Analysis of Whole Genome Sequences and Homology Modelling of a 3C Like Peptidase and a Non-Structural Protein of the Novel Coronavirus COVID-19 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 Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) of this family have emerged before and now the Novel COVID-19 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 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 Novel COVID-19 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 vaccine development. Our results showed that the tertiary structure of the polyprotein isolate COVID-19 _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 254 to 13480 in Frame 2 with 4409 amino acids) of the Novel COVID-19 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 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 inhibiters can be used for vaccine development for the Novel COVID-19.

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 Novel

Coronavirus COVID-19 has emerged in China. This repeated onslaught of these viruses goes to show that it can assume pandemic proportions at any time and at any place.

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

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

The World Health Organization (WHO), China Country Office was informed of cases of pneumonia of unknown aetiology in Wuhan City, Hubei Province, on 31 December 2019 (WHO 2020). A novel coronavirus currently termed COVID-19 was officially announced as the causative agent by Chinese authorities on 7 January 2020. As on 3 Feb 2020 China's National Health Commission reported that there are 20,438 confirmed cases in China, including 15 in Hong Kong and eight in Macao. The self-governing island of Taiwan reported 10 cases. The World Health Organization reported 319 confirmed cases in 23 countries outside China (WHO Situation Report 21 2020). This novel corona virus has been designated as novel coronavirus COVID-19.

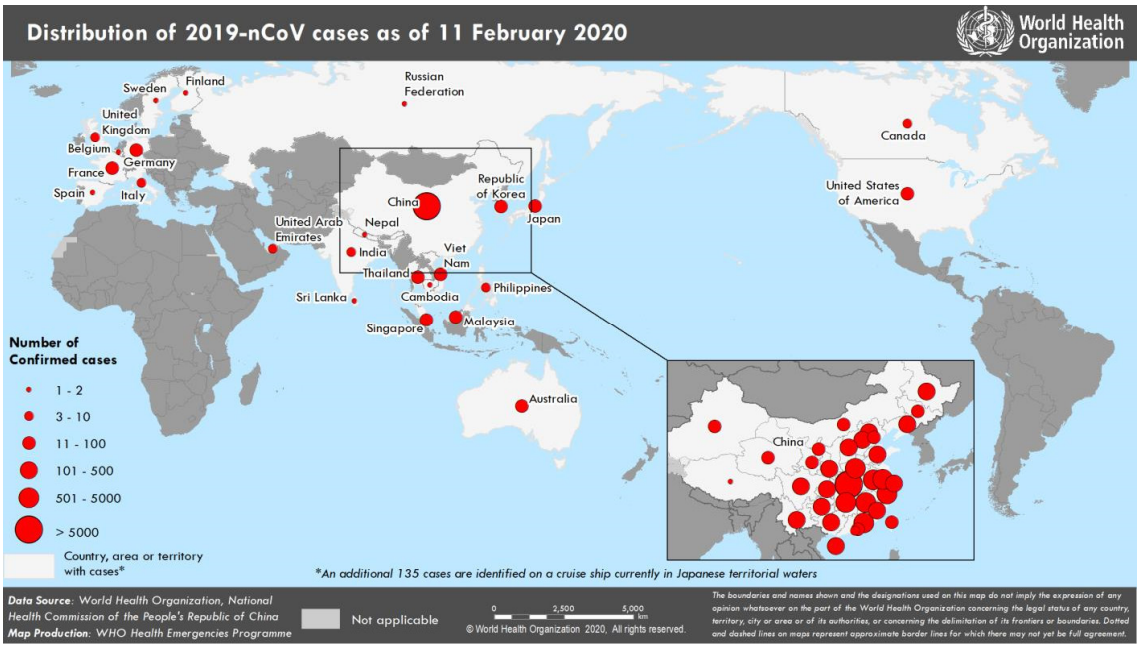


Fig.1 Countries, territories or areas with reported confirmed cases of COVID-19 , 3 February 2020 Source WHO (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2).

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 novel coronavirus COVID-19 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).

Here in this study we did a complete bioinformatic analysis, sequence alignment, comparison of multiple sequences of the Novel COVID-19 whole genome sequences, the Spike protein and the polyproteins for homology with known spike proteins and also analysed receptor binding sites for possible vaccine development.

Materials and Methods

Six complete viral genome sequences, seven polyproteins (RdRp region) and seven glycoproteins available on NCBI portal on 4 Feb 2020 were taken for analysis. The sequence details and GenBank accession numbers are listed in Table 1. Amongst the seven polyproteins, five are of Wuhan pneumonia virus isolate COVID-19 and two sequences are of Wuhan pneumonia virus isolate SI200040-SP. The seven Glycoproteins are of the same isolate, Wuhan pneumonia virus isolate COVID-19.

Table 1 List of available Wuhan seafood market pneumonia virus isolate sequences at NCBI

Genbank Accession Number	Title	Description
MN988713.1	Wuhan seafood market pneumonia virus isolate COVID-19 /USA-IL1/2020	Complete genome
MN938384.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-002a_2020	Complete genome

MN975262.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-005b_2020	Complete genome
MN985325.1	Wuhan seafood market pneumonia virus isolate COVID-19 /USA-WA1/2020	Complete genome
NC_045512.2	Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1	Complete genome
MN908947.3	Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1	Complete genome
MN938385.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-001_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN938386.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-004_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN975263.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-007a_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN975264.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-007b_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region

MN975265.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-007c_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN970003.1	Wuhan seafood market pneumonia virus isolate SI200040-SP orf1ab polyprotein, RdRP region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN970004.1	Wuhan seafood market pneumonia virus isolate SI200121-SP orf1ab polyprotein, RdRP region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN938387.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-001_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN938388.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-002b_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN938389.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-004_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN938390.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-005_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN975266.1	Wuhan seafood market pneumonia virus isolate COVID-	Glycoprotein

	19 _HKU-SZ-007a_2020 surface glycoprotein (S) gene, partial cds	
MN975267.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-007b_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN975268.1	Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-007c_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein

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

1. MN938385.1 Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-001_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds: 0 to 284: Frame 3 95 aa

2. MN970003.1 Wuhan seafood market pneumonia virus isolate SI200040-SP orf1ab polyprotein, RdRP region, (orf1ab) gene, partial cds: 2 to 289: Frame 2 96 aa

3. MN938387.1 Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-001_2020 surface glycoprotein (S) gene, partial cds: 1 to 105: Frame 1 35 aa

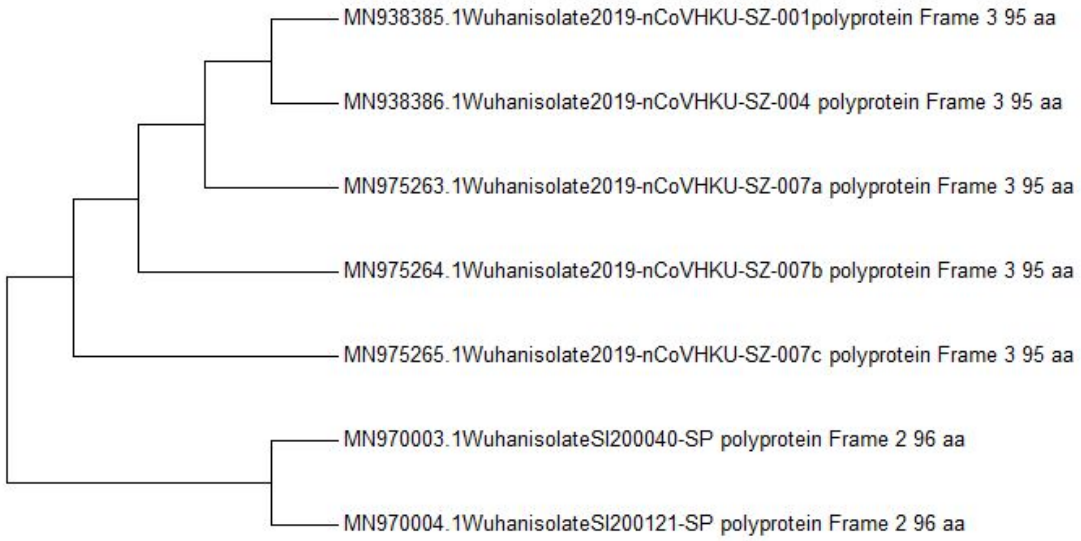
Expasy proteomics server (Gasteiger et al., 2003) was used to study the protein sequence and structural details. These peptides were studied for their physio-chemical properties using the tool Protparam (Gasteiger et al., 2005). The secondary structure analysis was done using Chou and Fasman algorithm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

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

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

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

Results and Discussion



The phylogenetic tree of the seven polyproteins is shown in Fig.2. It is seen that two polyproteins were distinctly different from the rest.

Fig.2 Phylogenetic tree of the seven polyproteins of Wuhan seafood market pneumonia virus isolate

The tertiary structure analysis of the isolate COVID-19 _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 and a 19.74 percent identity with a ABC-type uncharacterized transport system periplasmic component-like protein.

The Phylogenetic tree of the seven glycoproteins of the Wuhan seafood market pneumonia virus isolate is shown in Fig.3, it is seen that the glycoproteins are similar in all the isolates.

Fig.3 Phylogenetic tree of the seven polyproteins of Wuhan seafood market pneumonia virus isolate

Table 2 Tertiary Structure of Wuhan seafood market pneumonia virus isolate COVID-19 _HKU-SZ-001_2020 ORF1ab polyprotein alignment with templates

PDB Template	Gene	Identity
6nur.1.A	NSP12	98.947
1khv.1.A	RNA-directed RNA polymerase	8.97
1khv.2.A	RNA-directed RNA polymerase	8.97
5z6v.1.A	ABC-type uncharacterized transport system periplasmic component-like protein	19.74
6k1y.1.A	ABC-type uncharacterized transport system periplasmic component-like protein	19.74
2ckw.1.A	RNA-directed RNA polymerase	10.53

2uuw.1.A	RNA-directed RNA polymerase	10.67
2wk4.1.A	Protease-polymerase p70	10.67
2wk4.1.B	Protease-polymerase p70	10.67
2yan.1.A	Glutaredoxin-3	12.50
2yan.2.A	Glutaredoxin-3	12.50

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 complex es and also act as processivity clamp for RNA polymerase and primase (Fehr et al., 2016).

Multiple alignment of the Polyproteins of the Novel Coronavirus COVID -19 is shown in Fig.4

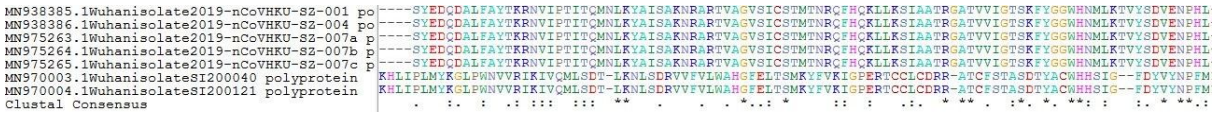


Fig.4 Multiple alignment of the Polyproteins of the Novel Coronavirus COVID -19

This protein as in SARS virus may be involved in the assembly of the coronavirus core RNA-synthesis machinery. This polyprotein can be taken as a template to design antiviral compounds. 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 primary structure parameters of the 7 polyproteins RdRp region of the Wuhan seafood market pneumonia virus isolate is given in Supplementary Table 3. 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.

Table 3 Physico-chemical properties of polyproteins of Novel Coronavirus 2019 n-CoV

Accession Number	MN938385.1	MN938386.1	MN975263.1	MN975264.1	MN975265.1	MN970003.1	MN970004.1
Reading Frame	3	3	3	3	3	2	2
Number of amino acids	95	95	95	95	95	96	96
Molecular weight	10640.22	10640.22	10640.22	10640.22	10640.22	11239.26	11239.26
Theoretical pI	9.87	9.87	9.87	9.87	9.87	8.9	8.9
Formula	$C_{472}H_{752}N_{134}O_{138}S_4$	$C_{472}H_{752}N_{134}O_{138}S_4$	$C_{472}H_{752}N_{134}O_{138}S_4$	$C_{472}H_{752}N_{134}O_{138}S_4$	$C_{472}H_{752}N_{134}O_{138}S_4$	$C_{516}H_{786}N_{132}O_{132}S_9$	$C_{516}H_{786}N_{132}O_{132}S_9$
Total number of atoms	1500	1500	1500	1500	1500	1575	1575
Extinction coefficients	12950	12950	12950	12950	12950	24200	24200
Instability index	20.51	20.51	20.51	20.51	20.51	29.66	29.66
Aliphatic index	80.11	80.11	80.11	80.11	80.11	89.27	89.27
Grand average of hydropathicity (GRAVY)	-0.264	-0.264	-0.264	-0.264	-0.264	0.161	0.161
Estimated half-life	1.9 hours (mammalian reticulocytes, in vitro).	1.9 hours (mammalian reticulocytes, in vitro).	1.9 hours (mammalian reticulocytes, in vitro).	1.9 hours (mammalian reticulocytes, in vitro).	1.9 hours (mammalian reticulocytes, in vitro).	1.3 hours (mammalian reticulocytes, in vitro).	1.3 hours (mammalian reticulocytes, in vitro).
	>20 hours (yeast, in vivo).	>20 hours (yeast, in vivo).	>20 hours (yeast, in vivo).	>20 hours (yeast, in vivo).	>20 hours (yeast, in vivo).	3 min (yeast, in vivo).	3 min (yeast, in vivo).
	>10 hours (*Escherichia coli*, *in vivo*).	>10 hours (*Escherichia coli*, *in vivo*).					

The isolates SI200040-SP orf1ab polyprotein and the isolate SI200121-SP orf1ab 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 expect one isolate SI200040-SP which had one amino acid more than the other polyproteins. The extinction coefficients of the two isolates SI200040-SP orf1ab polyprotein and the isolate SI200121-SP orf1ab 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 the 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 orf1ab polyprotein and the isolate SI200121-SP orf1ab polyprotein also showed shorter half-lives as compared to the other isolates indicating that they are susceptible to enzymatic degradation.

The homology model developed from the residues 254 to 13480 in Frame 2 with 4409 amino acids from the Complete genome sequence of the Wuhan seafood market pneumonia 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.

The homology models of the 4409 amino acid residues of the whole genome of the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1 with the ligand association with templates 2a5i and 3e9s are shown in Fig. 5 and Fig. 6 respectively. The model has similarity with the 3Clike proteinase and a papain-like protease/deubiquitinase protein which are known

antiviral drug targets. Ligand binding and their action is on viral replication and inactivation can be useful in stopping the viral replication (Baez-Santos et al 2015).

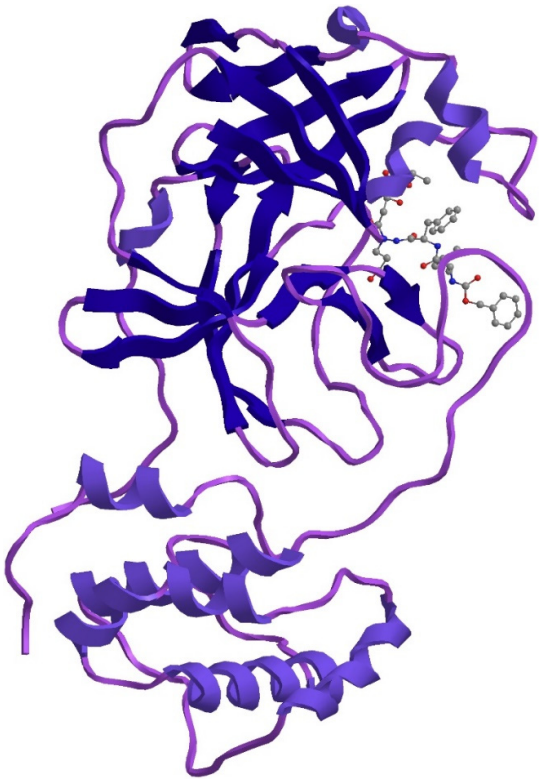


Fig. 5 Homology model with ligand binding of protein from 4409 amino acids 254 to 13480 in Frame 2 of the Complete genome sequence of the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp linear ss-RNA with 2a5i of the PDB database as template.

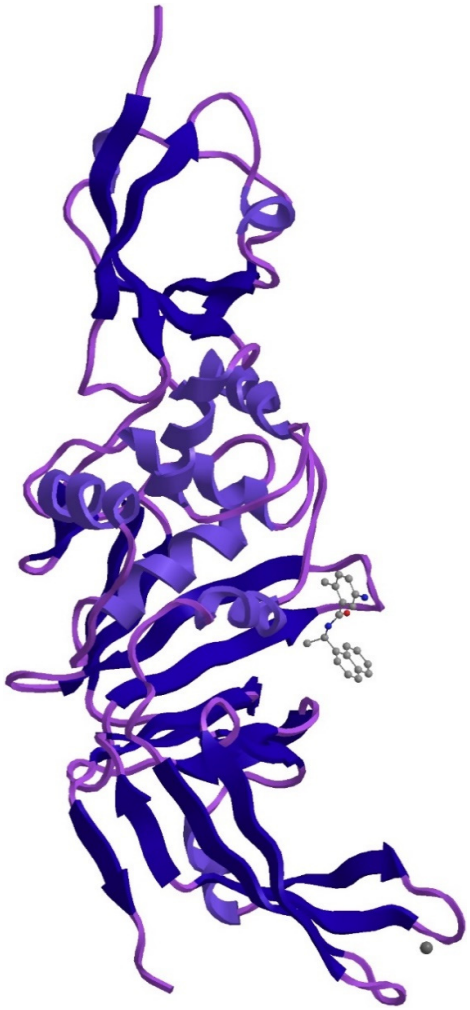


Fig. 6 Homology model with ligand binding of protein from 4409 amino acids 254 to 13480 in Frame 2 of the Complete genome sequence of the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp linear ss-RNA with 3e9s of the PDB database as template.

The important templates that aligned with this 4409 amino acid residues of the whole genome of the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1were 2a5i of the PDB database which is a crystal structure of SARS coronavirus main peptidase inhibited by an Aza-Peptide epoxide in the space group C2 (Lee et al 2005) and 3e9s of the PDB database which

is new class of papain-like protease/deubiquitinase which when combined with ligand GRL0617 acts as inhibitors blocking SARS virus replication (Ratia et al 2008). The model with template 2a5i of the PDB database shows that Aza-Peptide Epoxide (APE; kinact/Ki=1900(±400) M^{-1} s^{-1}) which is a known anti SARS agent can be used to develop a molecular target with irreversible inhibitor properties.

The protein ligand interaction analysis of the Novel Coronavirus C3 like peptidase and aza-peptide epoxide is shown in Fig.7

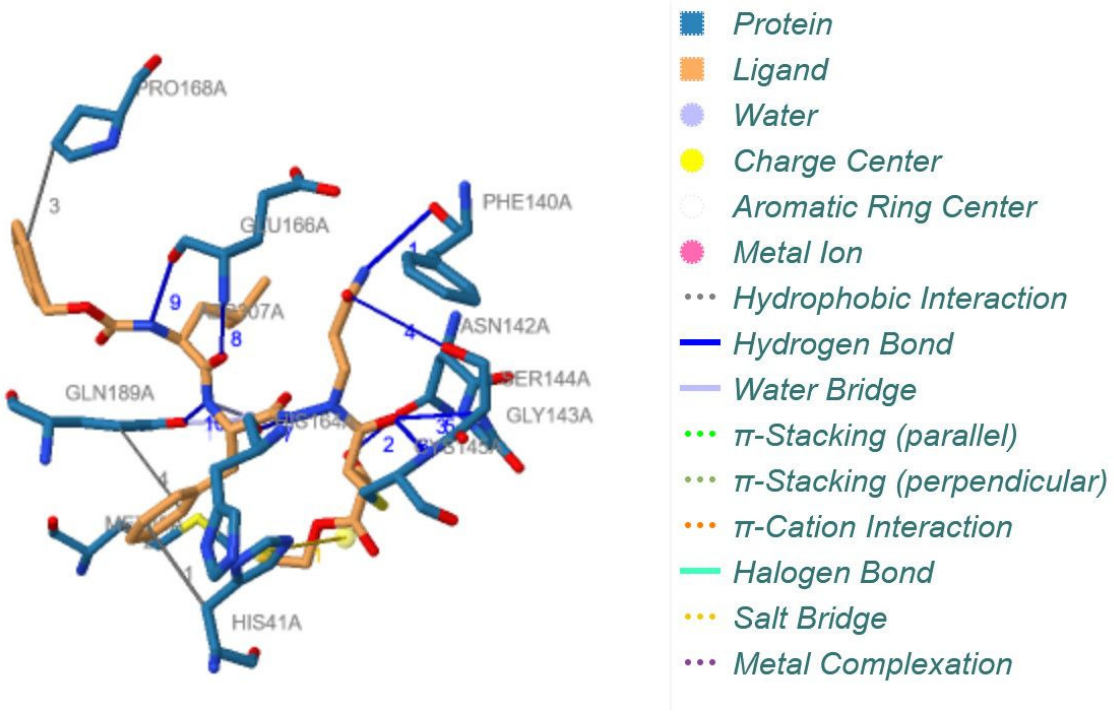


Fig.7 Protein Ligand interaction between the C3 like peptidase with aza-peptide epoxide

The substrate binding properties and structural and chemical complementarity of this Aza-Peptide Epoxide can be explored as an anti - Coronavirus COVID-19 agent. The APE which is ethyl (2S)-4-[(3-amino-3-oxo-propyl)-[[(2S)-2-[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-phenyl-propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate structure is shown in Fig.8.

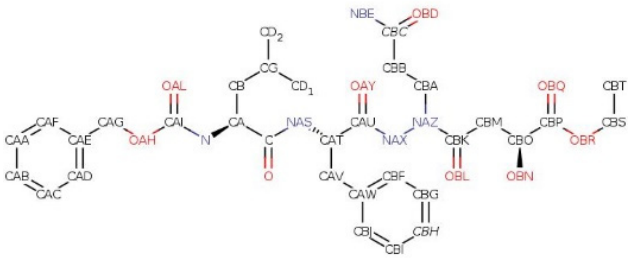


Fig. 8 Structure of Aza-Peptide Epoxide (APE) ethyl (2S)-4-[(3-amino-3-oxo-propyl)-[[(2S)-2-[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-phenyl-propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate with possible anti Coronavirus activity – (Source https://www.rcsb.org/ligand/AZP)

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

The two parts of the Main protein from the whole genome of the Novel Coronavirus COVID-19 aligned with two SAR proteins and the ligand binding sites were similar, the alignment positions, number of amino acids and ligand and the interacting residues is given in Table 3

The main protein with a sequence length of 5509aa of the Wuhan Corona Virus showing structural alignment with two other proteins of SARS-CoV is given in Table 4

252 **Table 4** Main Protein with a sequence length – 4409aa of Wuhan Corona Virus showing structural alignment with two other
253 proteins of SARS-CoV

Template ID	Template Title	Alignment Positions	Number of aa	Ligands	Interacting Residues
3e9s.1	A new class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication	1568-1882	315	TTT	Chain A: L.1729, G.1730, D.1731, P.1814, P.1815, Y.1831, Y.1835, Q.1836, Y.1840, T.1868
2a5i.1	Crystal structures of SARS coronavirus main peptidase inhibited by an aza-peptide epoxide in the space group C2	3268-3573	306	AZP	Chain A: T.3292, T.3293, H.3308, M.3316, Y.3321, F.3407, L.3408, N.3409, G.3410, S.3411, C.3412, H.3430, H.3431, M.3432, E.3433, P.3435, H.3439, D.3454, R.3455, Q.3456, T.3457, A.3458, Q.3459

254

255 The complete genome of MN908947.3 Wuhan seafood market pneumonia virus isolate
256 Wuhan-Hu-1 encodes a 4409aa long protein along with the other glycoproteins and
257 polyproteins. The homology modelling of this protein showed sequence and structural
258 alignment with two SARS proteases with structural accession numbers 3e9s.1 and 2a5i.1 at
259 positions 1568-1882 and 3268-3573 respectively. Reports suggests inhibition of virus
260 replication by TTT ligand and an aza-peptide epoxide inhibiting the main peptidase. The
261 structural similarity of these templates are 83% and 96% respectively. The multiple sequence
262 alignment shows complete conservation of the sequence suggesting a high degree of homology.

263 The protein ligand interaction analysis of the Novel Coronavirus non structural protein and
264 papain-like protease is shown in Fig. 9

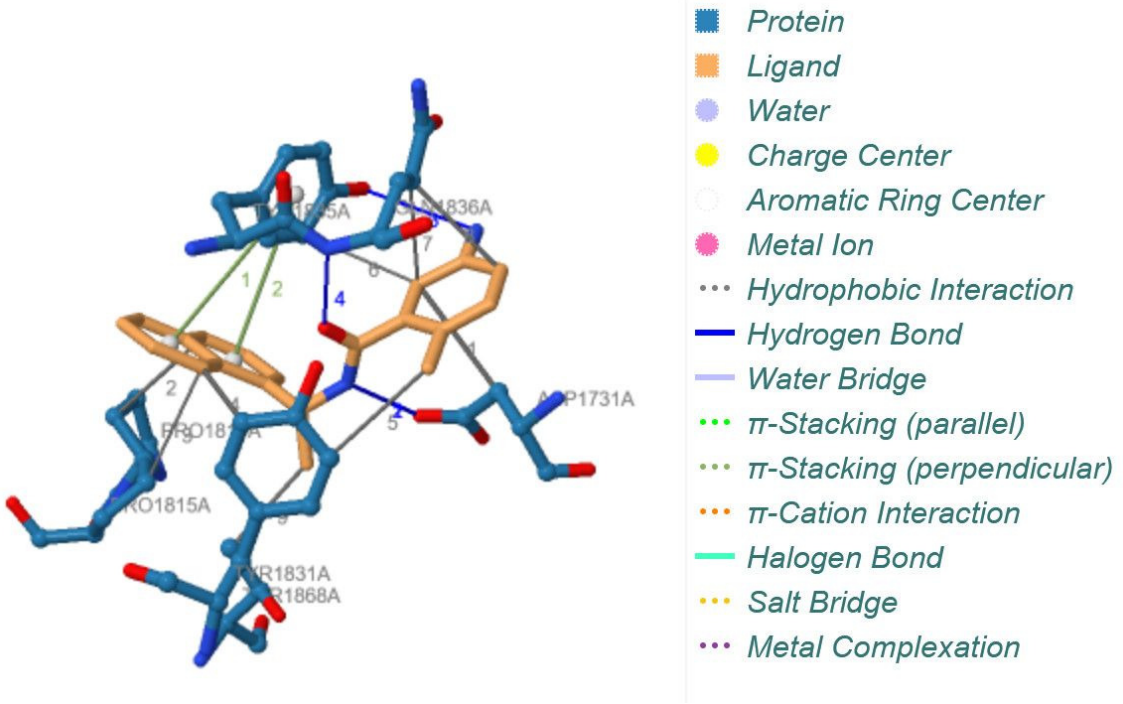


Fig.9 Protein Ligand interaction between the Novel Coronavirus non structural protein and papain-like protease

The Hydrophobic interaction, hydrogen bonding, salt bridges of the constructed model of the Novel Coronavirus protein from region 3268-3573 aa to ligand AZP is given in Suppl. Table 1, the Hydrophobic interaction, hydrogen bonding, salt bridges of the template 2a5i is given in Suppl. Table 2, when comparing both it is seen that the binding properties are the same expect for the presence of water bridge in the template 2a5i.

The Hydrophobic interaction, hydrogen bonding, π-Stacking of the constructed model of the Novel Coronavirus protein from region 1568-1882 aa to ligand Small molecule Noncovalent Lead Inhibitor is given in Suppl. Table 3, the Hydrophobic interaction, hydrogen bonding, π-Stacking of the template 3e9s is given in Suppl. Table 4, when comparing both it is seen that the binding properties are the same except or an addition π-Stacking at Tyr in the template 2a5i. This shows that there is high possibility of binding of the these antiviral compounds with the regions of Novel Coronavirus protein that is in homology with the SARS protein.

from the in case of template 2a5i and π-Stacking in the case of template 3e9s

The targeting of this part of the genome of the Novel Coronavirus COVID-19 with the antiviral compounds which have to shown to bind in the similar region of the SARS virus can have implication in the development of an effective antiviral compound against the Novel Coronavirus COVID-19 . The residues 254 to 13480 in Frame 2 with 4409 amino acids from the Complete genome sequence of the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1 shows homology with the SARS coronaviral proteases, papain-like protease (PLpro) and 3C-like protease (3CLpro), these proteins have the function of processing the viral polyprotein and also they perform the function of stripping ubiquitin and the ubiquitin-like interferon (IFN)-stimulated gene 15 (ISG15) from the hosts to facilitate coronavirus replication and help in evading immune response of the host, these inhibitors can also have a role in disrupting 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 Novel COVID-19 has homology with this and the binding sites for this in the structural protein of the Novel COVID-19 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 COVID-19.

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.

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.

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.

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

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

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

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

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

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

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