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[1] *, Divya Bhanu[1,2] and Anjani Alluri[3]

[1]ICAR - Central Research Institute for Dryland Agriculture

Santoshnagar, Hyderabad – 500059, India

Corresponding author email: arunshank@gmail.com

[2] Centre for Plant Molecular Biology, Osmania University, Hyderabad, India

[3] 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. This region was conserved in 98 genomes of SARS-CoV-2. 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. This region was conserved in 250 genomes of SARS-CoV-2. It is possible that these viral inhibiters 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

34 type of Coronavirus - Middle East Respiratory Syndrome (MERS-CoV) and now the SARS-

35 CoV-2 has emerged in China. This repeated onslaught of these viruses goes to show that it can

36 assume pandemic proportions at any time and at any place.

37 The family of viruses belonging to Coronaviridae mainly consist of virulent pathogens that

38 have a zoonotic property and this large family of corona viruses, have been known to be

39 circulating in animals including camels, cats and bats. It has been seen in the past that Severe

40 Acute Respiratory Syndrome associated coronavirus (SARS-CoV) and Middle East

41 Respiratory Syndrome-associated coronavirus (MERS-CoV) belonging to this family of

42 viruses can be transmitted from animals to humans and can cause respiratory diseases. Human

43 to human transmission on this virus has been a concern and due to this search for antiviral

44 compounds and vaccine development for this family of virus becomes the need of the hour.

45 The SARS was first seen in 2002 in Guangdong province of China, and later spread globally

46 and has caused close to about 8096 cases (WHO 2004, de Vit et al., 2016). In 2012, a novel

47 betacoronavirus, designated Middle East respiratory syndrome coronavirus or MERS-CoV

48 associated with severe respiratory disease in humans, emerged in the Arabian Peninsula (de

49 Wit et al., 2013).

50 The World Health Organization (WHO), China Country Office was informed of cases of

51 pneumonia of unknown aetiology in Wuhan City, Hubei Province, on 31 December 2019

52 (WHO 2020). A novel coronavirus currently termed SARS-COV-2 was officially announced

53 as the causative agent by Chinese authorities on 7 January 2020. As on 20 Feb 2020 China's

54 National Health Commission reported that there are 74,280 confirmed cases in China (Fig.1).

55 The World Health Organization reported 924 confirmed cases in 25 countries outside China

56 (WHO Situation Report 29 2020). This novel corona virus has been designated as SARS-CoV-

57 2.

58

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

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

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

82   We hypothesised that there can be some proteins in the large chuck of proteins in the SARS-

83   CoV-2 that could have homology with the Non-structural protein 3 (nsp3) SARS CoV and

84   these proteins can possibly have binding sites with ligands that can bind with known ligand

85   with antiviral properties.

86   Here in this study we did a complete bioinformatic analysis, sequence alignment, comparison

87   of multiple sequences of the SARS-CoV-2 whole genome sequences, the Spike protein and

88   the polyproteins for homology with known spike proteins and also analysed receptor binding

89   sites for possible vaccine development.

90   **Materials and Methods**

91   Six complete viral genome sequences, seven polyproteins (RdRp region) and seven

92   glycoproteins available on NCBI portal on 4 Feb 2020 were taken for analysis. The sequence

93   details and GenBank accession numbers are listed in Supplementary Table 1. Amongst the

94   seven polyproteins, five are of Wuhan pneumonia virus isolate SARS-COV-2 and two

95   sequences are of Wuhan pneumonia virus isolate SI200040-SP. The seven Glycoproteins are

96   of the same isolate, Wuhan pneumonia virus isolate SARS-COV-2.

97   The available polyproteins (RdRp region) and glycoprotein sequences were retrieved from

98   Genbank, NCBI (Benson et al., 2000). These sequences were translated to amino acid

99   sequences using sorted six frame translation with Bioedit (Hall et al., 2011). Multiple sequence

100   alignment of the translated protein sequences was performed and phylogenetic tree was

101   constructed using Mega-X (Kumar et al., 2018). The alignment shows that amongst the seven

102   polyproteins, five sequences were identical being from the same isolate and two other

103   sequences of the other isolate are identical. Similar analysis of the seven glycoproteins was

104   done, all the seven glycoprotein sequences were found to be identical. Therefore, further

105   analysis was carried out for three sequences.

106   1.   MN938385.1 SARS-CoV-2 virus isolate SARS-COV-2 _HKU-SZ-001_2020 ORF1ab

107   polyprotein, RdRp region, (orf1ab) gene, partial cds: 0 to 284: Frame 3 95 aa

108   2.   MN970003.1 SARS-CoV-2 virus isolate SI200040-SP orf1ab polyprotein, RdRP

109   region, (orf1ab) gene, partial cds: 2 to 289: Frame 2 96 aa

110   3.   MN938387.1 SARS-CoV-2 virus isolate SARS-COV-2 _HKU-SZ-001_2020 surface

111   glycoprotein (S) gene, partial cds: 1 to 105: Frame 1 35 aa

112   Expasy proteomics server (Gasteiger et al., 2003) was used to study the protein sequence and

113   structural details. These peptides were studied for their physio-chemical properties using the

114   tool Protparam (Gasteiger et al., 2005). The secondary structure analysis was done using Chou

115   and Fasman algorithm with CFSSP (Kumar, 2013). To generate the 3D structure from the fasta

116   sequence, homology modelling was performed and the templates were identified. The model

117   was built using the template with highest identity. Swiss-model (Schwede et al., 2003) was

118   used to build and validate the 3D model, structural assessment was also performed to validate

119   the model built.

120   Complete genome sequence of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank

121   Accession Number MN908947.3) which has 29903 bp ss-RNA   linear   was translated sorted

122   6 frame with minimum ORF of 20 with any start codon and the resultant protein sequence was

123   used for homology modelling, homology models where done with large chunks of proteins

124   21503 to 25381 in Frame 2 with 1293 amino acids, 13450 to 21552 in Frame 1 with 2701

125   amino acids and 254 to 13480 in Frame 2 with 4409 amino acids.

126   SWISS-MODEL server was used for homology modelling (Waterhouse et al 2018) where

127   computation was on ProMod3 engine which is based on Open Structure (Biasini et al 2013).

128   Structural information is extracted from the template, sequence alignment is used to define

129   insertions and deletions.

130   Protein ligand interaction profile with hydrogen bonding, hydrophobic interactions, salt bridges

131   and π-Stacking was done with PLIP server (Salentin et al., 2015)

132   **Results and Discussion**

133   The physico- chemical properties and primary structure parameters of the 7 polyproteins RdRp

134   region of the SARS-CoV-2 virus isolate is given in Table 1. RdRP forms an important part of

135   the viral genome where in the RNA viruses its function is to catalyze the synthesis of the RNA

136   strand complementary to a given RNA template.

137   The isolates SI200040-SP orf1ab polyprotein and the isolate SI200121-SP orf1ab polyprotein

138   had 2 reading frames as compared to the rest of the isolates which had 3 reading frames. The

139   presence of multiple reading frames suggests the possibility of overlapping genes as seen in

140   many virus and prokaryotes and mitochondrial genomes. This could affect how the proteins

141   are made. The number of amino acid residues in all the polyproteins were the same expect one

142   isolate SI200040-SP which had one amino acid more than the other polyproteins. The

143   extinction coefficients of the two isolates SI200040-SP orf1ab polyprotein and the isolate

144   SI200121-SP orf1ab polyprotein was much higher compared to the rest of the polyproteins.

145   The extinction coefficient is important when studying protein-protein and protein-ligand

146   interactions. The instability index of these two isolates was also high when compared to the

147   others indicating the that these two isolates are instable. Regulation of gene expression by

148   polyprotein processing is known in viruses and this is seen in many viruses that are human

149   pathogens (Yost et al 2013).

150   The isolates here like many other viruses may be using replication strategy which could involve

151   the translation of a large polyprotein with subsequent cleavage by viral proteases. The two

152   isolates SI200040-SP orf1ab polyprotein and the isolate SI200121-SP orf1ab polyprotein also

153   showed shorter half-lives as compared to the other isolates indicating that they are susceptible

154   to enzymatic degradation.

155   The tertiary structure analysis of the isolate SARS-CoV-2 _HKU-SZ-001_2020 ORF1ab

156   polyprotein is given in Table 2. It is seen that the polyprotein has a 98.94 percent identity with

157   PDB structure 6nur.1.A which is a hetero-1-2-1-mer. The polyprotein is an RNA directed RNA

158   polymerase. The protein is identical to the SARS-Coronavirus NSP12 bound to NSP7 and

159   NSP8 co-factors (Kirchdoerfer and Ward 2019). In SARS it is basically a nonstructural protein

160   with NSP12 being the RNA dependent RNA polymerase and the co factors NSP 7 and NSP 8

161   having the function of forming hexadecameric complexes and also act as processivity clamp

162   for RNA polymerase and primase (Fehr et al., 2016). This structure as in SARS CoV here in

163   SARS-CoV-2 may be involved in the machinery of core RNA synthesis and can be a template

164   for exploring antiviral properties.

165   The phylogenetic tree of the seven polyproteins is shown in Fig.2. It is seen that two

166   polyproteins were distinctly different from the rest. The Phylogenetic tree of the seven

167   glycoproteins of the SARS-CoV-2 virus isolate is shown in Fig.3, it is seen that the

168   glycoproteins are similar in all the isolates. Multiple alignment of the Polyproteins of the

169   SARS-CoV-2 is shown in Supplementary Fig.1.

170   This structure as in SARS CoV here in SARS-CoV-2 may be involved in the machinery of core

171   RNA synthesis and can be a template for exploring antiviral properties. Based on its functions

172   in the SARS CoV and its identity to the SARS-CoV-2, it is possible that it has the same

173   functions in SARS-CoV-2 an RNA polymerase which does de novo initiation and primer

174   extension with possible exonuclease activities, the activity itself being primer dependent useful

175   for understanding the mechanism of SARS-CoV-2 replication and can be used as an antiviral

176   target (Te Velthuis et al 2012; Te Velthuis et al 2010; Subissi et al 2014; Subissi et al 2014).

177

178   The polyprotein also has an identity of 19.74 percent with an ABC-type uncharacterized

179   transport system periplasmic component-like protein, this protein is known to be a substrate

180   binding protein and possible binding can be explored here (Bae et al 2019).

181   The homology model developed from the residues 254 to 13480 in Frame 2 with 4409 amino

182   acids from the Complete genome sequence of the SARS-CoV-2 virus isolate Wuhan-Hu-1

183   (Genbank Accession Number MN908947.3) which has 29903 bp with linear ss-RNA   linear

184   showed interesting template alignments, in all the model aligned with 50 templates from the

185   PDB database with most of them being replicase polyprotein 1ab which is a SARS-CoV

186   papain-like protease (Daczkowski 2017). The maximum similarity of 97.3 percent was with

187   template structure of a Nsp9 protein from SARS-coronavirus indicating that this novel

188   coronavirus has high degree of similarity with the SARS-coronavirus and this can be used for

189   gaining insights into vaccine development. Nsp 9 is an RNA binding protein and has an

190   oligosaccharide/oligonucleotide fold-like fold, this protein can have an important function in

191   the replication machinery of the virus and can be important when designing antiviral for this

192   virus (Egloff et al 2004).

193   Two models were developed, one from residues   3268 -3573 in Frame 2 with   306 amino

194   acids and the other from the part of the genome residues   1568-1882 in Frame 2 with 315

195   amino acids of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number

196   MN908947.3). The models had similarity with the 3C like proteinase and a papain-like

197   protease/deubiquitinase protein which are known antiviral drug targets. Ligand binding with

198   these proteins and their action is on viral replication and inactivation can be useful in stopping

199   the viral replication (Baez-Santos et al 2015).

200

201 The homology models of the 4409 amino acid residues of the whole genome of the SARS-

202 CoV-2 virus isolate Wuhan-Hu-1 with the ligand association with templates 2a5i and 3e9s are

203 shown in Fig. 4 and Fig. 5 respectively.

204 The statistics of structural comparison with PDB templates is given in Table 5, it is seen that

205 the proteins from the SARS-CoV-2 are significantly close to the proteins of SARS CoV and

206 the amino acid alignment in the biding region is the same in both the viruses.

207 The alignment of the 305 residues from 3268-3573 aa of the Novel Coronavirus COVI-19 with

208 the template 2a5i is shown in Fig.6 and the alignment of the 315 residues from 1568-1882 aa

209 of the Novel Coronavirus COVI-19 with the template 3e9s is shown in Fig.7.

210 A PSI-BLAST of a length of 306 amino acid residues 3268 -3573 in Frame 2 from the SARS-

211 CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) was conducted

212 to ascertain the conservation of these amino acids in 250 genome sequences of SARS-CoV-2

213 and it was found that there was a complete match in these genomes of the virus. The fact that

214 the region is conserved in all these SARS-CoV-2 sequences further emphasizes this ligand

215 interaction of Aza-Peptide epoxide with the protein can be used as an antiviral in SARS-CoV-

216 2. Similarly A PSI-BLAST of a length of 315 amino acid residues 3268 -3573 in Frame 2

217 with 315 amino acid residues 1568-1882 in Frame 2 from SARS-CoV-2 virus isolate Wuhan-

218 Hu-1 (Genbank Accession Number MN908947.3) was conducted to ascertain the conservation

219 of these amino acids in 250 genome sequences of SARS-CoV-2 and it was found that there

220 was a complete match in these genomes of the virus. The fact that the region is conserved in

221 all these SARS-CoV-2 sequences further emphasizes this ligand interaction of ligand

222 GRL0617 with the protein can be used as an antiviral in SARS-CoV-2.

223 The important templates that aligned with this 4409 amino acid residues of the whole genome

224 of the SARS-CoV-2 virus isolate Wuhan-Hu-1were 2a5i of the PDB database which is a

225 crystal structure of SARS coronavirus main peptidase inhibited by an Aza-Peptide epoxide in

226 the space group C2 (Lee et al 2005) and 3e9s of the PDB database which is new class of

227 papain-like protease/deubiquitinase which when combined with ligand GRL0617 acts as

228 inhibitors blocking SARS virus replication (Ratia et al 2008). The model with template 2a5i of

229 the PDB database shows that Aza-Peptide Epoxide (APE; kinact/Ki=1900($\pm$400) M^{-1} s^{-1})

230 which is a known anti SARS agent can be used to develop a molecular target with irreversible

231 inhibitor properties. The protein ligand interaction analysis of the Novel Coronavirus C3 like

232 peptidase and aza-peptide epoxide is shown in Fig.8. The substrate binding properties and

233 structural and chemical complementarity of this Aza-Peptide Epoxide can be explored as an

234 anti - Coronavirus SARS-COV-2 agent. The APE which is ethyl (2S)-4-[(3-amino-3-oxo-

235 propyl)-[[(2S)-2-[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-

236 phenyl-propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate structure is shown in Fig.9.

237

238 The model with template 3e9s of the PDB database shows that the Coronavirus viral protein

239 can have a ligand which is a papain-like protease (PLpro) that is known to be a potent inhibitor

240 of viral replication in SARS (Ratia et al 2008).

241 The two parts of the Main protein from the whole genome of the SARS-CoV-2 aligned with

242 two SAR proteins and the ligand binding sites were similar, the alignment positions, number

243 of amino acids and ligand and the interacting residues is given in Table 3

244 The complete genome of MN908947.3 SARS-CoV-2 virus isolate Wuhan-Hu-1 encodes a

245 4409aa long protein along with the other glycoproteins and polyproteins. The homology

246 modelling of this protein showed sequence and structural alignment with two SARS proteases

247 with structural accession numbers 3e9s.1 and 2a5i.1 at positions 1568-1882 and 3268-3573

248 respectively. Reports suggests inhibition of virus replication by TTT ligand and an aza-peptide

249 epoxide inhibiting the main peptidase. The structural similarity of these templates are 83% and

250 96% respectively. The multiple sequence alignment shows complete conservation of the

251 sequence suggesting a high degree of homology. The protein ligand interaction analysis of the

252 Novel Coronavirus non structural protein and papain-like protease is shown in Fig. 10.

253 The Comparison of Hydrophobic interaction, hydrogen bonding, salt bridges of the

254 constructed model of the Novel Coronavirus protein from region 3268-3573 aa to ligand AZP

255 with Hydrophobic interaction, hydrogen bonding, salt bridges of the template 2a5i is given in

256 Suppl. Table 2, when comparing both it is seen that the binding properties are the same expect

257 for the presence of water bridge in the template 2a5i.

258

259 The Comparison of Hydrophobic interaction, hydrogen bonding, π-Stacking of the constructed

260 model of the Novel Coronavirus protein from region 1568-1882 aa to ligand Small molecule

261 Noncovalent Lead Inhibitor with the Hydrophobic interaction, hydrogen bonding, π-Stacking

262 of the template 3e9s is given in Suppl. Table 3, when comparing both it is seen that the binding

263 properties are the same except or an additional π-Stacking at Tyr in the template 2a5i. This

264 shows that there is high possibility of binding of these antiviral compounds with the regions of

265 Novel Coronavirus protein that is in homology with the SARS protein.

266 Comparison of the hydrophobic interaction of the biding of the ligand AZP between the SARS-

267 CoV-2 protein and the template 2a5i of SARS CoV is shown in Fig.11 and the comparison of

268 the hydrophobic interaction of the biding of the ligand AZP between the SARS-CoV-2 protein

269 and the template 3e9s of SARS CoV is shown in Fig.12. It is seen that the interaction is the

270 same in both proteins with the same amino acids participating in the interaction indicating that

271 there is a possibility that these ligands with antiviral properties can bind to the new virus.

272   The similarity in the amin acids involved in the Hydrophobic interactions which are short range

273   interactions and have an important role in the affinities of the ligands and receptors shows that

274   the proteins of the SARS-CoV-2 may bind with the same affinity as seen in the SARS CoV

275   and this also shows a similar action of the ligand as seen in SARS CoV, indicating that these

276   ligands can be used as antivirals in the SARS-CoV-2.

277   The targeting of this part of the genome of the SARS-CoV-2 with the antiviral compounds

278   which have shown to bind in the similar region of the SARS virus can have implication in the

279   development of an effective antiviral compound against the SARS-CoV-2 . The SARS-CoV-2

280   shows homology with the SARS coronaviral proteases, papain-like protease (PLpro) and 3C-

281   like protease (3CLpro), these proteins have the function of processing the viral polyprotein and

282   also they perform the function of stripping ubiquitin and the ubiquitin-like interferon (IFN)-

283   stimulated gene 15 (ISG15) from the hosts to facilitate coronavirus replication and help in

284   evading immune response of the host, these inhibitors can also have a role in disrupting

285   signalling cascades in infected cells and protecting the uninfected cells.

286   The chemical GRL0617 is 5-Amino-2-methyl-N-[(1R)-1-(1-naphthalenyl)ethyl]benzamide

287   and is known to inhibit the papainlike protease that is present in SARS CoV . This protease is

288   a potential target for antiviral compounds (Chaudhuri et al., 2011). We found the SARS-CoV-

289   2 has homology with this and the binding sites for this in the structural protein of the SARS-

290   CoV-2 is the same (Table 4). This compound inhibits the enzyme that is required for the

291   cleavage of the viral protein from the virus in SARS CoV, it also cleaves ubiquitin and has a

292   structural homology with the Deubiquitinases (DUBs) of the Ubiquitin-Specific Proteases

293   Compound GRL0617 binds in the S4 and S3 enzyme subsite that gets the C terminal tail of the

294   Ubiquitin (King and Finley 2014; Schauer et al., 2019). Our results show that Aza-Peptide

295   Epoxide an irreversible protease inhibitor and GRL0617 a viral replication inhibitor can be

296   used to develop inhibitors of the Novel Coronavirus SARS-COV-2.

297

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 α/β-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 Rresearch, *28*(1), 15-18.

Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., Johner, N., Schenk, A.D., Philippsen, A. and Schwede, T., 2013. OpenStructure: an integrated software framework for computational structural biology. Acta Crystallographica Section D: Biological Crystallography, 69(5), pp.701-709.

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

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

Daczkowski, C.M., Dzimianski, J.V., Clasman, J.R., Goodwin, O., Mesecar, A.D. and Pegan, S.D., 2017. Structural insights into the interaction of coronavirus papain-like proteases

and interferon-stimulated gene product 15 from different species. Journal of Molecular Biology, 429(11), pp.1661-1683.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Luo, C.M., Wang, N., Yang, X.L., Liu, H.Z., Zhang, W., Li, B., Hu, B., Peng, C., Geng, Q.B., Zhu, G.J. and Li, F., 2018. Discovery of novel bat coronaviruses in south China that

365   use the same receptor as Middle East respiratory syndrome coronavirus. Journal of
366   Virology, 92(13), pp.e00116-18.

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

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

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

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

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

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

387   Te Velthuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H. and Snijder, E.J., 2010.

388   The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent.

389   Nucleic acids research, 38(1), pp.203-214.

390   Te Velthuis, A.J., van den Worm, S.H. and Snijder, E.J., 2012. The SARS-coronavirus nsp7+

391   nsp8 complex is a unique multimeric RNA polymerase capable of both de novo

392   initiation and primer extension. Nucleic acids research, 40(4), pp.1737-1747.

393   Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T.,

394   de Beer, T.A.P., Rempfer, C., Bordoli, L. and Lepore, R., 2018. SWISS-MODEL:

395   homology modelling of protein structures and complexes. Nucleic Acids Research,

396   46(W1), pp.W296-W303.

397   World   Health   Organization   (WHO)   2004.   [Accessed   11   Feb   2020]

398   https://www.who.int/csr/don/2004_05_18a/en/

399   World Health Organization (WHO). Coronavirus. Geneva: WHO; 2020 [Accessed 4 Feb

400   2020]. Available from: https://www.who.int/health-topics/coronavirus

401   Yost, S.A. and Marcotrigiano, J., 2013. Viral precursor polyproteins: keys of regulation from

402   replication to maturation. Current Opinion in Virology, 3(2), pp.137-142.

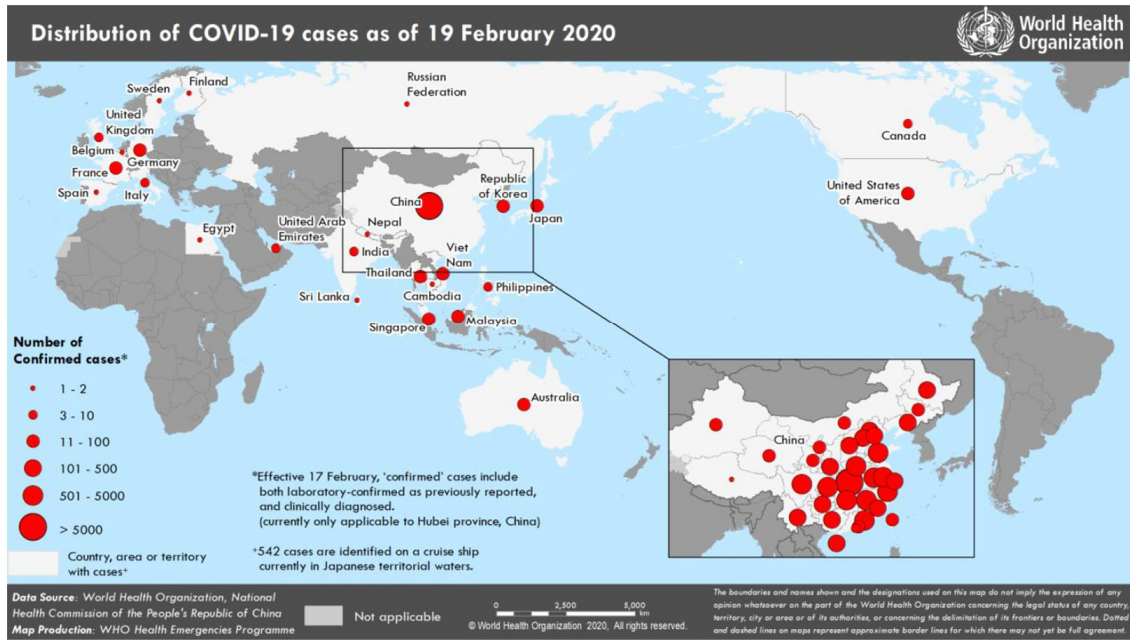403

404   **Figure Captions**

405

406 **Fig.1** Countries, territories or areas with reported confirmed cases of SARS-COV-2 , 3

407   February   2020   Source   WHO   (https://www.who.int/docs/default-

408   ource/coronaviruse/situation-reports/20200219-sitrep-30-covid-

409   19.pdf?sfvrsn=6e50645_2)

410

18

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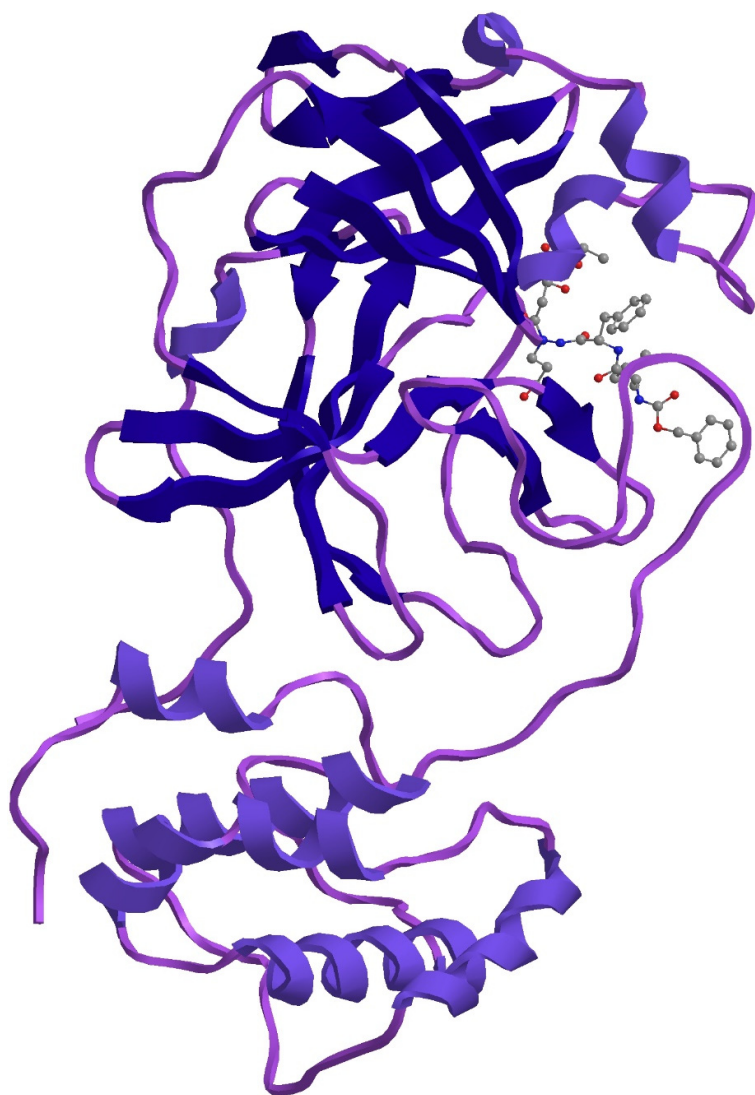
412   coronavirus 2 isolate virus isolates

413

414 **Fig.3** Phylogenetic tree of the seven glycoproteins of Severe acute respiratory syndrome

415   coronavirus 2 isolate virus isolates

416

417

418

419

Fig. 4 Homology model with ligand binding of protein from amino acids 3268 -3573 in Frame

2 with 306 amino acids of the Complete genome sequence of the SARS-CoV-2 virus

isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp

linear ss-RNA with 2a5i of the PDB database as template.

424

425

426

427

428
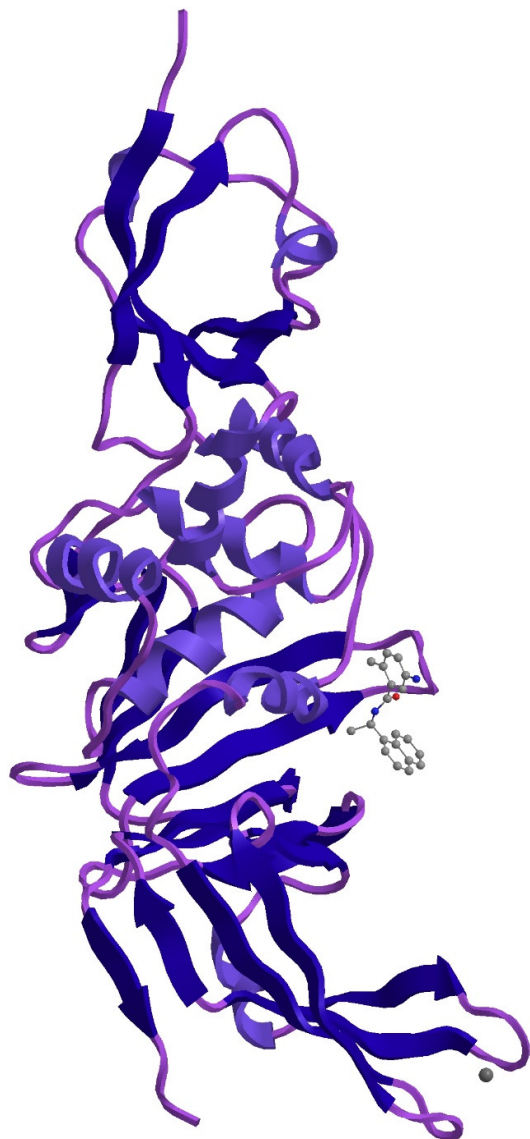
Fig. 5 Homology model with ligand binding of protein from residues 1568-1882 in Frame 2

with 315 amino acids of the Complete genome sequence of the SARS-CoV-2 virus

isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp

linear ss-RNA with 3e9s of the PDB database as template.

433

434

435

436

Fig. 6 Alignment of the 305 residues from 3268-3573 aa of the Novel Coronavirus COVI-19

with the template 2a5i

439

Fig.7 the alignment of the 315 residues from 1568-1882 aa of the Novel Coronavirus COVI-

19 with the template 3e9s
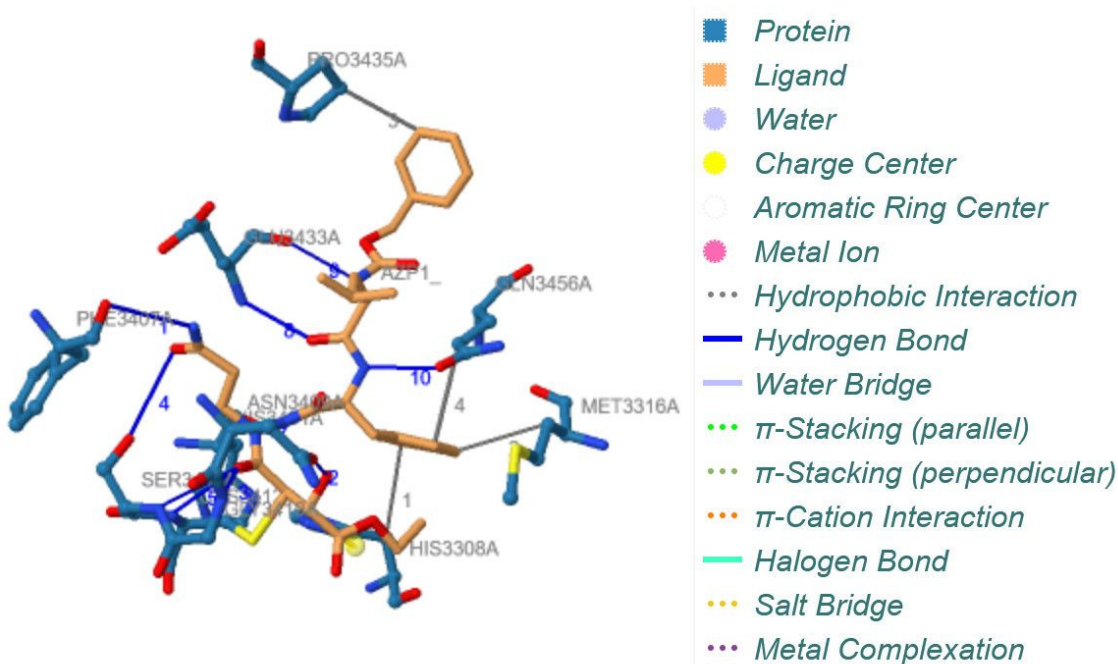
442

443   **Fig.8** Protein Ligand interaction between the C3 like peptidase with aza-peptide epoxide of the

444   model with the template 2a5i
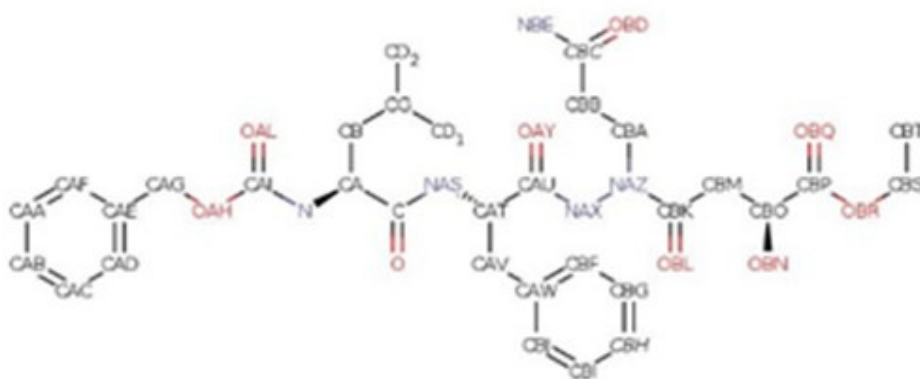
445

446   **Fig. 9** Structure of Aza-Peptide Epoxide (APE) ethyl (2S)-4-[(3-amino-3-oxo-propyl)-[[(2S)-

447   2-[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-phenyl-

448   propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate with possible anti Coronavirus

449   activity – (Source https://www.rcsb.org/ligand/AZP)

Legend:

- Protein
- Ligand
- Water
- Charge Center
- Aromatic Ring Center
- Metal Ion
- ⋯ Hydrophobic Interaction
- — Hydrogen Bond
- — Water Bridge
- ⋯ π-Stacking (parallel)
- ⋯ π-Stacking (perpendicular)
- ⋯ π-Cation Interaction
- — Halogen Bond
- ⋯ Salt Bridge
- ⋯ Metal Complexation

450
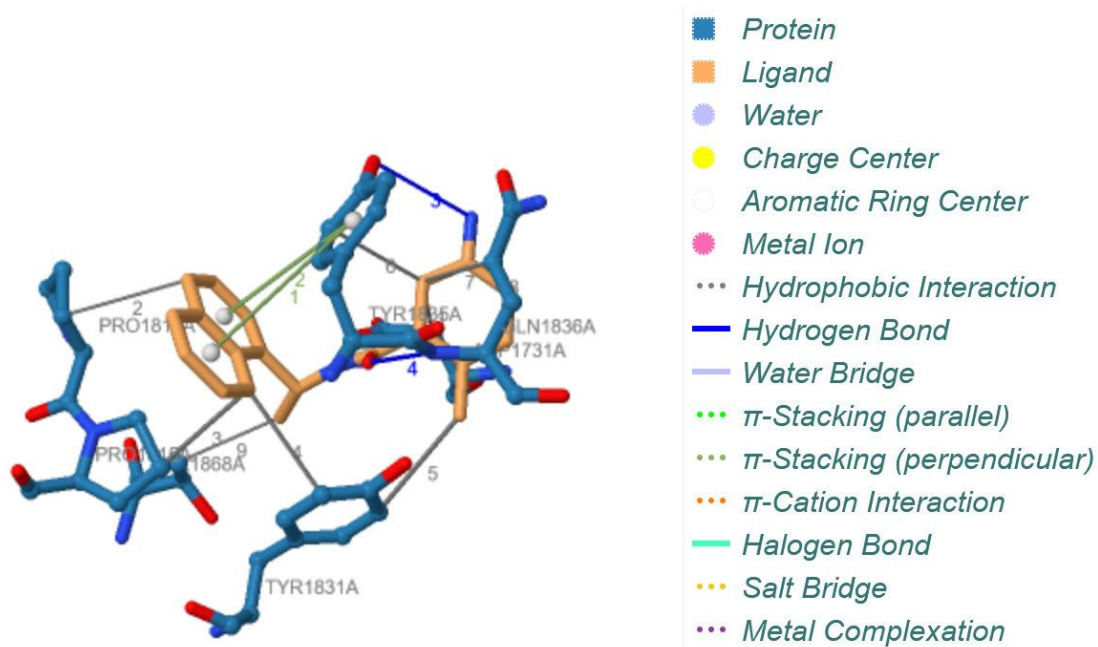
451   **Fig.10** Protein Ligand interaction between the Novel Coronavirus non structural protein and

452   papain-like protease of the model with the template 3e9s

453

454

455

456

457

458

459

460

461

462

463

2a5i_Template   2a5i_Covid

464
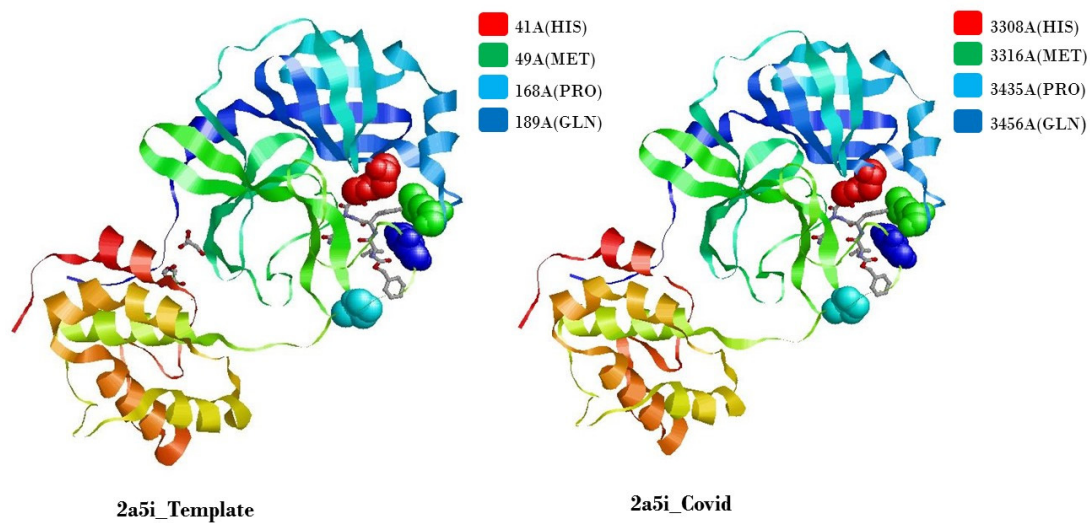
465   **Fig.11** Comparison of the hydrophobic interaction of the biding of the ligand AZP between the

466   SARS-CoV-2 protein and the template 2a5i of SARS CoV

467

468

469

470

471

472

473

474

475

476

477

478

3e9s_Template

3e9s_Covid
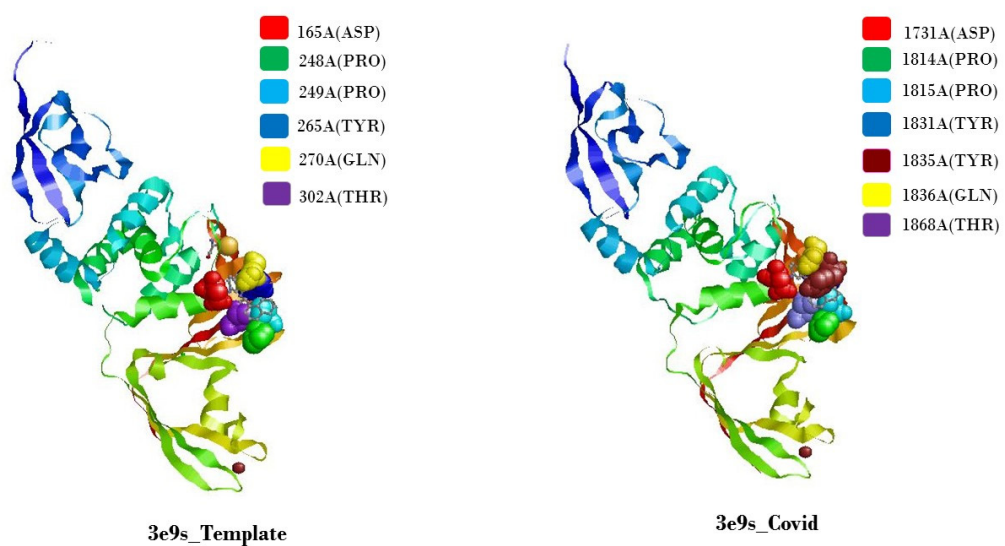
479

480   **Fig.12** comparison of the hydrophobic interaction of the biding of the ligand AZP between the

481   SARS-CoV-2 protein and the template 3e9s of SARS CoV

482

483

```
MN938385.1Wuhanisolate2019-nCoVHKU-SZ-001 po ----SYEDQDALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHL
MN938386.1Wuhanisolate2019-nCoVHKU-SZ-004 po ----SYEDQDALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHL
MN975263.1Wuhanisolate2019-nCoVHKU-SZ-007a p ----SYEDQDALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHL
MN975264.1Wuhanisolate2019-nCoVHKU-SZ-007b p ----SYEDQDALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHL
MN975265.1Wuhanisolate2019-nCoVHKU-SZ-007c p ----SYEDQDALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHL
MN970003.1WuhanisolateSI200040 polyprotein  KHLIPLMYKGLPWNVVRIKIVQMLSDT-LKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCLCDRR-ATCFSTASDTYACWHHSIG--FDYVYNPFM
MN970004.1WuhanisolateSI200121 polyprotein  KHLIPLMYKGLPWNVVRIKIVQMLSDT-LKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCLCDRR-ATCFSTASDTYACWHHSIG--FDYVYNPFM
Clustal Consensus                             .  :.  :   .: ::: :::  **  .   .  .*..:  *    :: :   .:. * ** . :*. *. **: :   :. * **.:
```

484

485   Supplementary Fig.1 Multiple alignment of the Polyproteins of Severe acute respiratory

486   syndrome coronavirus 2 isolate virus isolates

487

488

489