

# Whole Genome Sequence Analysis and Homology Modelling of a 3C Like Peptidase and a Non-Structural Protein 3 of the SARS-CoV-2 Shows Protein Ligand Interaction with an Aza-Peptide and a Noncovalent Lead Inhibitor with Possible Antiviral Properties

Arun K. Shanker<sup>1</sup> \*, Divya Bhanu<sup>1,2</sup> and Anjani Alluri<sup>3</sup>

<sup>1</sup>ICAR - Central Research Institute for Dryland Agriculture

Santoshnagar, Hyderabad – 500059, India

Corresponding author email: [arunshank@gmail.com](mailto:arunshank@gmail.com)

<sup>2</sup> Centre for Plant Molecular Biology, Osmania University, Hyderabad, India

<sup>3</sup> Advanced Post Graduate Centre, Acharya N.G.Ranga Agricultural University, Guntur, India

## Abstract

The family of viruses belonging to Coronaviridae mainly consist of virulent pathogens that have a zoonotic property, Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) of this family have emerged before and now the SARS-CoV-2 has emerged in China. Characterization of spike glycoproteins, polyproteins and other viral proteins from viruses are important for vaccine development. Homology modelling of these proteins with known templates offers the opportunity to discover ligand binding sites and explore the possible antiviral properties of these protein ligand complexes. Any information emerging from these protein models can be used for vaccine development. In this study we did a complete bioinformatic analysis, sequence alignment, comparison of multiple sequences and homology modelling of the SARS-CoV-2 whole genome sequences, the spike protein and the polyproteins for homology with known proteins, we also analysed receptor binding sites in these models for possible binding with ligands that exhibit antiviral properties. Our results showed that the tertiary structure of the polyprotein isolate SARS-CoV-2\_HKU-SZ-001\_2020 had 98.94 percent identity with SARS-Coronavirus NSP12 bound to NSP7 and NSP8 co-factors. Our results indicate that a part of the viral genome (residues 3268 -3573 in Frame 2 with 306 amino acids) of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) when modelled with template 2a5i of the PDB database had 96 percent identity with a 3C like peptidase of SARS-CoV which has ability to bind with Aza-Peptide Epoxide (APE) which is known for irreversible inhibition of SARS-CoV main peptidase. The part of the genome (residues 1568-1882 in Frame 2 with 315 amino acids) when modelled with template 3e9s of the PDB database had 82 percent identity with a papain-like protease/deubiquitinase which when complexed with ligand GRL0617 acts as inhibitor which can block SARS-CoV replication. It is possible that these viral inhibitors can be used for vaccine development for the SARS-CoV-2.

## Introduction

More than a decade has passed since the emergence human Coronavirus that caused Severe Respiratory Syndrome (SARS-CoV) and it is about 7 years since the emergence of another type of Coronavirus - Middle East Respiratory Syndrome (MERS-CoV) and now the SARS-

34 CoV-2 has emerged in China. This repeated onslaught of these viruses goes to show that it can  
35 assume pandemic proportions at any time and at any place.

36 The family of viruses belonging to Coronaviridae mainly consist of virulent pathogens that  
37 have a zoonotic property and this large family of corona viruses, have been known to be  
38 circulating in animals including camels, cats and bats. It has been seen in the past that Severe  
39 Acute Respiratory Syndrome associated coronavirus (SARS-CoV) and Middle East  
40 Respiratory Syndrome-associated coronavirus (MERS-CoV) belonging to this family of  
41 viruses can be transmitted from animals to humans and can cause respiratory diseases. Human  
42 to human transmission on this virus has been a concern and due to this search for antiviral  
43 compounds and vaccine development for this family of virus becomes the need of the hour.

44 The SARS was first seen in 2002 in Guangdong province of China, and later spread globally  
45 and has caused close to about 8096 cases (WHO 2004, de Wit et al., 2016). In 2012, a novel  
46 betacoronavirus, designated Middle East respiratory syndrome coronavirus or MERS-CoV  
47 associated with severe respiratory disease in humans, emerged in the Arabian Peninsula (de  
48 Wit et al., 2013).

49 The World Health Organization (WHO), China Country Office was informed of cases of  
50 pneumonia of unknown aetiology in Wuhan City, Hubei Province, on 31 December 2019  
51 (WHO 2020). A novel coronavirus currently termed SARS-CoV-2 was officially announced  
52 as the causative agent by Chinese authorities on 7 January 2020. As on 20 Feb 2020 China's  
53 National Health Commission reported that there are 74,280 confirmed cases in China (Fig.1).  
54 The World Health Organization reported 924 confirmed cases in 25 countries outside China  
55 (WHO Situation Report 29 2020). This novel corona virus has been designated as SARS-CoV-  
56 2.

57

58    Coronaviruses are RNA viruses and have large genomes structures and due to this they can  
59    have high error in replication as compared to host genomes. It is also known that various CoVs  
60    can do effective recombination of their genomes after infecting host cells (Luo et al 2018). This  
61    recombination can be a factor for their evolution to novel types which may have new animals  
62    as their intermediate hosts. These factors give the CoVs high adaptive ability and the capability  
63    to jump across species and have a relatively large host range.

64    Characterization of Spike glycoproteins from viruses are important for vaccine development.  
65    Any information coming from the protein model can be used for vaccine development. *In Silico*  
66    Epitope, polyprotein and spike protein-based peptide vaccine designing for infectious viruses  
67    is a way that can hasten the process of vaccine development. Spike (S) protein, polyprotein and  
68    other viral proteins of the SARS-CoV-2 as a target for the development of vaccines and  
69    therapeutics for the prevention and treatment of infection is an important approach. In the case  
70    of SARS-CoV, these proteins can mediate binding of the virus with its receptor and promotes  
71    the fusion between the viral and host cell membranes and virus entry into the host cell, hence  
72    peptides, antibodies, organic compounds and short interfering RNAs that interact with the spike  
73    protein can have a potential role in vaccine development (Du et al 2009).

74    There are multiple domain functions that are active in the replication of the coronavirus and  
75    these domains are present in a protein designated as Non-structural protein 3 (nsp3) which is  
76    the largest protein in the coronavirus genome (Chen et al 2015). 3C like protease (3CLpro) and  
77    Papain like Protease (PLpro) are two important class of proteases that are involved in the  
78    process of translation of the polypeptide from the genomic RNA to protein components that  
79    are required structurally or non-structurally for replication and packaging of new generation  
80    viruses (Liu et al 2020)

81 We hypothesised that there can be some proteins in the large chunk of proteins in the SARS-  
82 CoV-2 that could have homology with the Non-structural protein 3 (nsp3) SARS CoV and  
83 these proteins can possibly have binding sites with ligands that can bind with known ligand  
84 with antiviral properties.

85 Here in this study we did a complete bioinformatic analysis, sequence alignment, comparison  
86 of multiple sequences of the SARS-CoV-2 whole genome sequences, the Spike protein and  
87 the polyproteins for homology with known spike proteins and also analysed receptor binding  
88 sites for possible vaccine development.

## 89 **Materials and Methods**

90 Six complete viral genome sequences, seven polyproteins (RdRp region) and seven  
91 glycoproteins available on NCBI portal on 4 Feb 2020 were taken for analysis. The sequence  
92 details and GenBank accession numbers are listed in Supplementary Table 1. Amongst the  
93 seven polyproteins, five are of Wuhan pneumonia virus isolate SARS-COV-2 and two  
94 sequences are of Wuhan pneumonia virus isolate SI200040-SP. The seven Glycoproteins are  
95 of the same isolate, Wuhan pneumonia virus isolate SARS-COV-2.

96 The available polyproteins (RdRp region) and glycoprotein sequences were retrieved from  
97 Genbank, NCBI (Benson et al., 2000). These sequences were translated to amino acid  
98 sequences using sorted six frame translation with Bioedit (Hall et al., 2011). Multiple sequence  
99 alignment of the translated protein sequences was performed and phylogenetic tree was  
100 constructed using Mega-X (Kumar et al., 2018). The alignment shows that amongst the seven  
101 polyproteins, five sequences were identical being from the same isolate and two other  
102 sequences of the other isolate are identical. Similar analysis of the seven glycoproteins was  
103 done, all the seven glycoprotein sequences were found to be identical. Therefore, further  
104 analysis was carried out for three sequences.

105 1. MN938385.1 SARS-CoV-2 virus isolate SARS-COV-2 \_HKU-SZ-001\_2020 ORF1ab  
106 polyprotein, RdRp region, (orf1ab) gene, partial cds: 0 to 284: Frame 3 95 aa

107 2. MN970003.1 SARS-CoV-2 virus isolate SI200040-SP orf1ab polyprotein, RdRP  
108 region, (orf1ab) gene, partial cds: 2 to 289: Frame 2 96 aa

109 3. MN938387.1 SARS-CoV-2 virus isolate SARS-COV-2 \_HKU-SZ-001\_2020 surface  
110 glycoprotein (S) gene, partial cds: 1 to 105: Frame 1 35 aa

111 Expasy proteomics server (Gasteiger et al., 2003) was used to study the protein sequence and  
112 structural details. These peptides were studied for their physio-chemical properties using the  
113 tool Protparam (Gasteiger et al., 2005). The secondary structure analysis was done using Chou  
114 and Fasman algorithm with CFSSP (Kumar, 2013). To generate the 3D structure from the fasta  
115 sequence, homology modelling was performed and the templates were identified. The model  
116 was built using the template with highest identity. Swiss-model (Schwede et al., 2003) was  
117 used to build and validate the 3D model, structural assessment was also performed to validate  
118 the model built.

119 Complete genome sequence of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank  
120 Accession Number MN908947.3) which has 29903 bp ss-RNA linear was translated sorted  
121 6 frame with minimum ORF of 20 with any start codon and the resultant protein sequence was  
122 used for homology modelling, homology models were done with large chunks of proteins  
123 21503 to 25381 in Frame 2 with 1293 amino acids, 13450 to 21552 in Frame 1 with 2701  
124 amino acids and 254 to 13480 in Frame 2 with 4409 amino acids.

125 SWISS-MODEL server was used for homology modelling (Waterhouse et al 2018) where  
126 computation was on ProMod3 engine which is based on Open Structure (Biasini et al 2013).  
127 Structural information is extracted from the template, sequence alignment is used to define  
128 insertions and deletions.

Protein ligand interaction profile with hydrogen bonding, hydrophobic interactions, salt bridges and  $\pi$ -Stacking was done with PLIP server (Salentin et al., 2015)

## **Results and Discussion**

The physico- chemical properties and primary structure parameters of the 7 polyproteins RdRp region of the SARS-CoV-2 virus isolate is given in Table 1. RdRP forms an important part of the viral genome where in the RNA viruses its function is to catalyze the synthesis of the RNA strand complementary to a given RNA template.

The isolates SI200040-SP orflab polyprotein and the isolate SI200121-SP orflab polyprotein had 2 reading frames as compared to the rest of the isolates which had 3 reading frames. The presence of multiple reading frames suggests the possibility of overlapping genes as seen in many virus and prokaryotes and mitochondrial genomes. This could affect how the proteins are made. The number of amino acid residues in all the polyproteins were the same except one isolate SI200040-SP which had one amino acid more than the other polyproteins. The extinction coefficients of the two isolates SI200040-SP orflab polyprotein and the isolate SI200121-SP orflab polyprotein was much higher compared to the rest of the polyproteins. The extinction coefficient is important when studying protein-protein and protein-ligand interactions. The instability index of these two isolates was also high when compared to the others indicating that these two isolates are instable. Regulation of gene expression by polyprotein processing is known in viruses and this is seen in many viruses that are human pathogens (Yost et al 2013).

The isolates here like many other viruses may be using replication strategy which could involve the translation of a large polyprotein with subsequent cleavage by viral proteases. The two isolates SI200040-SP orflab polyprotein and the isolate SI200121-SP orflab polyprotein also showed shorter half-lives as compared to the other isolates indicating that they are susceptible to enzymatic degradation.

The tertiary structure analysis of the isolate SARS-CoV-2 \_HKU-SZ-001\_2020 ORF1ab polyprotein is given in Table 2. It is seen that the polyprotein has a 98.94 percent identity with PDB structure 6nur.1.A which is a hetero-1-2-1-mer. The polyprotein is an RNA directed RNA polymerase. The protein is identical to the SARS-Coronavirus NSP12 bound to NSP7 and NSP8 co-factors (Kirchdoerfer and Ward 2019). In SARS it is basically a nonstructural protein with NSP12 being the RNA dependent RNA polymerase and the co factors NSP 7 and NSP 8 having the function of forming hexadecameric complexes and also act as processivity clamp for RNA polymerase and primase (Fehr et al., 2016). This structure as in SARS CoV here in SARS-CoV-2 may be involved in the machinery of core RNA synthesis and can be a template for exploring antiviral properties.

The phylogenetic tree of the seven polyproteins is shown in Fig.2. It is seen that two polyproteins were distinctly different from the rest. The Phylogenetic tree of the seven glycoproteins of the SARS-CoV-2 virus isolate is shown in Fig.3, it is seen that the glycoproteins are similar in all the isolates. Multiple alignment of the Polyproteins of the SARS-CoV-2 is shown in Supplementary Fig.1.

This structure as in SARS CoV here in SARS-CoV-2 may be involved in the machinery of core RNA synthesis and can be a template for exploring antiviral properties. Based on its functions in the SARS CoV and its identity to the SARS-CoV-2, it is possible that it has the same functions in SARS-CoV-2 an RNA polymerase which does de novo initiation and primer extension with possible exonuclease activities, the activity itself being primer dependent useful for understanding the mechanism of SARS-CoV-2 replication and can be used as an antiviral target (Te Velthuis et al 2012; Te Velthuis et al 2010; Subissi et al 2014; Subissi et al 2014).

The polyprotein also has an identity of 19.74 percent with an ABC-type uncharacterized transport system periplasmic component-like protein, this protein is known to be a substrate binding protein and possible binding can be explored here (Bae et al 2019).

The homology model developed from the residues 254 to 13480 in Frame 2 with 4409 amino acids from the Complete genome sequence of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp with linear ss-RNA linear showed interesting template alignments, in all the model aligned with 50 templates from the PDB database with most of them being replicase polyprotein 1ab which is a SARS-CoV papain-like protease (Daczkowski 2017). The maximum similarity of 97.3 percent was with template structure of a Nsp9 protein from SARS-coronavirus indicating that this novel coronavirus has high degree of similarity with the SARS-coronavirus and this can be used for gaining insights into vaccine development. Nsp 9 is an RNA binding protein and has an oligosaccharide/oligonucleotide fold-like fold, this protein can have an important function in the replication machinery of the virus and can be important when designing antiviral for this virus (Egloff et al 2004).

Two models were developed, one from residues 3268 -3573 in Frame 2 with 306 amino acids and the other from the part of the genome residues 1568-1882 in Frame 2 with 315 amino acids of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3). The models had similarity with the 3C like proteinase and a papain-like protease/deubiquitinase protein which are known antiviral drug targets. Ligand binding with these proteins and their action is on viral replication and inactivation can be useful in stopping the viral replication (Baez-Santos et al 2015).



200 The homology models of the 4409 amino acid residues of the whole genome of the SARS-  
201 CoV-2 virus isolate Wuhan-Hu-1 with the ligand association with templates 2a5i and 3e9s are  
202 shown in Fig. 4 and Fig. 5 respectively.

203 The statistics of structural comparison with PDB templates is given in Table 5, it is seen that  
204 the proteins from the SARS-CoV-2 are significantly close to the proteins of SARS CoV and  
205 the amino acid alignment in the binding region is the same in both the viruses.

206 The alignment of the 305 residues from 3268-3573 aa of the Novel Coronavirus COVI-19 with  
207 the template 2a5i is shown in Fig.6 and the alignment of the 315 residues from 1568-1882 aa  
208 of the Novel Coronavirus COVI-19 with the template 3e9s is shown in Fig.7.

209 The important templates that aligned with this 4409 amino acid residues of the whole genome  
210 of the SARS-CoV-2 virus isolate Wuhan-Hu-1 were 2a5i of the PDB database which is a  
211 crystal structure of SARS coronavirus main peptidase inhibited by an Aza-Peptide epoxide in  
212 the space group C2 (Lee et al 2005) and 3e9s of the PDB database which is new class of  
213 papain-like protease/deubiquitinase which when combined with ligand GRL0617 acts as  
214 inhibitors blocking SARS virus replication (Ratia et al 2008). The model with template 2a5i of  
215 the PDB database shows that Aza-Peptide Epoxide (APE;  $k_{inact}/K_i=1900(\pm 400) \text{ M}^{-1} \text{ s}^{-1}$ )  
216 which is a known anti SARS agent can be used to develop a molecular target with irreversible  
217 inhibitor properties. The protein ligand interaction analysis of the Novel Coronavirus C3 like  
218 peptidase and aza-peptide epoxide is shown in Fig.8. The substrate binding properties and  
219 structural and chemical complementarity of this Aza-Peptide Epoxide can be explored as an  
220 anti - Coronavirus SARS-COV-2 agent. The APE which is ethyl (2S)-4-[(3-amino-3-oxo-  
221 propyl)-[(2S)-2-[[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-  
222 phenyl-propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate structure is shown in Fig.9.

223

224 The model with template 3e9s of the PDB database shows that the Coronavirus viral protein  
225 can have a ligand which is a papain-like protease (PLpro) that is known to be a potent inhibitor  
226 of viral replication in SARS (Ratia et al 2008).

227 The two parts of the Main protein from the whole genome of the SARS-CoV-2 aligned with  
228 two SAR proteins and the ligand binding sites were similar, the alignment positions, number  
229 of amino acids and ligand and the interacting residues is given in Table 3

230 The complete genome of MN908947.3 SARS-CoV-2 virus isolate Wuhan-Hu-1 encodes a  
231 4409aa long protein along with the other glycoproteins and polyproteins. The homology  
232 modelling of this protein showed sequence and structural alignment with two SARS proteases  
233 with structural accession numbers 3e9s.1 and 2a5i.1 at positions 1568-1882 and 3268-3573  
234 respectively. Reports suggests inhibition of virus replication by TTT ligand and an aza-peptide  
235 epoxide inhibiting the main peptidase. The structural similarity of these templates are 83% and  
236 96% respectively. The multiple sequence alignment shows complete conservation of the  
237 sequence suggesting a high degree of homology. The protein ligand interaction analysis of the  
238 Novel Coronavirus non structural protein and papain-like protease is shown in Fig. 10.

239 The Comparison of Hydrophobic interaction, hydrogen bonding, salt bridges of the  
240 constructed model of the Novel Coronavirus protein from region 3268-3573 aa to ligand AZP  
241 with Hydrophobic interaction, hydrogen bonding, salt bridges of the template 2a5i is given in  
242 Suppl. Table 2, when comparing both it is seen that the binding properties are the same expect  
243 for the presence of water bridge in the template 2a5i.

244

245 The Comparison of Hydrophobic interaction, hydrogen bonding,  $\pi$ -Stacking of the constructed  
246 model of the Novel Coronavirus protein from region 1568-1882 aa to ligand Small molecule  
247 Noncovalent Lead Inhibitor with the Hydrophobic interaction, hydrogen bonding,  $\pi$ -Stacking

248 of the template 3e9s is given in Suppl. Table 3, when comparing both it is seen that the binding  
 249 properties are the same except or an additional  $\pi$ -Stacking at Tyr in the template 2a5i. This  
 250 shows that there is high possibility of binding of these antiviral compounds with the regions of  
 251 Novel Coronavirus protein that is in homology with the SARS protein.

252 Comparison of the hydrophobic interaction of the binding of the ligand AZP between the SARS-  
 253 CoV-2 protein and the template 2a5i of SARS CoV is shown in Fig.11 and the comparison of  
 254 the hydrophobic interaction of the binding of the ligand AZP between the SARS-CoV-2 protein  
 255 and the template 3e9s of SARS CoV is shown in Fig.12. It is seen that the interaction is the  
 256 same in both proteins with the same amino acids participating in the interaction indicating that  
 257 there is a possibility that these ligands with antiviral properties can bind to the new virus.

258 The similarity in the amino acids involved in the Hydrophobic interactions which are short range  
 259 interactions and have an important role in the affinities of the ligands and receptors shows that  
 260 the proteins of the SARS-CoV-2 may bind with the same affinity as seen in the SARS CoV  
 261 and this also shows a similar action of the ligand as seen in SARS CoV, indicating that these  
 262 ligands can be used as antivirals in the SARS-CoV-2.

263 The targeting of this part of the genome of the SARS-CoV-2 with the antiviral compounds  
 264 which have shown to bind in the similar region of the SARS virus can have implication in the  
 265 development of an effective antiviral compound against the SARS-CoV-2. The SARS-CoV-2  
 266 shows homology with the SARS coronaviral proteases, papain-like protease (PLpro) and 3C-  
 267 like protease (3CLpro), these proteins have the function of processing the viral polyprotein and  
 268 also they perform the function of stripping ubiquitin and the ubiquitin-like interferon (IFN)-  
 269 stimulated gene 15 (ISG15) from the hosts to facilitate coronavirus replication and help in  
 270 evading immune response of the host, these inhibitors can also have a role in disrupting  
 271 signalling cascades in infected cells and protecting the uninfected cells.

The chemical GRL0617 is 5-Amino-2-methyl-N-[(1R)-1-(1-naphthalenyl)ethyl]benzamide and is known to inhibit the papainlike protease that is present in SARS CoV . This protease is a potential target for antiviral compounds (Chaudhuri et al., 2011). We found the SARS-CoV-2 has homology with this and the binding sites for this in the structural protein of the SARS-CoV-2 is the same (Table 4). This compound inhibits the enzyme that is required for the cleavage of the viral protein from the virus in SARS CoV, it also cleaves ubiquitin and has a structural homology with the Deubiquitinases (DUBs) of the Ubiquitin-Specific Proteases. Compound GRL0617 binds in the S4 and S3 enzyme subsite that gets the C terminal tail of the Ubiquitin (King and Finley 2014; Schauer et al., 2019). Our results show that Aza-Peptide Epoxide an irreversible protease inhibitor and GRL0617 a viral replication inhibitor can be used to develop inhibitors of the Novel Coronavirus SARS-COV-2.

## References

- Bae, J.E., Kim, I.J., Kim, K.J. and Nam, K.H., 2018. Crystal structure of a substrate-binding protein from *Rhodothermus marinus* reveals a single  $\alpha/\beta$ -domain. Biochemical and Biophysical Research Communications, 497(1), pp.368-373.
- Baez-Santos, Y.M., John, S.E.S. and Mesecar, A.D., 2015. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. Antiviral Research, 115, pp.21-38.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., & Wheeler, D. L. (2000). GenBank. Nucleic Acids Research, 28(1), 15-18.
- Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., Johnner, N., Schenk, A.D., Philippsen, A. and Schwede, T., 2013. OpenStructure: an integrated software

295 framework for computational structural biology. *Acta Crystallographica Section D:*  
 296 *Biological Crystallography*, 69(5), pp.701-709.

297 Chaudhuri, R., Tang, S., Zhao, G., Lu, H., Case, D.A. and Johnson, M.E., 2011. Comparison  
 298 of SARS and NL63 papain-like protease binding sites and binding site dynamics:  
 299 inhibitor design implications. *Journal of molecular biology*, 414(2), pp.272-288.

300 Chen, Y., Savinov, S.N., Mielech, A.M., Cao, T., Baker, S.C. and Mesecar, A.D., 2015. X-ray  
 301 structural and functional studies of the three tandemly linked domains of non-structural  
 302 protein 3 (nsp3) from murine hepatitis virus reveal conserved functions. *Journal of*  
 303 *Biological Chemistry*, 290(42), pp.25293-25306.

304 Daczkowski, C.M., Dzimianski, J.V., Clasman, J.R., Goodwin, O., Mesecar, A.D. and Pegan,  
 305 S.D., 2017. Structural insights into the interaction of coronavirus papain-like proteases  
 306 and interferon-stimulated gene product 15 from different species. *Journal of Molecular*  
 307 *Biology*, 429(11), pp.1661-1683.

308 de Wit, E., Rasmussen, A.L., Falzarano, D., Bushmaker, T., Feldmann, F., Brining, D.L.,  
 309 Fischer, E.R., Martellaro, C., Okumura, A., Chang, J. and Scott, D., 2013. Middle East  
 310 respiratory syndrome coronavirus (MERS-CoV) causes transient lower respiratory tract  
 311 infection in rhesus macaques. *Proceedings of the National Academy of Sciences*,  
 312 110(41), pp.16598-16603.

313 de Wit, E., van Doremalen N., D. Falzarano, V. J. Munster, SARS and MERS: recent insights  
 314 into emerging coronaviruses. *Nat Rev Microbiol* 14, 523-534 (2016).

315 Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.J. and Jiang, S., 2009. The spike protein of SARS-  
 316 CoV—a target for vaccine and therapeutic development. *Nature Reviews*  
 317 *Microbiology*, 7(3), pp.226-236.

318 Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J.,  
319 Gorbalenya, A.E., Cambillau, C. and Canard, B., 2004. The severe acute respiratory  
320 syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding  
321 subunit unique in the RNA virus world. *Proceedings of the National Academy of*  
322 *Sciences*, 101(11), pp.3792-3796.

323 Fehr, A.R. and Perlman, S., 2015. Coronaviruses: an overview of their replication and  
324 pathogenesis. In *Coronaviruses* (pp. 1-23). Humana Press, New York, NY.

325 Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003).  
326 ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic*  
327 *Acids Research*, 31(13), 3784-3788.

328 Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005).  
329 Protein identification and analysis tools on the ExPASy server. In *The proteomics*  
330 *protocols handbook* (pp. 571-607). Humana press.

331 Hall, T., Biosciences, I., & Carlsbad, C. (2011). BioEdit: an important software for molecular  
332 biology. *GERF Bull Biosci*, 2(1), 60-61.

333 King, R.W. and Finley, D., 2014. Sculpting the proteome with small molecules. *Nature*  
334 *chemical biology*, 10(11), p.870.

335 Kirchdoerfer, R.N. and Ward, A.B., 2019. Structure of the SARS-CoV nsp12 polymerase  
336 bound to nsp7 and nsp8 co-factors. *Nature Communications*, 10(1), pp.1-9.

337 Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular  
338 evolutionary genetics analysis across computing platforms. *Molecular Biology and*  
339 *Evolution*, 35(6), 1547-1549.

340 Kumar, T. A. (2013). CFSSP: Chou and Fasman secondary structure prediction server. Wide  
341 Spectrum, 1(9), 15-19.

342 Lee, T.W., Cherney, M.M., Huitema, C., Liu, J., James, K.E., Powers, J.C., Eltis, L.D. and  
343 James, M.N., 2005. Crystal structures of the main peptidase from the SARS coronavirus  
344 inhibited by a substrate-like aza-peptide epoxide. Journal of Molecular Biology, 353(5),  
345 pp.1137-1151.

346 Liu, W., Morse, J.S., Lalonde, T. and Xu, S., 2020. Learning from the Past: Possible Urgent  
347 Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by  
348 2019-nCoV. ChemBioChem.

349 Luo, C.M., Wang, N., Yang, X.L., Liu, H.Z., Zhang, W., Li, B., Hu, B., Peng, C., Geng, Q.B.,  
350 Zhu, G.J. and Li, F., 2018. Discovery of novel bat coronaviruses in south China that  
351 use the same receptor as Middle East respiratory syndrome coronavirus. Journal of  
352 Virology, 92(13), pp.e00116-18.

353 Ratia, K., Pegan, S., Takayama, J., Sleeman K., Coughlin, M., Baliji, S., Chaudhuri, R., Fu,  
354 W., Prabhakar, B.S., Johnson, M.E. and Baker, S.C., 2008. A noncovalent class of  
355 papain-like protease/deubiquitinase inhibitors blocks SARS virus replication.  
356 Proceedings of the National Academy of Sciences, 105(42), pp.16119-16124.

357 Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F. and Schroeder, M., 2015. PLIP: fully  
358 automated protein–ligand interaction profiler. Nucleic acids research, 43(W1),  
359 pp.W443-W447

360 Schauer, N.J., Magin, R.S., Liu, X., Doherty, L.M. and Buhrlage, S.J., 2019. Advances in  
361 Discovering Deubiquitinating Enzyme (DUB) Inhibitors. Journal of medicinal  
362 chemistry.

363 Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: an automated  
 364 protein homology-modeling server. *Nucleic Acids Research*, 31(13), 3381-3385.

365 Subissi, L., Imbert, I., Ferron, F., Collet, A., Coutard, B., Decroly, E. and Canard, B., 2014.  
 366 SARS-CoV ORF1b-encoded nonstructural proteins 12–16: replicative enzymes as  
 367 antiviral targets. *Antiviral research*, 101, pp.122-130.

368 Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly,  
 369 E., Snijder, E.J., Canard, B. and Imbert, I., 2014. One severe acute respiratory syndrome  
 370 coronavirus protein complex integrates processive RNA polymerase and exonuclease  
 371 activities. *Proceedings of the National Academy of Sciences*, 111(37), pp.E3900-  
 372 E3909.

373 Te Velhuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H. and Snijder, E.J., 2010.  
 374 The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent.  
 375 *Nucleic acids research*, 38(1), pp.203-214.

376 Te Velhuis, A.J., van den Worm, S.H. and Snijder, E.J., 2012. The SARS-coronavirus nsp7+  
 377 nsp8 complex is a unique multimeric RNA polymerase capable of both de novo  
 378 initiation and primer extension. *Nucleic acids research*, 40(4), pp.1737-1747.

379 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T.,  
 380 de Beer, T.A.P., Rempfer, C., Bordoli, L. and Lepore, R., 2018. SWISS-MODEL:  
 381 homology modelling of protein structures and complexes. *Nucleic Acids Research*,  
 382 46(W1), pp.W296-W303.

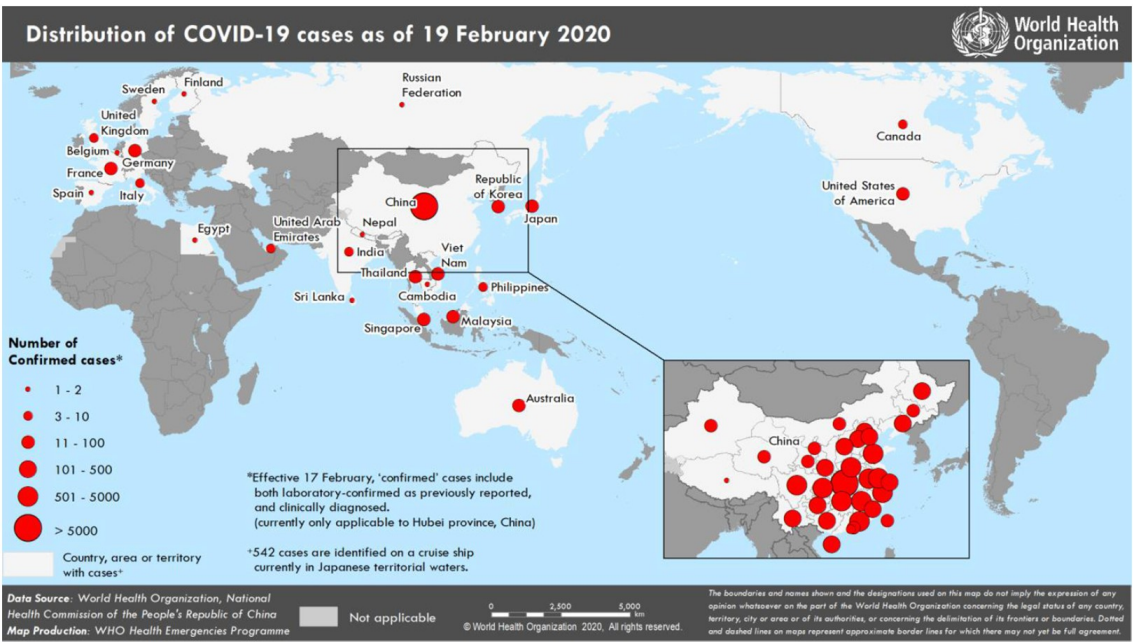
383 World Health Organization (WHO) 2004. [Accessed 11 Feb 2020]  
 384 [https://www.who.int/csr/don/2004\\_05\\_18a/en/](https://www.who.int/csr/don/2004_05_18a/en/)



World Health Organization (WHO). Coronavirus. Geneva: WHO; 2020 [Accessed 4 Feb 2020]. Available from: <https://www.who.int/health-topics/coronavirus>

Yost, S.A. and Marcotrigiano, J., 2013. Viral precursor polyproteins: keys of regulation from replication to maturation. Current Opinion in Virology, 3(2), pp.137-142.

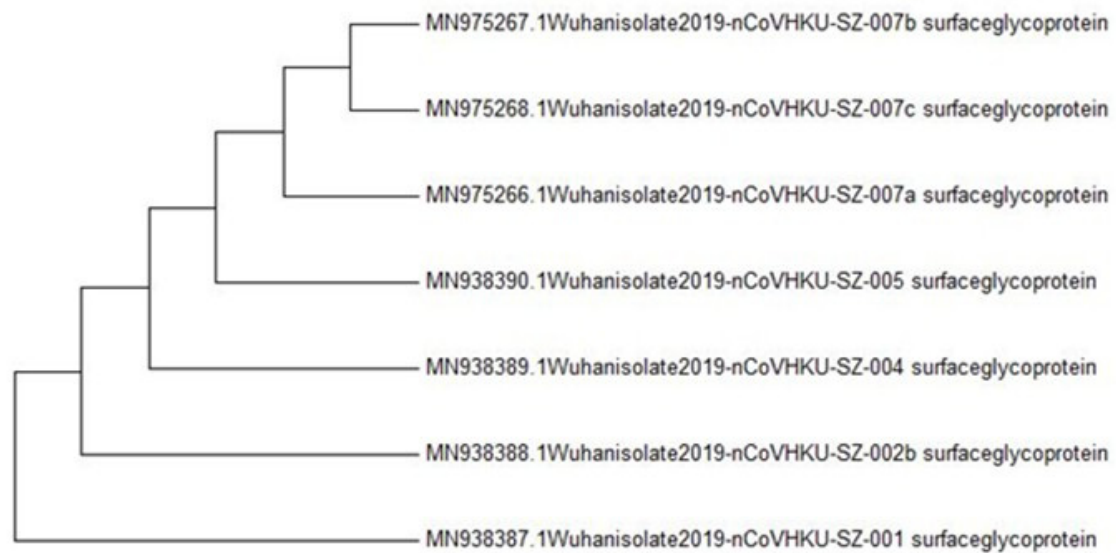
**Figure Captions**



**Fig.1** Countries, territories or areas with reported confirmed cases of SARS-COV-2 , 3 February 2020 Source WHO ([https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200219-sitrep-30-covid-19.pdf?sfvrsn=6e50645\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200219-sitrep-30-covid-19.pdf?sfvrsn=6e50645_2))

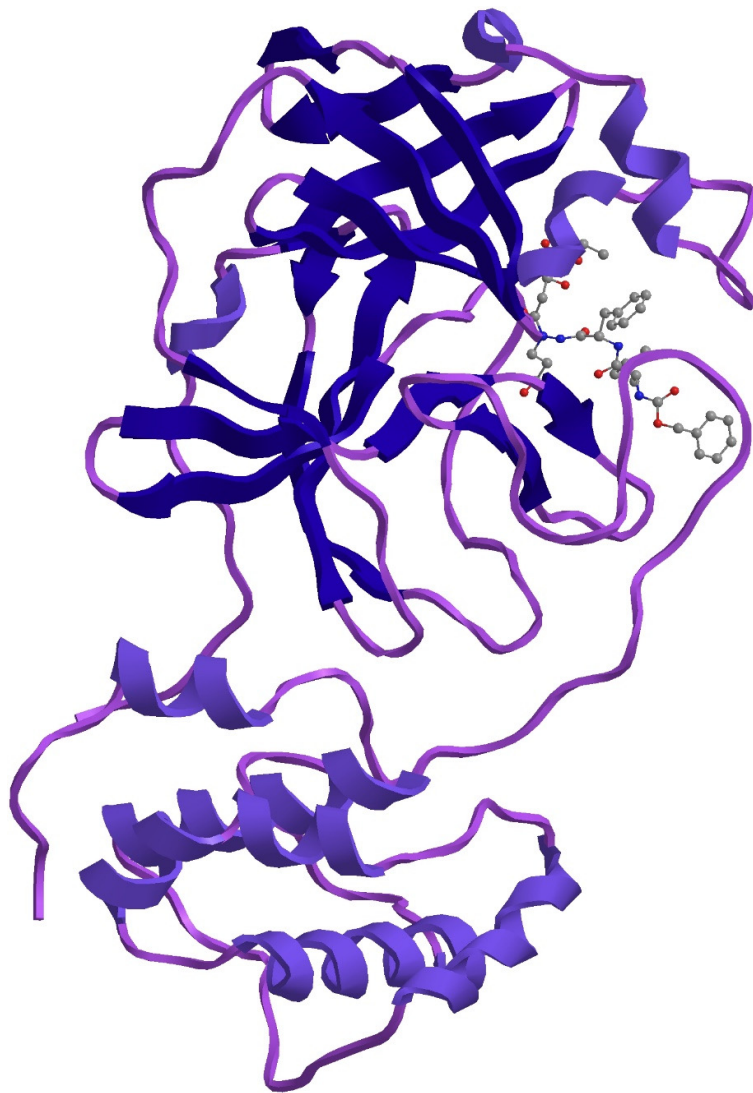


**Fig.2** Phylogenetic tree of the seven polyproteins of Severe acute respiratory syndrome coronavirus 2 isolate virus isolates



**Fig.3** Phylogenetic tree of the seven glycoproteins of Severe acute respiratory syndrome coronavirus 2 isolate virus isolates

403



404

405

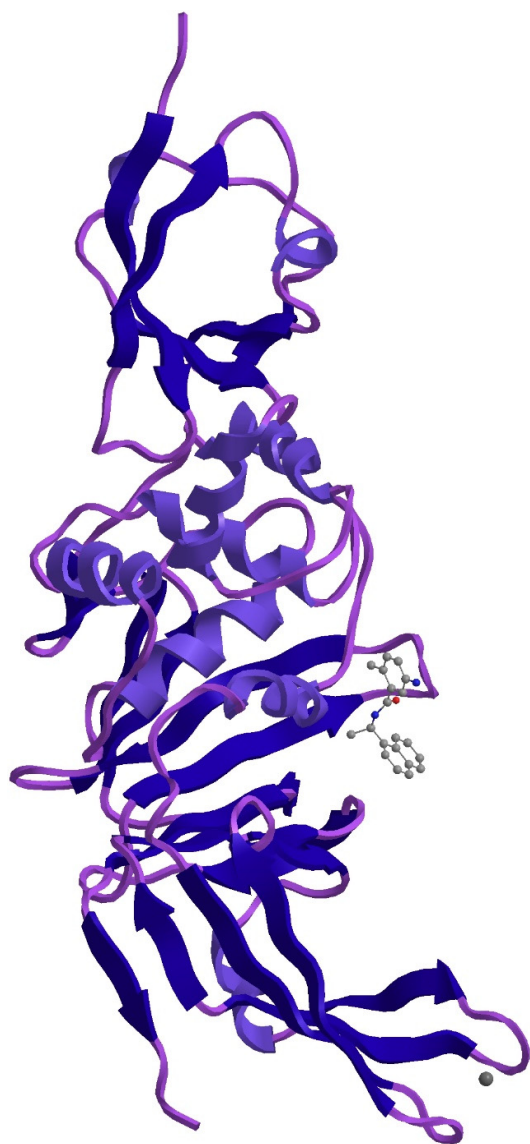
406 **Fig. 4** Homology model with ligand binding of protein from amino acids 3268 -3573 in Frame  
407 2 with 306 amino acids of the Complete genome sequence of the SARS-CoV-2 virus  
408 isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp  
409 linear ss-RNA with 2a5i of the PDB database as template.

410

411

412

413



414

415 **Fig. 5** Homology model with ligand binding of protein from residues 1568-1882 in Frame 2  
416 with 315 amino acids of the Complete genome sequence of the SARS-CoV-2 virus  
417 isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp  
418 linear ss-RNA with 3e9s of the PDB database as template.

419

420

421

```

Seqres  SGFRKMAFPSSGKVEGCMVQVTCGTTTLNGLWDDTVYCPRHVICIAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGH 80
2a5i.1. (AB) SGFRKMAFPSSGKVEGCMVQVTCGTTTLNGLWDDTVYCPRHVICIAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGH 80

Seqres  SMQNCCLRLKYDTSNPKTFKYKFFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIIDYDCVSFC 160
2a5i.1. (AB) SMQNCCLRLKYDTSNPKTFKYKFFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIIDYDCVSFC 160

Seqres  YMHHEMLPTGVHAGTDLEGKFYGFVDRQTAQAAGTDTITLNVLANLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNMF 240
2a5i.1. (AB) YMHHEMLPTGVHAGTDLEGKFYGFVDRQTAQAAGTDTITLNVLANLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNMF 240

Seqres  PLTQDHVDILGFLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTPFDDVVRQCSGVTFQ 306
2a5i.1. (AB) PLTQDHVDILGFLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTPFDDVVRQCSGVTFQ 306

```

423 **Fig. 6** Alignment of the 305 residues from 3268-3573 aa of the Novel Coronavirus COVI-19  
 424 with the template 2a5i

```

Seqres  ASMEVKTIKVFTTYDNTNLHTQLVDMSTYGGQFGFTYLDGADVTIKIKPHYNHEGKTFEVLPSDDTLRSEAFEYYHTLDESFLGR 85
3e9s.1.A ASMEVKTIKVFTTYDNTNLHTQLVDMSTYGGQFGFTYLDGADVTIKIKPHYNHEGKTFEVLPSDDTLRSEAFEYYHTLDESFLGR 85

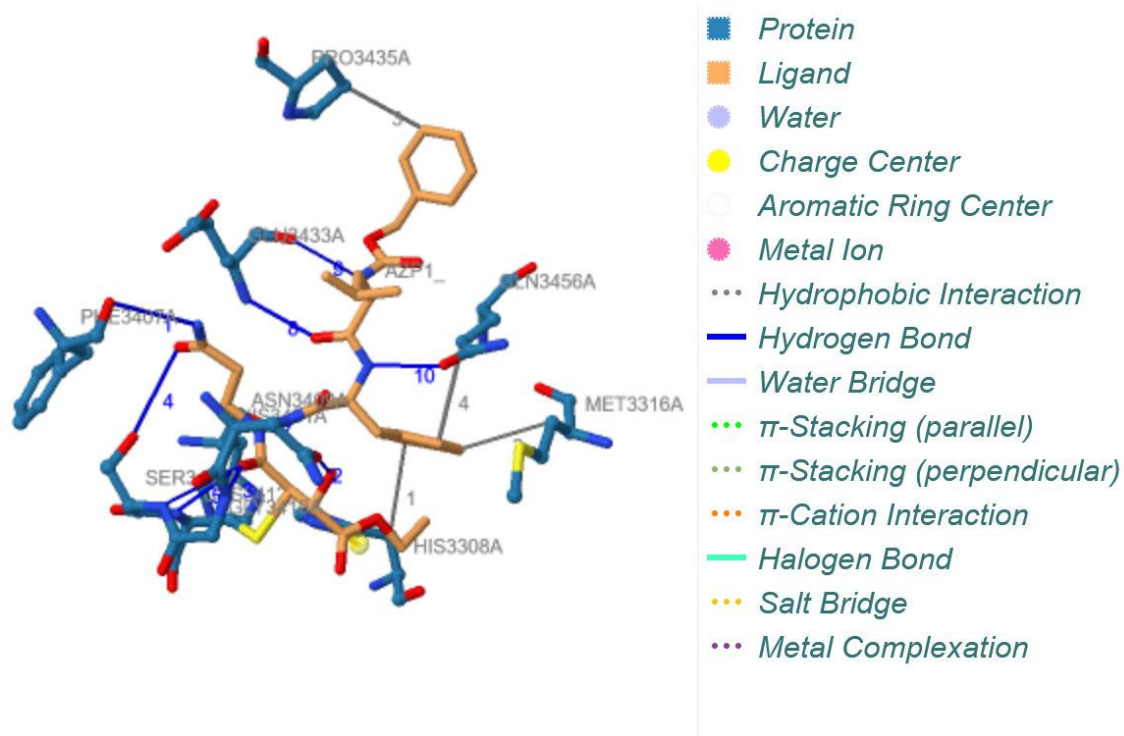
Seqres  YMSALNHTKKWKFFQVGGLTSLKWADNNCYLSSVLLALQQLLEVKNAPALQEAAYRARAGDAANFCALILAYSNKTVGELGDVRE 170
3e9s.1.A YMSALNHTKKWKFFQVGGLTSLKWADNNCYLSSVLLALQQLLEVKNAPALQEAAYRARAGDAANFCALILAYSNKTVGELGDVRE 170

Seqres  TMTHLLQHANLESARKVNLNVVCKKCGQKTTTLTGVEAVMYMGTLSYDNLKTGVSIQVCVGRDATQYLVQQESSFVMM$APPAEYK 255
3e9s.1.A TMTHLLQHANLESARKVNLNVVCKKCGQKTTTLTGVEAVMYMGTLSYDNLKTGVSIQVCVGRDATQYLVQQESSFVMM$APPAEYK 255

Seqres  LQOGTFLCANEYTGNYQCCHYTHITAKETLYRIDGAHLTKMSEYKGPVTDVFKETSXTTIR 318
3e9s.1.A LQOGTFLCANEYTGNYQCCHYTHITAKETLYRIDGAHLTKMSEYKGPVTDVFKETSXTTIR 317

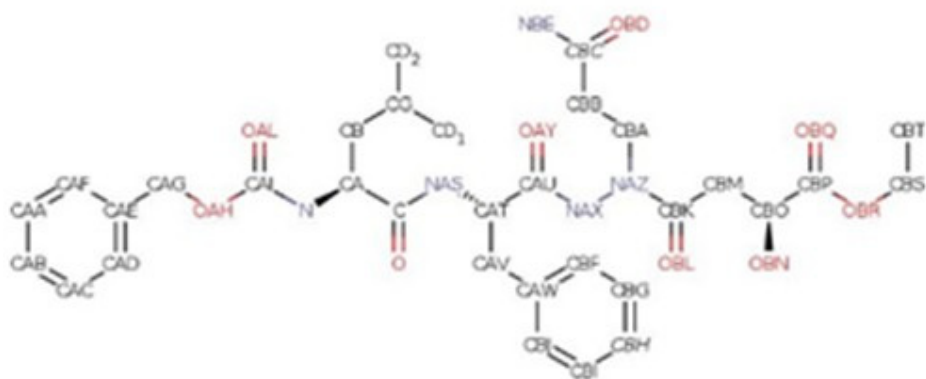
```

426 **Fig.7** the alignment of the 315 residues from 1568-1882 aa of the Novel Coronavirus COVI-  
 427 19 with the template 3e9s



428

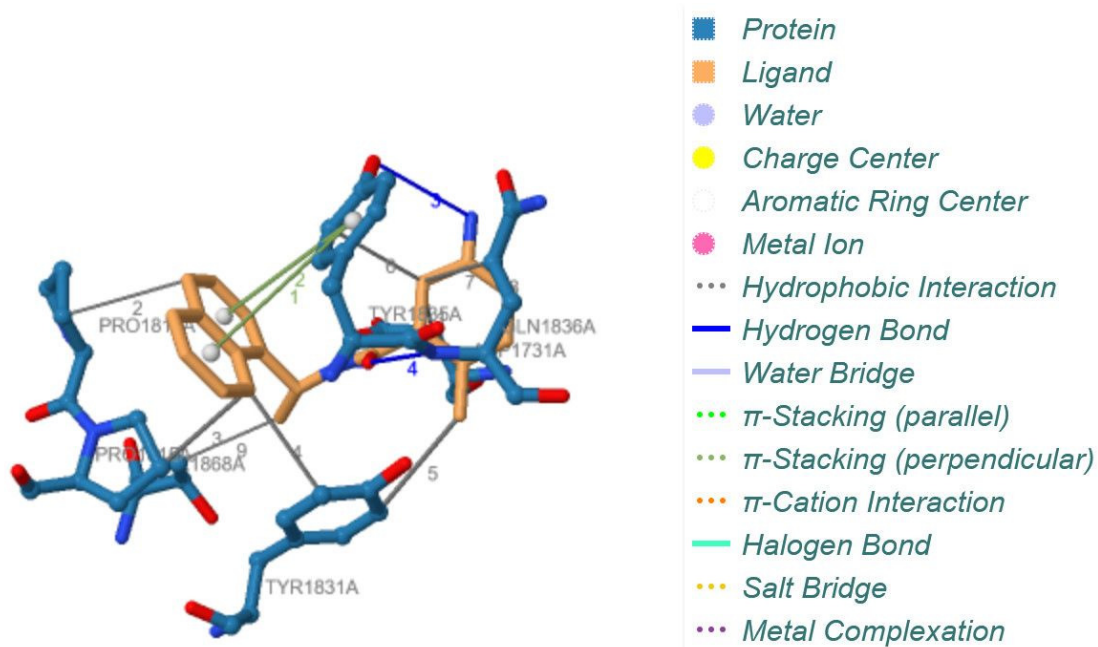
429 **Fig.8** Protein Ligand interaction between the C3 like peptidase with aza-peptide epoxide of the  
 430 model with the template 2a5i



431

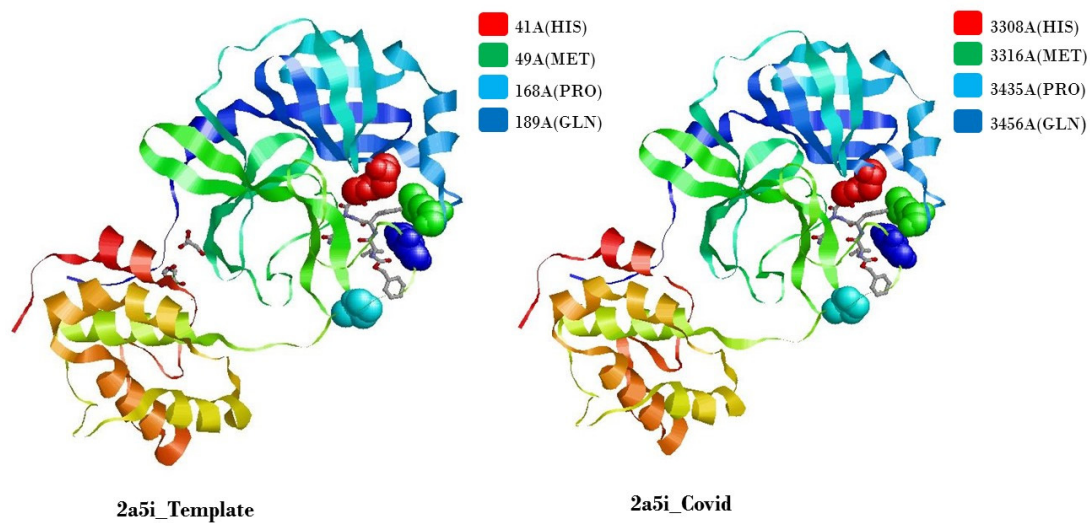
432 **Fig. 9** Structure of Aza-Peptide Epoxide (APE) ethyl (2S)-4-[(3-amino-3-oxo-propyl)-[[[(2S)-  
 433 2-[[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-phenyl-  
 434 propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate with possible anti Coronavirus  
 435 activity – (Source <https://www.rcsb.org/ligand/AZP>)





**Fig.10** Protein Ligand interaction between the Novel Coronavirus non structural protein and papain-like protease of the model with the template 3e9s



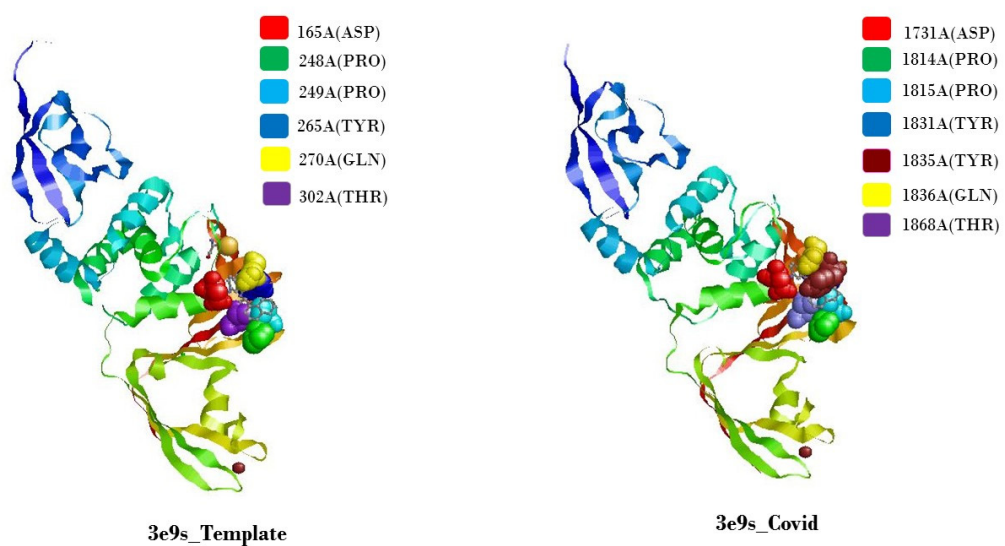


**Fig.11** Comparison of the hydrophobic interaction of the binding of the ligand AZP between the SARS-CoV-2 protein and the template 2a5i of SARS CoV

462

463

464



465

466 **Fig.12** comparison of the hydrophobic interaction of the binding of the ligand AZP between the  
467 SARS-CoV-2 protein and the template 3e9s of SARS CoV

468

469

