SARS-CoV-2 mutations in Brazil: from genomics to clinical conditions

L. F. S. M. Timmers,[1,2]* J. V. Peixoto,[3] R. G. Ducati,[1] J. F. R. Bachega,[4] L. M. Pereira,[5] R. A. Caceres,[4,6,7] F. Majolo,[1] G. L. da Silva,[8] D. B. Anton,[1] M. I. Goettert,[1,2] O. A. Dellagostin,[9] J. A. P. Henriques,[1,2] L. L. Xavier,[10] S. Laufer[11]

Affiliations

[1]Graduate Program in Biotechnology, Universidade do Vale do Taquari - Univates, Lajeado, RS, Brazil.
[2]Graduate Program in Medical Sciences, Universidade do Vale do Taquari - Univates, Lajeado, RS, Brazil.
[3]Graduate Program in Cellular and Molecular Biology, Federal University of Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.
[4]Department of Pharmacosciences, Federal University of Health Sciences of Porto Alegre - UFCSPA, Porto Alegre, RS, Brazil.
[5]Laboratory of Molecular Microbial Ecology, Federal University of Rio de Janeiro - UFRJ, Rio de Janeiro, RJ, Brazil.
[6]Graduate Program in Biosciences, Federal University of Health Sciences of Porto Alegre - UFCSPA, Porto Alegre, RS, Brazil.
[7]Graduate Program in Health Sciences, Federal University of Health Sciences of Porto Alegre - UFCSPA, Porto Alegre, RS, Brazil.
[8]Laboratory of Acarology, Tecnovates, Universidade do Vale do Taquari - Univates, Lajeado, RS, Brazil.
[9]Graduate Program in Biotechnology, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas - UFPel, Pelotas, RS, Brazil
[10]Laboratory of Cell and Tissue Biology, Pontifical Catholic University of Rio Grande do Sul - PUCRS, Porto Alegre, RS, Brazil.
[11]Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, University of Tübingen, Tübingen, Germany.

*Corresponding author: Luis F. S. M. Timmers (luis.timmers@univates.br)

Abstract

Due to the high rate of transmissibility, Brazil became the new COVID-19 outbreak epicenter, being a reference region to monitor how SARS-CoV-2 is mutating and spreading. Here, we combined genomic and structural biology analysis evaluate genomes isolated from different regions of Brazil. We showed that the most prevalent mutations were located in the S, N, ORF3a and ORF6 genes, which are involved in different stages of viral life cycle and its interaction with the host cells. Structural analysis brought to light the positions of these mutations on protein structures, helping studies of structure-based drug discovery and vaccine development.

Introduction

Early into 2020, a new coronavirus infectious disease (COVID-19) began to rapidly spread around the globe, pushing the World Health Organization (WHO) to declare this outbreak a public health emergency of international concern *(1)*. Following the first reported case, in Wuhan (on December 19th, 2019), more than 106 million people became infected

and 2.3 million died due to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) *(2)*. The etiologic agent of this disease is an enveloped single-stranded, positive-sense RNA virus, member of the betacoronavirus genus, presenting a genome of 29,903 kilobases in length *(3-5)*. Its genome organization consists of an ORF1ab encoding 16 non-structural proteins (NSPs), a Surface glycoprotein (S), an Envelope protein (E), a Membrane protein (M) and a Nucleocapsid protein (N) *(6)*. SARS-CoV-2 has been gaining massive attention from scientific researchers worldwide, who are putting a great amount of effort into elucidating the mechanisms behind its transmissibility and deciphering how it can be defeated. Distinct socio-economic impacts have been experienced by nations, mainly due to particular political actions in connection with the outbreak. China, Spain, Italy, USA and Brazil have assumed leading roles during the pandemic, as they were major epicenters at some point. Therefore, extra attention should be given to these countries, perhaps by monitoring the changes in the SARS-CoV-2 genome due to the high rate of transmissibility.

In order to monitor the presence of new mutations in the SARS-CoV-2 genome, we applied genomic and structural biology analysis to evaluate lineages isolated from different regions of Brazil, and mapped where such mutations are located in the protein structures. This information could help understand the impact of these mutations on the stability of the viral proteins, the efficacy of vaccines and to monitor how different are the viruses when compared to other regions.

Results and Discussion

Genomic and structural analysis have brought to light the diversity of the SARS-CoV-2 genome. A large number of mutations has been identified within different regions of the world, and the identification of such mutation sites could potentially help the scientific community to build bridges between molecular and disease manifestations. A spike glycoprotein specific mutation (D614G) was the first mutation described to come across advantage/election under positive selection, being associated with a higher ability of the virus to infect cells *(7)*. However, this was not the only mutation identified, as new mutations have been continuously observed, and high rates of transmissibility were reported in countries such as USA, Italy, South Africa and Spain, which could be associated with a higher number of mutations in the SARS-CoV-2 genome. In addition, the identification of new mutations has been receiving increasing attention due to the evaluation of vaccine efficacy. As more data is being gathered on viral behavior, we proceeded to annotate and carry out a structural characterization on the distribution of mutations present in genomes isolated from different regions in order to understand if there are novel mutations when compared to other countries. We also analyzed the presence of mutations in the spike glycoprotein, and performed analysis on all Brazilian sequences presented in the GISAID database.

Spike glycoprotein

The spike glycoprotein is responsible for interacting with the human Angiotensin-converting enzyme 2 (hACE2) driving the process of cell invasion *(8-10)*. It can be divided

into three functional domains: N-terminal domain (NTD), receptor-binding domain (RBD) and S2 domain. Due to its importance in the process of cell recognition and vaccine development, RBD has attracted more attention when it comes to mutations. According to our analysis, we identify the presence of seven mutations in the interface between RBD and hACE2, which are T470A, A475V, S477N, V483F, E484K, N501Y and G502N (**Figure 1A**). In addition, combinations of two mutations (S477N+E484K, E484K+N501Y and V483F+E484K) were also observed in sequences from Rio Grande do Sul, Amazonas and São Paulo. Among these, A475V has been associated with decreased sensitivity to neutralizing monoclonal antibodies (*11*). Also, S477N, N484K and N501Y were reported to strengthen binding with hACE2 (*12,13*). Four out of seven identified mutations are located at a flexible loop of RBD that could be important for binding stabilization with hACE2. In addition, mutations in RBD are important to be monitored due to its possible interferences in the efficacy of recombinant spike protein vaccines. Regarding the other domains, 15 point mutations (L5F, S12F, V16F, N74K, S155I, D198Y, D614G, A647S, A684V, M731I, L878S, F1109L, V1129L, V1176F and K1191N) were observed (**Figure 1B**). Five of these mutations (L5F, S12F, V16F, N74K and S155I) have surface exposure. Although these residues are not in the RBD domain, the proximity could have an impact on cell interaction. Korber and coworkers (*7*) previously identified the L5F mutation in 13 different countries, and our analysis shows that Brazil also presents this particular mutation. In addition, mutations on hydrophobic residues in the signal peptide sequence (L5F, S12F, V16F) have been associated to the entry of Spike glycoprotein into the endoplasmic reticulum for viral folding and assembly (*14,15*). According to our site-wise evolutionary rate analysis (Supplementary materials), these positions are more likely to mutate and among these D614G was previously described to increase infectivity rates (*7*). In addition, five of these positions were predicted as structure stabilizer mutations, V16F, A647S, M731I, L878S and V1129L, presenting ΔΔG values of 2.35, 1.18, 1.18, 4.55 and 1.03, respectively. A647S and V1129L are located at the subunit interface, being potentially involved in the stabilization of the quaternary structure, whereas V16F, M731I and L878S are involved in inter-domain interactions that could present a promotion in tertiary structure stability. In addition to the structural impacts, some mutations require special attention when the goal is to develop vaccines; L5F and K1191N are in sites recognized by Human Leukocyte Antigen (HLA) according to Campbell et al. (*16*).

ORF8 protein

The ORF8 protein, a member of the betacoronavirus NS8 family, is composed of 121 amino acids. Its function has been related to viral pathogenicity or replication acting on the host interferon pathway (*17-20*). Members of this family present two conserved motifs ranging from 34EDPCP38 and 88INCQ91. Interestingly, SARS-CoV-2 has a mutation on E34D, changing the sequence motif to 34DDPCP38, which may not reflect on a larger impact on protein structure as both amino acids share the same physicochemical properties. On the other hand, we identified six amino acid positions with higher evolutionary rates (P36S, V62L, S69P, I76F, L84S and R115H); V62L was identified in eleven other countries

and L84S has been widely observed throughout the world (**Figure 2A**). Zhang and coworkers *(21)* suggest that the ORF8 protein is able to interfere with the antigen presentation, affecting the process of recognition and elimination of the virus by cytotoxic T cells.

We carried out docking and molecular dynamics simulations to understand how ORF8 protein and MHC-I interact. **Figure 2B** presents the residues involved in these protein-protein interactions. Our analysis showed that 11 residues of the ORF8 protein (H28, Y46, Y53, Q72, I74, D75, N78, 79Y, 80T, E106 and H112) are involved in hydrogen bonding to the MHC-I molecule. In order to gain insight and to evaluate the stability of the ORF8:MHC-I interaction, we applied molecular dynamics simulations. We specifically analyzed the most representative conformations obtained during the simulation time. Here, cluster 1 presents values of 2.38 nm and 0.28 nm, whereas cluster 2 shows 2.34 and 0.40 of radius of gyration and root mean square deviation, respectively (Supplementary Materials). Analysis of protein flexibility shows that the recognition region of the ORF8:MHC-I complex is stable during the simulation time, supporting docking experiments on the binding site (Supplementary Materials). Although we did not observe larger conformational changes between both clusters, it is interesting to highlight that residues ranging from Y73 to V81 are maintaining hydrogen bonds with the MHC-I molecule. This region is close to two mutation sites (I76F and L84S) observed in isolates from different regions of Brazil (and throughout the world); L84S belongs to the Cytotoxic T Lymphocyte (CTL) linear epitope predicted to be recognized by HLA.

Envelope protein

The envelope (E) protein is a small protein composed of 76-109 amino acids, highly conserved among coronaviruses. E protein functions as an ion-channeling viroporin, disrupting host membranes and allowing viral release. It also interacts with other CoV and host proteins, participating in viral assembly and budding. It is well known that E protein interacts with membrane (M) protein to form the envelope *(22-25)*. E protein can be detected in abundance inside infected cells during replication, but only a small portion is incorporated into the viral envelope *(25)*. Inside the host cells, E protein is located mainly in the Golgi, endoplasmic reticulum (ER) and ERGIC (ER-Golgi intermediate compartment) - location of intracellular trafficking between ER and Golgi complex. There, it acts on viral assembly and budding, being essential for virion release and, therefore, considered critical for replication *(26)*.

The E protein is composed of three main domains, the N-terminal (NT), C-terminal (CT) and transmembrane domain (TMD) *(25,27)*. TMD, located from residues 10 to 35 *(28)*, is responsible for the construction of the ion channel. TMD also possesses hydrophobic regions that interact with membrane lipids anchoring the protein into the membrane *(23)*. The CT domain is a key region for the interaction with M protein, where this association promotes membrane curvature assuring the envelope formation *(25)*. Furthermore, the CT D-L-L-V motif, also called PDZ-binding motif (PBM), interacts with the PDZ domain of

PALS1 (Protein Associated with *Caenorhabditis elegans* Lin-7 protein 1), disrupting tight junctions in epithelial cells by relocation of the protein, and, therefore, altering cell polarity and permeability *(29)*. We have identified three point mutations in the E protein sequence, T9I, V52E and P71L, assigned in **Figure 3A**. T9I is located at the negatively charged NT adjacent to TMD (residues 10-35). This mutation indicates a change between threonine (hydrophilic) and isoleucine (hydrophobic). It is characteristic of viroporins to present positively charged residues in the structure, usually lysine and arginine, which assures the anchoring to the membrane by interacting with negatively charged phospholipids *(23,30)*. The NT function is still not clear, but NT of viroporins are probably important for ER targeting, together with the TMD hydrophobic region *(31)*. The substitution of threonine by isoleucine could indicate an enhanced power of interaction with the membrane lipids due to its hydrophobic characteristic and its localization; this could also modify positively the capacity of membrane attachment and ER targeting by E protein. The V52E and P71L mutations are located in the CT domain. V52E shows a change from valine, a hydrophobic residue, to glutamic acid, a hydrophilic and slightly bigger residue. V52E is in a position in the CT which interacts with the membrane lipids *(32)* and could alter the interaction with the membrane. P71L is adjacent to the D-L-L-V motif (residues 72-75), a region that binds to host protein PALS1, facilitating infection. P71L is a mutation that changes a hydrophobic residue into another hydrophobic residue, and has been previously reported in samples collected in the USA *(22,24)*. It probably promotes none to little effect in the protein structure. Although it is possible that the P71L mutation could have some effect on the binding dynamics of the D-L-L-V motif due to its adjacent position, it is less likely to be a deleterious effect considering how frequently this point mutation is found.

Nucleocapsid protein

The SARS-CoV-2 nucleocapsid (N) protein is a multifunctional RNA-binding protein involved in essential processes regarding viral RNA transcription and replication. This protein is composed of 419 amino acids and its structure can be divided into three main domains: (i) N-terminal domain (NTD), where the binding of RNA molecules takes place, (ii) a linker domain rich on serine and arginine residues (SR-rich linker), and (iii) a C-terminal domain (CTD) (**Figure 3B**), which is also involved in RNA binding modulation. Besides being involved in RNA binding modulation, CTD is also required for the process of protein oligomerization, an important step to protect the viral genomic RNA. The role of the N protein to the viral life cycle and host infection has been discussed by many research groups, who have suggested its involvement on the process of impairment of reproductive and hematopoietic systems via abolishment of pluripotency in human induced pluripotent stem cells *(33)*, evasion of immune system *(34)*, and signaling pathway *(35,36)*. Due to its involvement in different cellular processes, mutations on the N protein should be closely monitored. In addition, nuclear localization signals on the N protein were suggested to be associated with high case fatality rates by coronaviruses *(37)*, and the accumulation of positive charges could present a contribution to viral pathogenicity; however, we observed no mutations on positive residues in our analysis. Overall, we identified 21 mutation sites,

two (T76I and P80S) located at the RNA-binding domain, six (S188P, S197L, R203K, G204R, G238C and G243C) located at the (SR)-rich linker, six (A251V, E253D, A254T, I292T, T296I and S327L) located at CTD, and seven (P13L, I15V, P20L, G25D, R32H, S37P and A398V) located at intrinsic disorder regions. It is interesting to note that the N protein presents the highest number of mutational sites together with ORF3a. Among these mutations, P13L, S188P, S197L, R203K and G204R were widely observed worldwide. The Brazilian sequences P13L, I15V, P20L, G25D, P80S, G238C, G243C and I292T show high evolutionary rates when compared to the other sequence positions. Analysis of the impact of these mutations on protein structure suggests that P80S, A251V and I292T are highly stabilizing, whereas E253D and S327L could be associated with destabilizing mutations. Another important information is that P13L and P80S were predicted to belong to sites of the CTL linear epitopes *(16)*. Therefore, the high variability of these regions should be taken into account during the process of vaccine development. Furthermore, the 168-208 region was suggested to be critical for viral assembly and the maturation process to SARS-CoV *(38)*, so the impact of S188P, S197L, R203K and G204R mutations should be carefully evaluated.

Membrane protein

The membrane (M) protein is composed of 221 amino acids in three transmembrane helices and a cytoplasmic C-terminal domain (**Figure 3C**) and is associated with viral assembly. Specific phosphorylation sites in the M protein have been suggested to be involved in interactions with the host's cells *(36)*. According to our results, the M protein presents lower sequence variability, which is in agreement with previous work *(39)*, suggesting that this protein could be associated with housekeeping functions. The mutations are concentrated in the cytoplasmic C-terminal domain at positions P132A, T175M and R186C, and another one at the third transmembrane helix (V70F). Site-wise evolutionary analysis suggests that positions P132A and R186C show higher rates of variability, where P132A seems to be a high stabilizer, whereas T175M a destabilizing mutation regarding the tertiary structure. In addition, T175M and R186C mutations are located at the protein surface, which could lead to an impact on protein-protein association since M and N proteins are expected to interact in order to complete viral assembly.

OFR7a protein

The accessory ORF7a is a type I transmembrane protein composed of 121 amino acids and organized in a transmembrane helix and a luminal domain. Together with nsp1 and nsp3c, ORF7a was suggested to interact with the innate immune system binding directly to BST-2 *(40)*. Due to its possible mechanism, acting as a virulence factor, any mutation should be carefully analyzed. According to our results, six mutations were observed in sequences isolated from different regions of Brazil, L9P, L10I, E22D, P34S, P84L and V93F (Supplementary Materials). All mutations are located at the luminal domain; however, due to the lack of structural information, we were unable to evaluate the impact on the protein structure for L9P, L10I, P84L and V93F. It is important to highlight that L10I and V93F

belong to sites of CTL linear epitopes *(16)*, and should be avoided during the vaccine development process. P34S was observed as a high stabilizer mutation, since the presence of a polar side chain could allow hydrogen bond formation to residues H73 and/or Y75.

ORF6 protein

ORF6 is a small protein composed of 61 residues whose structure is not experimentally known. Among the 27 viral proteins in the coronavirus, the ORF6 protein demonstrated a stronger suppression on primary interferon production and interferon signaling, corroborating previous studies including those on SARS-CoV that characterized the antagonist potential of that protein. Yuen and coworkers pointed out the ORF6 deletion as an alternative for the production of an attenuated virus-based vaccine *(41)*. Due to its importance in innate immune suppression, we evaluated the interaction of ORF6 against the transcription factor IRF3 using docking and MD simulations. Our results suggest that this interaction is stable during the simulation time, and is mainly guided by hydrophobic interactions maintained by residues at the C-terminal of ORF6 (R20, V24, Q56, M58, E59 and D61) (Supplementary materials). In addition, site-wise evolutionary rate analysis shows that positions E13D, V24F, I32F, I33T and D53Y are more variable (Supplementary materials). **Figure 4** shows the position of these mutations in the model structure of ORF6. Among these positions, it is important to highlight V24F, which could have impact on the interaction ORF6:IRF3, as this position is located at the interface, and D53Y, which was predicted to belong to CTL linear epitopes *(16)*.

ORF3a protein

The ORF3a protein is an ion transporter composed of 275 amino acids that was associated with different processes during the viral life cycle, such as replication and virulence *(42-44)*. Its importance in viral pathogenesis was confirmed by Tan and coworkers, who observed that the ORF3a protein up-regulates fibrinogen secretion, which could promote a cytokine storm *(45)*; however, this may be a result of a cooperative effect based on the interaction among other viral proteins. In addition, this protein was also sufficient to activate NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3) inflammasome by promoting TNF receptor-associated factor 3 (TRAF3)-dependent ubiquitination of caspase recruitment domain *(44)*, being an important sensor for the inflammatory stage of COVID-19. Our analysis revealed 21 mutations (T9I, A23V, T24I, T34K, A54V, Q57H, K67N, R68I, A72T, L94I, L108F, A110V, R134L, D155Y, G172C, G196V, T217I, P240L, E241A, S253P and T270I) on the ORF3a protein, where three positions (G172C, G196V and P240L) presented high evolutionary rates (Supplementary materials). **Figure 5** shows these mutations on the structure of the ORF3a protein. These mutations are widely spread through the protein structure, six mutations being located at the transmembrane region and another one at the di-acidic motif. It is important to highlight that none of these mutations were observed at the TRAF3-binding motif. However, mutations A23V, T24I and T34K are close to this motif and the impact on ORF3a:TRAF3 interaction should be further investigated. Three mutations (A54V, Q57H and L108F) are at the

interface of the subunits and may have effects on protein dimerization. In addition, mutations Q57H, A110V, R134L and P240L are predicted to belong to sites of CTL linear epitopes *(16)*.

ORF1a polyprotein

The ORF1a polyprotein is cleaved during the virus life cycle into 11 non-structural proteins, which are the leader protein, NSP2, NSP3, NSP4, 3CL-like protease, NSP6, NSP7, NSP8, NSP9, NSP10 and NSP11. Among these, 3C-like protease is the main target for antiviral development against SARS-CoV-2. Sequence analysis shows that the leader protein, NSP6, NSP9, NSP10 and NSP11 present no mutations when compared to the first genome sequence from Wuhan. However, the other non-structural proteins present at least one mutation. The 3CL-like protease presents two point mutations at position K90R and L205V, and are associated with positive selection. It is important to highlight that no mutation is observed near to the active site of the main protease. NSP7 also shows one mutation at position L71F, which has a high frequency in the Brazilians' genomes. NSP8 presents three point mutations (R51C, L95F and A194V), where A194V belongs to sites of CTL linear epitopes *(16)*; this position was associated with positive selection. Therefore, since the functions of NSP7 and NSP8 were involved in the process of making new copies of viral RNA, any mutation in these proteins should be carefully evaluated. NSP4 presents two mutations, A307Y and H313Y, which were not predicted to be involved on sites of linear epitopes. Site-wise evolutionary rate analysis shows that NSP2 and NSP3 proteins accounted for most of the mutations in ORF1a, nine being observed in NSP2 and 11 in NSP3 (Supplementary materials). Among all these mutations, only A194V of the NSP2 was described as being associated with positive selection and to belong to CTL linear epitopes *(16)*.

SARS-CoV-2 and the inflammatory process

In general, as cells become infected by viral particles, transcription factors (including NF-κB) start to stimulate the expression of IFN-I, a mechanism activated by our innate immune system against viruses. The secreted IFN molecules bind to their specific receptors, initiating the activation of the JAK/STAT pathway and inducing the translocation of transcription factors (responsive to IFN) to the nucleus, establishing an antiviral state *(46)*. The SARS-CoV-2 infection mechanism, however, uses a selective process that prevents the induction of the IFN-I system, inactivating/blocking antagonistic members responsive for the viral replication system *(47)* (**Figure 6**).

SARS-CoV-2 ORF6, ORF8, N and ORF3 are potent INF antagonists acting on the early stages of infection. A recent study indicated that their relation to the delayed release of IFNs would hinder the host's antiviral response and benefit viral replication *(47)*. These proteins are not only capable of inhibiting the IFN-I signaling pathway via NF-κB, but are also related to the inflammation activation mechanism *(47)*. ORF6 displays a critical antagonistic activity, avoiding the antiviral innate immune response, inhibiting interferon β (IFN-β) production and

blocking the expression of STAT1-activated genes to SARS-CoV *(47,48)*. ORF8 has been considered one of the most relevant proteins related to viral replication in humans, as it is able to inhibit the expression of IFN-β *(47,49)*. The SARS-CoV-2 N protein acts as an IFN-I antagonist regulating the synthesis and signaling of this route. SARS-CoV has an additional important function that is critical in its antagonism to IFN-I during the viral life cycle: in the initiation stage, the N protein encapsulates the viral genomic RNA, protecting and allowing viral replication *(47)* (**Figure 6**). Apoptosis, a predominant type of programmed cell death, acts as an important antiviral defense mechanism of the host to control viral infection, being directed to the regulation of inflammatory response. ORF3, which is known for its pro-apoptotic activity, could be potentially associated with a reduction in apoptosis-mediated antiviral defense in infected cells, which could confer an advantage for SARS-CoV-2, as the infection can be relatively mild or even asymptomatic during the early stages, thus allowing the virus to spread more widely *(50)*. Therefore, this process is in many ways related to the induction of NF-kB, interfering in the IFN-1 route. NF-kB is a transcription factor for the formation of pro-inflammatory cytokines, so when the NF-kB activation pathway is inhibited by the virus, another pro-inflammatory response, such as caspase-1, affects the transcription of inflammatory cytokine genes *(42,50)* (**Figure 6**).

Another group of proteins which has been associated with SARS-CoV-2 infection and has gained attention from researchers is composed of mitogen-activated protein kinases (MAPKs), enzymes that affect the IFN-I pathway and act directly on inflammatory elements *(36)*. Through SARS-CoV-2 infection and pro-inflammatory cytokines resulting from the action of macrophages and other antigen-presenting cells (APCs), activation of the p38/MAPK pathway acts on NF-kB, STAT pathway and promotes greater expression for pro-inflammatory cytokines such as IL-6 and TNF-α *(36,51)*. Experimental evidence has shown that in cured SARS-CoV-2 infected patients the interferon signal was just mildly activated when compared to severe and moderate infections, and that suppression of the signal pathway related to mRNA translation of the virus occurred. Therefore, the down-regulation of the MAPK signaling pathway can be indicative of patient recovery *(52)*.

In Vero E6 cells (primate kidney cell line), SARS-CoV-2 infection activated the p38/MAPK signaling pathway and the production of various cytokines, and shutdown mitotic kinases, causing cell cycle arrest. Also, there was an upregulation of several components of the p38/MAPK signaling pathway and an increase in phosphorylation of substrates of the p38 pathway in places where regulation occurs 24 hours after infection. Therefore, this may be reflecting the more advanced stage of viral infection, replication and egress. Among the substrates, we can highlight heat shock protein beta-1 (HSPB1), cAMP response element-binding protein (CREB), transcription factor-1 (ATF-1) and signal transducer and activator of transcription 1-alpha/beta (STAT1) *(36)* (**Figure 6**). It is still unclear how the inhibition of p38/MAPK could suppress the production of cytokines and impair viral replication during SARS-CoV-2 infection. However, it is possible that the inhibition of p38/MAPK presents multiple mechanisms related to the pathogenesis of COVID-19 *(36)*. Thus, the systematic

elimination of IFN modulating functions of the virus becomes an attractive approach for the development of therapies against SARS-CoV-2.

The processes described above corroborate with an alteration in the inflammatory process together with viral replication *(53)*, triggering the patient's clinical condition. It is important to mention that the elements of inactivation of viral IFN antagonists are an approach for acquiring live attenuated vaccines, and that it has already been studied for influenza A (H1N1) and yellow fever viruses *(54,55)*.

SARS-CoV-2 and vaccine development

Several vaccines are being developed against SARS-Cov-2, being currently on clinical trials or already approved for emergency use. There are three main types of SARS-CoV-2 vaccines: inactivated vaccines (Sinovac), RNA vaccines (Pfizer/BioNTech and Moderna) and non-replicating vector vaccines (Oxford/AstraZeneca and Gamaleya Research Institute). The well established technology of inactivated vaccines was used to produce the CoronaVac designed by Sinovac. This vaccine is composed of inactivated viruses from viral propagation in cells infected with SARS-CoV-2 (*Zhang, Zeng, Pan et al 2021*). The clinical trials with CoronaVac demonstrated that it is well tolerated and showed 50.38% efficacy in phase 3, being approved for emergency use in China, Brazil, Indonesia and Chile *(56)*.

Vaccines produced based on mRNA or adenovirus vector technologies, and designed with the gene encoding the SARS-CoV-2 Spike glycoprotein, are presenting higher efficacy than inactivated vaccines, ranging from 62% to 95% efficacy *(57,58)*. However, since the emergence of SARS-CoV-2 variants, with mutations in the RBD of the Spike glycoprotein, uncertainties have taken place regarding the effectiveness of vaccines against these variants. SARS-CoV-2 variants include the B.1.1.7 from the United Kingdom, the B.1.351 from South Africa and P.1 - branch off the B.1.1.28 lineage, from Brazil. The B.1.1.7 variant has eight mutations located in the Spike glycoprotein, including N501Y in RBD domain *(59)*, while the B.1.351 variant has nine mutations located in the same protein, including E484K and N501Y *(60)*. The P.1 lineage contains K417T, E484K and N501Y mutations in the RBD of the Spike glycoprotein *(61)*. Specially for those vaccines designed before the appearance of these variants, mutations in the RBD of the Spike glycoprotein could reflect in loss of vaccine effectiveness. In addition, new mutations are being identified in the RBD such as T470A, A475V, S477N and G502N.

mRNA-based SARS-CoV-2 vaccines exhibit a tolerability and safety profile *(62)*. There are four types of vaccines developed by Pfizer/BioNTech: BNT162a1, BNT162b1, BNT162b2 and BNT162c2. In two of them, the sequence of Spike glycoprotein is included, while the others contain an optimized RBD. The BNT162b2 is a vaccine of nucleoside-modified mRNA (modRNA) with the full-length of SARS-CoV-2 Spike glycoprotein, encapsulated in lipid nanoparticles. Although this vaccine showed a 95% efficacy (in phase 3) in preventing Covid-19 *(58)*, its efficacy is not known against new SARS-CoV-2 variants. The analysis of the sera from individuals vaccinated with one dose of BNT162b2 demonstrated that mutations N501Y, A570D and 69/70 deletion combined did not appear to influence the efficacy of the vaccine *(63)*. Similar, small effects on neutralization by sera produced by two BNT162b2 doses were observed in SARS-CoV-2 containing the key Spike glycoprotein mutations: N501Y; or 69/70 deletion, N501Y and D614G combination; or E484K, N501Y and D614G combination, indicating that the efficacy of the vaccine may be not be affected by new

variants containing these mutations *(64)*. However, ten of 15 individuals who received the first dose of BNT162b2 vaccine after three weeks had decrease in efficacy of antibodies against the full B.1.1.7 variant (fold change varying between 1.8 and 6.1, with a median of 3.85-fold) *(63)*. Moreover, an additional decrease in the neutralization of vaccine-elicited antibodies was observed against the B.1.1.7 variant, which presents the mutation E484K *(63)*.

The vaccine mRNA-1273 developed by Moderna is a LNP-encapsulated mRNA encoding for a full-length Spike glycoprotein. In phase 3 clinical trials, the mRNA-1273 vaccine demonstrated 94.1% efficacy for preventing Covid-19 *(57)*. Studies revealed that this vaccine may also be effective against B.1.1.7 and B.1.351 strains *(65,66)*, but with a 6-fold decrease in neutralizing levels against the B.1.351 variant from South Africa *(65)*. Nonetheless, these neutralizing titer levels are still high and the two doses of mRNA-1273 are expected to be protective against the new SARS-CoV-2 strains *(65)*. Reduction of approximately 2-fold in the neutralization against variant B.1.1.7 was also observed in serum samples from individuals who received the mRNA-1273 vaccine *(67)*. SARS-CoV-2 variants carrying the mutations E484K, N501Y or the combination of K417N + E484K + N501Y might be responsible for affecting the effectiveness of the mRNA-1273 vaccine. Individuals who received two doses of mRNA-1273 showed a significant decrease in neutralization activity against variants encoding E484K (1 to 3-fold reduction), N501Y (1.3 to 2.5-fold reduction) and the K417N + E484K + N501Y combination (1.1 to 3-fold reduction) *(68)*.

Differently from mRNA vaccines, the Covishield (Oxford/AstraZeneca) and Sputnik V (Gamaleya Research Institute) vaccines are based on adenovirus vectors. Covishield (ChAdOx1 nCoV-19) was developed with an engineered chimpanzee adenovirus containing the genetic sequence of the Spike glycoprotein, and it showed 82.4% efficacy after two doses *(69)*. The Covishield vaccine appears to have similar effectiveness against the B.1.1.7 strain of SARS-CoV-2 *(70)*. However, the viral neutralisation by sera induced by the Covishield vaccine might be significantly reduced against the B.1.351 variant from South Africa, when compared to the original virus strain, providing a minimal protection *(71)*. Due to its lack of efficacy against the B.1.351 variant, Covishield should be carefully evaluated and monitored in Brazil, since mutations T470A, A475V, S477N, V483F, E484K, N501Y, G502N and a combination of two mutations (S477N+E484K, E484K+N501Y and V483F+E484K) have been identified and could also have impact on vaccine efficacy. Moreover, the P.1 variant from Brazil also contains K417T, E484K and N501Y mutations in the Spike glycoprotein *(61)*. However, no study has been performed to evaluate the effectiveness of the Covishield vaccine in relation to mutations isolate and P.1 variant.

Recently, the company Novavax released preliminary data on clinical trials with its NVX-CoV2373 vaccine, a protein-based vaccine that uses a recombinant SARS-CoV-2 nanoparticle composed of trimeric Spike glycoprotein and a Matrix-M1 adjuvant. This vaccine appeared to be safe and induced immune responses against Covid-19 *(72)*. Preliminary results already include data related to the new SARS-CoV-2 variants. The vaccine is 95.6% effective against the original variant of SARS-CoV-2, while it has 85.6% protection against variants B.1.1.7 and 60% protection against B.1.351 *(73)*.

Conclusion

According to our analysis, SARS-CoV-2 genomes isolated from different regions of the country present a high number of mutations distributed through several viral proteins. We observed a disproportionality in the origin of these sequences throughout the Brazilian territory. It is worth mentioning that Brazil is a continental size country represented by diverse

customs and population profiles. In 537 out of 627 genomes it is possible to check its origin. São Paulo (SP) and Rio de Janeiro (RJ) have the highest concentration of the sequenced genomes, 218 and 120, respectively, while other states present only one sequenced genome (Supplementary materials). It is possible that the 90 genomes whose origin was not confirmed belong to one of these states, although the information deposited in the database is not clear.

The mechanism by which the virus can develop, replicate and manifest itself with varying degrees of intensity within different Brazilian regions is still an impasse, because of our admixed populations, socio-economic variability and the influence of the environment, such as climatic conditions *(74-76)*. Due to the high number of cases, Brazil has become an important reference point to monitor the appearance of new mutations and to evaluate the efficacy of vaccines that are under clinical trials. It is important to highlight the high number of mutations in structural proteins such as S and N, as they have been used as targets for vaccine development and for molecular-based detection kits. Among the mutations observed in the SARS-CoV-2 genome, we should be aware of the appearance of (not single, but) double mutations at the interface of hACE2:RBD as observed in individuals from Rio Grande do Sul, Amazonas and São Paulo. These combined mutations could be a new challenge for the efficacy of vaccines. Another important consideration is the lack of public policies and the lack of adherence to social distancing by the Brazilian population, which could increase the probability of the emergence of new variants. Overall, sequence analysis revealed that the mutations observed in different regions of Brazil are positioned at sites of CTL linear epitopes and should be further investigated regarding their possible impacts on the development of new vaccines. In addition, we only described the impact of mutations in the proteins as they were isolated; however, the assembly process of the virion requires protein-protein interactions that are difficult to evaluate based on computational methods. Based on our analysis, we were able to highlight SARS-CoV-2 proteins with the highest evolutionary rates and point out where their mutations are located in the protein structure. This data shows how structural biology, combined with genomics, can be applied to better understand the viral variability, and may be useful in studies of structural-based drug discovery and vaccine development. In addition, the clinical practice will be a mirror for these genetic mutations of the virus where different patterns of clinical manifestations will probably be observed. Important clinical implications must be understood about admixture populations within geographical regions from different parts of the world for the interpretation and design of clinical trials. This heterogeneity influences the practice of clinical genetics, genomic medicine and drug prescriptions when compared to homogeneous populations.

Materials and Methods

Genome data, Open Reading Frames (ORF) prediction and functional annotation

The complete sequences were obtained from the GISIAD database. The viral's longest ORF (coding sequences and amino acids) was obtained according to the Markov Model with TransDecoder *(77)*. The parameters for prediction of Long ORFs and transcripts were modified to keep viral sequences with more than 15 amino acids' length. To refine and identify all ORF with functional significance, all genomes were annotated according to the sequence of SARS-CoV-2 from Wuhan (NC_045512). The sequences were annotated with OmicsBox

using the proteome of SARS-CoV-2 from Wuhan as reference and the tools from EggNOGg mapper, InterProScan and Gene Ontology.

Measure of amino acid rates

   In order to calculate site-specific evolutionary rates for all annotated sequences, we used the Likelihood Estimation of Individual Site Rates (LEISR) method, implemented into the HyPhy program, following the protocol described by Sydykova et al. *(78)*. All sequence alignments and phylogenetic trees were performed using the MAFFT online program with default parameters.

Crystallographic structures and comparative modeling

   In order to identify where the mutations are located in the protein structures, we used the PDB IDs 6VSB (Spike glycoprotein), 6YUN and 6VYO (Nucleocapsid), 6XDC (ORF3a) and 6W37 (ORF7a). We built the models for the sequences that do not have solved structures and for the mutations observed near the active sites and subunit interfaces using the MODELLER 9v23 program. The comparative modeling protocol consisted of the generation of 10 models, out of which one model for each gene sequence was selected. All models were submitted to the DOPE energy scoring function *(79)* implemented in the MODELLER 9v23 program aiming to select the best structures. MOLPROBITY webserver *(80)* and PROCHECK *(81)* were used to verify and validate the stereo chemical quality of the models.

Molecular dynamics simulations

   The molecular mechanics (MM) and molecular dynamics (MD) simulations of the ORF8:MHC-I and ORF6:IRF3 enzyme complexes were carried out using the Amber99SB force field *(82)* implemented in GROMACS 2016.3 *(83)*. Periodic bound conditions were applied, and the number of particles, pressure and temperature were maintained constant (NPT ensemble) during the production phase. Periodic boundary boxes of the ORF8:MHC-I and ORF6:IRF3 systems were composed of 24,538 and 20,857 water molecules, respectively. A V-rescale *(84)* thermostat was employed to maintain the system at constant temperature using a coupling time of 0.1 ps, and the Berendsen barostat *(85)* was applied to ensure the system pressure was maintained at 1 bar. The LINCS algorithm *(86)* was implemented to constrain all the covalent bonds involving hydrogen atoms so that the systems could evolve in a time step of 2 fs. Van der Waals interactions were computed using a 12.0 Å cutoff. The particle mesh Ewald (PME) method was used to calculate electrostatic contributions in a grid with 1.2 Å spacing. The macromolecule was fully solvated using the TIP3P model *(87)* in a cubic box extending 10.0 Å from the macromolecule surfaces. The systems were submitted to steepest-descent energy minimization up to a tolerance of 1,000 kJ mol-1 nm-1 to remove close contacts of Van der Waals forces. The equilibration phase lasted for 5 ns followed by the 100 ns of production phase. All MD analyses were performed with the GROMACS programs and Geo-Measures plugin *(88)*.

Protein-protein docking simulations

      In order to evaluate the best conformation and binding affinities of the ORF8:MHC-I and ORF6:IRF3 complexes, we applied a flexible protein-protein approach implemented in HADDOCK program *(89, 90)*. The (active) residues involved on protein-protein interactions were predicted according to the CPORT analysis, where the active residues for ORF8 were M1, L7, G8, I10, T11, T12, A15, F16, E19, P30, Y31, V32, V33, D34, D35, P36, C37, P38, I39, H40, F41, Y42, S43, K44, R48, R52, P56, L57, E59, V62, Y73, D75, I76, Y79, T80, S103, F104, Y105, E106, D107, F108, L109, Y111, F120 and I121, while for ORF6 were M1, F2, F7, A12, E13, L15, L16, I17, M19, R20, T21, F22, K23, V24, S25, I26, W27, N28, D30, Y31, I32, N34, L35, I36, I37, L40, I60 and D61. The docking protocol in HADDOCK comprises three main steps, which are (i) rigid body energy minimization, (ii) simulated annealing and (iii) refinement in explicit solvent. The best complexes were selected for further analysis based on HADDOCK score.

H2: Supplementary Materials

References and Notes

1. WHO. WHO announces COVID-19 outbreak a pandemic. http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic.

2. Worldometer. COVID-19 Coronavirus pandemic. https://www.worldometers.info/coronavirus/ accessed April 5th 2020.

3. A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A Gulyaeva, B.L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L.L.M. Poon, D.V. Samborskiy, I.A. Sidorov, I. Sola, J. Ziebuhr, The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5, 536-544 (2020).

4. X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, Z. Qian, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat. Commun. 11, 1620 (2020).

5. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F.Z. L. Wang, T.H., H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E.C. Holmes, G.F. Gao, G. Wu, W.

Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395, 565-574 (2020).

6. F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.HG. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E.C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China. Nature 579, 265-269 (2020).

7. B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, on behalf of the Sheffield COVID-19 Genomics Group, C. McDanal, L.G. Perez, H. Tan, A. Moon-Walker, S.P. Whelan, C.C. LaBranche, E.O. Saphire, D.C. Montefiori, Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell (2020).

8. Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K. Yuen, Q. Wang, H. Zhou, J. Yan, J. Qui, Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 181, 894–904 (2020).

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

10. S. Laha, J. Chakraborty, S. Das, S.K. Manna, S. Biswas, R. Chatterjee. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. Infection, Genetics and Evolution 85, 104445 (2020).

11. L. Wang, L. Wang, H. Zhuang. Profiling and characterization of SARS-CoV-2 mutants' infectivity and antigenicity. Sig Transduct Target Ther **5,** 185 (2020).

12. X. Xie, Y. Liu, J. Liu, X. Zhang, J. Zou, C. R. Fontes-Garfias, H. Xia, K. A. Swanson, M. Cutler, D. Cooper, V. D. Menachery, S. C. Weaver, P. R. Dormitzer, P. Y. Shi. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K, and N501Y variants by BNT162b2 vaccine-elicited sera. Nature Medicine. https://doi.org/10.1038/s41591-021-01270-4 (2021).

13. A. Singh, G. Steinkellner, K. Köchl, K. Gruber, C. C. Gruber. Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2., 13 November 2020, PREPRINT. https://doi.org/10.21203/rs.3.rs-106969/v2

14. X. Zhan, Y. Zhang, X. Zhou, K. Huang, Y. Qian, Y. Leng, L. Yan, B. Huang, Y. He, Molecular Evolution of SARS-CoV-2 Structural Genes: Evidence of Positive Selection in

Spike Glycoprotein. BioRxiv2020.06.25.170688; doi: https://doi.org/ 10.1101/2020.06.25.170688 (2020).

15. S. Nelson-Sathi, P.K. Umasankar, E. Sreekumar, R. Nair, I. Joseph, S.R.C. Nori, J.S. Philip, R. Prasad, K.V. Navyasree, S.T. Ramesh, H. Pillai, S. Ghosh, T.R.S. Kumar, M.R. Pillai. Structural and Functional Implications of Non-synonymous Mutatio.ns in the Spike protein of 2,954 SARS-CoV-2 Genomes. BioRxiv 2020.05.02.071811; doi: https://doi.org/ 10.1101/2020.05.02.071811 (2020).

16. K.M. Campbell, G. Steiner, D.K. Wells, A. Ribas, A. Kalbasi, Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. BioRxiv 2020.03.30.016931; doi: https://doi.org/ 10.1101/2020.03.30.016931 (2020).

17. R.A. Khailanya, M. Safdarb, M. Ozaslanc. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 19, 100682 (2020).

18. Jin-Yan Li, Ce-Heng Liao, Qiong Wang, Yong-Jun Tan, Rui Luo, Ye Qiu, Xing-YiG. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Research. Volume 286, September (2020).

19. K. Keng, Y. Choi, M.R.A. Welkers, D.Z.L. Chan, S. Shen, S.G. Lim, W. Hong, Y. Tan, The human severe acute respiratory syndrome coronavirus (SARS-CoV) 8b protein is distinct from its counterpart in animal SARS-CoV and down-regulates the expression of the envelope protein in infected cells. Virology 354, 132-42 (2006).

20. P.Y.P. Law, Y. Liu, H. Geng, K.H. Kwan, M.M. Waye, Y. Ho, Expression and functional characterization of the putative protein 8b of the severe acute respiratory syndrome-associated coronavirus. FEBS Lett. 580, 3643-8 (2006).

21. Y. Zhang, J. Zhang, Y. Chen, B. Luo, Y. Yuan, F. Huang, T. Yang, F. Yu, J. Liu, B. Liu, Z.Song, J. Chen, T. Pan,X. Zhang, Y. Li, R. Li, W. Huang, F. Xiao, H. Zhang. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. BioRxiv 2020.05.24.111823; doi: https://doi.org/10.1101/2020.05.24.111823 (2020).

22. 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).

23. D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge. Virology J., 16, 1-22 (2019).

24. S.S. Hassan, P.P. Choudhury, B. Roy, SARS-CoV2 envelope protein: non-synonymous mutations and its consequences. 112, 3890-92 (2020).

25. L. Kuo, K.R. Hurst, P.S. Masters, Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function. J. Virol. 81, 2249-2262 (2007).

26. J.L. Nieto-Torres, M.L. DeDiego, E. Álvarez, J.M. Jiménez-Guardeño, J.A. Regla-Nava, M. Llorente, L. Krremer, S. Shuo, L. Enjuanes, Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology 415, 69-82 (2011).

27. S. Srinivasan, H. Cui, Z. Gao, M. Liu, S. Lu, W. Mkandawire, O. Narykov, M. Sun, D. Korkin, Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. Viruses 12, 360 (2020).

28. K. Pervushin, E. Tan, K. Parthasarathy, X. Lin, F.L. Jiang, D. Yu, A. Vararattanavech, T.W. Soong, D.X. Liu, J. Torres, Structure and inhibition of the SARS coronavirus envelope protein ion channel. PLoS Pathog. 5, e1000511 (2009).

29. K. Teoh, Y. Siu, W. Chan, M.A. Schlüter, C. Liu, J.S.M. Peiris, R. Bruzzone, B. Margolis, B. Nal, The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. Mol Biol Cell. 21, 3838-52 (2010).

30. M.E. Gonzalez, L. Carrasco, Viroporins. FEBS Lett. 552, 28-34 (2003).

31. T. Suzuki, Y. Orba, Y. Okada, Y. Sunden, T. Kimura, S. Tanaka, K. Nagashima, W.W. Hall, H. Sawa, The Human Polyoma JC Virus Agnoprotein Acts as a Viroporin. PLOS Pathog. 6, e1000801 (2010).

32. Y. Li, W. Surya, S. Claudine, J. Torres, Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. J Biol Chem., 289, 12535-49 (2014).

33. Z. Lin, F. Qian, M. Jinlian, , Z. Lishi, Q. Yu, C. Tian, C. Zhenhua, W. Ping, L. Bin, The Nucleocapsid Protein of SARS-CoV-2 Abolished Pluripotency in Human Induced Pluripotent Stem Cells. BioRxiv 2020.03.26.010694; doi: https://doi.org/10.1101/2020.03.26.010694 (2020).

34. J. Mu, J. Xu, L. Zhang, T. Shu, D. Wu, M. Huang, Y. Ren, X. Li, Q. Geng, Y. Xu, Y. Qiu, X. Zhou, SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference in cells. Sci China Life Sci. Letter to the Editor, 1-4 (2020).

35. D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, T.A. Tummino, R. Hüttenhain, R.M. Kaake, A.L.

Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B.J. Polacco, H. Braberg, J.M. Fabius, M. Eckhardt, M. Soucheray, M.J. Bennett, M. Cakir, M.J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. MacKain, L. Miorin, E. Moreno, Z.Z.C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I.T. Kirby, J.E. Melnyk, J.S. Chorba, K. Lou, S.A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C.J.P. Mathy, T. Perica, K.B. Pilla, S.J. Ganesan, D.J. Saltzberg, R. Rakesh, X. Liu, S.B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X.P. Huang, Y. Liu, S.A. Wankowicz, M. Bohn, M. Safari, F.S. Ugur, C. Koh, N.S. Savar, Q.D. Tran, D. Shengjuler, S.J. Fletcher, M.C. O'Neal, Y. Cai, J.C.J. Chang, D.J. Broadhurst, S. Klippsten, P.P. Sharp, N.A. Wenzell, D. Kuzuoglu-Ozturk, H.Y. Wang, R. Trenker, J.M. Young, D.A. Cavero, J. Hiatt, T.L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R.M. Stroud, A.D. Frankel, O.S. Rosenberg, K.A. Verba, D.A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H.S. Malik, D.G. Fujimori, T. Ideker, C.S. Craik, S.N. Floor, J.S. Fraser, J.D. Gross, A. Sali, B.L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K.M. Shokat, B.K. Shoichet, N.J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583, 459-468 (2020).

36. M. Bouhaddou, D. Memon, B. Meyer, K.M. White, V.V. Rezelj, M.C. Marrero, B.J. Polacco, J.E. Melnyk, S. Ulferts, R.M. Kaake, J. Batra, A.L. Richards, E. Stevenson, D.E. Gordon, A. Rojc, K. Obernier, J.M. Fabius, M. Soucheray, L. Miorin, E. Moreno, C. Koh, Q.D. Tran, A. Hardy, R. Robinot, T. Vallet, B.E. Nilsson-Payant, C. Hernandez-Armenta, A. Dunham, S. Weigang, J. Knerr, M. Modak, D. Quintero, Y. Zhou, A. Dugourd, A. Valdeolivas, T. Patil, Q. Li, R. Hüttenhain, M. Cakir, M. Muralidharan, M. Kim, G. Jang, B. Tutuncuoglu, J. Hiatt, J.Z. Guo, J. Xu, S. Bouhaddou, C.J.P. Mathy, A. Gaulton, E.J. Manners, E. Félix, Y. Shi, M. Goff, J.K. Lim, T. McBride, M.C. O'Neal, Y. Cai, J.C.J. Chang, D.J. Broadhurst, S. Klippsten, E. De Wit, A.R. Leach, T. Kortemme, B. Shoichet, M. Ott, J. Saez-Rodriguez, B.R. tenOever, R.D. Mullins, E.R. Fischer, G. Kochs, R. Grosse, A. García-Sastre, M. Vignuzzi, J.R. Johnson, K.M. Shokat, D.L. Swaney, P. Beltrao, N.J. Krogan. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. Cell 8674, 30811-4 (2020).

37. A.B. Gussow, N. Auslander, G. Faure, Y. Wolf, F. Zhang, E.V. Koonin, Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. Proc Natl Acad Sci USA 117, 15193-15199 (2020).

38. R. He, A. Leeson, M. Ballantine, A. Andonov, L. Baker, F. Dobie, Y. Li, N. Bastien, H. Feldmann, U. Strocher, S. Theriault, T. Cutts, J. Cao, T.F. Booth, F.A. Plummer, S. Tyler, X. Li, Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res. 105, 121-5 (2004).

39. S. L Aha, J. Chakraborty, S. Das, S.K. Manna, S. Biswas, R. Chatterjee, Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. Infect Genet Evol. 85, 104445 (2020).

40. J.K. Taylor, C.M. Coleman, S. Postel, J.M. Sisk, J.G. Bernbaum, T. Venkataraman, E.J. Sundberg, M.B. Frieman, Severe Acute Respiratory Syndrome Coronavirus ORF7a Inhibits Bone Marrow Stromal Antigen 2 Virion Tethering through a Novel Mechanism of Glycosylation Interference. J. Virol. 89, 11820-33 (2015).

41. C.K. Yuen, J.Y. Lam, W.M. Wong, L.F. Mak, X. Wang, H. Chu, J.P. Cai, D.Y. Jin, K.K.W. To, J.F.W. Chan, K.Y. Yuen, K.H. Kok, SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. Emerging Microbes & Infections. 9, 1418-1428, doi:10.1080/22221751.2020.1780953

42. E. Issa, G, Merhi, B. Panossian, T. Salloum, S. Tokajiana, SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. Msystems. 5, e00266. https://doi.org/10.1128/mSystems.00266-20 (2020).

43. 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. Cellular & Molecular Immunology. 17, 881-883; https://doi.org/10.1038/s41423-020-0485-9 (2020).

44. 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. 8. 8865-8877. doi: 10.1096/fj.201802418R. (2019)

45. Y.J. Tan, P.Y. Tham, D.Z. Chan, C.F. Chou, S. Shen, B.C. Fielding, T.H. Tan, S.G. Lim, W. Hong, The severe acute respiratory syndrome coronavirus 3a protein up-regulates expression of fibrinogen in lung epithelial cells. J. Virol. 79, 10083-10087 (2005).

46. M. Catanzaro, F. Fagiani, M. Racchi, E. Corsini, S. Govoni, C. Lanni, Immune response in COVID-19: addressing a pharmacological challenge by targeting pathways triggered by SARS-CoV-2. Signal Transduction and Targeted Therapy. 5, 1-10 (2020)

47. J.Y. Li, C.H. Liao, Q. Wang, Y.J. Tan, R. Luo, Y. Qiu, X.Y. Ge. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Res. 286, 198074 (2020)

48. M. Frieman, B. Yount, M. Heise, S.A. Kopecky-Bromberg, P. Palese, R.S. Baric, Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. J. Virol. 81, 9812-9824 (2007).

49. Y. Zhang, J. Zhang, Y. Chen, B. L., Y. Yuan, F. Huang, T. Yang, F. Yu, J. Liu, B. Liu, Z. Song, J. Chen, T. Pan, X. Zhang, Y. Li, R. Li, W. Huang, F. Xiao, and H. Zhang, The ORF8 Protein of SARS-CoV2 Mediates Immune Evasion through Potently Downregulating MHC-I. bioRxiv (2020).

50. 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, 881-883 (2020)

51. M. Merad, J.C. Martin, Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol 20, 355-362 (2020).

52. L. Huang, Y. Shi, B. Gong, L. Jiang, X. Liu, J. Yang, J. Tang, C. You, Q. Jiang, B. Long, T. Zeng, M. Luo, F. Zeng, F. Zeng, S. Wang, X. Yang, Z. Yang, Blood single cell immune profiling reveals the interferon-MAPK pathway mediated adaptive immune response for COVID-19. MedRxiv (2020)

53. Y. Du, L. Xin, Y. Shi, T.-H. Zhang, N.C. Wu, L. Dai, D. Gong, G. Brar, S. Shu, J. Luo, W. Reiley, Y.-W. Tseng, H. Bai, T.-T. Wu, J. Wang, Y. Shu, R. Sun, Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. *Science*. 359, 290-296 (2018).

54. M. Laurent-Rolle, J. Morrison, R. Rajsbaum, J.M.L. Macleod, G. Pisanelli, A. Pham, J. Ayllon, L. Miorin, C. Martinez, B.R. tenOever, A. García-Sastre, The interferon signaling antagonist function of yellow fever virus NS5 protein is activated by type I interferon. Cell Host Microbe. 16, 314-327 (2014).

55. Y. Konno, I. Kimura, K. Uriu, M. Fukushi, T. Irie, Y. Koyanagi, S. Nakagawa, K. Sato, SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is further increased by a naturally occurring elongation variant. BioRxiv (2020).

56. Y. Zhang, G. Zeng, H. Pan, C. Li, Y. Hu, K. Chu, W. Han, Z. Chen, R. Tang, W. Yin, X. Chen, Y. Hu, X. Liu, C. Jiang, J. Li, M. Yang, Y. Song, X. Wang, Q. Gao, F. Zhu. Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine in healthy adults aged 18–59 years: a randomised, double-blind, placebo-controlled, phase 1/2 clinical trial. Lancet Infect Dis. 2, 181-192 (2021).

57. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, Diemert D, Spector SA, Rouphael N, Creech CB, McGettigan J, Khetan S, Segall N, Solis J, Brosz A, Fierro C, Schwartz H, Neuzil K, Corey L, Gilbert P, Janes H, Follmann D, Marovich M, Mascola J, Polakowski L, Ledgerwood J, Graham BS, Bennett H, Pajon R, Knightly C, Leav B, Deng W, Zhou H, Han S, Ivarsson M, Miller J, Zaks T; COVE Study Group. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. New England Journal of Medicine. 384(5), 403-416 (2020).

58. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper D, Frenck RW Jr, Hammitt LL, Türeci Ö, Nell H, Schaefer A, Ünal S, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Jansen KU, Gruber WC; C4591001 Clinical Trial Group. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine. 383(27):2603-2615.

59. P. Wang, L. Liu, S. Iketani, Y. Luo, Y. Guo, M. Wang, J. Yu, B. Zhang, P. D. Kwong, B. S. Graham, J. R. Mascola, J. Y. Chang, M. T. Yin, M. Sobieszczyk, C. A. Kyratsous, L. Shapiro, Z. Sheng, M. S. Nair, Y. Huang, D. D. Ho. Increased Resistance of SARS-CoV-2 Variants. BioRxiv (2021).

60. H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G. V. Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, M. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C. K. Wibmer, B. T. Sewell, J. Lourenço, L. C. J. Alcantara, S. L. K. Pond, S. Weaver, D. Martin, R. J. Lessells, J. N. Bhiman, C. Williamson, T. de Oliveira. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv (2020)

61. N. R. Faria, I. M. Claro, D. Candido, L. A. M. Franco, P. S. Andrade, T. M. Coletti, C. A. M. Silva, F. C. Sales, E. R. Manuli, R. S. Aguiar, N Gaburo, C. C. Camilo, N. A. Fraiji, M. A. E. Crispim, M. P. S. S. Carvalho, A. Rambaut, N. Loman, O. G. Pybus, E. C. Sabino, on behalf of CADDE Genomic Network. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (2021)

62. M. J. Mulligan, K. E. Lyke, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, K. Neuzil, V. Raabe, R. Bailey, K. A. Swanson, P. Li, K. Koury, W. Kalina, D. Cooper, C. Fontes-Garfias, P.-Y. Shi, O. Türeci, K. R. Tompkins, E. E. Walsh, R. Frenck, A. R. Falsey, P. R. Dormitzer, W. C. Gruber, U. Şahin, K. U. Jansen. Phase 1/2 Study to Describe the Safety and

Immunogenicity of a COVID-19 RNA Vaccine Candidate (BNT162b1) in Adults 18 to 55 Years of Age: Interim Report. medRxiv (2020).

63. D. A. Collier, A. De Marco, I. A.T.M. Ferreira, B. Meng, R. Datir, A. C. Walls, S. A. Kemp S, J. Bassi, D. Pinto, C. S. Fregni, S. Bianchi, M. A. Tortorici, J. Bowen, K. Culap, S. Jaconi, E. Cameroni, G. Snell, M. S. Pizzuto, A. F. Pellanda, C. Garzoni, A. Riva, The CITIID-NIHR BioResource COVID-19 Collaboration, A. Elmer, N. Kingston, B. Graves, L. E. McCoy, K. G. C. Smith, J. R. Bradley, J. J. Thaventhiran, L. L. Ceron-Gutierrez, G. Barcenas-Morales, H. W. Virgin, A. Lanzavecchia, L. Piccoli, R. Doffinger, M. Wills, D. Veesler, D. Corti, R. K. Gupta. SARS-CoV-2 B.1.1.7 escape from mRNA vaccine-elicited neutralizing antibodies medRxiv (2020).

64. Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, Cooper D, Menachery VD, Weaver SC, Dormitzer PR, Shi PY. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K, and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med. (2021) Epub ahead of print.

65. Moderna. Moderna COVID-19 Vaccine Retains Neutralizing Activity Against Emerging Variants First Identified in the U . K . and the Republic of South Africa. https://www.biospace.com/article/releases/moderna-covid-19-vaccine-retains-neutralizing-activity-against-emerging-variants-first-identified-in-the-u-k-and-the-republic-of-south-africa/ (2021).

66. K. Wu, A. P. Werner, J. I. Moliva, M. Koch, A. Choi, G. B. E. Stewart-Jones, H. Bennett, S. Boyoglu-Barnum, W. Shi, B. S. Graham, A. Carfi, K. S. Corbett, R. A. Seder, D. K. Edwards. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv (2021).

67. X. Shen, H. Tang, C. McDanal, K. Wagh, W. Fischer, J. Theiler, H. Yoon, D. Li, B. F. Haynes, K. O. Sanders, S. Gnanakaran, N. Hengartner, R. Pajon, G. Smith, F. Dubovsky, G. M. Glenn, B. Korber, D. C. Montefiori. SARS-CoV-2 Variant B.1.1.7 is Susceptible to Neutralizing Antibodies Elicited by Ancestral Spike Vaccines. bioRxiv (2021).

68. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Silva JD, Xu J, Colbert RA, Patel R, Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R, Hatziioannou T, Bieniasz PD, Nussenzweig MCmRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature. (2021) Epub ahead of print.

69. M. Voysey, S. A. C. Clemens, S. A. Madhi, L. Y. Weckx, P. M. Folegatti, P. K. Aley, B. J. Angus, V. Baillie, S. L. Barnabas, Q. E. Bhorat, S. Bibi, C. Briner, P. Cicconi, E. Clutterbuck, A. M. Collins, C. Cutland, T. Darton, K. Dheda, A. D. Douglas, C. J. A. Duncan, K. R.

W .Emary, K. Ewer, A. Flaxman, L. Fairlie, S. N. Faust, S. Feng, D. M. Ferreira, A. Finn, E. Galiza, A. L. Goodman, C. M. Green, C. A. Green, M. Greenland, C. Hill, H. C. Hill, I. Hirsch, A. Izu, D. Jenkin, S. Kerridge, A. Koen, G. Kwatra, R. Lazarus, V. Libri, P. J. Lillie, N. G. Marchevsky, R. P. Marshall, A. V. A. Mendes, E. P. Milan, A. M. Minassian, A. C. McGregor, Y. F. Mujadidi, A. Nana, S. D. Payadachee, D. J. Phillips, A. Pittella, E. Plested. Single dose administration , and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine. SSRN (preprint with The Lancet). (2021).

70. K. R. W. Emary, T. Golubchik, P. K. Aley et al. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 VOC 202012/01 (B.1.1.7). Lancet Preprints. (2021).

71. ChAdOx1 nCov-19 provides minimal protection against mild-moderate COVID-19 infection from B.1.351 coronavirus variant in young South African adults. https://www.ox.ac.uk/news/2021-02-07-chadox1-ncov-19-provides-minimal-protection-against-mild-moderate-covid-19-infection?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news#

72. A Keech C., Albert G., Cho I., Robertson A., Reed P., Neal S., Plested J. S., Zhu M., Cloney-Clark S., Zhou H., Smith G., Patel N., Frieman M. B., Haupt R. E., Logue J., McGrath M., Weston S., Piedra P. A., Desai C., Callahan K., Lewis M., Price-Abbott P., Formica N., Shinde V., Fries L., Lickliter J. D., Griffin P., Wilkinson B., Glenn G. M. Phase 1–2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine. New England Journal of Medicine. 383, 2320-2332 (2020).

73. M. Elisabeth. Covid-19: Novavax vaccine efficacy is 86% against UK variant and 60% against South African variant. BMJ. 372 (2021).

74. G. Suarez-Kurtz, S.D. Pena, Pharmacogenomics in the Americas: the impact of genetic admixture. Curr Drug Targets. 7, 1649-58 (2006)

75. D.E. Jonas, H.L. McLeod, Genetic and clinical factors relating to warfarin dosing. Trends Pharmacol Sci. 30, 375-86 (2009)

76. S.D. Pena, G. Di Pietro, M. Fuchshuber-Moraes, J.P. Genro, M.H. Hutz, F.S.G. Kehdy, F. Kohlrausch, L.A. Magno, R.C. Montenegro, M.O. Moraes, M.E. de Moraes, M.R. de Moraes, E.B. Ojopi, J.A. Perini, C. Racciopi, A.K. Ribeiro-Dos-Santos, F. Rios-Santos, M.A. Romano-Silva, V.A. Sortica, G. Suarez-Kurtz, The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. PLoS One. 6, e17063 (2011)

77. S. Tang, A. Lomsadze, M. Borodovsky, Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 43, e78 (2015).

78. D.K. Sydykova, B.R, Jack, S.J. Spielman, C.O. Wilke, Measuring evolutionary rates of proteins in a structural contexto. 6, 1845 (2017).

79. M. Shen, A. Sali, Statistical potential for assessment and predicition of protein structures. Protein Sci. 15, 2507-24 (2006).

80. V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 66, 12-21 (2010).

81. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, Procheck: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291 (1993).

82. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 65, 712-725 (2006).

83. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free. J Comput Chem. 26, 1701–1718 (2005).

84. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. J Chem Phys. 126, 014101 (2007).

85. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath. J Chem Phys. 81, 3684–3690 (1984).

86. B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations. J Comput Chem. 18, 1463-1472 (1998).

87. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water. J Chem Phys. 79, 926-935 (1983).

88. L.P. Kagami, G.M. das Neves, L.F.S.M. Timmers, R.A. Caceres, V.L. Eifer-Lima, Geo-Measures: A Pymol plugin for protein structure ensembles analysis. Comput Biol Chem. 87, 107322 (2020).

89. C. Dominguez, R. Boelens, A.M.J.J. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. J Am Chem Soc. 125, 1731-1737 (2003).
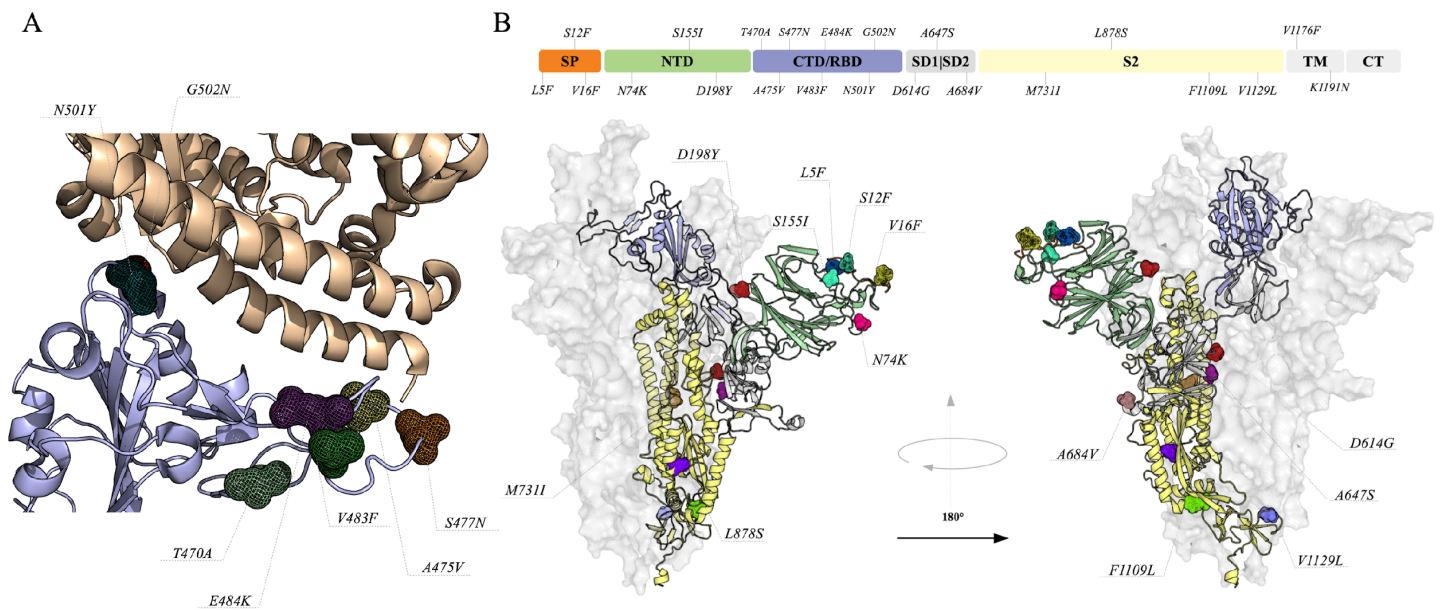
90. G.C.P. van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries, A.M.J.J. Bonvin, The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. J Mol Biol. 428, 720-725 (2016).

Acknowledgments

Figures and Tables

Fig. 1. **Mutations on the structure of Spike glycoprotein.** (A) Protein-protein interface of the Receptor binding domains (CTD/RBD) and hACE2. Tertiary structures of proteins are represented as cartoon and the mutation sites are highlighted in different colors and shown as mesh. (B) The quaternary structure is shown as surface and cartoon, the mutation sites are highlighted in different colors and shown as mesh. To better show the position of each mutation, we divided the structure in seven domains that are colored as follows; Signal peptide (SP) in orange, N-terminal (NTD) in green, CTD/RBD in purple, Subdomains 1 and 2 (SD1/SD2) in gray, S2 domain in yellow, Transmembrane (TM) and C-terminal (CT) in gray. Mutations at the CTD/RBD are not highlighted in the main structure, since they are presented interacting with hACE2.
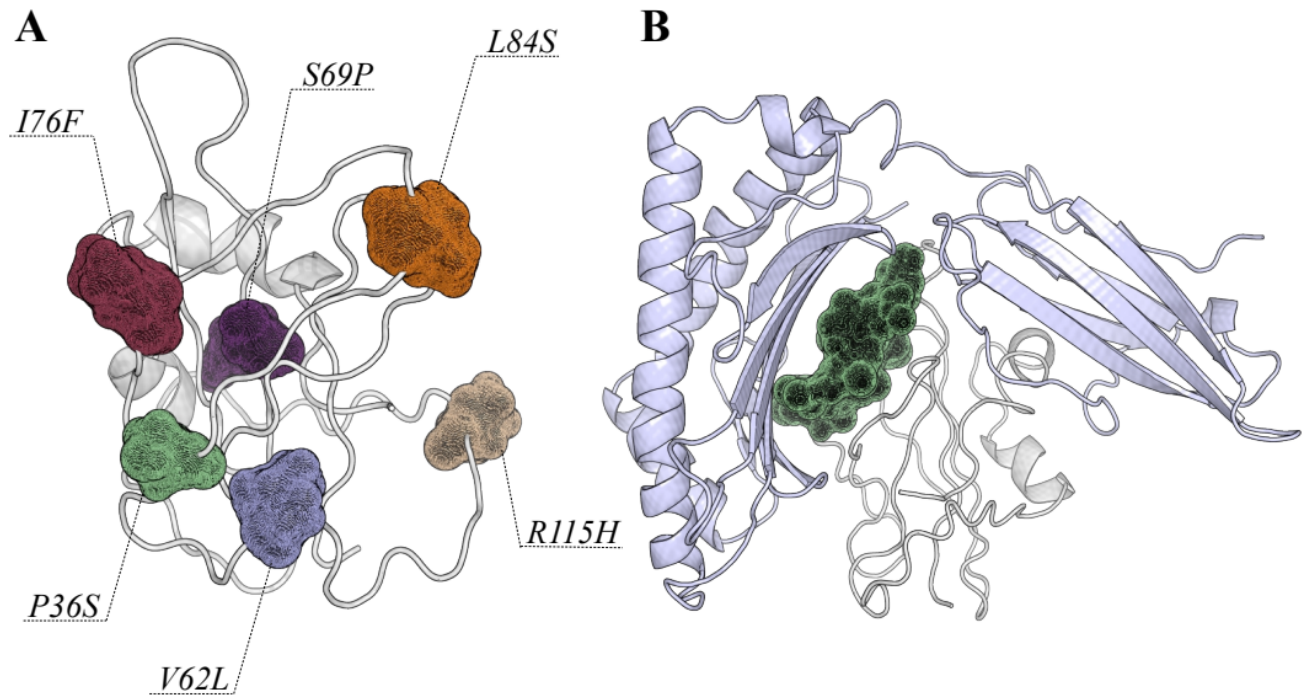
Fig. 2. ORF8 protein and its interaction with MHC-I. (**A**) Shows the position of each mutation observed on the SARS-CoV-2 genomes isolated from different regions. The ORF8 structure is shown as cartoon, and the amino acids positions are shown as mesh. (**B**) The most stable conformation of the ORF8:MHC-I complex predicted by the HADDOCK program. MHC-I and ORF8 are shown as cartoon and colored in light blue and gray, respectively. The region from Y73 to V81 is shown as mesh, to highlight where the viral protein is recognized by the host. It is important to observe that mutations on these regions could directly impact on protein-protein interactions.
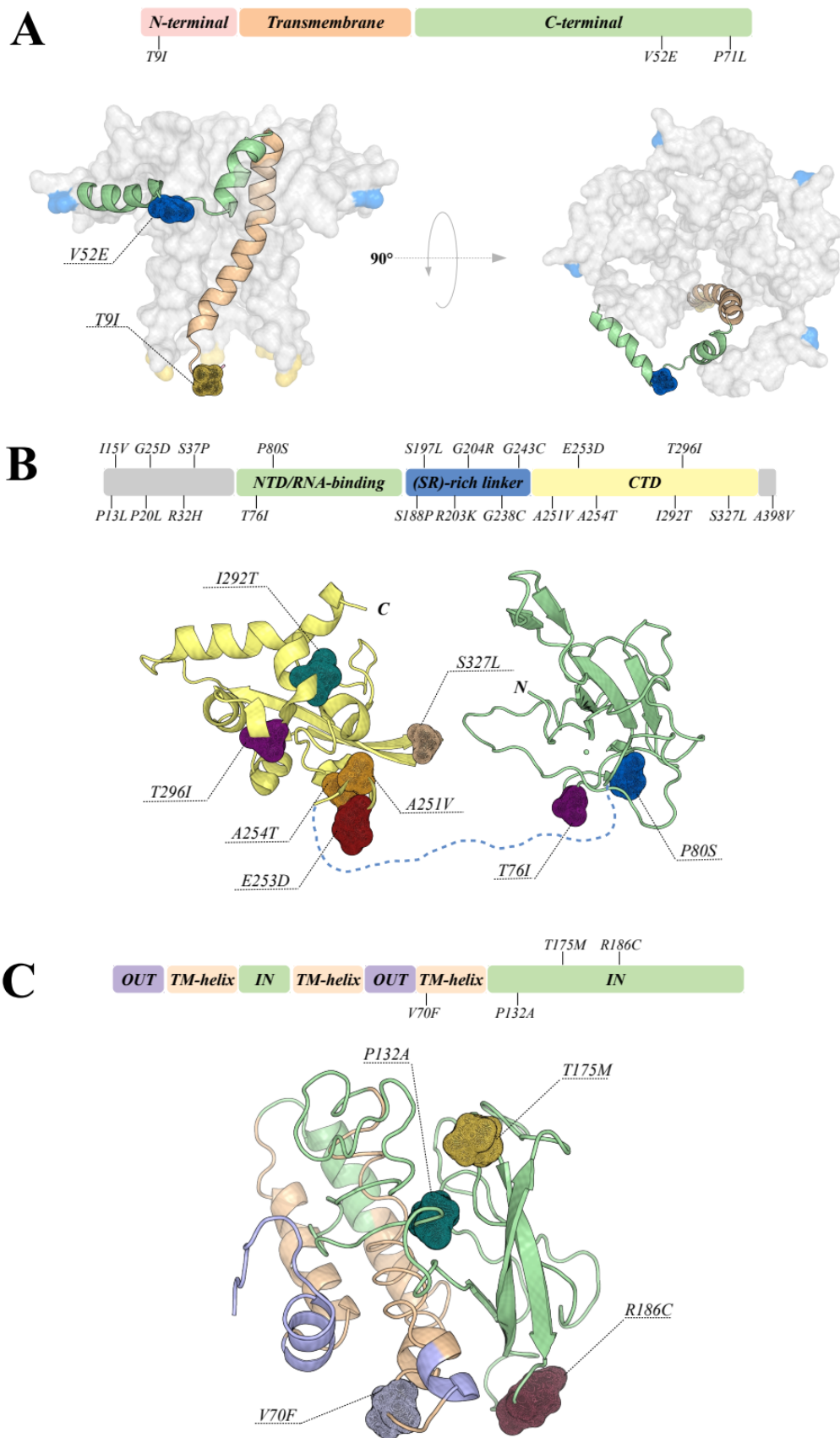
Fig. 3. Quatenary structure of the Envelope protein and Tertiary structures of Nucleocapsid and Membrane proteins. (**A**) The modeled structure is shown as surface and cartoon. The regions involved in mutations are highlighted in mesh. The E protein was divided into three domains and colored as follows; N-terminal in light pink, Transmembrane in orange, and C-terminal in green. The left figure shows the influence of the mutations in

the formation of the viroporin. (**B**) Presentation of crystal structures from the RNA-binding and C-terminal domains of the Nucleocapsid. RNA-binding and C-terminal domains are colored in yellow and green, respectively. All mutation sites are shown as mesh. The linker domain is shown as a dashed line, since it does not have an experimentally determined structure. (**C**) Membrane protein is divided by domains, where most of the mutations are located at the luminal domain. Overall structure is shown as cartoon, and positions involved in mutations are highlighted as mesh.
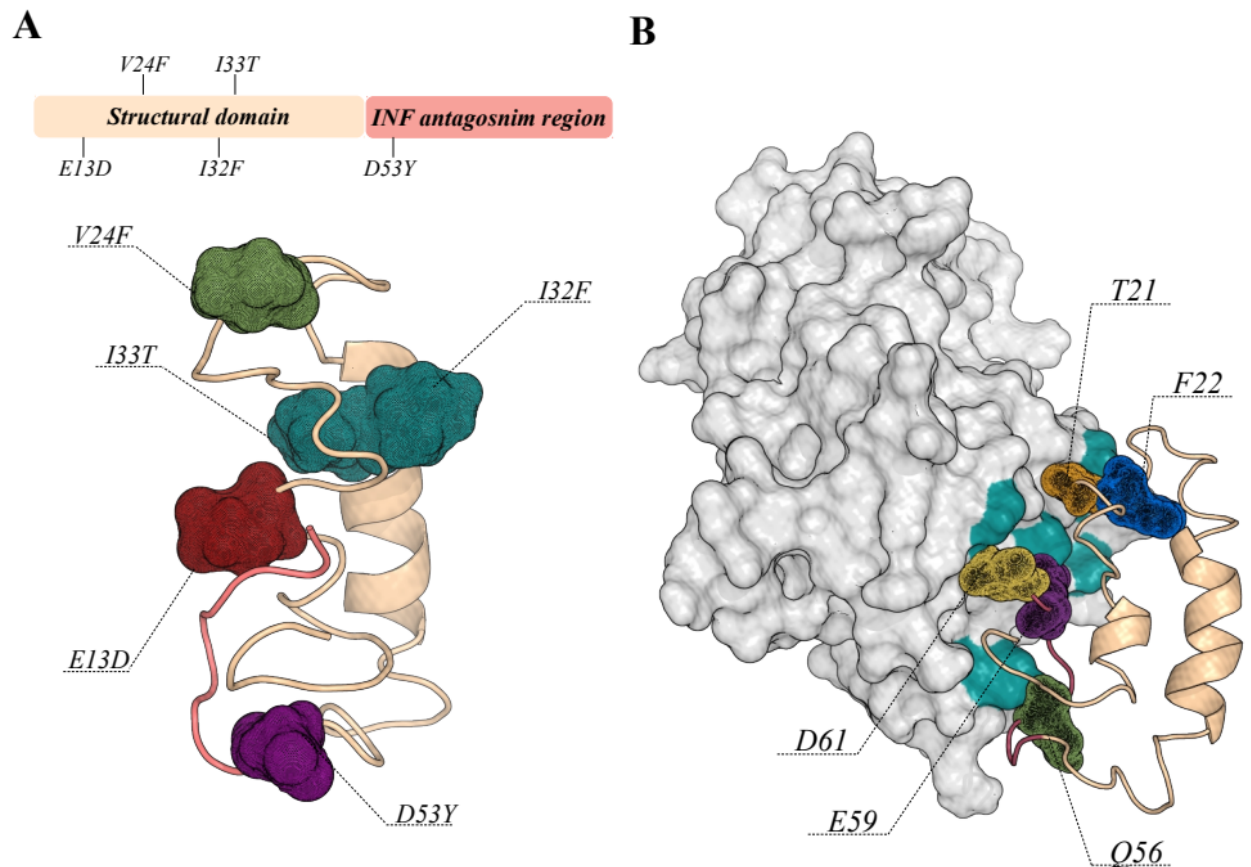
Fig. 4. ORF6 tertiary structure and its interaction with the interferon regulatory factor 3. (**A**) Shows the ORF6 structure divided into a structural domain (light orange) and an interferon-binding region (light red). The mutations observed in the genome of SARS-CoV-2 isolated from different regions of Brazil are shown as mesh. (**B**) The predicted binding pose of ORF6:IRF3 complex. IRF3 structure is shown as surface, and the amino acids highlighted in green (Y260, H290, S351, Q354, and R361) are involved in hydrogen bonds with ORF6 residues.
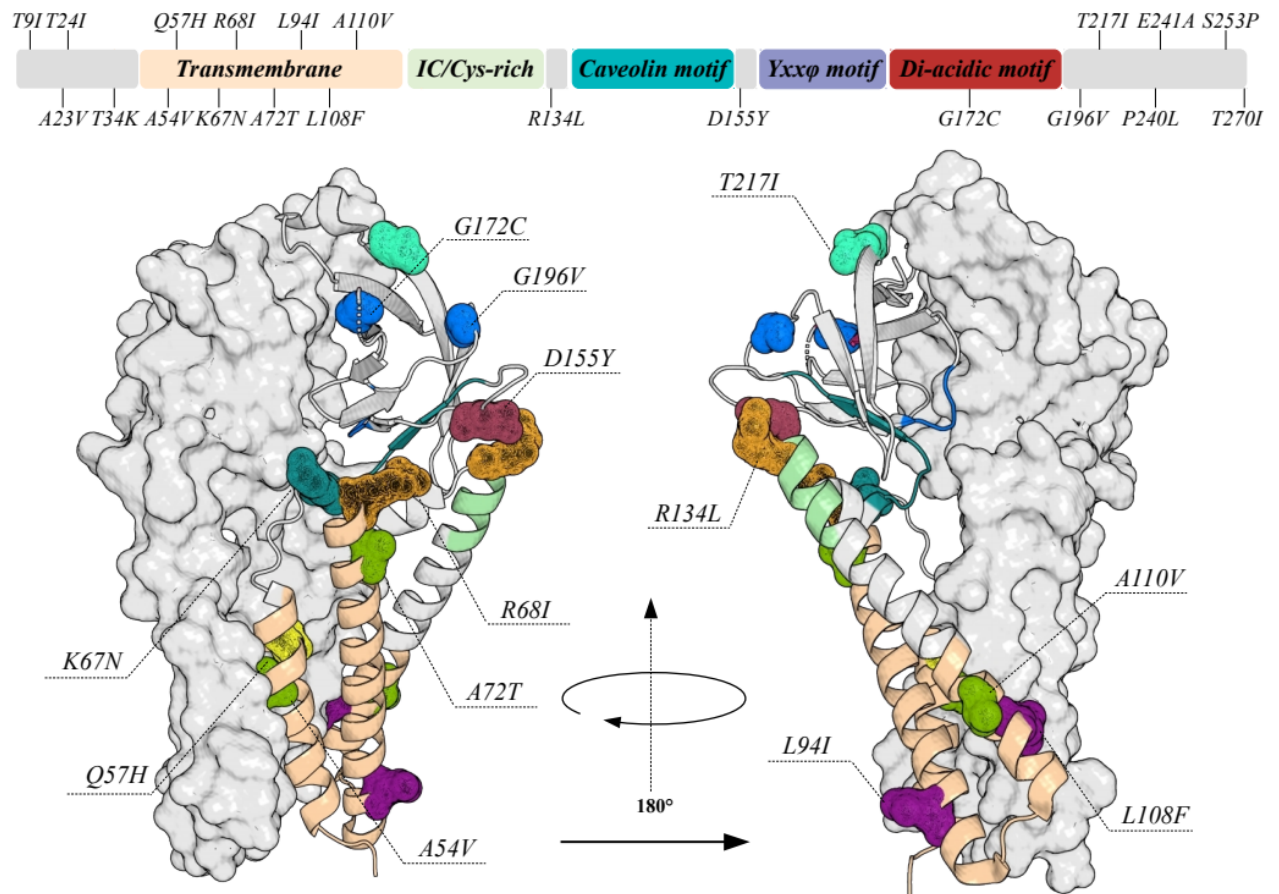
Fig. 5. The quaternary structure of the ORF3a protein. The domains of the ORF3a are described by the 2D scheme and its colors are highlighted on the structure. The position where mutations were identified are shown as mesh. It is interesting to observe that most of the mutations are located in the transmembrane region. The tertiary structures are presented as cartoon and surface.
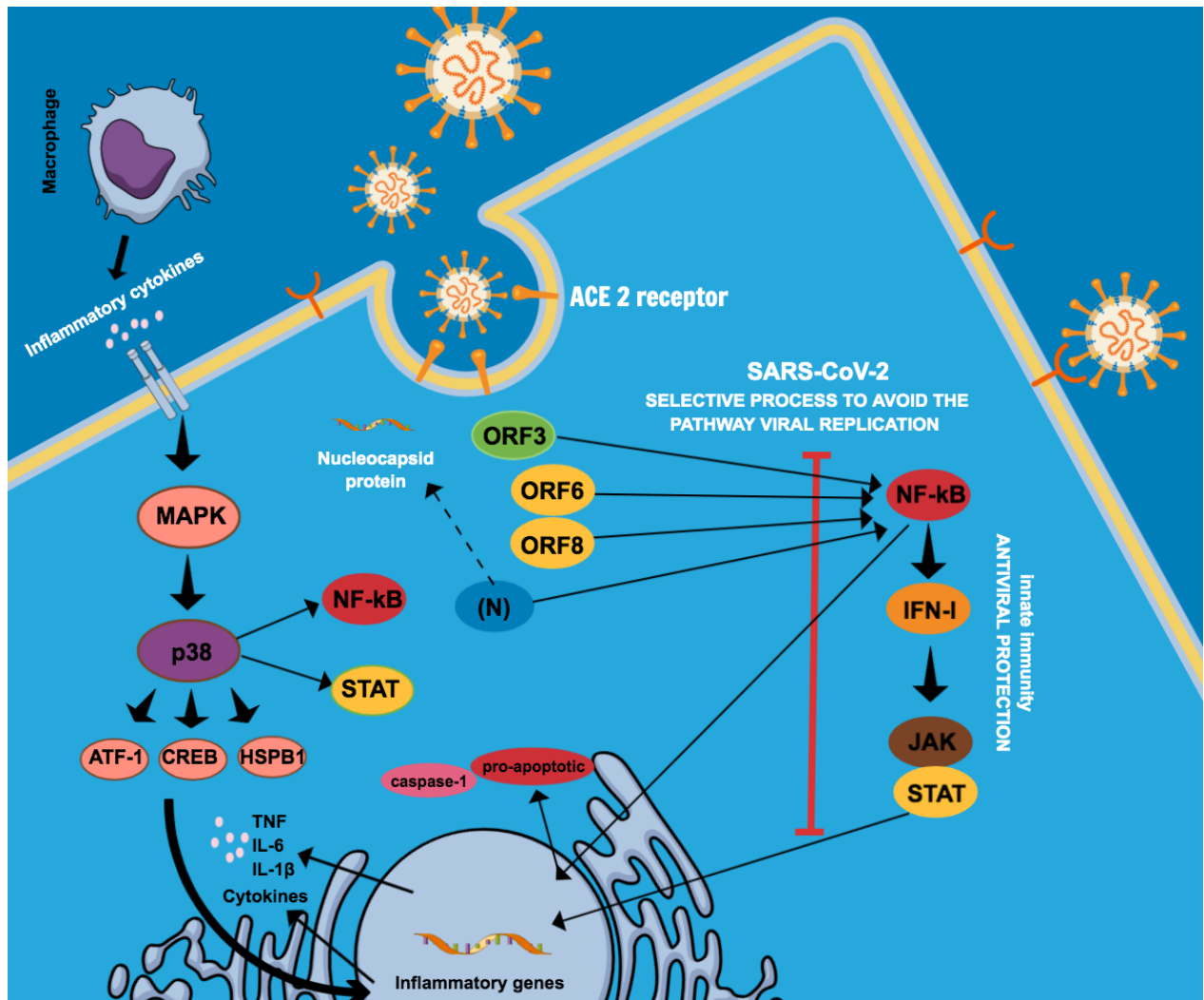
Fig. 6. Schematic representation of the SARS-CoV-2 viral infection mechanism through preventing the induction of the IFN-I pathway. The virus inactivates antagonistic members responsive to the viral replication system: SARS-CoV-2 ORF6, ORF8, N and ORF3 inhibit the IFN-I signaling pathway, via NF-κB, being also related to the inflammation activation mechanism. Regarding SARS-CoV, (N) encapsulates the virus genomic RNA protecting and allowing, in the initial stage, the viral replication. ORF3 has been mentioned for its pro-apoptotic activity and is probably associated with a reduction in apoptosis-mediated antiviral defense in infected cells. When the NF-kB activation pathway is inhibited by the virus, another pro-inflammatory response, such as caspase-1, affects the transcription of genes inflammatory cytokines. Mitogen-activated protein kinases (MAPKs) act on the IFN-I pathway and directly on inflammatory elements. From SARS-CoV-2 infection and pro-inflammatory cytokines resulting from the action of macrophages and other antigen-presenting cells (APCs), activation of the p38/MAPK pathway occurs acting on the NF-kB, STAT pathway, thus promoting high expression for cytokines pro-inflammatory, such as IL-6 and TNF-α.