), e00266-20, https://doi.org/10.1128/mSystems.00266-20

145) Pascarella S, Ciccozzi M, Zella D, Bianchi M, Benedetti F, Benvenuto D, Broccolo F, Cauda R, Caruso A, Angeletti S, Giovanetti M, Cassone A. SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate? J Med Virol. 2021 Dec;93(12):6551-6556. doi: 10.1002/jmv.27210. Epub 2021 Jul 27. PMID: 34260088; PMCID: PMC8426736

146) Pascarella S, Ciccozzi M, Bianchi M, Benvenuto D, Cauda R, Cassone A. The electrostatic potential of the Omicron variant spike is higher than in Delta and Delta-plus variants: A hint to higher transmissibility? J Med

Virol. 2022 Apr;94(4):1277-1280. doi: 10.1002/jmv.27528. Epub 2021 Dec 27. PMID: 34914120

147) COVID-19 Reinfection Study from South Africa Yields Ominous Data about Omicron | Science | AAAS. Accessed on December 7, 2021. https://www.science.org/content/article/covid-19-reinfection-study-south-africa-yields-ominous-data-about-omicron

148) Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol. 2022 Apr;94(4):1641-1649. doi: 10.1002/jmv.27526. Epub 2021 Dec 27. PMID: 34914115

149) Organization W.H. Worls Health Organization; 2022. Enhances Responses to Omicron SARS-CoV-2 variant. Technical brief and Priority For Member States. Updated #6:21.

150) Pascarella S, Ciccozzi M, Bianchi M, Benvenuto D, Giovanetti M, Cauda R, Cassone A. Shortening Epitopes to Survive: The Case of SARS-CoV-2 Lambda Variant. Biomolecules. 2021 Oct 10;11(10):1494. doi: 10.3390/biom11101494. PMID: 34680128; PMCID: PMC8533401

151) B. Lorenz, J. Leuck, D. Kohl, W. E. Muller, H. C. Schroder, Anti-HIV-1 activity of inorganic polyphosphates. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 14, 110–118 (1997)

152) Ferrucci V, Kong DY, Asadzadeh F, Marrone L, Boccia A, Siciliano R, Criscuolo G, Anastasio C, Quarantelli F, Comegna M, Pisano I, Passariello M, Iacobucci I, Monica RD, Izzo B, Cerino P, Fusco G, Viscardi M, Brandi S, Pierri BM, Borriello G, Tiberio C, Atripaldi L, Bianchi M, Paolella G, Capoluongo E, Castaldo G, Chiariotti L, Monti M, De Lorenzo C, Yun KS, Pascarella S, Cheong JH, Kim HY, Zollo M. Long-

chain polyphosphates impair SARS-CoV-2 infection and replication. Sci Signal. 2021 Jul 6;14(690):eabe5040. doi: 10.1126/scisignal.abe5040. PMID: 34230209; PMCID: PMC8432949

153) https://covdb.stanford.edu/susceptibility-data/table-mab-susc/

154) Bartlam, M.,Chen, C.,Cong, L.,Fu, J.,Rao, Z.,Tang, H.,Wei, L.,Wu, W.,Xu, Y.,Yang, A.,Ye, Z. IBV nsp2 is an endosome-associated protein and viral pathogenicity factor To be Published.

155) Saikatendu KS,Joseph JS,Subramanian V, Clayton T, Griffith M, Moy K, Velasquez J, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1"-phosphate dephosphorylation by a conserved domain of nsP3. Structure 2005;13(11):1665-75

156) Zhou A , Li S , Khan FA , Zhang S . Autophagy postpones apoptotic cell death in PRRSV infection through Bad-Beclin1 interaction. Virulence 2016; 7 (2):98–109 pages

157) Tang JW, Cheung JL, Chu IM, Sung JJ, Peiris M, Chan PK. The Large 386-nt Dele- tion in SARS-associated coronavirus: evidence for quasispecies? J Infect Diseas 2006; 194 (6):808–13 https://doi.org/. doi: 10.1086/507044

158) Xu W, Wang M, Yu D, Zhang X. Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. Front Immunol. 2020 Sep 4;11:565278. doi: 10.3389/fimmu.2020.565278. PMID: 33013929; PMCID: PMC7509417

159) EFSA Panel on Food Additives and Flavourings (FAF), M. Younes, G. Aquilina, L. Castle, K.-H. Engel, P. Fowler, M. J. F. Fernandez, P. Fürst,

R. Gürtler, T. Husøy, W. Mennes, P. Moldeus, A. Oskarsson, R. Shah, I. Waalkens-Berendsen, D. Wölfle, P. Aggett, A. Cupisti, C. Fortes, G. Kuhnle, I. T. Lillegaard, M. Scotter, A. Giarola, A. Rincon, A. Tard, U. Gundert-Remy, Re-evaluation of phosphoric acid-phosphates - di-, tri- and polyphosphates (E 338-341, E 343, E 450-452) as food additives and the safety of proposed extension of use. EFSA J. 17, e05674 (2019)

160)   Li J, Ma X, Guo S, Hou C, Shi L, Zhang H, Zheng B, Liao C, Yang L, Ye L, He X. A Hydrophobic-Interaction-Based Mechanism Triggers Docking between the SARS-CoV-2 Spike and Angiotensin-Converting Enzyme 2. Glob Chall. 2020 Oct 15;4(12):2000067. doi: 10.1002/gch2.202000067. PMID: 33173592; PMCID: PMC7646041

161)   VanBlargan, L.A., Errico, J.M., Halfmann, P.J. et al. An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. Nat Med 28, 490–495 (2022). https://doi.org/10.1038/s41591-021-01678-y

162)   Touret, F., Baronti, C., Pastorino, B. et al. In vitro activity of therapeutic antibodies against SARS-CoV-2 Omicron BA.1, BA.2 and BA.5. Sci Rep 12, 12609 (2022). https://doi.org/10.1038/s41598-022-16964-z

LIST OF PUBLICATIONS

1. Angeletti S, Benvenuto D, **Bianchi M**, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. J Med Virol. 2020; 92:584–588. https://doi.org/10.1002/jmv.25719588

2. Domenico Benvenuto, Silvia Angeletti, Marta Giovanetti, **Martina Bianchi**, Stefano Pascarella, Roberto Cauda, Massimo Ciccozzi, Antonio Cassone "Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy" April 2020, Journal of Infection 81(1). 10.1016/j.jinf.2020.03.058

3. Marco Ciotti, Silvia Angeletti, Marilena Minieri, Marta Giovannetti, Domenico Benvenuto, Stefano Pascarella, Caterina Sagnelli, **Martina Bianchi**, Sergio Bernardini, Massimo Ciccozzi "COVID-19 Outbreak: An Overview" Chemotherapy, (2020), pp. 1-9-9. https://doi.org/10.1159/000507423

4. Ciccozzi M, Benvenuto D, Giovanetti M, **Bianchi M**, Pascarella S, Angeletti S. Response to Ribeiro da Silva et al.,"Role of Nonstructural Proteins in the Pathogenesis of SARS-CoV-2" J Med Virol. 2020. https://doi.org/10.1002/jmv.25858

5. **Martina Bianchi**, Domenico Benvenuto, Marta Giovanetti, Silvia Angeletti, Massimo Ciccozzi, Stefano Pascarella "Sars-CoV-2 Envelope and Membrane proteins: structural differences linked to virus characteristics?" BioMed Research International 2020, 4389089. DOI: 10.1155/2020/4389089

6. Domenico Benvenuto, Ayse Banu Demir, Marta Giovanetti, **Martina Bianchi**, Silvia Angeletti, Stefano Pascarella, Roberto Cauda, Massimo Ciccozzi, Antonio Cassone "Evidence for mutations in SARS-CoV-2 Italian isolates potentially affecting virus transmission" J Med Virol. 2020; 1-6. https://doi.org/10.1002/jmv.26104

7. Prezioso C, **Bianchi M**, Obregon F, Ciotti M, Sarmati L, Andreoni M, Palamara AT, Pascarella S, Moens U, Pietropaolo V. Structural Analysis of Merkel Cell Polyomavirus (MCPyV) Viral Capsid Protein 1 (VP1) in HIV-1 Infected Individuals. Int J Mol Sci. 2020 Oct 27;21(21):7998. doi: 10.3390/ijms21217998. PMID: 33121182; PMCID: PMC7663277.

8. **Bianchi M**, Borsetti A, Ciccozzi M, Pascarella S. SARS-Cov-2 ORF3a: Mutability and function. Int J Biol Macromol. 2021 Feb 15; 170:820-826. doi: 10.1016/j.ijbiomac.2020.12.142. Epub 2021 Jan 8. PMID: 33359807; PMCID: PMC7836370.

9. Zella D, Giovanetti M, Cella E, Borsetti A, Ciotti M, Ceccarelli G, D'Ettorre G, Pezzuto A, Tambone V, Campanozzi L, Magheri M, Unali F, **Bianchi M**, Benedetti F, Pascarella S, Angeletti S, Ciccozzi M. The importance of genomic analysis in cracking the coronavirus pandemic. Expert Rev Mol Diagn. 2021 Jun;21(6):547-562. doi: 10.1080/14737159.2021.1917998. Epub 2021 Apr 28. PMID: 33849359; PMCID: PMC8095159.

10. Pascarella S, Ciccozzi M, Zella D, **Bianchi M**, Benedetti F, Benvenuto D, Broccolo F, Cauda R, Caruso A, Angeletti S, Giovanetti M, Cassone A. SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate? J Med Virol. 2021 Jul 14:10.1002/jmv.27210. doi: 10.1002/jmv.27210. Epub ahead of print. PMID: 34260088; PMCID: PMC8426736.

11. Ferrucci V, Kong DY, Asadzadeh F, Marrone L, Boccia A, Siciliano R, Criscuolo G, Anastasio C, Quarantelli F, Comegna M, Pisano I, Passariello M, Iacobucci I, Monica RD, Izzo B, Cerino P, Fusco G, Viscardi M, Brandi S, Pierri BM, Borriello G, Tiberio C, Atripaldi L, **Bianchi M**, Paolella G, Capoluongo E, Castaldo G, Chiariotti L, Monti M, De Lorenzo C, Yun KS, Pascarella S, Cheong JH, Kim HY, Zollo M. Long-chain polyphosphates impair SARS-CoV-2 infection and replication. Sci Signal. 2021 Jul

6;14(690): eabe5040. doi: 10.1126/scisignal.abe5040. PMID: 34230209; PMCID: PMC8432949.

12. Pascarella, S.; **Bianchi, M.**; Giovanetti, M.; Benvenuto, D.; Borsetti, A.; Cauda, R.; Cassone, A.; Ciccozzi, M. The Biological Properties of the SARS-CoV-2 Cameroon Variant Spike: An Intermediate between the Alpha and Delta Variants. Pathogens 2022, 11, 814. https://doi.org/10.3390/pathogens11070814

13. Pascarella S, Ciccozzi M, **Bianchi M**, Benvenuto D, Giovanetti M, Cauda R, Cassone A. Shortening Epitopes to Survive: The Case of SARS-CoV-2 Lambda Variant. Biomolecules. 2021 Oct 10;11(10):1494. doi: 10.3390/biom11101494. PMID: 34680128; PMCID: PMC8533401.

14. Pascarella S, Ciccozzi M, **Bianchi M**, Benvenuto D, Cauda R, Cassone A. The electrostatic potential of the Omicron variant spike is higher than in Delta and Delta-plus variants: A hint to higher transmissibility? J Med Virol. 2022 Apr;94(4):1277-1280. doi: 10.1002/jmv.27528. Epub 2021 Dec 27. PMID: 34914120.

15. Pascarella S, **Bianchi M**, Giovanetti M, Narzi D, Cauda R, Cassone A, Ciccozzi M. The SARS-CoV-2 Mu variant should not be left aside: It warrants attention for its immuno-escaping ability. J Med Virol. 2022 Jun;94(6):2479-2486. doi: 10.1002/jmv.27663. Epub 2022 Feb 28. PMID: 35174519; PMCID: PMC9088528.

16. Pascarella S, Ciccozzi M, **Bianchi M**, Benvenuto D, Cauda R, Cassone A. The value of electrostatic potentials of the spike receptor binding and N-terminal domains in addressing transmissibility and infectivity of SARS-CoV-2 variants of concern. J Infect. 2022 May;84(5):e62-e63. doi: 10.1016/j.jinf.2022.02.023. Epub 2022 Feb 23. PMID: 35218789; PMCID: PMC8864947.

17. Boccia A, Tufano R, Ferrucci V, Sepe L, **Bianchi M**, Pascarella S, Zollo M, Paolella G. SARS-CoV-2 Pandemic Tracing in Italy Highlights Lineages with Mutational Burden in Growing Subsets. Int J Mol Sci. 2022 Apr 9;23(8):4155. doi: 10.3390/ijms23084155. PMID: 35456974; PMCID: PMC9029933.

ACKNOWLEDGMENTS

APPENDIX

SCIENTIFIC PUBLICATIONS MOST CITED

RESEARCH ARTICLE

COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis

Silvia Angeletti[1] | **Domenico Benvenuto**[2] | **Martina Bianchi**[3] | **Marta Giovanetti**[4] | **Stefano Pascarella**[3] | **Massimo Ciccozzi**[2]

[1]Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome, Italy

[2]Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy

[3]Department of Biochemical Sciences "A. Rossi Fanelli", University of Rome "La Sapienza", Rome, Italy

[4]Flavivirus Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

Correspondence

Silvia Angeletti, Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, 00128 Rome, Italy.
Email: s.angeletti@unicampus.it

Abstract

Last December 2019, a new virus, named novel Coronavirus (COVID-2019) causing many cases of severe pneumonia was reported in Wuhan, China. The virus knowledge is limited and especially about COVID-2019 pathogenesis. The Open Reading Frame 1ab (ORF1ab) of COVID-2019 has been analyzed to evidence the presence of mutation caused by selective pressure on the virus. For selective pressure analysis fast-unconstrained Bayesian approximation (FUBAR) was used. Homology modelling has been performed by SwissModel and HHPred servers. The presence of transmembrane helical segments in Coronavirus ORF1ab non structural protein 2 (nsp2) and nsp3 was tested by TMHMM, MEMSAT, and MEMPACK tools. Three-dimensional structures have been analyzed and displayed using PyMOL. FUBAR analysis revealed the presence of potential sites under positive selective pressure ($P < .05$). Position 723 in the COVID-2019 has a serine instead a glycine residue, while at aminoacidic position 1010 a proline instead an isoleucine. Significant ($P < .05$) pervasive negative selection in 2416 sites (55%) was found. The positive selective pressure could account for some clinical features of this virus compared with severe acute respiratory syndrome (SARS) and Bat SARS-like CoV. The stabilizing mutation falling in the endosome-associated-protein-like domain of the nsp2 protein could account for COVID-2019 high ability of contagious, while the destabilizing mutation in nsp3 proteins could suggest a potential mechanism differentiating COVID-2019 from SARS. These data could be helpful for further investigation aimed to identify potential therapeutic targets or vaccine strategy, especially in the actual moment when the epidemic is ongoing and the scientific community is trying to enrich knowledge about this new viral pathogen.

KEYWORDS

epidemiology, infection, pandemics, pathogenesis, protein-protein interaction analysis, research and analysis methods

1 | INTRODUCTION

A novel Coronavirus, the COVID-2019, first appeared in Wuhan, China, last December 2019, spreading in other provinces/regions of China and in many countries other continents.[1] The epidemic originated probably from bat after viral mutation in the spike glycoprotein, as recently suggested,[2] began human-to-human transmission. The rapid spread of epidemic generated fear leading China authorities to restrict people movement to and from Wuhan in China, where the first start of epidemic was reported. As of 12 February

2020, 45 206 cases have been documented with 44 687 cases in Mainland China, including 1117 deaths and 5123 recovered https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6. The emergence of such a novel, highly virulent pathogen warrants rapid investigation of its etiology and evolution to control the impact on human health.

Knowledge about COVID-2019 is still incomplete, many questions have raised and many answers are needed first of all regarding its pathogenicity, its ability to change, how many people will get sick from each infected person, the so-called R0 and when infection will be preventable or treatable.[3] In the last period where many researchers are intensively studying the mechanism of COVID-2019 replication, pathogenicity, and therapeutic strategies, the present study has been realized. The aim was to provide information about how quickly the virus could potentially increase its genetic variability, with important implications for disease progression and drug or vaccine development. At this aim the Open Reading Frame 1ab (ORF1ab) of COVID-2019 has been analyzed to evidence the presence of mutation caused by selective pressure on the virus and their influence on viral ability to infect human host promoting epidemic spread.

2 | MATERIALS and METHODS

2.1 | Sequence dataset

The ORF1ab of 15 COVID-2019 sequences have been downloaded from GISAID (https://www.gisaid.org/) and GenBank (http://www.ncbi.nlm.nih.gov/genbank/) databanks. A dataset has been built using the five sequences of the severe acute respiratory syndrome (SARS) virus and five sequences from Bat SARS-like virus sharing the highest sequence similarity to the COVID-2019 sequence (Table 1). The pairwise percentage of similarity has been calculated using Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi); duplicated sequences have been removed from the dataset. The 25 sequences have been aligned using a multiple sequence alignment multiple alignment using fast fourier transform online tool[4] and manually edited using Bioedit program v7.0.5.[5]

2.2 | Selective pressure analysis

The selective pressure analysis was focused on the polyprotein ORF1ab because it differs from the most similar bat *Coronavirus* (QHR63299) for only 103 amino acid residues, 64 of them are conservative changes. In particular, non structural protein 2 (nsp2) differs from bat *Coronavirus* for 11 residues while nsp3 for 64 residues of which 44 are conservative changes.

Adaptive Evolution Server (http://www.datamonkey.org/) was used to find eventual sites under of positive or negative selection pressure. At this purpose the following tests has been used: fast-unconstrained Bayesian approximation (FUBAR).[6] These tests allowed to infer the site-specific pervasive selection, the episodic diversifying selection across the region of interest and to identify

episodic selection at individual sites.[7] Statistically significant positive or negative selection was based on $P < .05$.

2.3 | Structural modelling

Homology modelling has been attempted with SwissModel[8] and HHPred[9] servers. Models for ORF1ab nsp2 and nsp3 proteins available at the I-Tasser web site (corresponding to codes QHD43415_2 and QHD43415_3)[10] have been considered. PDB Proteins structurally close to the target have been evaluated using the TM-score[11] while the RAMPAGE[12] online tool has been used to assess the folding quality of the model.

To test for the presence of transmembrane helical segments in Coronavirus ORF1ab nsp2 and nsp3, TMHMM,[13] MEMSAT,[14] and

TABLE 1 Accession numbers, virus type, and the archive where they have been taken from

Accession number	Virus	Sequences archive
EPI_ISL_403933	2019-nCoV	GISAID
EPI_ISL_403934	2019-nCoV	GISAID
EPI_ISL_403936	2019-nCoV	GISAID
EPI_ISL_403962	2019-nCoV	GISAID
EPI_ISL_402132	2019-nCoV	GISAID
EPI_ISL_402130	2019-nCoV	GISAID
EPI_ISL_404895	2019-nCoV	GISAID
EPI_ISL_404253	2019-nCoV	GISAID
EPI_ISL_402125	2019-nCoV	GISAID
EPI_ISL_402124	2019-nCoV	GISAID
EPI_ISL_403930	2019-nCoV	GISAID
EPI_ISL_402120	2019-nCoV	GISAID
EPI_ISL_402129	2019-nCoV	GISAID
EPI_ISL_404228	2019-nCoV	GISAID
EPI_ISL_403931	2019-nCoV	GISAID
MG772933.1	Bat SARS-like	GeneBank
KY417146.1	Bat SARS-like	GeneBank
KT444582.1	Bat SARS-like	GeneBank
KY417147.1	Bat SARS-like	GeneBank
DQ084199.1	Bat SARS-like	GeneBank
AY559093.1	SARS	GeneBank
JX163925.1	SARS	GeneBank
GU553365.1	SARS	GeneBank
JQ316196.1	SARS	GeneBank
AY714217.1	SARS	GeneBank

Abbreviations: 2019-nCoV, novel coronavirus; SARS, severe acute respiratory syndrome.

MEMPACK[15] online tools have been used. Three-dimensional structures have been analyzed and displayed using PyMOL.[16]

3 | RESULTS

3.1 | Selective pressure analysis

Regarding the FUBAR analysis performed on the ORF1ab region, the presence of potential sites under positive selective pressure have been found ($P < .05$), in particular: on the amino acidic position 501 the COVID-2019 has a glutamine residue, the Bat SARS-like coronavirus has a threonine residue and the SARS virus has an alanine residue. At position 723 in the COVID-2019 there is a serine residue while the Bat SARS-like virus and the SARS virus have a glycine residue. On the aminoacidic position 1010, the COVID-2019 has a proline residue, the Bat SARS-like coronavirus has a histidine residue and the SARS virus has an isoleucine residue. Significant ($P < .05$) pervasive negative selection in 2416 sites (55%) has been evidenced and confirmed by FUBAR analysis.

3.2 | Structural modelling

To map the structural variability of the ORF1ab region of the virus and its sites under selection pressure, homology modelling has been attempted. Unfortunately, neither SwissModel nor HHPred found suitable templates for the amino acid region containing the sites under selective pressure. For that reason, the corresponding models available on the I-Tasser web site has been used. Moreover, some regions of the nsp2 and nsp3 proteins structurally homologous to other known viral proteins have been identified through HHpred analysis and have been mapped within the ORF1ab nsp2 and nsp3 sequences (Figure 1).

The results of the analysis suggest the presence of a segment within the nsp2 and the nsp3 regions that has no evident homologous structures. In an attempt to structurally characterize as far as possible these regions, TMHMM, MEMSAT, and MEMPACK analyses have been utilized and have shown the presence of several potential trans-membrane helices (Figure 1). In particular, our transmembrane helices were predicted by MEMSAT in nsp2 while six helices were predicted by MEMSTA and TMHMM in nsp3 (Figure 1).

Referring to the amino acids under positive selective pressure found using the FUBAR analysis: the amino acid in position 501 (position 321 of the nsp2 protein), the corresponding site in the Bat SARS-like coronavirus has an apolar amino acid while the SARS and COVID-2019 has a polar amino acid. It can be speculated, that due to its side chain length, polarity, and potential to form H-bonds the glutamine amino acid may confer higher stability to the protein. The mutations fall within the protein nsp2 on the region homologous to the endosome-associated protein similar to the avian infectious bronchitis virus (PDB 3ld1) that plays a key-role in the viral
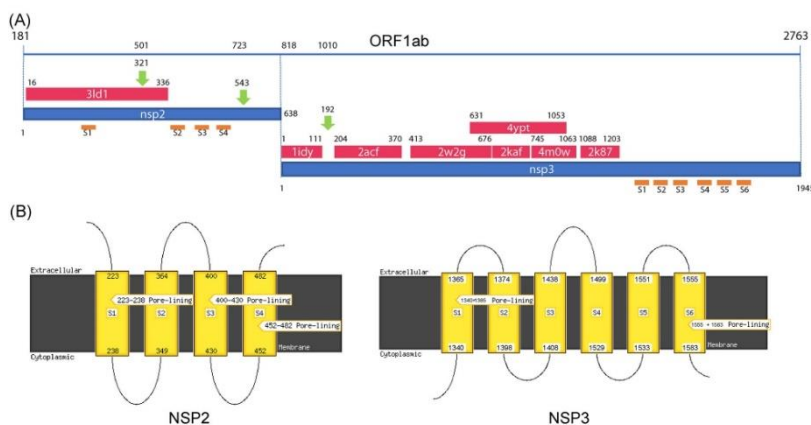

FIGURE 1 A, HHpred mapping of the homologous structures onto the ORF1ab sequence shown as a blue line on the top of the panel. Numbering above the line refers to the entire ORF. Red and Blue strips represent the PDB homologous structures and the nsp2 and nsp3 sequences, respectively. PDB codes are reported within the corresponding red stripes. Numbering below the blue line is relative to each single nsp. Orange lines indicate approximately the positions of the transmembrane helices predicted by MEMSAT. Label refers to panel B. B, diagram of the topology of predicted transmembrane helices. Number refer to the corresponding nsp sequences. nsp, non structural protein; ORF, open reading frame
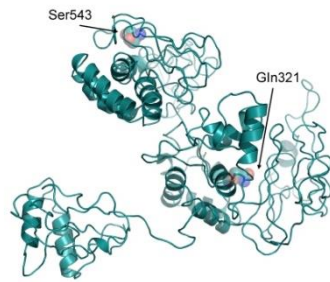
III

FIGURE 2 I-Tasser model of the COVID-2019 nsp2. Residues under positive selective pressure with a $P < .05$ are shown as sticks and transparent spheres and are marked by the corresponding labels. COVID-2019, novel Coronavirus; nsp2, non structural protein-2

FIGURE 3 I-Tasser model of the COVID-2019 nsp3. The residue under positive selective pressure with a $P < .05$ is shown as sticks and transparent spheres and is marked by the corresponding label. COVID-2019, novel Coronavirus; nsp, non structural protein

pathogenicity. (Figure 2) In the nsp2 structure model available at the I-Tasser site, this position appears to be exposed to the solvent.

As for the residue in position 723 (543 in the nsp3 protein), the COVID-2019 sequence displays a Ser replacing for Gly in Bat SARS-like and SARS coronaviruses. In this case, it may be argued that this substitution could increase local stiffness of the polypeptide chain both for steric effect (at variance with Ser, Gly has no side chain) and for ability of Ser side chain to form H-bonds. Moreover, Ser can act as a nucleophile in determined structural environments, such as those of enzyme active sites. Within the I-Tasser model, this position is predicted to have a low solvent accessibility (Figure 2).

Regarding the amino acid in position 1010 (corresponding to position 192 of the nsp3 protein), the homologous region of the Bat SARS-like coronavirus and SARS virus have a polar and an apolar amino acid, respectively, while the COVID-2019 has proline. In this case, it may be speculated that due to the steric bulge and stiffness of the proline, the molecular structure of the COVID-2019 may undergo a local conformation perturbation compared with the proteins of the other two viruses. In Nsp3, the mutation falls near the protein similar to a phosphatase present also in the SARS coronavirus (PDB code 2acf) playing a key-role in the replication process of the virus in infected cells[17] (Figure 3). In the I-Tasser model, the position is partially accessible to the solvent. It should be emphasized that all these considerations are speculative and they need to be substantiated by the availability of the experimental crystallographic structure of the corresponding proteins.

4 | DISCUSSION

The COVID-2019 ongoing epidemic is worrying worldwide for its high contagiosity. From its first appearance in Wuhan, China, about 1 month ago, the virus infected thousands people with new cases number rapidly growing every day. For this acceleration in

human-to-human transmission in China but with evident spreading also in other countries, World Health Organization declared the epidemic a global health emergency.[18,19]

Many questions are open and need an answer, of these the most frequent is how much this virus can be dangerous and how much it differs from SARS virus which epidemic scared all the world some years ago. In this study some interesting findings have been evidenced to support and fill gaps in knowledge about the new COVID-2019 that is still causing infection all over the world.[20,21]

The positive selective pressure in this protein could justify some clinical features of this virus compared with SARS and Bat SARS-like CoV.[22] First which are the probably most common sites undergoing to an aminoacidic change, providing an insight of some important proteins of the COVID-2019 that are involved in the mechanism of viral entry and viral replication. This data can contribute for a better understanding of how this virus acts in its pathogenicity. Furthermore, to identify a potential molecular target is fundamental to follow the molecular evolution of the virus suggesting some interesting sites for potential therapy or vaccine.

The structural similarity of the region in which falls the positive selective pressure as so as the stabilizing mutation falling in the endosome-associated-protein-like domain of the nsp2 protein, could explain why this virus is more contagious than SARS. The destabilizing mutation happening near the phosphatase domain of the nsp3 proteins could suggest a potential mechanism differentiating COVID-2019 from SARS.

The results of this study could fill some gaps about COVID-2019 knowledge especially in the actual moment when the epidemic is ongoing and the scientific community is trying to enrich knowledge about this new viral pathogen. During epidemic, all strength has to be done to enforce virus fight. This can be achieved by understanding

IV

the main drivers for pathogen appearance, spreading, and supremacy on human defense.

ORCID

Silvia Angeletti http://orcid.org/0000-0002-7393-8732
Domenico Benvenuto http://orcid.org/0000-0003-3833-2927
Massimo Ciccozzi http://orcid.org/0000-0003-3866-9239

REFERENCES

1. Hui DS, I Azhar E, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis.* 2020;91:264-266.
2. Benvenuto D, Giovannetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. *J Med Virol.* 2020;92(4):455-459. https://doi.org/10.1002/jmv.25688
3. Liu J, Zheng X, Tong Q, et al. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. *J Med Virol.* 2020. https://doi.org/10.1002/jmv.25709
4. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019;20(4):1160-1166.
5. Hall TA. BioEdit A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95-98.
6. Murrell B, Moola S, Mabona A, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* 2013;30(5):1196-1205.
7. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012;8(7):e1002764.
8. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-W303.
9. Zimmermann L, Stephens A, Nam SZ, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol.* 2018;430(15):2237-2243.
10. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. *Nature Methods.* 2015;12(1):7-8.
11. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins.* 2004;57(4):702-710.
12. Lovell SC. Structure validation by Calpha geometry: phi, psi, and Cbeta deviation. *Proteins.* 2003;50(3):437-450.
13. Moller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics.* 2001;17(7):646-653.
14. Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics.* 2009;10:159.
15. Nugent T, Ward S, Jones DT. The MEMPACK alpha-helical transmembrane protein structure prediction server. *Bioinformatics.* 2011;27(10):1438-1439.
16. Schrödinger LLC. *The [PyMOL] Molecular Graphics System.* Version 1.80 LLC, New York, NY. 2015.
17. Saikatendu KS, Joseph JS, Subramanian V, et al. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1''-phosphate dephosphorylation by a conserved domain of nsP3. *Structure.* 2005;13(11):1665-1675.
18. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395(10223):P497-P506. https://doi.org/10.1016/S0140-6736(20)30183-5
19. Zhu N, Zhang D, Wang W, et al. China novel coronavirus investigating and research team. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020;382:727-733. https://doi.org/10.1056/NEJMoa2001017
20. The Lancet. Emerging understandings of 2019-nCoV. *Lancet.* 2020;395(10221):311.
21. Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019-nCoV in Italy: where they come from? *J Med Virol.* 2020. https://doi.org/10.1002/jmv.25699
22. Guarner J. Three emerging coronaviruses in two decades. *Am J Clin Pathol.* 2020:aqaa029. https://doi.org/10.1093/ajcp/aqaa029

V

Research Article

Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics?

Martina Bianchi [ID],[1] Domenico Benvenuto [ID],[2] Marta Giovanetti [ID],[3] Silvia Angeletti [ID],[4] Massimo Ciccozzi [ID],[2] and Stefano Pascarella [ID][1]

[1]Department of Biochemical sciences "A Rossi Fanelli", Sapienza University of Rome, 00185 Rome, Italy
[2]Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy
[3]Flavivirus Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
[4]Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome, Italy

Correspondence should be addressed to Stefano Pascarella; stefano.pascarella@uniroma1.it

The Coronavirus Disease 2019 (COVID-19) is a new viral infection caused by the severe acute respiratory coronavirus 2 (SARS-CoV-2). Genomic analyses have revealed that SARS-CoV-2 is related to Pangolin and Bat coronaviruses. In this report, a structural comparison between the Sars-CoV-2 Envelope and Membrane proteins from different human isolates with homologous proteins from closely related viruses is described. The analyses here reported show the high structural similarity of Envelope and Membrane proteins to the counterparts from Pangolin and Bat coronavirus isolates. However, the comparisons have also highlighted structural differences specific of Sars-CoV-2 proteins which may be correlated to the cross-species transmission and/or to the properties of the virus. Structural modelling has been applied to map the variant sites onto the predicted three-dimensional structure of the Envelope and Membrane proteins.

1. Introduction

COVID-19 has become a planetary emergency which is seriously threatening human health [1, 2]. Development of effective therapeutic and prevention strategies is significantly hampered also by the lack of detailed structural information on virus proteins, although several crystallographic structures of Sars-CoV-2 proteins are now available [3–5]. In this report, a structural comparison between the Sars-CoV-2 surface proteins from different isolates with homologous proteins from closely related viruses such as those from Bat and Pangolin is described. This work has been focussed onto the Envelope (E) and Membrane (M) proteins that, along with the Spike, form the virus protein interface to the external environment. The Spike glycoprotein has been already extensively studied, and a few crystallographic structures are available in the Protein Data Bank [3–6]; consequently, this protein has not been specifically addressed within this note. Identification of local structural differences, even

minimal, to the closest virus proteins may indicate the mutations that enabled Sars-CoV-2 to cross species and/or to acquire its peculiar pathogenic properties [7, 8]. Indeed, a number of examples have been reported in the scientific literature suggesting how even single point mutations in virus proteins can significantly alter their biology and pathogenesis [9, 10]. Therefore, comparative studies may shed light on the molecular mechanisms through which an epidemic of epizootic origin can emerge and may also suggest molecular targets for therapeutics or reverse vaccinology experiments.

2. Material and Methods

Nucleotide and protein sequences have been taken from GenBank [11] data repository. Blast suite [12] has been used for databank searches; Jalview [13] and MAFFT [14] have been used for multiple sequence display and alignment, respectively. Transmembrane helix prediction has been obtained by TMHMM [15], MEMSAT [16], and Protter
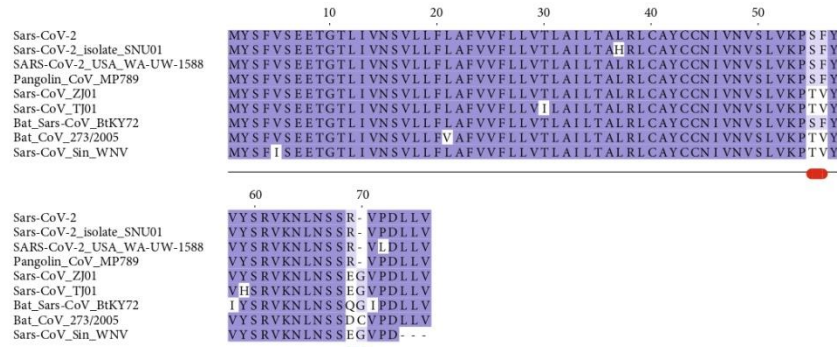
FIGURE 1: Multiple sequence alignment among Sars-CoV-2 Envelope protein variants and a set of the most similar homologous proteins. The sequence labelled Sars-CoV-2 corresponds to the reference sequence identified by the RefSeq code YP_009724392. Red lines below the alignment indicate the changed sites discussed in the text. Blu background denotes conserved alignment positions.

[17]. Cd-hit program [18] has been used for sequence clustering. Homology modelling relied on Swiss-Model [19], Modeller [20], or HHpred [21] and structure display and analysis on Open-Source PyMOL [22]. When necessary, I-Tasser [23] has been used as an alternative source of ab initio homology models.

3. Results

3.1. Databank Searches and Structure Modelling. From the GenBank repository, 797 complete genomes of Sars-CoV-2 have been collected (the full list is reported in Supplementary Data). The TblastN program has been used to extract the sequences of E and M proteins from each genome. To remove redundancy within each E and M protein set, cd-hit clustering has been applied at 100% sequence identity level: identical sequences have been assigned to the same group for which only one representative has been considered for further analysis. The Sars-CoV-2 E and M protein sets have been grouped into three and seven clusters, respectively. This finding suggests that within the 797 genomes three and seven variants of the E and M proteins can be observed, respectively. E and M homologous proteins from closely related virus have been retrieved from the GenBank using the TblastN tool.

3.2. Envelope Protein. The E protein is conserved across β-coronaviruses. Only three variants have been found in the Sars-CoV-2 E set collected. Sequence comparisons show that the Sars-CoV-2 E protein from the reference genome (RefSeq code YP_009724392) is identical to the sequences from Pangolin CoV MP798 and Bat CoV CoVZXC21, CoVZC45, and RaTG13 isolates. The multiple sequence alignment reported in Figure 1 demonstrates that a distinguishing feature of Sars-2-CoV E variants is the presence of

Arg at position 69 that substitutes Glu, Gln, Asp in other homologous Sars-CoV E proteins. This site is followed by a deletion in position 70 corresponding to Gly or Cys in the other proteins. Sars-CoV-2 E sequences differ from the homologous proteins also at positions 55-56, where the dyad Ser-Phe replaces Thr-Val (except in Bat coronavirus isolate BtKY72, accession code KY352407). Variants of the Sars-CoV-2 E protein differ at positions 37 and 72 where His substitutes a Leu and Leu replaces a conserved Pro, respectively. The size of each Envelope variant cluster is reported in Table 1 along with accession codes and definitions of the isolates. A homology model of the E protein has been built with Modeller using as a template the pentameric ion channel structure of Sars-CoV protein identified by the PDB code 5X29. This sequence shares 91% identity to Sars-CoV-2 E protein and covers the segment encompassed by positions 8-65. Figure 2 displays the structure of the homology model of the Sars-CoV-2 E protein assembled as a pentameric viroporin-like protein. Figure 2 displays also the position of the variant sites onto the three-dimensional model. Prediction of the transmembrane helices is difficult in a short protein. Therefore, transmembrane topology cannot be assigned reliably. Likewise, experiments have not clarified definitively which portions of the E protein are exposed to the external or internal side of the virus membrane [24].

3.3. Membrane Glycoprotein. The M glycoprotein is conserved across the β-coronaviruses. However, seven variants of Sars-CoV-2 M protein were identified in the collected set, while only three variants were observed for the E protein (Figure 3). The multiple sequence alignment shows a remarkable similarity (98% identity) among the Sars-CoV-2 M variants and the sequences from Bat and Pangolin isolates. However, a difference at the N-terminal position (Figure 3) can be observed: the insertion of a Ser residue at position 4

TABLE 1: Size of the variant clusters of the Sars-CoV-2 Envelope and Membrane proteins.

| Variant	Cluster size (no. of sequences)	Accession code	Definition
Envelope			
YP_009724392 (reference)	795		
His37	1	MT03980	Korea/SNU01/2020
Leu72	1	MT293206	USA/WA-UW-1588/2020
Membrane			
YP_009724393 (reference)	773		
Ser2	1	MT291836	CHN/Wuhan_IME-BJ07/2020
Gly3	1	MT325626	USA/SC_3572/2020
Val57, Arg89	1	MT127115	VIE/NIHE/2020
		MT293184	USA/WA-UW-1297/2020
Ile70	3	MT326166	USA/WA-UW-1735/2020
		MT293211	USA/WA-UW-1591/2020
Ser85	1	MT326167	USA/WA-UW-1753/2020
		MT326093	USA/WA-UW-1775/2020
Met175	2	MT246451	USA/WA-UW-194/2020

(a)   (b)
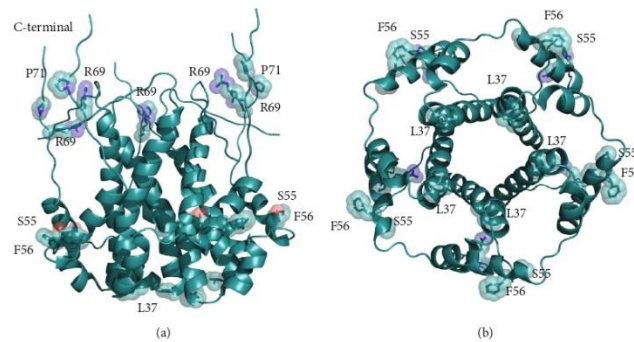
FIGURE 2: Three-dimensional model of the viroporin-like tetrameric assembly of the E protein from Sars-CoV-2 represented as a cartoon model. Residues corresponding to the mutated sites indicated in Figure 1 are displayed as transparent space-filling spheres and labelled with the amino acid one-letter code. The C-terminal segments of the model are reported for completeness. However, they convey no structural information due to lack of a corresponding segment in the structural template used in homology modelling. Structure in panel (b) is rotated by approximately 180° along the x axis with respect to the orientation shown in panel (a).

of human Sars-CoV-2 seems to be a unique feature of this protein. In the corresponding position, the RaTG13 Bat M protein displays a deletion, while Bat CoVZXC21, CoVZC45, and Pangolin MP789 proteins have an Asp residue. The seven M protein variants differ at positions 2, 3, 57, 70, 85, 89, and 175. The size of each Membrane variant cluster is reported in Table 1 along with accession codes and definitions of the isolates. Noteworthy, the protein from the Sars-CoV-2 NIHE isolate (accession code MT127115) possesses an Arg instead of a conserved Gly at position 89 (Figure 3). The mutation occurs within a predicted transmembrane

helix and, if confirmed, may have a significant impact on the protein properties (Figure 3).

The three-dimensional model of the M protein has been taken from the I-Tasser server (code QHD43419) as other methods failed to find any suitable template. However, it should be mentioned that HHpred found a weak local affinity, albeit below the statistical significance level, to 4N31, a peptidase-like protein from *Streptococcus pyogenes* essential for pilus polymerisation. Figure 4 displays the positions of the variant sites onto the model structure. This model has been predicted by ab initio techniques. Therefore, it should
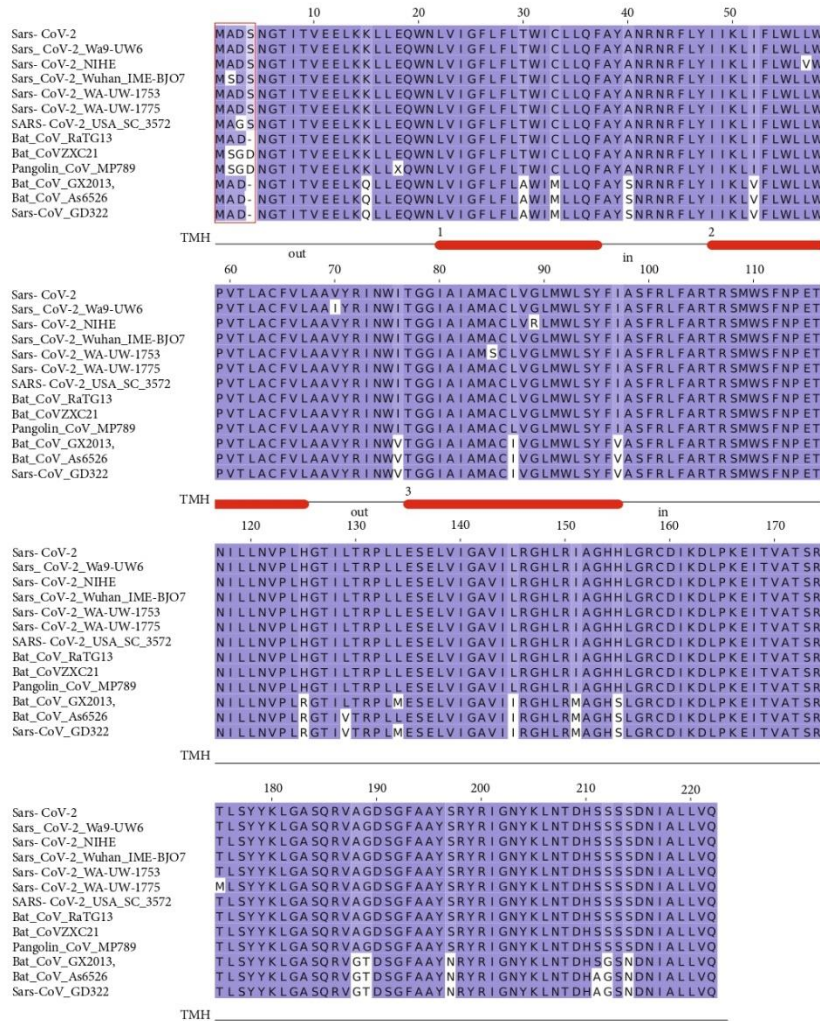
FIGURE 3: Multiple sequence alignment among Sars-CoV-2 M protein variants and a set of most similar homologous proteins. The sequence label Sars-CoV-2 indicates the reference sequence identified by the RefSeq code YP_009724393. Red box indicates the variant sites at the N-terminal discussed in the text. Numbered red bars under the multiple alignment mark the prediction of transmembrane helices. The location of the connect loop with respect to the virion surface is indicated as "in" or "out". Blu background denotes conserved alignment positions.

IX

FIGURE 4: I-Tasser model of the Membrane protein represented as cartoon model. Variant positions are displayed as transparent space-filling spheres and labelled with the amino acid one-letter code.
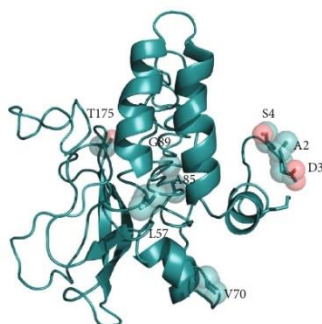
be considered with great caution and should be treated as a low-resolution approximation of the real structure. According to the prediction of the transmembrane helix topology, the N- and C-terminal portions of the M protein are exposed outside and inside the virus particle, respectively (Figure 4).

4. Discussion

Previous studies pointed out that E and M proteins could be important for viral entry, replication, and particle assembly within the human cells [24, 25]. According to the accepted theories, the current COVID-19 pandemic has been caused by the cross-species transmission of a β-coronavirus normally hosted by Bats and, perhaps, Pangolin to humans [3, 26]. In this paper, E and M proteins from 797 Sars-CoV-2 genomes have been compared to the counterparts taken from the most closely related virus also to evaluate the potential role of amino acid mutations in the epizootic origin of COVID-19. E protein is a minor component of the virus membrane though it is deemed to be important for many stages of virus infection and replication [24, 25]. Sequence comparison has shown that this protein is identical to the counterparts of specific Bat and Pangolin coronavirus isolates, even though the Sars-CoV-2 sequence seems to possess specific modifications and characteristics with respect to other Sars CoVs. In particular, Arg69, a positively charged amino acid, replaces Glu or Gln residues, negatively charged and neutral, respectively, in the homologous CoV proteins. Moreover, a deletion specific to Sars-CoV-2 proteins flanks this position. Unfortunately, it is not possible to predict reliably whether the sites of these modifications are exposed to the internal or external side of the membrane. In any case, the substitution and the deletion appear a rather drastic change and may have a significant impact on conformational properties and possibly on protein-protein interactions. Further structural studies are needed. However, it may be hypothesized that these changes can also affect the olig-

omerization process necessary to form a transmembrane ion channel.

It has been demonstrated that M protein is more prevalent within the virus membrane, and it is deemed to be important for the budding process of coronaviruses. Indeed, during the process of virus particle assembly, this protein interacts with the Nucleocapsid, Envelope, Spike, and Membrane glycoprotein itself [25]. Moreover, in Alphacoronaviruses, it has been demonstrated that this protein cooperates with the Spike during the cell attachment and entry [27]. Therefore, mutations occurring at the N-terminus region, which is exposed to the virus surface, could play a key role in the host cell interaction.

In conclusion, the analyses here reported show the structural similarity of E and M proteins to the counterparts from Pangolin and Bat coronavirus isolates. At the same time, comparisons have highlighted structural differences specific of Sars-CoV-2 proteins which may be correlated to the cross-species transmission and/or to the properties of the virus. Although further studies are needed, it is clear that these amino acid variations have been important for the virus evolutionary history, and the results may hint at how similar mutations within the coronavirus family can lead in the next years to other epizootic epidemic events similar to the one that we are experiencing these days.

Data Availability

All sequence data are available in the GenBank repository. The complete list is available in the Supplementary Materials.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

Supplementary Materials

The supplementary data consist of an Excel file containing the list of the 797 Sars-CoV-2 genomes retrieved from GenBank and analysed in the article. For each genome, the sequence header is displayed as found in the GenBank. The header reports the accession code in the first field of each line after the ">" character. *(Supplementary Materials)*

References

[1] D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, and M. Ciccozzi, "The 2019-new coronavirus epidemic: evidence for virus evolution," *Journal of Medical Virology*, vol. 92, no. 4, pp. 455–459, 2020.

[2] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus Disease-2019 (COVID-19): the epidemic and the challenges," *International Journal of Antimicrobial Agents*, vol. 55, no. 3, article 105924, 2020.

[3] D. Benvenuto, M. Giovanetti, M. Salemi et al., "The Global Spread of 2019-NCoV: A Molecular Evolutionary Analysis," *Pathogens and Global Health*, vol. 114, no. 2, pp. 64–67, 2020.

[4] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell*, vol. 181, no. 2, pp. 281–292.e6, 2020.

[5] D. Wrapp, N. Wang, K. S. Corbett et al., "Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation," *Science*, vol. 367, no. 6483, pp. 1260–1263, 2020.

[6] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, "Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2," *Science*, vol. 367, no. 6485, pp. 1444–1448, 2020.

[7] S. Angeletti, D. Benvenuto, M. Bianchi, M. Giovanetti, S. Pascarella, and M. Ciccozzi, "COVID-2019: the role of the Nsp2 and Nsp3 in its pathogenesis," *Journal of Medical Virology*, vol. 92, no. 6, pp. 584–588, 2020.

[8] W. Ji, W. Wang, X. Zhao, J. Zai, and X. Li, "Cross-species transmission of the newly identified coronavirus 2019-NCoV," *Journal of Medical Virology*, vol. 92, no. 4, pp. 433–440, 2020.

[9] N. M. André, B. Cossic, E. Davies, A. D. Miller, and G. R. Whittaker, "Distinct mutation in the feline coronavirus spike protein cleavage activation site in a cat with feline infectious peritonitis-associated Meningoencephalomyelitis," *Journal of Feline Medicine and Surgery Open Reports*, vol. 5, no. 1, 2019.

[10] Y. Sakai, K. Kawachi, Y. Terada, H. Omori, Y. Matsuura, and W. Kamitani, "Two-amino acids change in the Nsp4 of SARS coronavirus abolishes viral replication," *Virology*, vol. 510, pp. 165–174, 2017.

[11] D. A. Benson, M. Cavanaugh, K. Clark et al., "GenBank," *Nucleic Acids Research*, vol. 46, no. D1, pp. D41–D47, 2018.

[12] S. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.

[13] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, "Jalview version 2–a multiple sequence alignment editor and analysis workbench," *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009.

[14] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772–780, 2013.

[15] Y. Chen, P. Yu, J. Luo, and Y. Jiang, "Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT," *Mammalian Genome*, vol. 14, no. 12, pp. 859–865, 2003.

[16] L. J. McGuffin, K. Bryson, and D. T. Jones, "The PSIPRED protein structure prediction server," *Bioinformatics*, vol. 16, no. 4, pp. 404-405, 2000.

[17] U. Omasits, C. H. Ahrens, S. Müller, and B. Wollscheid, "Protter: interactive protein feature visualization and integration with experimental proteomic data," *Bioinformatics*, vol. 30, no. 6, pp. 884–886, 2014.

[18] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "CD-HIT: accelerated for clustering the next-generation sequencing data," *Bioinformatics*, vol. 28, no. 23, pp. 3150–3152, 2012.

[19] A. Waterhouse, M. Bertoni, S. Bienert et al., "SWISS-MODEL: homology Modelling of protein structures and complexes," *Nucleic Acids Research*, vol. 46, no. W1, pp. W296–W303, 2018.

[20] B. Webb and A. Sali, "Protein structure modeling with MODELLER," *Methods in Molecular Biology*, vol. 1654, pp. 39–54, 2017.

[21] L. Zimmermann, A. Stephens, S.-Z. Nam et al., "A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core," *Journal of Molecular Biology*, vol. 430, no. 15, pp. 2237–2243, 2018.

[22] L. L. C. Schrodinger, *The AxPyMOL Molecular Graphics System, Version 2.0*, Schrödinger, LLC., 2015.

[23] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER suite: protein structure and function prediction," *Nature Methods*, vol. 12, no. 1, pp. 7-8, 2015.

[24] D. Schoeman and B. C. Fielding, "Coronavirus envelope protein: current knowledge," *Virology Journal*, vol. 16, no. 1, p. 69, 2019.

[25] E. A. J. Alsaadi and I. M. Jones, "Membrane binding proteins of coronaviruses," *Future Virology*, vol. 14, no. 4, pp. 275–286, 2019.

[26] R. Lu, X. Zhao, J. Li et al., "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *The Lancet*, vol. 395, no. 10224, pp. 565–574, 2020.

[27] A. Naskalska, A. Dabrowska, A. Szczepanski, A. Milewska, K. P. Jasik, and K. Pyrc, "Membrane protein of human coronavirus NL63 is responsible for interaction with the adhesion receptor," *Journal of Virology*, vol. 93, no. 19, pp. 455–459, 2019.

SARS-Cov-2 ORF3a: Mutability and function

Martina Bianchi [a], Alessandra Borsetti [b], Massimo Ciccozzi [c], Stefano Pascarella [a,*]

[a] Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, 00185 Rome, Italy
[b] National HIV/AIDS Research Center, Istituto Superiore di Sanità, Rome, Italy
[c] Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy

ABSTRACT

In this study, analysis of changes of SARS-CoV-2 ORF3a protein during pandemic is reported. ORF3a, a conserved coronavirus protein, is involved in virus replication and release. A set of 70,752 high-quality SARS-CoV-2 genomes available in GISAID databank at the end of August 2020 have been scanned. All ORF3a mutations in the virus genomes were grouped according to the collection date interval and over the entire data set. The considered intervals were: start of collection-February, March, April, May, June, July and August 2020. The top five most frequent variants were examined within each collection interval. Overall, seventeen variants have been isolated. Ten of the seventeen mutant sites occur within the transmembrane (TM) domain of ORF3a and are in contact with the central pore or side tunnels. The other variant sites are in different places of the ORF3a structure. Within the entire sample, the five most frequent mutations are V13L, Q57H, Q57H + A99V, G196V and G252V. The same analysis identified 28 sites identically conserved in all the genome isolates. These sites are possibly involved in stabilization of monomer, dimer, tetramerization and interaction with other cellular components. The results here reported can be helpful to understand virus biology and to design new therapeutic strategies.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Coronavirus Disease (COVID-19) became almost suddenly, though not unexpectedly, a serious threat to human health [1–3]. The etiological agent of the disease is the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2), an enveloped positive-sense RNA coronavirus with genome size approximately of 30,000 bases. Phylogenetic analysis has revealed that SARS-CoV-2 is distinct from SARS-CoV (79% sequence similarity) that in 2002 caused an outbreak of atypical and severe, often lethal, pneumonia in Guangdong province, China [4]. The coronaviruses are promiscuous and can be hosted by several species. The SARS-CoV-2 genome has about 96.2% and 91% sequence similarity with bat SARS-related coronavirus (SARS-CoV RaTG13) and pangolin CoV respectively, suggesting zoonotic origin of SARS-CoV-2 [5]. Indeed, it is has been proposed that the current pandemic has been ignited by a cross-species virus transmission from Pangolin and/or Bat to humans, at Wuhan, China [2,6–8]. However, the debate about this issue is still going on among the scientific community.

Like many viruses, the CoV evolves and adapts to the host through accumulation of synonymous and non-synonymous mutations [9] generated by several mechanisms including fidelity of RNA-dependent-RNA-polymerase [10]. It is known that even single mutations in specific proteins can change pathogenicity of these viruses [11,12].

In this context, it is useful to study the changes of the viral proteins of its proteome. Indeed, modification of specific virus proteins considered promising targets may put at risk the efficacy of drugs or vaccines. Moreover, study of the conserved/variable protein regions can provide structure-function hints that may help to determine the function of yet uncharacterized proteins.

Here the attention has been focussed onto the protein ORF3a as it is deemed to be involved in critical aspects of virus pathogenicity [13] and a three-dimensional structure has been recently made available by means of cryo-electron microscopy (cryo-EM) experiments [14]. ORF3a possesses an N-terminal, a transmembrane and a C-terminal domain folded as 8-strand β-barrel. ORF3a of both SARS-CoV and SARS-CoV-2 have been described to contain different functional domains linked to virulence, infectivity, and virus release [15].

In fact, ORF3a is a viroporin, an integral membrane protein able to function as an ion channel that may promote virus release [15–17]. Moreover, this protein interacts with caveolin potentially regulating different phase of viral cycle [18]. ORF3a presents also a TRAF3-binding motif that activates the NLRP3 inflammasome and it is a potent stimulator of pro–IL-1β gene transcription [19], and in animal models of SARS-CoV infection, genomic deletion of ORF3a reduced virus replication [20]. Importantly, significant CD4[+] and CD8[+] T cell responses to SARS-CoV-2 in infected individuals were directed against ORF3a [21].

* Corresponding author at: Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, 00185 Rome, Italy.
  E-mail address: Stefano.Pascarella@uniroma1.it (S. Pascarella).

Analysis of ORF3a nucleotide and protein sequences can predict their ability to alter viral cycle and therefore yields important insights into the biology of the virus.

In this study, a software workflow able to carry out a quick, systematic and repeatable screening of the SARS-CoV-2 genome isolates to detect protein mutations was utilized to scan 70,752 high-quality SARS-CoV-2 genomes available in GISAID databank [16] at the end of August 2020. Our aim was to identify ORF3a mutations over time and to assess the mutated amino acid residues identified as critical for protein activity and to gauge the likely effect of the changes.

The results of the screening suggest that ORF3a is hit by many mutations but only a few of them are observed with a frequency of at least 0.5%. Moreover, the same analysis pointed out the sites that apparently never mutated during the period considered and that can play crucial functional and structural roles. These indications help to prioritize experimental studies aimed at deciphering the function of ORF3a and assess it as a potential therapeutic target.

2. Materials and methods

The Refseq ORF3a protein denoted by code YP_009724391 has been taken as the reference (wild type) sequence. The collection of the ORF3a protein sequences coded by different SARS-CoV-2 genome isolates has been carried out using this workflow:

a. SARS-CoV-2 genome sequences have been downloaded as FASTA format from GISAID repository at www.gisaid.org [22]. Since the quality of the deposited sequences is not uniform, only complete sequences deposited with a high degree of coverage has been downloaded using the filters provided by the GISAID server.
b. The file containing the genomic sequences has been converted into a BLAST-formatted database with the "makeblastdb" tool [23].
c. The "tblastn" tool searches a protein sequence within a translated nucleotide sequence database. The reference ORF3a sequence has been used as a query to retrieve the other ORF3a coding sequences from the SARS-CoV-2 genomes. Incomplete sequences or sequences containing ambiguous codons (resulting in undetermined residues) have been eliminated. This step relied on the tools available in the EMBOSS suite [24] and on Linux bash shell scripts.
d. The clustering software "cd-hit" [25] has been applied to remove redundancies. Identical ORF3a sequences have been clustered and one representative sequence has been designated by the software. Each cluster contains all the sequences of one ORF3a variant. As a matter of fact, the ORF3a sequences belonging to different clusters differ for at least one residue.
e. The representative ORF3a variants have been multiply aligned to the reference protein with the program MAFFT [26].
f. A R script has been written within the Rstudio environment to scrutinize the multiple sequence alignments and collecting mutation statistics and for graphical output. The R script utilized input and output functions from the bio3d package [27].

Multiple sequence alignments display and editing relied on Jalview [28]. PyMOL and Chimera have been used for structure display and analysis. PyMOL plugin Caver 3.0.3 [29] has been utilized to study tunnels inside the protein structure. DynaMut [30] and Duet [31] have been used to predict the effect of point mutations on protein stability. Logos have been produced with the server WebLogo [32].

3. Results

ORF3a mutations were recorded in virus genomes grouped according to the collection date interval and over the entire data set. Intervals considered are indicated in Table 1: beginning of collection-February, March, April, May, June, July and August 2020. The total number of selected genomes was 70,752. In each time period, the number of all the

Table 1
Data set utilized.

Time collection interval	Total number of genomes	No. of different ORF3a variants
Start to Feb 2020	2257	68
March	21,521	356
April	18,316	436
May	8141	294
June	10,769	356
Jul	6338	233
Aug	3410	158

different ORF3a variants has been reported. Frequency is defined as the number of replicas of a single variant found in the data set considered. For example, if the ORF3a variant 1 is found 100 times in 1000 genomes collected, its frequency is 0.1.

3.1. Mutant sites

Attention was focussed onto the most prevalent mutations: only the top five most frequent variants have been considered within each collection interval. Overall, seventeen variants have been observed fulfilling this criterion (Table 2). Most of the ORF3a variants possess a single point mutation while four variants are distinguished by co-occurrence of two mutations, one of which is always Q → H in position 57 of the reference ORF3a sequence. Only the variant Q57H has a frequency constantly high in all the collection periods whereas the frequency of the other mutations fluctuates.

Availability of the three-dimensional structure of a large portion of the SARS-CoV-2 ORF3a protein, enables mapping of the mutations onto the structure and formulation of considerations on possible structural or functional effects.

Fig. 1 reports the ORF3a structure on which the positions of the most frequently observed mutations are indicated. Ten of the seventeen mutant sites occur within the transmembrane (TM) domain of ORF3a. Four of these variants contain the mutation Q57H paired with another amino acidic change (A99V, S58N, Y264C, G172V). In two cases, the associated mutations are in the extracellular domain (G172V and Y264C). The other seven mutations are found at the N-terminal intracellular portion or in the extracellular β-barrel domain (Table 3 and Fig. 1).

As shown by cryo-EM, the ORF3a dimer is characterized by a central pore in the transmembrane domain connected to six tunnels (three for each subunit) close to the barrel domain and that opens into the cytosolic space (Supplementary Material Fig. 1). Through application of Caver 3.0.3, the residues close to tunnels were identified. Among these, five mutant sites in the transmembrane portion lay in a position lining the central pore or the tunnels connecting it to the cytosolic compartment: L46F, A54S, Q57H, K75N and R126S (Table 3 and Fig. 1).

To test whether mutations at these sites may influence channel shape, the mutant sites of ORF3a were modelled by Chimera 1.14. The impact of these mutations on the tunnel geometry delineated by Caver 3.0.3 was visually assessed. No significant alteration of the shape of tunnels was detected.

However, the most frequent mutations observed from February to August (Table 2) related to the tunnel, may influence other properties of the ORF3a. For example, Q57H is ubiquitous and is consistently the most frequent ORF3a substitution described in the literature [33]. Although this site is lining the transmembrane tunnel, no significant modification of the channel geometry could be observed. This finding is coherent with the reported results of the site-directed mutagenesis experiments that demonstrated no alteration of the channel properties [14]. DynaMut attributes to this mutation a stabilizing effect while Duet predicts a marginal destabilization.

Similarly, L46F is a relatively rare change isolated mainly in India in June 2020 that creates an aromatic interaction at the

XIII

Table 2
Most frequent mutations observed in each time interval and in the entire data set and corresponding geographical location.

Position	Mutation	Start-Feb	March	April	May	Jun	Jul	Aug	All
–	Reference	75.9	56.4	59.4	55.1	69.1	75.0	57.8	60.2
13	V → L		1.2 Ubiquitous	2.6 UK	1.5 UK, USA				1.6 Ubiquitous
14	T → I			0.6 Canada, UK					
46	L → F					0.5 India, USA			
54	A → S				0.4 UK	0.7 UK			
57	Q → H	5.0 Ubiquitous	25.7 Ubiquitous	24.2 Ubiquitous	29.0 Ubiquitous	14.9 Ubiquitous	9.8 Ubiquitous	11.8 Ubiquitous	22.6 Ubiquitous
57 99	Q → H A → V	1.0 Netherlands	0.7 Ubiquitous						0.4 Ubiquitous
57 58	Q → H S → N	0.7 Netherlands				0.8 Netherlands		1.5 Netherlands	
57 264	Q → H Y → C				1.0 USA				
57 172 75	Q → H G → V K → N							2.0 USA 3.4 UK	
108	L → F						1.0 UK		
126	R → S					0.9 South Africa	1.6 South Africa		
196	G → V		2.3 Ubiquitous						0.9 Ubiquitous
207	F → L						0.7 UK		
223	T → I							8.6 UK	
251	G → V	11.9 Ubiquitous	9.1 Ubiquitous	4.1 UK	0.6 Europe, USA				5.2 Ubiquitous
257	N → S	0.5 Australia				0.9 Australia	1.4 Australia		

end of the transmembrane helix 1 (TM1), already described as deleterious by other authors [15]. The aromatic interaction between the two Phe aromatic rings, one from each subunit, can stabilize the structure as predicted by DynaMut (Table 4) but may create a steric constriction at the mouth of the central pore (Fig. 1).
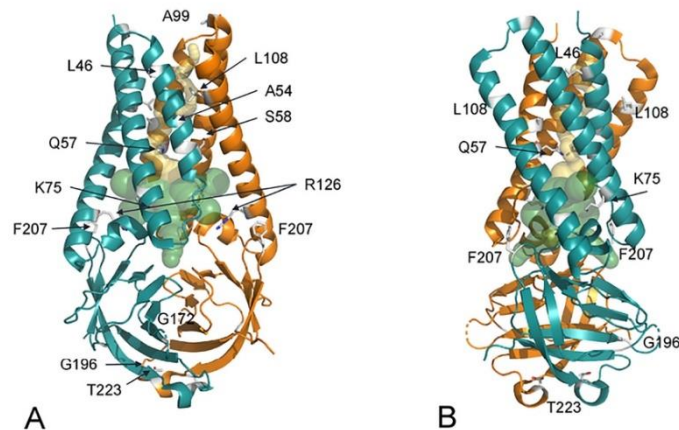

Fig. 1. (A) ORF3a dimer represented as ribbon model. The two subunits are colored in orange and deep teal. Variant sites are labelled and the corresponding side chains reported as grey sticks. Transparent internal spheres indicate the transmembrane channel (yellow) and the tunnels connecting to the extracellular environment (green). (B) is rotated approximately 90° along the y axis with respect to (A). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

XIV

Table 3
ORF3a mutant sites.

Position[a]	Mutation	Structural features[b]
13	V → L	Not available
14	T → I	Not available
46[*]	L → F	Central pore lining. Aromatic interaction at the interface between TM1 of the two subunits.
54[*]	A → S	Central pore lining interface TM1-TM3 of the other subunit
57[*]	Q → H	Intersection between central pore and lower tunnel. Interface TM1-TM1 of the other subunit
57[*]	Q → H	Loop connecting TM1 e TM2 on the extracellular side
99[*]	A → V	
57[*]	Q → H	Interface TM1-TM3 of the other subunit.
58[*]	S → N	
57[*]	Q → H	C-terminal side of β3.
172	G → V	
57[*]	Q → H	Position 264 not available
264	Y → C	
75[*]	K → N	Intersection among central pore and upper, lower, intersubunit tunnels. Within TM2.
108[*]	L → F	Interface TM1-TM3
126[*]	R → S	Intersubunit tunnel lining; TM3
196	G → V	Loop connecting β5 and β6. Exposed to the solvent
207	F → L	Loop connecting β6 and β7. Exposed to the solvent
223	T → I	Loop connecting β7 and β8. Hydrophobic interaction with β8 of the other subunit
251	G → V	Not available
257	N → S	Not available

[a] Asterisks mark mutations within the transmembrane domain.
[b] Not available indicates that the corresponding spatial coordinates are not available in the PDB file.

A54S, at the inter-subunit interface TM1-TM3′ (prime denotes the other subunit), is relatively rare and it has been observed in the top five frequencies only in April and May isolates mostly in UK. It lines the central transmembrane tunnel and appears to be destabilizing (Table 4).

The variant R126S emerged mainly in June and July isolates from South Africa. This mutation removes a positive charge in proximity of the lower tunnel and may facilitate cation transit and/or alter selectivity. This mutation is predicted to be destabilizing (Table 4). The substitution K75N appears relatively frequent in August and was isolated exclusively in UK. It is potentially interesting because occurs in proximity of the intersection of the tunnels connecting the transmembrane pore to the extracellular environment. In this case also, removal of a positive charge can influence cation transport and/or selectivity. This substitution is predicted to be destabilizing although not at the level of R126S.

The other mutations are in places not directly connected with tunnels. Considering the order of temporal appearance of mutations during the pandemic, the double mutant Q57H + A99V has been isolated mainly in European countries at the beginning of the pandemic and overall, it is one of the five top variants observed over the entire period.

Table 4
Predicted effect of mutations on stability.

Variants	ΔΔG (kcal/mol)	
	DynaMut[a]	DUET[a]
L46F	**0.874**	−0.821
Q57H	**0.429**	−0.503
K75N	−0.559	−0.186
A99V	−0.269	−0.152
A54S	−0.439	−2.112
S58N	−0.176	−0.904
G172V	**0.538**	**0.149**
L108F	−0.136	−1.153
R126S	−2.024	−3.073
G196V	**0.173**	**0.483**
F207L	−0.063	−0.34
T223I	−0.286	−0.117

[a] Boldfaced numbers indicate stabilization.

The A99V mutation is predicted to be only marginally destabilizing (Table 4).

The substitution G196V showed a peak frequency in March but, overall, it is among the top five more frequent variants. It has been isolated all over the world. Its effect is predicted to be structurally stabilizing (Table 4) which may explain its success. Mutation L108F was detected in July in UK with a relatively high frequency. Substitution of a hydrophobic residue with an aromatic side chain destabilizes the interaction between TM3 and TM1′ (Table 4). F207L replace an aromatic residue with Leu at the interface with the lipid bilayer. Therefore, it can influence interaction with the membrane. It is relatively rare since it has occurred in the top five frequencies only in July and it has been isolated only in UK. The mutation is predicted to be destabilizing (Table 4). S58N at the inter-subunit interface TM1-TM3′ is present in a double mutant Q57H + S58N relatively frequent in August, isolated primarily in Netherlands. The analyses conducted by DynaMut and DUET servers predicted that S58N substitution reduces ORF3a stability (Table 4). The variant Q57H + G172V emerged in the top five relative frequencies in the isolates collected exclusively in the USA in August. The substitution G172V may contribute the stabilization of the β-barrel by increasing the hydrophobic interactions while decreasing local flexibility. Indeed, DynaMut and DUET predicts a stabilizing effect of the mutation. T223I emerged in the top five relative frequencies in the isolates collected in August from UK. The mutation T223I occurs in the loop connecting β7 to β8 and it is predicted to be destabilizing (Table 4).

Four mutant sites are in N-terminal, loop or C-terminal portions for which no structural information is reported in the coordinate set. Among these mutations, the variant G251V is consistently highly frequent and has been isolated all over the world. Unfortunately, no structural consideration can be drawn. By analogy with G172V, it may be speculated that this mutation stabilizes the β-barrel by adding hydrophobic interaction.

3.2. Conserved sites

Scanning of the ORF3a variants indicates the sites that are identically conserved in the entire data set considered. Twenty-eight positions, the 10% of ORF3a sequence, so far, are identically conserved in all isolates. The positions are listed in Table 5 along with notations about the structural environments and the possible roles. Four sites are in structural regions not visible by the cryo-EM analysis and their spatial coordinates are not available. Twelve sites are in the transmembrane domain and twelve in the β-sandwich cytosolic domain (Fig. 2). Four sites (L71, F79, L139 and Y141) are involved in pore and tunnel stabilization (Table 5). Eight sites (Q80, L84, P138, F146, I169, L203, Y212, and L214) are involved in intra-monomer interactions that assure structural stability and consistency (Fig. 2). Four conserved positions (Q116, T164, T170 and V228) are at the dimer interface. In this case also it can be assumed that they are essential for dimer stability. Interestingly, one of the conserved sites, K132, is close to the putative tetramerization surface suggesting that also this residue may contribute to the tetramerization interface, as suggested by other authors [15]. Noteworthy is the conservation of C133 and C157. Residue C133 is the most important for homodimerization and is conserved between different species [15]. The conservation of the two residues strongly supports this observation and suggests that C157 also is essential for ORF3a structural stability and function. Distance between the sulphur atoms of the two cysteine is 3.9 Å that is not compatible with the presence of a disulfide bridge. However, flexibility of the loop bearing C157 may allow, in certain circumstances, the formation of a bond. Other conserved sites are exposed at different locations. E102 is exposed at the extracellular side. I124 is exposed to the bilayer interface. C200 is exposed to the intracellular compartment and may have, for its reactivity, a role in the interaction with other cellular components. S209 and E226 are also exposed to the intracellular compartment. Interestingly, the conserved Y141 and F146 belong to the Domain IV described by Issa et al. [15] that is deemed to be involved in interaction with caveolin.

XV

Table 5
Identically conserved positions in ORF3 sequence.

Position[a]	Residue	Features[b]
28	F	Not available. Predicted to occur in α-helix.
70	Q	N-terminal side of TM1; partly buried
71*	L	Lining of the lower tunnel. N-terminal side of TM1; buried. Interacts with Y141 of the other chain
79*	F	Mouth of upper tunnel. Within TM-1; exposed to the surface in contact with the lipid bilayer
84*	L	TM2; interaction with TM1 L52
102*	E	Exposed in a loop connecting TM2 and 3 on the extracellular side
116*	Q	Buried in TM3; interaction with TM1'
124*	I	TM3; exposed to the lipid bilayer
132*	K	C-term side of TM3; partly buried. Proximal to the tetramerization interface
133*	C	C-terminal of TM3
138*	P	Short helix in the cytosolic domain. Packs against F146. Buried
139*	L	Upper tunnel. Partly buried; interacts with L127 of TM3
141*	Y	Lining the lower tunnel mouth. Partly buried; hydrophobic interaction with L71 in TM2. Interaction with caveolin.
146	F	Buried in β1. Interacts with P138. Interaction with caveolin.
157	C	Loop connecting β1 and β2. Buried.
164	T	Loop connecting β2 and β3. Partly buried. Dimer interface.
169	I	Buried in β3. Hydrophobic interaction with L147, I167, Y184
170	T	β3 at the interface with β3 of the other chain
200	C	Exposed on the surface of the cytosolic domain in β6
203	L	Buried in β6. Interacts with Y212
209	S	Exposed to the surface of the cytosolic domain in β7
212	Y	Buried in β7. Interacts with L203
214	L	Partially buried in β7
226	E	Exposed on the surface of the cytosolic domain
228	V	β8 at the interface with β8 of the other chain
243	H	Not available. Predicted in β-sheet
248	T	Not available. Predicted in β-sheet
249	I	Not available. Predicted in β-sheet

[a] Asterisks mark residues of the transmembrane domain.
[b] Not available indicates that the corresponding spatial coordinates are not available in the PDB file.

A multiple sequence alignment of 66 homologous ORF3a (Supplementary Material Table 1 and Supplementary Material Fig. 2) from other coronaviruses has been calculated to assess whether the SARS-CoV-2 unmutated positions are conserved in other species. Results were outlined as a Logo (Fig. 3). Most of the residues are identically or virtually identically conserved: L84, E102, Q116, K132, C133, P138, L139, Y141, F146, C157, T164, 170T, 212Y, L214, 248T, I249. Seven positions display conservative substitutions, namely conserve the physical-chemical characteristics of the site: Q70, L71, F79, I124, I169, L203, E226.
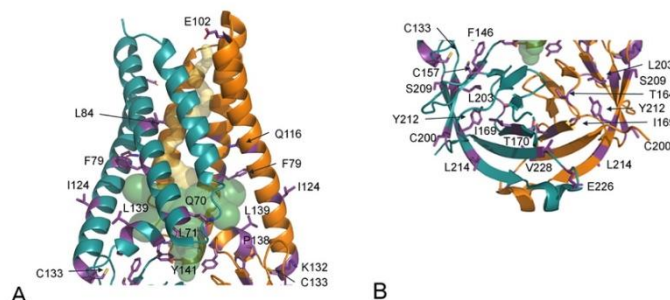

Fig. 3. Logo representation of the conservation of SARS-CoV-2 identical sites among other ORF3a from different Coronaviruses. X-axis numbering refers to the sequence positions in the ORF3a reference protein. Pile height is proportional to the information content of the site while letter height indicates frequency of the residue in the corresponding alignment column. Color indicates physical-chemical properties. The Logo was built using the alignment reported in Supplementary Fig. 2 using the site WebLogo. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Five positions contain drastic substitutions: F28, C200, S209, V228 and H243 (Fig. 3).

4. Discussion

SARS-CoV-2 is seriously threatening global health and it is claiming many lives. To effectively fight this pathogen, it is important to understand its evolution and the mechanism of adaptation to the host. This knowledge will also pave the way to face future epidemics of zoonotic origin. Proteins coded by the virus genome are the effectors of biological function. Pathogenesis and host adaptation depend in an intricate way on the changes accumulated by the virus proteins that may lose or acquire function that alter SARS-CoV-2 properties. In this work we focused onto the ORF3a, a membrane protein whose three-dimensional structure is available, involved in crucial steps of virus replication and pathogenesis [18–21]. The genomes deposited in GISAID up to August 2020 were analysed to record the ORF3a mutations within a space-temporal frame. Possible roles of non-synonymous mutations on protein functional domains were determined. In general, this protein appears rather conserved. Indeed, mutation rate is moderate in coronaviruses [34].

Different SARS-Cov-2 isolates display mutations in all sites of the ORF3a sequence except for 28 positions found identically conserved in all samples considered (Table 3). The size of the data set we utilized is large and the results we obtained can be considered robust and stable, unlikely to change significantly soon. Moreover, to avoid inclusion of possible statistical noise, only the top five most frequent substitutions were considered.

Fig. 2. ORF3a dimer represented as ribbon model. The two subunits are colored in orange and deep teal. Conserved sites are labelled and the corresponding side chains are reported as violet sticks. Transparent internal spheres indicate the transmembrane channel (yellow) and the tunnels connecting to the extracellular environment (green). (A) Trans-membrane domain; (B) extracellular domain. The protein is oriented as in Fig. 1A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

XVI

In general, the most frequent mutations found do not influence significantly the central pore topography. Most of the seventeen mutations were isolated only in specific pandemic periods after which their frequencies decreased. Considering the entire sample, the five most frequent mutations are V13L, Q57H, Q57H + A99V, G196V and G252V. According to the predictions, G196V is stabilizing the protein. By analogy, the mutations V13L and G252V, for which lack of spatial coordinates hinders predictions, can have also a similar effect. The stabilization can explain the relative success of these variants. Q57H is stabilizing according to DynaMut and slightly destabilizing from Duet calculations. Its ubiquitous prevalence in the virus population suggests that the mutation confers the virus an advantage which may also be connected to the stabilization of the central pore. The other mutations are predicted to be destabilizing and tend to disappear from the virus isolate population.

However, two mutations (K75N and R126S) remove two basic and positively charged residues from the proximity of the pore. Lack of positive charges may facilitate translocation of cations and/or alter pore selectivity. The two mutations have emerged in July and August. Continuous monitoring of SARS-Cov-2 ORF3a evolution will indicate whether these changes can attribute the virus any advantage and become frequent in the population as observed for the change Q57H.

The same analysis provided information on the ORF3a conserved sites. Conservation of a site is often a strong marker of critical functional relevance [35]. In this study, only identically conserved positions were considered. The position and the role of these sites are rather heterogeneous. They are involved in pore, monomer and dimer stabilization and in tetramerization. Four conserved sites are exposed to the intra- or extra-cellular environment. This pattern suggests possible and essential interactions with other cellular components. This concept is reinforced by the analysis of the conservation of the SARS-CoV-2 positions in homologous ORF3a sequences. E102 is exposed to the extracellular surface. It can be speculated that it may be involved in recognition with host or virus factors. Conservation of K132, C133, and C157 suggests that the dimerization and tetramerization functions are an essential structural feature of ORF3a. Conservation of Y141 and F146 corroborate their role in interaction with caveolin.

Interesting are also the ORF3a positions conserved in the SARS-CoV-2 isolates that are variable in the other coronaviruses. For example, F28 seems to be unique to SARS-CoV-2. Likewise, C200 and Ser209 exposed to the cytosolic side are conserved only in a few other SARS-CoV-2 from pangolin or bats. This pattern points to functions specific to SARS-CoV-2 possibly connected to its peculiar pathogenicity, contagiousness and ability to cross-species transmission.

Systematic in-silico analysis of the evolution of SARS-CoV-2 genome and proteome is a powerful tool to provide elements to understand virus biology and pathogenesis and to guide the design of specific experiments or therapeutic strategies. The observations and the hypotheses here reported can be experimentally tested, for example, by site-directed mutagenesis and other experimental protocols. Moreover, the relative conservation of the ORF3a extra- and intracellular domains suggests possible target for vaccine design.

CRediT authorship contribution statement

Martina Bianchi: Methodology, Software, Investigation, Writing – Original Draft, Visualization.
Alessandra Borsetti: Conceptualization, Validation, Methodology, Writing – Review & Editing.
Massimo Ciccozzi: Conceptualization, Validation, Methodology.
Stefano Pascarella: Conceptualization, Methodology, Software, Writing – Original Draft.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2020.12.142.

References

[1] Y.Z. Zhang, E.C. Holmes, A genomic perspective on the origin and emergence of SARS-CoV-2, Cell 181 (2020) 223–227, https://doi.org/10.1016/j.cell.2020.03.035.

[2] D. Benvenuto, M. Giovanetti, M. Salemi, M. Prosperi, C De Flora, L.C. Junior Alcantara, S. Angeletti, M. Ciccozzi, The global spread of 2019-nCoV: a molecular evolutionary analysis, Pathog. Glob. Health. (2020) 1–4, https://doi.org/10.1080/20477724.2020.1725339.

[3] M. Ciotti, S. Angeletti, M. Minieri, M. Giovannetti, D. Benvenuto, S. Pascarella, C. Sagnelli, M. Bianchi, S. Bernardini, M. Ciccozzi, COVID-19 outbreak: an overview, Chemotherapy (2020)https://doi.org/10.1159/000507423.

[4] J. Xu, S. Zhao, T. Teng, A.E. Abdalla, W. Zhu, L. Xie, Y. Wang, X. Guo, Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV, Viruses 12 (2020) 244, https://doi.org/10.3390/v12020244.

[5] D. Kaul, An overview of coronaviruses including the SARS-2 coronavirus - molecular biology, epidemiology and clinical implications, Curr. Med. Res. Pract. 10 (2020) 54–64, https://doi.org/10.1016/j.cmrp.2020.04.001.

[6] D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, M. Ciccozzi, The 2019-new coronavirus epidemic: evidence for virus evolution, J. Med. Virol. 92 (2020) 455–459, https://doi.org/10.1002/jmv.25688.

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

[8] M. Bianchi, D. Benvenuto, M. Giovanetti, S. Angeletti, M. Ciccozzi, S. Pascarella, Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? Biomed. Res. Int. 2020 (2020) 1–6, https://doi.org/10.1155/2020/4389089.

[9] T. Phan, Genetic diversity and evolution of SARS-CoV-2, Infect. Genet. Evol. 81 (2020), 104260, https://doi.org/10.1016/j.meegid.2020.104260.

[10] R. Sanjuán, P. Domingo-Calap, Mechanisms of viral mutation, Cell. Mol. Life Sci. 73 (2016) 4433–4448, https://doi.org/10.1007/s00018-016-2299-6.

[11] M.L. DeDiego, J.L. Nieto-Torres, J.M. Jimenez-Guardeño, J.A. Regla-Nava, C. Castaño-Rodriguez, R. Fernandez-Delgado, F. Usera, L. Enjuanes, Coronavirus virulence genes with main focus on SARS-CoV envelope gene, Virus Res. 194 (2014) 124–137, https://doi.org/10.1016/j.virusres.2014.07.024.

[12] M.R. Denison, R.L. Graham, E.F. Donaldson, L.D. Eckerle, R.S. Baric, Coronaviruses, RNA Biol. 8 (2011) 270–279, https://doi.org/10.4161/rna.8.2.15013.

[13] Y. Ren, T. Shu, D. Wu, J. Mu, C. Wang, M. Huang, Y. Han, X.Y. Zhang, W. Zhou, Y. Qiu, X. Zhou, The ORF3a protein of SARS-CoV-2 induces apoptosis in cells, Cell. Mol. Immunol. 17 (2020) 881–883, https://doi.org/10.1038/s41423-020-0485-9.

[14] D.M. Kern, B. Sorum, C.M. Hoel, S. Sridharan, J.P. Remis, D.B. Toso, S.G. Brohawn, Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs, BioRxiv Prepr. Serv. Biol. (2020)https://doi.org/10.1101/2020.06.17.156554.

[15] E. Issa, G. Merhi, B. Panossian, T. Salloum, S. Tokajian, SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis, MSystems. 5 (2020), e00266-20, https://doi.org/10.1128/mSystems.00266-20.

[16] W. Lu, B.J. Zheng, K. Xu, W. Schwarz, L. Du, C.K.L. Wong, J. Chen, S. Duan, V. Deubel, B. Sun, Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 12540–12545, https://doi.org/10.1073/pnas.0605402103.

[17] C. Scott, S. Griffin, Viroporins: structure, function and potential as antiviral targets, J. Gen. Virol. 96 (2015) 2000–2027, https://doi.org/10.1099/vir.0.000201.

[18] K. Padhan, C. Tanwar, A. Hussain, P.Y. Hui, M.Y. Lee, C.Y. Cheung, J.S.M. Peiris, S. Jameel, Severe acute respiratory syndrome coronavirus Orf3a protein interacts with caveolin, J. Gen. Virol. 88 (2007) 3067–3077, https://doi.org/10.1099/vir.0.82856-0.

[19] K.-L. Siu, K.-S. Yuen, C. Castaño-Rodriguez, Z.-W. Ye, M.-L. Yeung, S.-Y. Fung, S. Yuan, C.-P. Chan, K.-Y. Yuen, L. Enjuanes, D.-Y. Jin, Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC, FASEB J. 33 (2019) 8865–8877, https://doi.org/10.1096/fj.201802418R.

[20] C. Castaño-Rodriguez, J.M. Honrubia, J. Gutiérrez-Álvarez, M.L. DeDiego, J.L. Nieto-Torres, J.M. Jimenez-Guardeño, J.A. Regla-Nava, R. Fernandez-Delgado, C. Verdia-Báguena, M. Queralt-Martín, G. Kochan, S. Perlman, V.M. Aguilella, I. Sola, L. Enjuanes, Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis, MBio 9 (2018)https://doi.org/10.1128/mBio.02325-17.

[21] A. Grifoni, D. Weiskopf, S.I. Ramirez, J. Mateus, J.M. Dan, C.R. Moderbacher, S.A. Rawlings, A. Sutherland, L. Premkumar, R.S. Jadi, D. Marrama, A.M. de Silva, A. Frazier, A.F. Carlin, J.A. Greenbaum, B. Peters, F. Krammer, D.M. Smith, S. Crotty, A. Sette, Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals, Cell 181 (2020) 1489–1501.e15, https://doi.org/10.1016/j.cell.2020.05.015.

XVII

[22] Y. Shu, J. McCauley, GISAID: global initiative on sharing all influenza data - from vision to reality, Euro Surveill. Bull. Eur. Sur Les Mal. Transm. = Eur. Commun. Dis. Bull. 22 (2017)https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

[23] S. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402, https://doi.org/10.1093/nar/25.17.3389.

[24] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite, Trends Genet. 16 (2000) 276–277, https://doi.org/10.1016/s0168-9525(00)02024-2.

[25] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics. 22 (2006) 1658–1659, https://doi.org/10.1093/bioinformatics/btl158.

[26] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[27] B.J. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3d: an R package for the comparative analysis of protein structures, Bioinformatics 22 (2006) 2695–2696, https://doi.org/10.1093/bioinformatics/btl461.

[28] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview version 2-a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (2009) 1189–1191, https://doi.org/10.1093/bioinformatics/btp033.

[29] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, J. Damborsky, CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures, PLoS Comput. Biol. 8 (2012), e1002708, https://doi.org/10.1371/journal.pcbi.1002708.

[30] C.H.M. Rodrigues, D.E.V. Pires, D.B. Ascher, DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability, Nucleic Acids Res. 46 (2018) W350–W355, https://doi.org/10.1093/nar/gky300.

[31] D.E.V. Pires, D.B. Ascher, T.L. Blundell, DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach, Nucleic Acids Res. 42 (2014) W314–W319, https://doi.org/10.1093/nar/gku411.

[32] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190, https://doi.org/10.1101/gr.849004.

[33] S. Liu, J. Shen, S. Fang, K. Li, J. Liu, L. Yang, C.-D. Hu, J. Wan, Genetic spectrum and distinct evolution patterns of SARS-CoV-2, Front. Microbiol. 11 (2020) 2390, https://doi.org/10.3389/fmicb.2020.593548.

[34] Z. Zhao, H. Li, X. Wu, Y. Zhong, K. Zhang, Y.P. Zhang, E. Boerwinkle, Y.X. Fu, Moderate mutation rate in the SARS coronavirus genome and its implications, BMC Evol. Biol. 4 (2004) 1–9, https://doi.org/10.1186/1471-2148-4-21.

[35] J.A. Capra, M. Singh, Predicting functionally important residues from sequence conservation, Bioinformatics 23 (2007) 1875–1882, https://doi.org/10.1093/bioinformatics/btm270.

XVIII

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate?

Stefano Pascarella[1] | Massimo Ciccozzi[2] | Davide Zella[3] |
Martina Bianchi[1] | Francesca Benedetti[3] | Domenico Benvenuto[2] |
Francesco Broccolo[4] | Roberto Cauda[5] | Arnaldo Caruso[6] | Silvia Angeletti[7] |
Marta Giovanetti[8] | Antonio Cassone[9]

[1]Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Rome, Rome, Italy

[2]Medical Statistic and Molecular Epidemiology Unit, University of Biomedical Campus, Rome, Italy

[3]Department of Biochemistry and Molecular Biology, Institute of Human Virology and Global Virus Network Center, University of Maryland School of Medicine, Baltimore, Maryland, USA

[4]Department of Clinical Medicine and Prevention, University of Milano-Bicocca, Milan, Italy

[5]Istituto Clinica di Malattie Infettive, Università Cattolica del Sacro Cuore, Rome, Italy

[6]Department of Microbiology and Virology, Spedali Civili, Brescia, Italy

[7]Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome, Italy

[8]Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Rio de Janeiro, Brazil

[9]Center of of Genomics, Genetics and Biology, Siena, Italy

Correspondence

Massimo Ciccozzi, Medical Statistic and Molecular Epidemiology Unit, University of Biomedical Campus, Via Álvaro del Portillo, 21, 00128 Rome, Italy.
Email: M.ciccozzi@unicampus.it

Silvia Angeletti, Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Via Álvaro del Portillo, 21, 00128 Rome, Italy.
Email: S.Angeletti@unicampus.it

Abstract

Lineage B.1.617+, also known as G/452R.V3 and now denoted by WHO with the Greek letters δ and κ, is a recently described SARS-CoV-2 variant under investigation first identified in October 2020 in India. As of May 2021, three sublineages labeled as B.1.617.1 (κ), B.1.617.2 (δ), and B.1.617.3 have been already identified, and their potential impact on the current pandemic is being studied. This variant has 13 amino acid changes, three in its spike protein, which are currently of particular concern: E484Q, L452R, and P681R. Here, we report a major effect of the mutations characterizing this lineage, represented by a marked alteration of the surface electrostatic potential (EP) of the receptor-binding domain (RBD) of the spike protein. Enhanced RBD-EP is particularly noticeable in the B.1.617.2 (δ) sublineage, which shows multiple replacements of neutral or negatively charged amino acids with positively charged amino acids. We here hypothesize that this EP change can favor the interaction between the B.1.617+ RBD and the negatively charged ACE2, thus conferring a potential increase in the virus transmission.

Stefano Pascarella and Massimo Ciccozzi contributed equally to this study.

XIX

1 | INTRODUCTION

The coronavirus disease 19 (COVID-19), caused by the new coronavirus SARS-CoV-2, continues to spread worldwide, with more than 163 million infections and about 3.5 million deaths as of May 17, 2021 (www.who.int). To fight this dreadful disease, several safe and efficacious vaccines against SARS-CoV-2 are being used with remarkable effectiveness in some countries and limited availability in others. In particular, the capacity of some countries to halt SARS-CoV-2 spread is still limited due to inadequate resources and vaccination infrastructures.[1,2] In this scenario, several SARS-CoV-2 variants have been identified and have become a global threat. Some of them have been classified as variants of concern (VOCs) due to their mutations in the S1 subunit of the spike (S) protein, particularly in its receptor-binding domain (RBD).[3-5] One of them, identified as B.1.1.7 (α), also known as the UK variant, bears a substitution of asparagine with tyrosine on the position 501 and deletion of two amino acids in the position 69–70 of the S1 subunit. This variant has quickly spread in several European countries to become globally dominant.[5] Other VOCs have been isolated in South Africa and Brazil and have been studied for their enhanced contagiousness and resistance to neutralization by antibodies from convalescent and vaccine-recipient subjects[6-8] (Table 1). Quite recently, a new variant under investigation (VUI) has been isolated from Maharashtra, India, in a setting of the highly diffusive epidemic with devastating proportions. This variant, identified as B.1.617, carries several nonsynonymous mutations. Two of them, the E484Q (or the P478K) and the L452R, are located in the RBD region, and they are critical sites for the binding with ACE2. Initial data suggest these mutations could confer increased transmission and immune evasion.[9-11] Currently, B.1.617 comprises three subvariants, B.1.617.1-3, with different distribution of the mutations P478K and E484Q (Public Health England). Here, we focus on biochemical and biophysical changes conferred to the B.1.617+ VUI by the P478K and E484Q mutations. We then compare these changes with other VOCs to establish whether and to what extent those amino acid changes can influence the interaction of the spike protein with ACE2, thus potentially affecting SARS-CoV-2 transmission and immune-escape properties.

2 | METHODS

To perform a robust analysis three different three-dimensional structures of SARS-CoV-2 spike glycoprotein have been downloaded from Protein Data Bank with the following characteristics:

1. SARS-CoV-2 RBD in complex with neutralizing antibody CC12.3 (PDB code: 6XC4, chain A),
2. SARS-CoV-2 RBD in complex with ACE2 (PDB code: 6M0J, chain E),
3. SARS-CoV-2 RBD in complex with ACE2-B0AT1 (PDB code: 6M17, chain F).

The DynaMut server[12] has been used to predict the impact of mutations on protein stability analyzing the folding free energy (ΔΔG) and the vibrational movement (ΔΔS), two crucial characteristics of the function and the molecular recognition of the protein. DynaMut automatically provides also results from DUET analysis, a method that combines two complementary approaches (mCSM and SDM) in a consensus prediction of ΔΔG.[13] DUET results also have been considered in combination with DynaMut for the characterization of the mutants. PyMol (PyMol, version 2.4) suite was utilized for in-silico mutagenesis and its Adaptive Poisson Boltzmann Solver (APBS) plugin[14] has been used to calculate the electrostatic potential

TABLE 1 Results of DynaMut and DUET analysis

Mutant sites	DynaMut						DUET		
	ΔΔG (kcal/mol)[a]			ΔΔS (kcal/mol K)[b]			ΔΔG (kcal/mol)[a]		
	6M17	6M0J	6XC4	6M17	6M0J	6XC4	6M17	6M0J	6XC4
N501Y	0.089	−0.272	0.089	−0.162	0.138	−0.312	−0.297	−0.474	−0.391
K417N	−0.399	−0.651	−1.174	0.487	0.659	0.347	−0.513	−1.295	−0.990
K417T	−0.152	−0.566	−0.832	0.198	0.507	0.193	−0.854	−1.343	−1.119
E484K	0.087	−0.101	−0.109	−0.075	0.171	0.336	0.656	0.128	0.348
E484Q	0.399	−0.644	−0.755	−0.084	0.151	0.189	0.099	−0.438	−0.319
L452R	−0.417	−0.319	−0.462	0.150	0.059	−0161	−0.548	−0.741	−0.661
T478K	−0.334	1.003	0.257	0.111	−0.385	−0.152	0.109	−0.024	0.037

[a]Free energy difference (ΔΔG) between wild-type and mutant structure. Negative values indicate destabilization.
[b]Vibrational entropy difference (ΔΔS) between wild-type and mutant structure. Positive values indicate increased flexibility.

XX

of the wild-type and VOCs SARS-CoV-2 spike receptor-binding domain (RBD) and evaluate potential differences in terms of molecular interaction with ACE2 receptor. The results have been reported within a range between −5 and +5 kT/e.

3 | RESULTS

3.1 | Protein stability

In silico prediction of the mutation impact on the RBD stability has been carried out with DynaMut. Three alternative RBD structures denoted by the PDB codes 6M17, 6M0J, and 6XC4 have been tested. These structures display small differences in the conformation of loops, especially in the one inside the receptor-binding motif (RBM). According to the parameters of our in silico experiments, the output of DynaMut for mutant sites within loops differs depending on the starting loop conformation. For this reason, we used a normalized procedure, whereby the same mutation has been tested in each of the three different RBD structures. The effect on stability, de- or stabilization, has been evaluated following a majority criterion over the results of DynaMut and DUET. Detailed results have been reported in Table 1.

Starting with position 501 within a loop in the RDB interacting with the ACE2 receptor, we note that the mutation N501Y does not show any clear structural effect as the negative ΔΔG values are very close to 0.0 kcal/mol in DynaMut. At variance with DynaMut, DUET

consistently assigns a destabilizing effect. For the majority rule, this mutation should be considered destabilizing (Table 1). Also, position 417 is within the interface α-helix where Lys interacts with ACE2 Asp30. However, in this case, our results predict that both mutants (K417N and K417T) could destabilize the protein and increase local flexibility. Similarly, Glu484 is in an interfacial loop and interacts with ACE2 Lys31. Our data predict that mutation E484K is stabilizing, although data from DynaMut and DUET do not match. E484Q exerts a destabilizing effect, again with no strong consensus from the two methods. We note that glutamic acid is a polar, negatively charged, hydrophilic amino acid while lysine is a basic, charged and partly aliphatic amino acid, and its ε-amino group often participates in hydrogen bonding, salt bridges, and covalent interactions. Both mutants E484K and E484Q would likely increase local molecular flexibility. L452 is in a short β-strand, and it is exposed to the solvent. Apparently, it does not interact directly with ACE2. Mutation L452R is predicted to be destabilizing with increased local flexibility. Position 478 is in a loop in the proximity of the ACE2 although not in direct contact with it. Mutation T478K is stabilizing and is predicted to decrease local protein flexibility.

3.2 | Surface and interface analysis

The entire set of mutations in positions 417, 452, 478, 484, and 501 are in the spike RBD at the interface with the ACE2 N-terminal helix (Figure 1). However, we note that the mutant positions 452, 478,
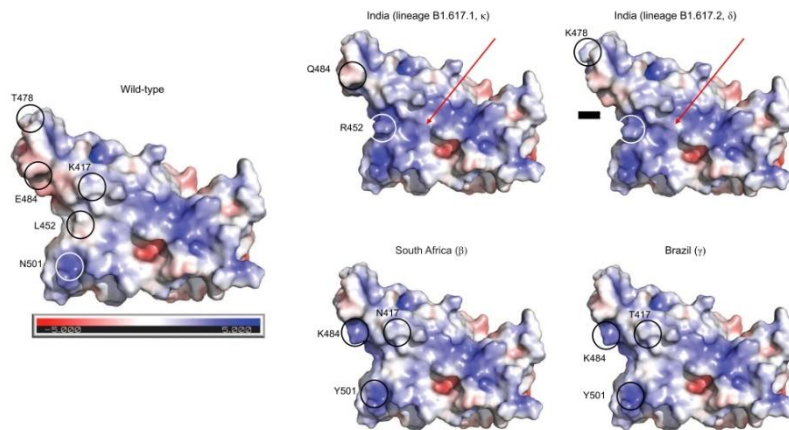

FIGURE 1 Comparison between the wild-type and variant spike receptor-binding domains (RBDs). Protein surface is colored according to the electrostatic potential. Color scale ranges from −5 kT/e (red) to +5 kT/e (blue) as reported by the bar under the wild-type RBD. Position of the mutant sites is indicated by a circle and an attached label. Red arrows mark the area of increased positive potential in the RBD Indian variants

XXI

484, and 501 are within the RBM, containing residues that bind to ACE2. In contrast, mutant position 417 is located outside the motif.[15] According to our analysis, the mutations in positions 417, 484, and 501 might increase the spike binding affinity with the ACE2 receptor. In particular, the Tyr replacing Asn501 may form an aromatic interaction with ACE2 Tyr41, a hydrogen bond with ACE2 D38, and a potential cation–π interaction with ACE2 Lys353. In addition, substitutions of Glu484 with Lys or Gln may form hydrophobic interactions to Ile472, Gly482, Phe486, Cys488, and Tyr489. Our data also indicate that replacing Lys or Gln with Glu484 abolishes the interfacial salt bridge between Glu484 and ACE2 Lys31. Due to the fact that Lys417 is solvent-exposed and forms salt-bridge interactions with Asp30 of ACE2, the replacement of Thr/Asn with Lys417 could abolish this interaction. Moreover, although the mutations in positions 452 and 478 are within the receptor-binding motif, our analysis does not show a direct interaction with ACE2.

3.3 | Electrostatic potential

We note that a major, global effect of the mutations characterizing the Indian variants is represented by the alteration of the RBD surface electrostatic potential. In particular, in the lineage B.1.617.1 (κ) the uncharged and hydrophobic residue Leu452 changes to the positively charged residue Arg, and the negatively charged residue

Glu484 is replaced by the uncharged residue Gln. In contrast, the B.1.617.2 (δ) lineage shares the same mutation in position 452, but it has another mutation in position 478 where the neutral residue Thr changes to the positively charged Lys. The presence of two positively charged residues in the variant B.1.617.2 (δ) increases the positive electrostatic potential surface (Figure 2).

4 | DISCUSSION

Given that SARS-CoV-2 variants are characterized by evolutionary and genetic changes accumulated in the genome, the use of new and improved phylodynamic techniques for the study of how epidemiological, immunological, and evolutionary processes contribute to shaping viral phylogenies is extremely important and useful. Here, using a well-structured software workflow, we provide evidence of a strong system able to carry out a quick, systematic, and reproducible screening of the SARS-CoV-2 genome isolates. Protein mutations were identified by scanning high-quality SARS-CoV-2 genomes variants downloaded from the GISAID databank.[16] We reasoned that the most likely health-threatening properties of SARS-CoV-2 VOC rely on fine biochemical and biophysical changes that eventually impact the RBD interaction with the ACE2 receptor present on the host's cell surface. We thus characterized B.1.617+ SARS-CoV-2 VUI using in-silico methods capable of predicting the effect of mutations on S-RBD protein stability and its electrostatic potential. The
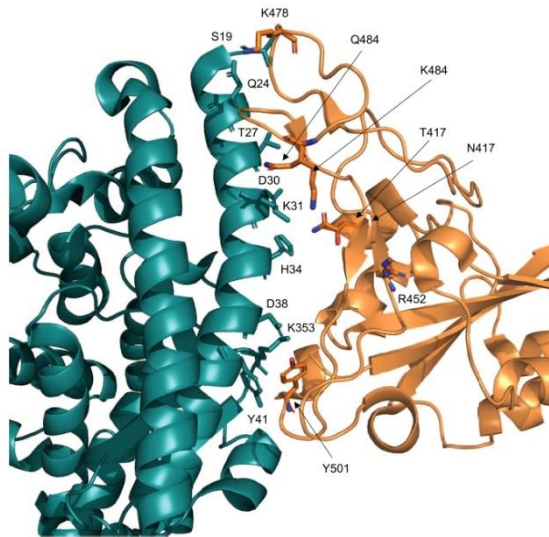

FIGURE 2 Ribbon model of the interface between ACE2 (deep teal) and receptor-binding domain (RBD) (orange). Sidechains of relevant residues are displayed as stick models and labeled. The two mutations at the RBD sites 417 and 484 have been simultaneously displayed

B.1.617+ mutations have been investigated by comparing most of the already known mutations of previously reported VOCs.[17] To further justify our assumption, we note that the intensely investigated D614G substitution of the spike protein, early reported in Italian isolates,[18-20] and subsequently attributed with increased virus transmissibility[21] was found to enhance the protein torsional ability and potentially affecting its stability.[21] Regarding the commonly denominated Indian variants, these constitute the new SARS-CoV-2 lineage B.1.617, which is actually composed of a family of three subvariants, namely B.1.617.1 (κ), B.1.617.2 (δ), and B.1.617.3. This lineage, which emerged in India in October 2020, has since spread to other countries, particularly in the United Kingdom, in settings with a high density of resident immigrants from India. Data from Public Health England registry shows that the subvariant B.1.617.2 (δ) has become epidemiologically prevalent.[22] Due to its subvariants composition, the variant has been designed as B.1.617+ by the WHO.[23]

A dichotomic behavior is observed with variant B.1.617.2 (δ). In contrast, it lacks mutation E484Q which is present in the other two lineages and was initially suspected to confer a certain degree of resistance to antibody neutralization. In contrast, this subvariant has a mutation at site 478 where a lysine replaces the proline. Of note, variant B.1.617.2 (δ) is indeed characterized by a major shift toward increased positive electrostatic potential because of three amino acid changes from negative or neutral to clearly positive charge, as shown in Section 3.

Subvariants B.1.617.1 (κ) and B.1.617.3 have both the double mutations E484Q and the L452R. Although initially believed to enhance the antibody-escape potential, it has been shown that the B.1.617.1 (κ) subvariant is pretty neutralized by the majority of sera from convalescent individuals and all sera from vaccinated subjects.[24] Nonetheless, this variant has been shown to be more pathogenic than the B.1.1.7 (α) variant in an experimental model of SARS-CoV-2 infection in hamsters.[25]

Our data indicate that most of the mutants are predicted to destabilize the RBD structure, except for E484K and T478K. It is conceivable that destabilization alters binding affinity to ACE2 and to neutralizing antibodies. As noticed above, the influence of the mutations on the spike surface properties is evident in the B.1.617+ lineage, particularly in the B.1.617.2 (δ) lineage subvariant, where the positive electrostatic surface potential is markedly increased. This may favor RBD interaction with the negatively charged ACE2,[26] which, in turn, would then increase affinity for the ACE2 receptor. All of these changes have the potential to eventually modify infectivity, pathogenicity, and virus spread. Regarding differential binding to neutralizing antibodies, previous studies suggested that VOCs RBD changes in the electrostatic potential surface could induce SARS-CoV-2 antibody evasion, and even single amino acid changes that marginally affect ACE2 affinity could greatly influence nAbs affinity.[27] Several factors have been demonstrated to affect the impact of VOCs. For example, it has been observed an increased effect at pH associated with nasal secretions (from 5.5 to 6.5).[28] For this reason, additional experiments both in vitro and in vivo are needed to establish the biological significance of SARS-CoV-2 mutations and how the interactions between mutations and local cellular microenvironment influence the clinical outcome and the transmission dynamics of this virus.

AUTHOR CONTRIBUTIONS

Stefano Pascarella and Martina Bianchi performed the in-silico analysis and proposed data interpretation. Antonio Cassone and Massimo Ciccozzi proposed addressing the issue of this study, wrote the initial draft of the manuscript, and contributed to writing the final version of the manuscript. Stefano Pascarella, Martina Bianchi, Antonio Cassone, Massimo Ciccozzi, Davide Zella, Francesca Benedetti, Francesco Broccolo, Roberto Cauda, Arnaldo Caruso, Silvia Angeletti, Marta Giovanetti, and Domenico Benvenuto discussed the data, their interpretation, and contributed to writing the final version of the manuscript. All Authors have seen and approved the initial draft and the revised, final text of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in bioRxiv at https://submit.biorxiv.org/.

ORCID

Stefano Pascarella http://orcid.org/0000-0002-6822-4022
Massimo Ciccozzi http://orcid.org/0000-0003-3866-9239
Davide Zella http://orcid.org/0000-0001-5576-5770
Martina Bianchi http://orcid.org/0000-0002-2547-5999
Francesca Benedetti http://orcid.org/0000-0001-8720-090X
Domenico Benvenuto http://orcid.org/0000-0003-3833-2927
Silvia Angeletti http://orcid.org/0000-0002-7393-8732
Marta Giovanetti http://orcid.org/0000-0002-5849-7326
Antonio Cassone http://orcid.org/0000-0002-0434-0761

REFERENCES

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020;395:P470-P473. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30185-9/fulltext
2. Mullard A. How COVID vaccines are being divided up around the world. Nature; 2020. https://doi.org/10.1038/d41586-020-03370-6
3. Fontanet A, Autran B, Lina B, Kieny MP, Abdol Karim SS, Shridar D. SARS-CoV-2 variants and ending the COVID-19 pandemic. Lancet. 2021;397:952-954.
4. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. Biochem Biophys Res Commun. 2020;525:135-140.
5. Galloway SE, Paul P, MacCannell DR, et al. Emergence of SARS-CoV-2 B.1.1.7 lineage—United States, December 29, 2020–January 12, 2021. MMWR Morb Mortal Wkly Rep. 2021;70(3):95-99.

XXIII

6. Voysey M, Clemens S, Madhi SA, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet.* 2020;397:99-111.

7. Volz E, Mishra S, Chand M, et al. Transmission of SARS-CoV-2 lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. *medRxiv.* 2021. https://doi.org/10.1101/2020.12.30.20249034

8. Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science.* 2021; 372:eabg3055. https://doi.org/10.1126/science.abg3055

9. Ferreira I, Datir R, Papa G, et al. SARS-CoV-2 B.1.617 emergence and sensitivity to vaccine-elicited antibodies. *bioRxiv.* 2021. https://doi.org/10.1101/2021.05.08.443253

10. Hoffmann M, Hofmann-Winkler H, N Krüger, et al. SARS-CoV-2 variant B.1.617 is resistant to Bamlanivimab and evades 2 antibodies induced by infection and vaccination. *bioRxiv.* 2021. https://doi.org/10.1101/2021.05.04.442663

11. Yadav PD, Mohandas S, Shete AM, et al. SARS CoV-2 variant B.1.617.1 is highly pathogenic in hamsters than B.1 variant. *bioRxiv.* 2021. https://doi.org/10.1101/2021.05.05.442760

12. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 2018;46(W1):W350-W355. https://doi.org/10.1093/nar/gky300

13. Pires DEV, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research,* 42:W314-W319. https://doi.org/10.1093/nar/gku411

14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA.* 2001;98(18):10037-10041. https://doi.org/10.1073/pnas.181342398

15. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020; 581(7807):215-220. https://doi.org/10.1038/s41586-020-2180-5

16. Bianchi M, Borsetti A, Ciccozzi M, Pascarella S. SARS-Cov-2 ORF3a: mutability and function. *Int J Biol Macromol.* 2021;170:820-826. https://doi.org/10.1016/j.ijbiomac.2020.12.142

17. Tegally H, Wilkinson E, Giovanetti M, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature.* 2021;592:438-443. https://doi.org/10.1038/s41586-021-03402-9

18. Benvenuto D, Demir AB, Giovanetti M, et al. Evidence for mutations in SARS-CoV- Italian isolates potentially affecting virus transmission. *J Med Virol.* 2020;92:2232-2237.

19. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell.* 2020;182:812-827.

20. Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med.* 2020;18(1):179. https://doi.org/10.1186/s12967-020-02344-6

21. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature.* 2020;592:116-121. https://doi.org/10.1038/s41586-020-2895-3

22. GOV.UK. Variants: distribution of cases data, 20 May 2021. https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data

23. WHO. Emergencies preparedness, response: SARS-CoV-2 variants; 2021. https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/

24. Doria-Rose N, Suthar MS, mRNA-1273 Study Group, et al. Antibody persistence through 6 months after the second dose of mRNA-1273 vaccine for Covid-19. *N Engl J Med.* 2021;384:2259-2261. https://doi.org/10.1056/NEJMc2103916

25. Abdelnabi R, Boudewijns R, Foo CS, et al. Comparative infectivity and pathogenesis of emerging SARS-CoV-2 variants in Syrian 2 hamsters. *bioRxiv.* 2021.

26. Amin SA, Jha T. Fight against novel coronavirus: A perspective of medicinal chemists. *Eur J Med Chem.* 2020;201:112559. https://doi.org/10.1016/j.ejmech.2020.112559

27. Ferraz M, Moreira E, Coêlho DF, Wallau G, Lins R. SARS-CoV-2 VOCs immune evasion from previously elicited neutralizing antibodies is mainly driven by lower cross-reactivity due to spike RBD electrostatic surface changes. *ChemRxiv.* 2020. https://doi.org/10.26434/chemrxiv.14343743.v1

28. Morton SP, Phillips JL. Computational electrostatics predict variations in SARS-CoV-2 spike and human ACE2 interactions. *bioRxiv.* 2020. https://doi.org/10.1101/2020.04.30.071175

XXIV