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], Anjani Alluri[3] and Samriddhi Gupta[4]

[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

[4] Department of Biochemistry, School of Life Sciences, University of Hyderabad, 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. 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. Docking profile with 9 different conformations of the ligand with the protein model using Autodock Vina showed an affinity of -7.1 Kcal/mol. This region was conserved in 831 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. Docking profile with 9 different conformation of the ligand with the protein model using Autodock Vina showed an affinity of -7.9 Kcal/mol. This region was conserved in 831 genomes of SARS-CoV-2. It is possible that these ligands can be antivirals of 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-CoV-2 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 SARS-COV-2 was officially announced as the causative agent by Chinese authorities on 7 January 2020. As on 24 April 2020 the World

Health Organization reported 2544792 confirmed cases globally (WHO Situation Report 94 2020). This novel corona virus has been designated as SARS-CoV-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 and for designing antiviral drug candidates *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)

The main protease in the SARS virus is the key enzyme for processing of polyprotiens of the virus. This has been the main target for antivirals in the past in SARS-CoV and we hypothesize that as this has high homology with the main protease of SARS-CoV-2, the same protein can be a target for anitivirals in this virus as well. It has been known that viral replication can be blocked by inhibiting this protein (Anand et al 2003). The nonexistence of this proteins in humans makes it an even more attractive antiviral target as there can be no cytotoxity to humans.

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

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

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 Supplementary Table 1. Amongst the seven polyproteins, five are of Wuhan pneumonia virus isolate SARS-COV-2 and two

sequences are of Wuhan pneumonia virus isolate SI200040-SP. The seven Glycoproteins are of the same isolate, Wuhan pneumonia virus isolate SARS-COV-2.

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

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

3. MN938387.1 SARS-CoV-2 virus isolate SARS-COV-2 _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. Swiss-model (Schwede et al., 2003) was

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

Complete genome sequence of the SARS-CoV-2 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.

Alignments of the 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 was done with 831 genomes of the SARS-CoV-2 and found that they were identical.

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)

Blind Docking with ligands of the homology models generated from template 2a5i of the PDB database and template 3e9s of the PDB database and the also the templates 2a5i and 3e9s was done by docking the ligands to the whole surface of a protein using Autodock Vina The ligands and the protein molecules were taken in PDBQT format for docking. The preparation of the ligand and protein files was done by converting the SDF format to PDB format using OpenBabel. The ligands where then prepared by detecting their root and Torsion tree. The proteins were prepared by deleting all heteroatoms, water molecules, polar Hydrogen were

added. In order to know where the ligand would bind optimally the grid was specified to be the whole protein. The Docking was done in Autodock Vina (Trott and Olson 2010) that considered 9 different conformations of the ligand in each docking. The Docked results were obtained as PDBQT and txt. Files.

The docking was analyzed by opening the PDB form of the protein in Pymol along with the PDBQT file of the most suitable ligand conformation result. The complex obtained was saved as a PDB file. This PDB file is then viewed in DS Visualizer where the Ligand Interactions were observed with the corresponding amino acids with the kind of interactions and the distance between them. The software used for the process was Pymol, Autodock Vina, Discovery Studio Visualizer.

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.

Table 1 Physico-chemical properties of polyproteins of SARS-CoV-2 virus isolates

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 tertiary structure analysis of the isolate SARS-CoV-2 _HKU-SZ-001_2020 ORF1ab polyprotein is given in Table 2.

Table 2 Tertiary Structure of SARS-CoV-2 virus isolate SARS-COV-2 _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

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.1.

Fig.1 Phylogenetic tree of the seven polyproteins of Severe acute respiratory syndrome coronavirus 2 isolate virus isolates

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.2,

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

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 two parts of the Main protein from the whole genome of the SARS-CoV-2 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

Table 3 Main Protein with a sequence length – 4409aa of SARS-CoV-2 Virus showing structural alignment with two other 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

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).

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

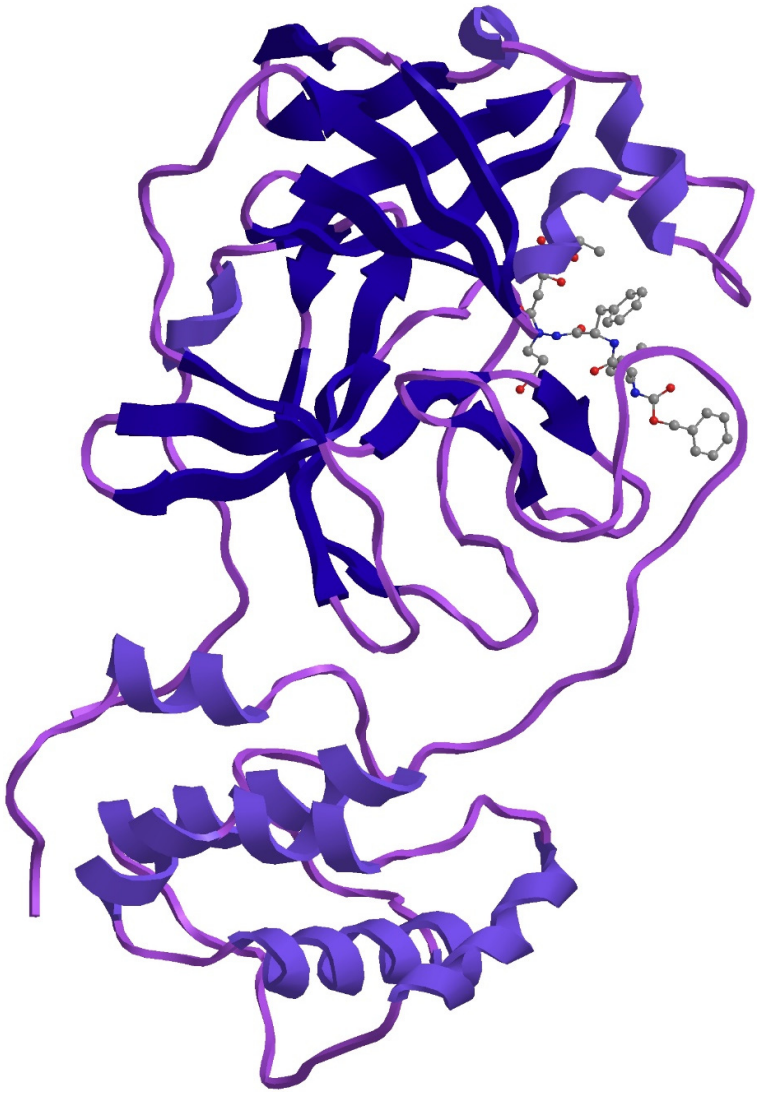


Fig. 3 Homology model with ligand binding of protein from amino acids 3268 -3573 in Frame 2 with 306 amino acids of the Complete genome sequence of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp linear ss-RNA with 2a5i of the PDB database as template.

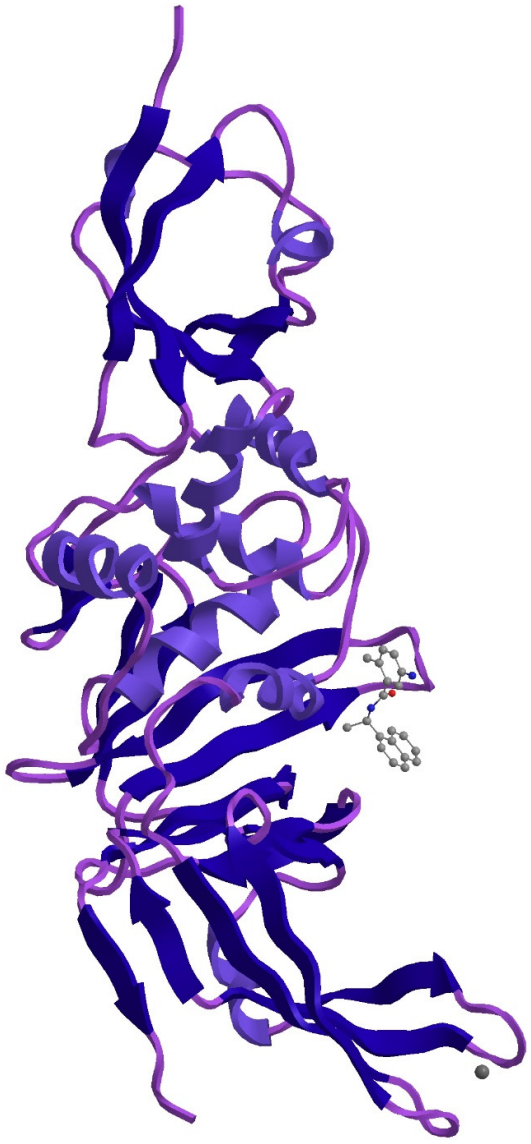


Fig. 4 Homology model with ligand binding of protein from residues 1568-1882 in Frame 2 with 315 amino acids of the Complete genome sequence of the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) which has 29903 bp linear ss-RNA with 3e9s of the PDB database as template.

The statistics of structural comparison with PDB templates is given in Table 4, it is seen that the proteins from the SARS-CoV-2 are significantly close to the proteins of SARS CoV and the amino acid alignment in the biding region is the same in both the viruses.

Table 4 Statistics of structural comparison with PDB templates

Structure	Template	Similarity	p-Value	No. of equivalent positions	RMSD	Raw Score
3e9s_covid	3e9s	Significantly Similar	0.00e+00	314	0.10	935.61
2a5i_covid	2a5i	Significantly Similar	0.00e+00	306	0.08	911.72

The alignment of the 305 residues from 3268-3573 aa of the Novel Coronavirus COVI-19 with the template 2a5i is shown in Fig.5 and the alignment of the 315 residues from 1568-1882 aa of the Novel Coronavirus COVI-19 with the template 3e9s is shown in Fig.6.

Fig. 5 Alignment of the 305 residues from 3268-3573 aa of the Novel Coronavirus COVI-19 with the template 2a5i

Fig.6 the alignment of the 315 residues from 1568-1882 aa of the Novel Coronavirus COVI-19 with the template 3e9s

A PSI-BLAST of a length of 306 amino acid residues 3268 -3573 in Frame 2 from the SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) was conducted to ascertain the conservation of these amino acids in 250 genome sequences of SARS-CoV-2 and it was found that there was a complete match in these genomes of the virus. The fact that the region is conserved in all these SARS-CoV-2 sequences further emphasizes this ligand interaction of Aza-Peptide epoxide with the protein can be used as an antiviral in SARS-CoV-2. Similarly A PSI-BLAST of a length of 315 amino acid residues 3268 -3573 in Frame 2 with 315 amino acid residues 1568-1882 in Frame 2 from SARS-CoV-2 virus isolate Wuhan-Hu-1 (Genbank Accession Number MN908947.3) was conducted to ascertain the conservation of these amino acids in 250 genome sequences of SARS-CoV-2 and it was found that there was a complete match in these genomes of the virus. The fact that the region is conserved in all these SARS-CoV-2 sequences further emphasizes this ligand interaction of ligand GRL0617 with the protein can be used as an antiviral in SARS-CoV-2.

The important templates that aligned with this 4409 amino acid residues of the whole genome of the SARS-CoV-2 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.

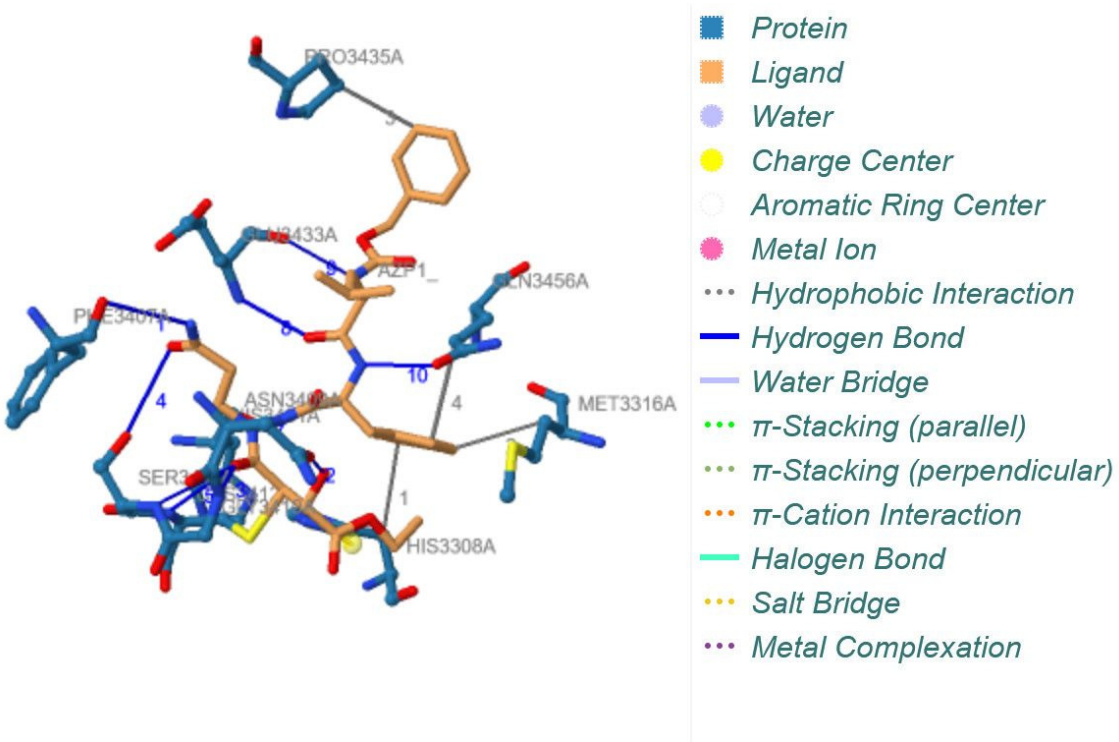


Fig.7 Protein Ligand interaction between the C3 like peptidase with aza-peptide epoxide of the model with the template 2a5i

The substrate binding properties and structural and chemical complementarity of this Aza-Peptide Epoxide can be explored as an anti - Coronavirus SARS-COV-2 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.

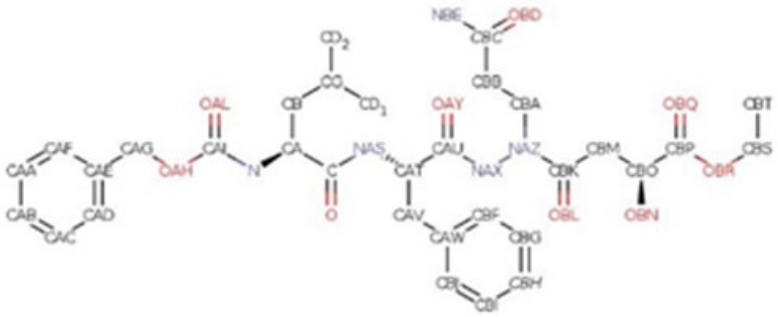


Fig. 8 Structure of Aza-Peptide Epoxide (APE) ethyl (2S)-4-[(3-amino-3-oxo-propyl)-[[(2S)-2-[[(2S)-4-methyl-2-phenylmethoxycarbonylamino-pentanoyl]amino]-3-phenyl-propanoyl]amino]amino]-2-hydroxy-4-oxo-butanoate with possible anti Coronavirus activity – (Source https://www.rcsb.org/ligand/AZP)

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 complete genome of MN908947.3 SARS-CoV-2 virus isolate Wuhan-Hu-1 encodes a 4409aa long protein along with the other glycoproteins and polyproteins. The homology modelling of this protein showed sequence and structural alignment with two SARS proteases with structural accession numbers 3e9s.1 and 2a5i.1 at positions 1568-1882 and 3268-3573 respectively. Reports suggests inhibition of virus replication by TTT ligand and an aza-peptide epoxide inhibiting the main peptidase. The structural similarity of these templates are 83% and 96% respectively. The multiple sequence alignment shows complete conservation of the sequence suggesting a high degree of homology. The protein ligand interaction analysis of the Novel Coronavirus non structural protein and papain-like protease is shown in Fig. 9.

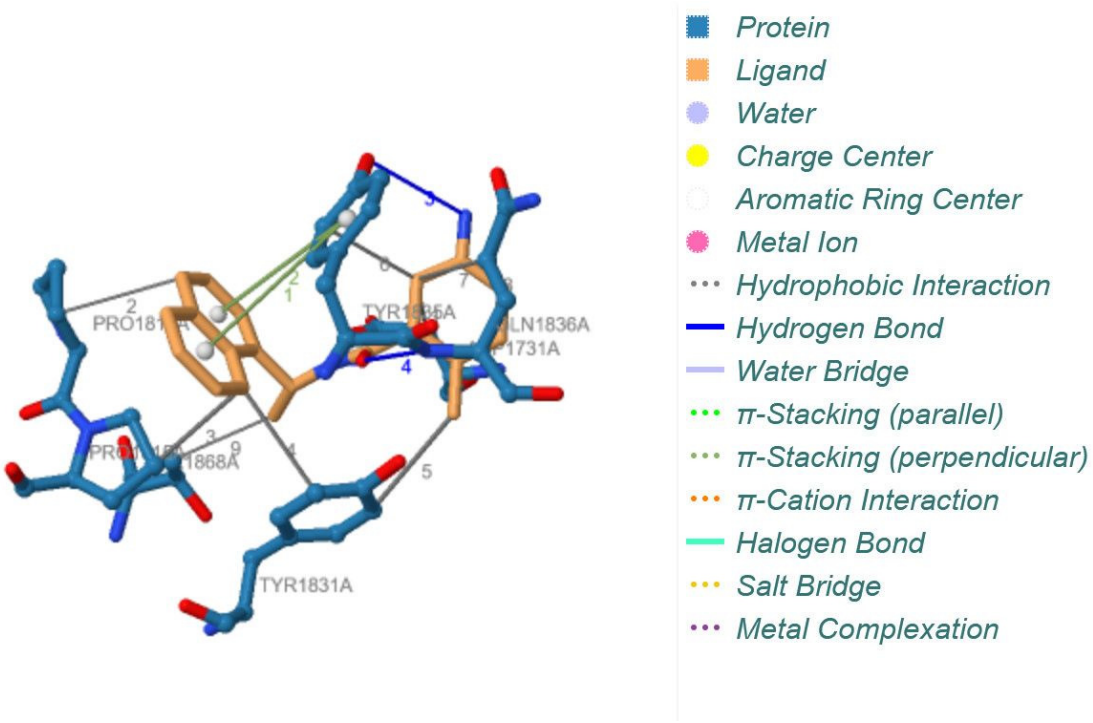


Fig.9 Protein Ligand interaction between the Novel Coronavirus non structural protein and papain-like protease of the model with the template 3e9s

The Comparison of Hydrophobic interaction, hydrogen bonding, salt bridges of the constructed model of the Novel Coronavirus protein from region 3268-3573 aa to ligand AZP with 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 Comparison of 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 with the Hydrophobic interaction, hydrogen bonding, π-Stacking of the template 3e9s is given in Suppl. Table 3, when comparing both it is seen that the binding properties are the same except or an additional π-Stacking at Tyr in the template 2a5i. This shows that there is high possibility of binding of these antiviral compounds with the regions of Novel Coronavirus protein that is in homology with the SARS protein.

Comparison of the hydrophobic interaction of the biding of the ligand AZP between the SARS-CoV-2 protein and the template 2a5i of SARS CoV is shown in Fig.10 and the comparison of the hydrophobic interaction of the biding of the ligand AZP between the SARS-CoV-2 protein and the template 3e9s of SARS CoV is shown in Fig.11. It is seen that the interaction is the same in both proteins with the same amino acids participating in the interaction indicating that there is a possibility that these ligands with antiviral properties can bind to the new virus.

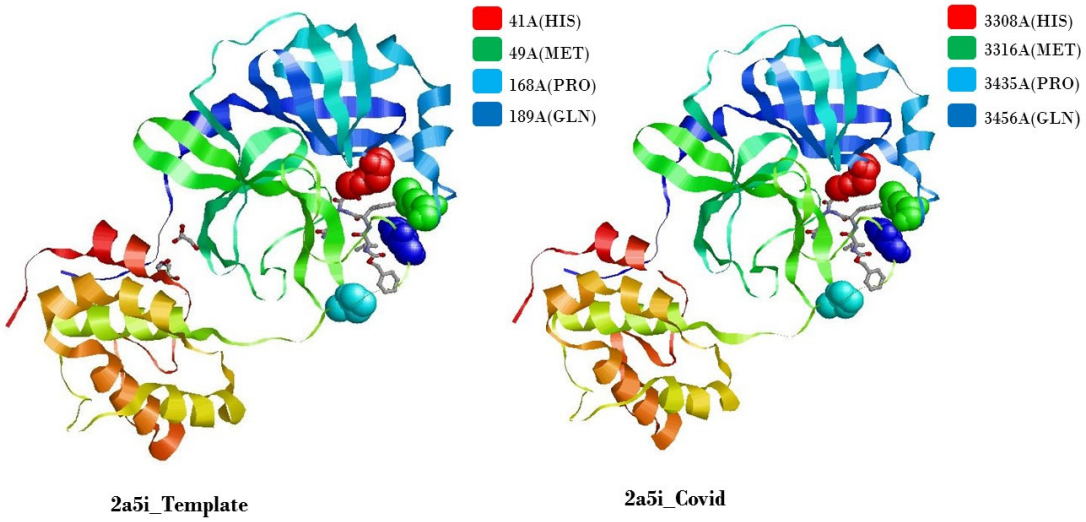


Fig.10 Comparison of the hydrophobic interaction of the biding of the ligand AZP between the SARS-CoV-2 protein and the template 2a5i of SARS CoV

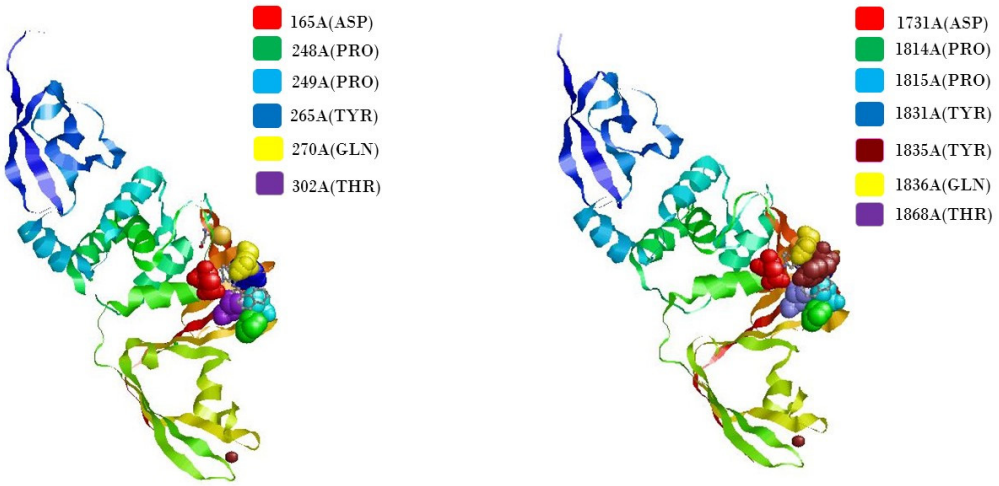


Fig.11 Comparison of the hydrophobic interaction of the binding of the ligand between the SARS-CoV-2 protein and the template 3e9s of SARS CoV

449

450 The protein ligand interaction obtained via Docking can offer us important information in
451 determining the effectiveness of the binding in terms of its antiviral properties in the homology
452 models obtained using 3C like peptidase(2A5I) and the papain-like protease/deubiquitinase
453 protein(3E9S) as templates of SARS virus.

454 We used AutoDock Vina which uses a function which has an empirical and knowledge based
455 powerful hybrid scoring, the software employs an optimized search which is iterated till a
456 considerably accepted solution is found for the minimum-energy docking conformations
457 (Hassan et al 2017). The comparison of interaction of GRL0617 with the amino acid residues
458 of PLPro and the model obtained using the template 3e9s is shown in Fig. 12a and Fig. 12 b
459 respectively.

460

461

462

463

464

465

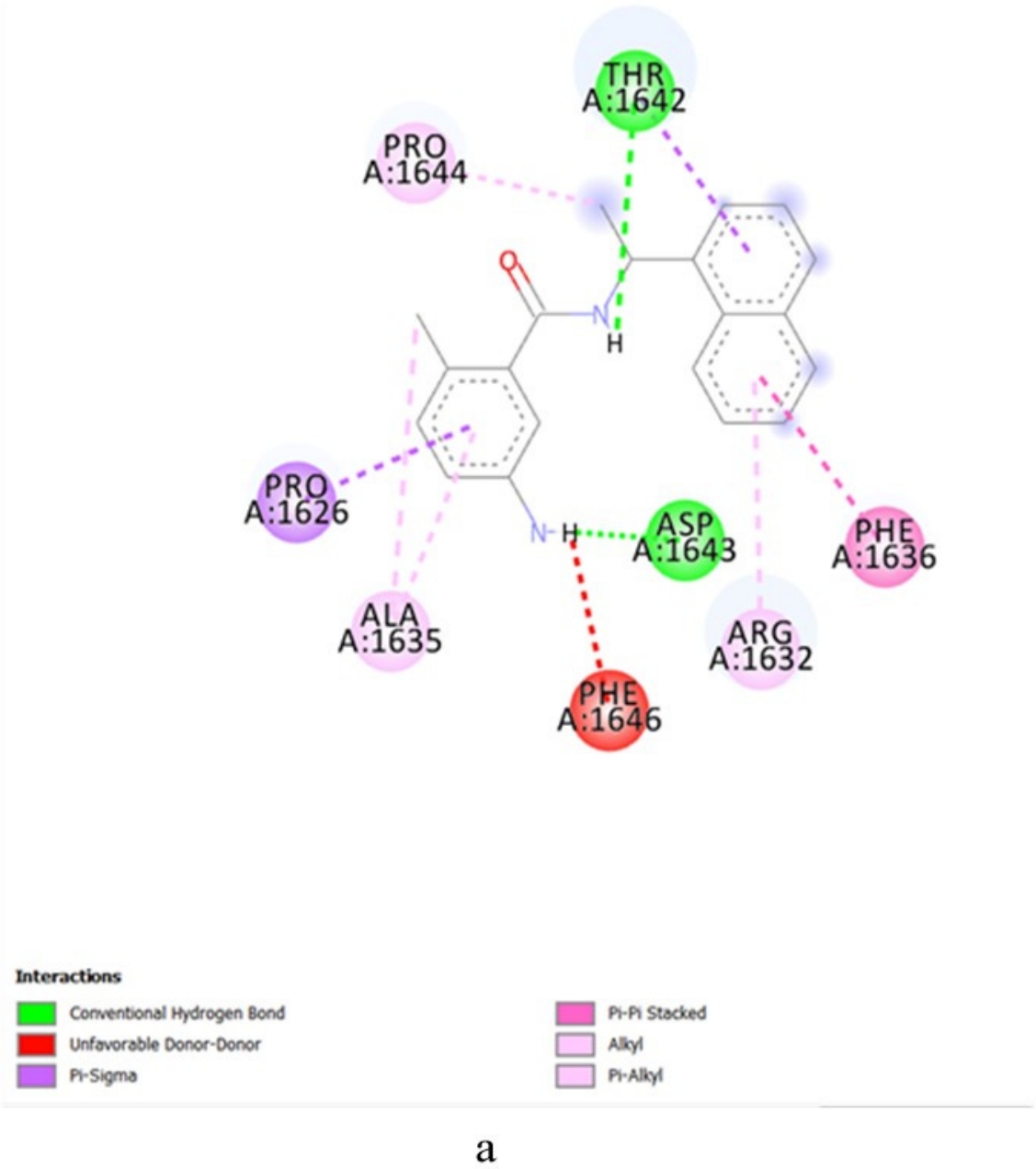
466

467

468

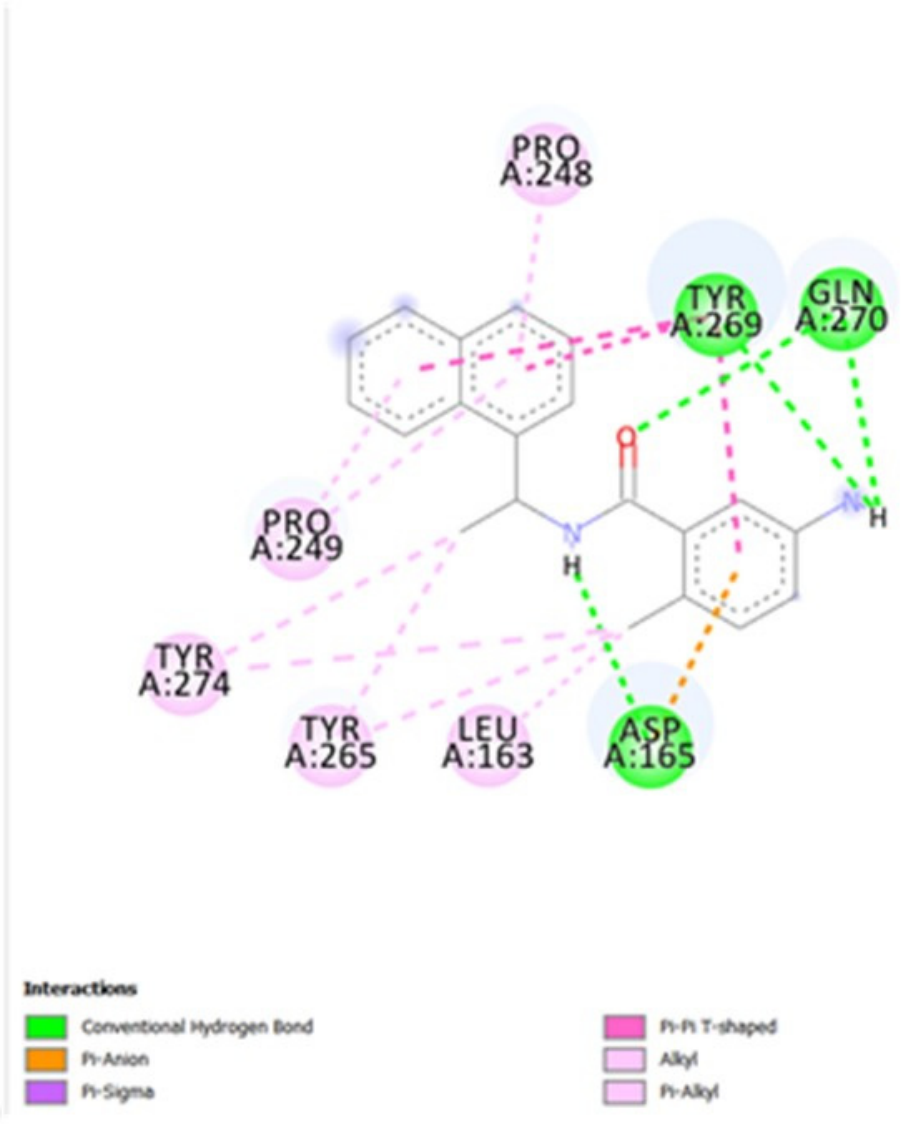
469

470



a

Fig. 12 a Interaction Profile of GRL0617 with amino acid residues of the homology Model of SARS-CoV-2 papain-like protease



b

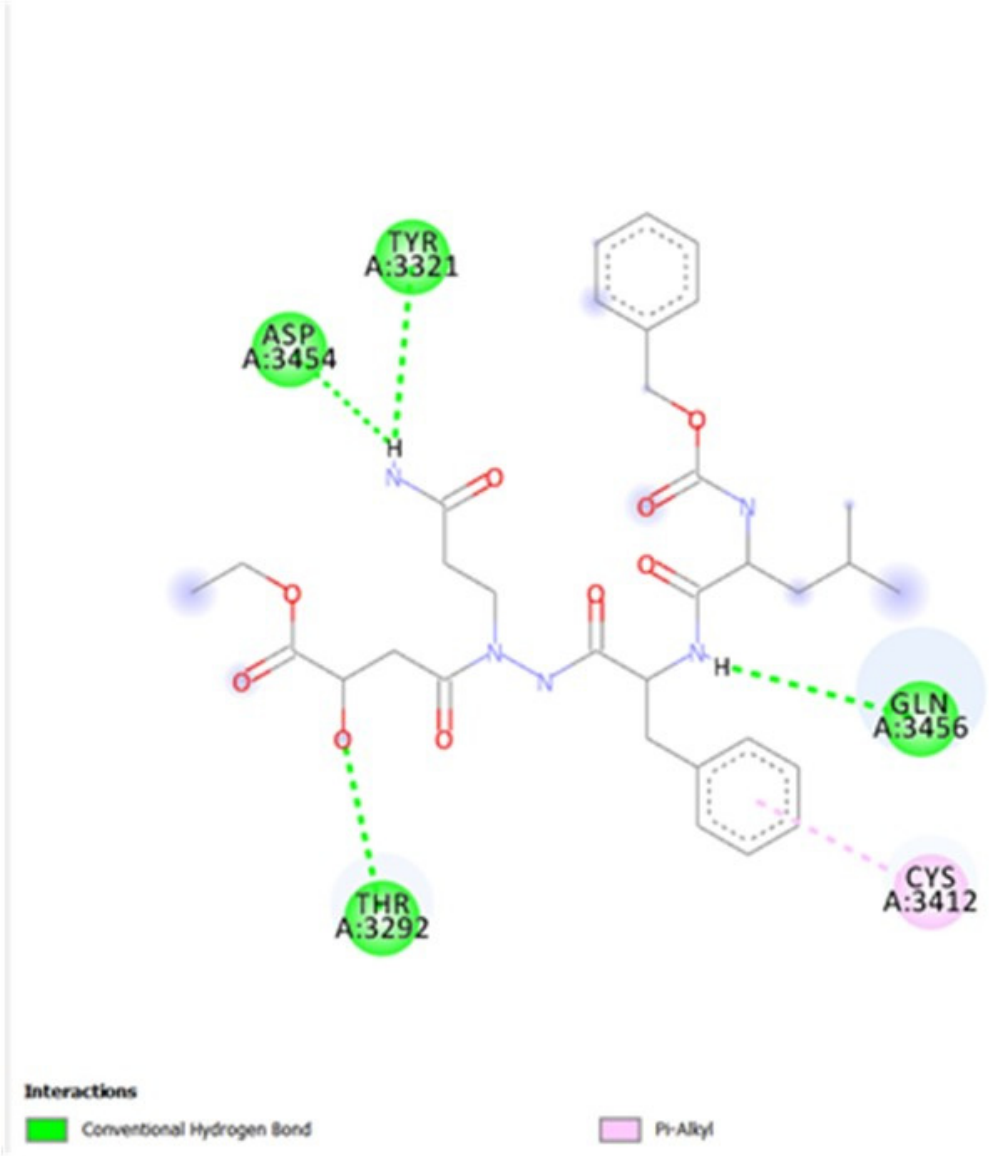
Fig. 12 b Interaction Profile of GRL0617 with amino acid residues of the template 3e9s

Both show eight interacting amino acids, few of which exhibits multiple interactions. The complex with PLPro shows very high affinity i.e. 10.2Kcal/mol as compared to the Complex with the model whichshows lesser affinity i.e.7.9Kcal/mol. The comparison of conserved amino acids show Asp1643 in the homology model and Asp165 in the template both of which show H bond at a distance of 2.60 and 2.07 respectively. Additionally, Asp165 shows Pi-Sigma bond at a distance of 3.53 and Pi-Anion at a distance of 4.39, this accounts for the stronger affinity in PLPro as against the homology model. In the case of Pro1644 in the homology model the Alkyl bond at a distance of 4.70 whereas the template shows a Pi-Alkyl bond at a distance of 5.04. Pi-Alkyl bond being stronger than Alkyl bond. Similarly, Pro1632 in the homology model shows Pi-Alkyl bond at a distance of 5.06 and the PLPro shows 2 Pi-Alkyl bonds at a distance of 4.31 and 4.72, two Pi-Alkyl bonds at a close distance accounts for stronger affinity in the template. Ala1635 in the homology model and Leu163 in the PLPro both are hydrophobic amino acids and show Alkyl bond at a distance of 3.80 and4.25 respectively. Ala1635 additionally exhibits Pi-Alkyl bond at a distance of 4.24. Thr1642 in the homology model and Gln270 in the PLPro both exhibit H bond. However, Gln270 exhibits 2 H bond at a distance of 2.83 and 2.74 via its –NH group. Thr1642 exhibits 1 H bond at a distance of 2.62 via its –OH and 1 Pi-Sigma bond at a distance of 3.87. Phe1636 in the homology model and Tyr269 template are both Aromatic amino acids. They both show Pi-Pi interactions. Phe1636 exhibits Pi-Pi stacking at a distance of 5.60 whereas Tyr269 exhibits 3 Pi-Pi T shaped bonds at a distance of 5.06, 5.25 and 5.44. It also exhibits an additional H bonding at a distance of 3.07.

506 The comparison of interaction of AZP with the amino acid residues of 3CL Pro and the model

507 obtained using the the template 2a5i is shown in Fig. 13 a and Fig. 13 b respectively.

508

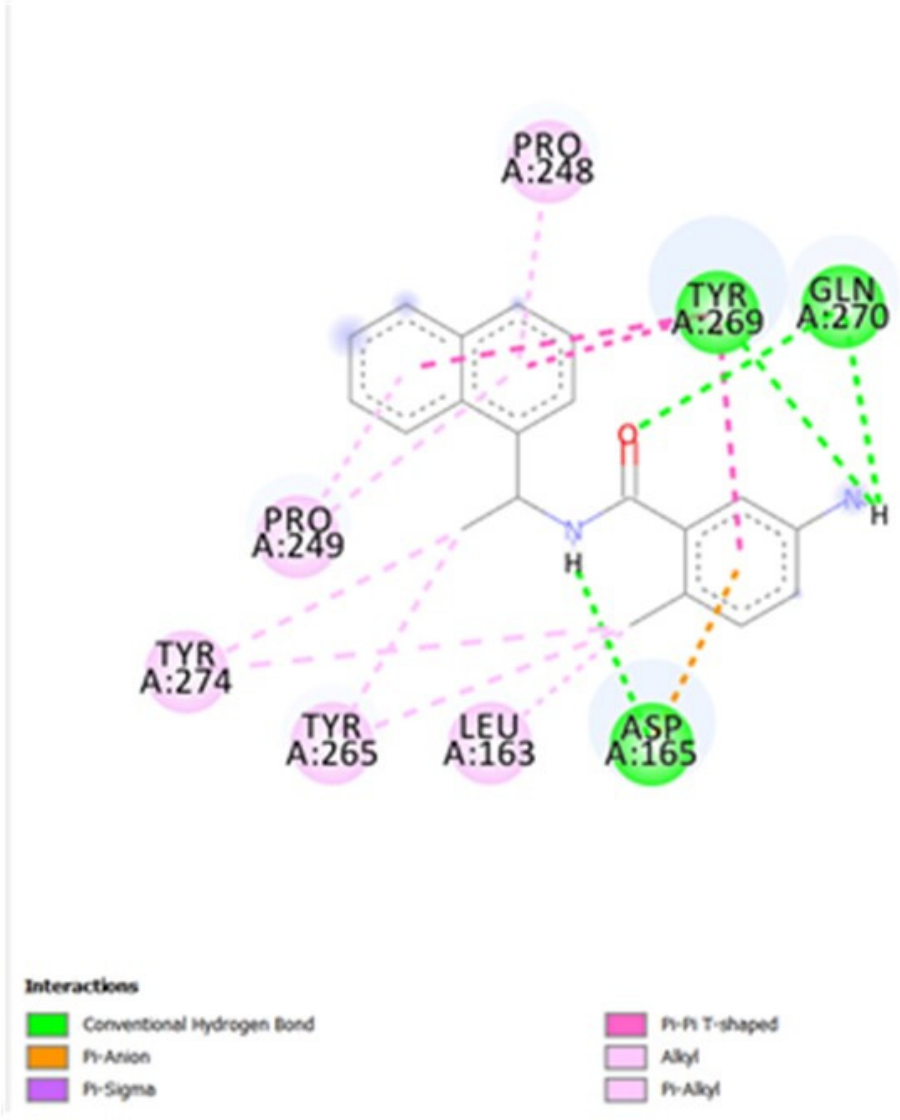


a

509

510 **Fig. 13 a** Interaction Profile of ligand AZP with amino acid residues of the homology Model

511 of SARS-CoV-2 3C like Protease



b

Fig. 13 b Interaction Profile of ligand AZP with amino acid residues of the template 2a5i

Both show five interacting amino acids, the conserved amino acids are Gln3456 in the homology model and Gln110 in 3CLPro both show H bond at a distance of 2.40 and 2.39 respectively. The similarities present are Thr3292 in the homology model and Ser158 in 3CLPro both show H bond at a distance 2.78 and 2.71 respectively. Both of them have –OH group that participates in the H bond. Tyr3321 in the homology model and Lys102 in 3CLPro both show H bond at a distance of 2.73 and 2.97 respectively which is reflective of the electronegative group i.e. participating. –OH in the former and –NH in the latter. Cys3412 in the homology model and Val297 in 3CLPro both show Pi-Alkyl bond at a distance of 4.82 and 5.30 respectively. Asp3454 in the homology model and Phe294 in 3CLPro exhibit a H bond at a distance of 1.97, the latter exhibits a Pi-Pi Stacking at a distance of 4.51. This is responsible for the slightly higher affinity of AZP to 3CLPro than AZP to the model. The former having an affinity (Kcal/mol) of -7.4 and the latter having -7.1.

However, it is also interesting to note that even though Alignment studies showed 83% and 96% identity in case of Model 1(obtained using the template 3E9S) and Model 2(obtained using the template 2A5I) with PlPro and 3CLPro respectively, the binding cavity interactions/milieu were very similar in the 2nd case inspite of not much conserved amino acid residues,and in the former case, the binding cavity showed certain similarity in terms of the cavity milieu, however the intensity varied due to multiple, additional stability.

We were able to see the difference in the protein ligand interaction in both the models by docking of these ligands to the whole surface of a protein. As we had no prior knowledge of the target pocket. As the docking involved several runs and energy calculations for arriving at a favorable protein-ligand complex, the interactions observed shows that the interaction profile of ligand AZP with amino acid residues of the homology Model of SARS-CoV-2 3C like

Protease showed an affinity of -7.1 Kcal/mol and Interaction Profile of GRL0617 with amino acid residues of the homology Model of SARS-CoV-2 papain-like protease showed an affinity of – 7.9 Kcal/mol.

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

The targeting of this part of the genome of the SARS-CoV-2 with the antiviral compounds which have shown to bind in the similar region of the SARS virus can have implication in the development of an effective antiviral compound against the SARS-CoV-2 . The SARS-CoV-2 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 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

566 Compound GRL0617 binds in the S4 and S3 enzyme subsite that gets the C terminal tail of the
567 Ubiquitin (King and Finley 2014; Schauer et al., 2019). Our results show that Aza-Peptide
568 Epoxide an irreversible protease inhibitor and GRL0617 a viral replication inhibitor can be
569 used to develop inhibitors of the Novel Coronavirus SARS-COV-2.

570

571 **References**

572 Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. and Hilgenfeld, R., 2003. Coronavirus
573 main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science,
574 300(5626), pp.1763-1767.

575 Bae, J.E., Kim, I.J., Kim, K.J. and Nam, K.H., 2018. Crystal structure of a substrate-binding
576 protein from *Rhodothermus marinus* reveals a single α/β-domain. Biochemical and
577 Biophysical Research Communications, 497(1), pp.368-373.

578 Baez-Santos, Y.M., John, S.E.S. and Mesecar, A.D., 2015. The SARS-coronavirus papain-like
579 protease: structure, function and inhibition by designed antiviral compounds. Antiviral
580 Research, 115, pp.21-38.

581 Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., & Wheeler, D. L.
582 (2000). GenBank. Nucleic Acids Rresearch, *28*(1), 15-18.

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

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

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

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

598 de Wit, E., Rasmussen, A.L., Falzarano, D., Bushmaker, T., Feldmann, F., Brining, D.L.,
599 Fischer, E.R., Martellaro, C., Okumura, A., Chang, J. and Scott, D., 2013. Middle East
600 respiratory syndrome coronavirus (MERS-CoV) causes transient lower respiratory tract
601 infection in rhesus macaques. Proceedings of the National Academy of Sciences,
602 110(41), pp.16598-16603.

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

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

608 Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J.,
609 Gorbalenya, A.E., Cambillau, C. and Canard, B., 2004. The severe acute respiratory
610 syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding

611 subunit unique in the RNA virus world. Proceedings of the National Academy of
612 Sciences, 101(11), pp.3792-3796.

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

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

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

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

623 Hassan, N.M., Alhossary, A.A., Mu, Y. and Kwoh, C.K., 2017. Protein-ligand blind docking
624 using QuickVina-W with inter-process spatio-temporal integration. Scientific reports,
625 7(1), pp.1-13.

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

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

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

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

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

639 Liu, W., Morse, J.S., Lalonde, T. and Xu, S., 2020. Learning from the Past: Possible Urgent
640 Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by
641 2019-nCoV. ChemBioChem.

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

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

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

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

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

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

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

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

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

672 Trott, O. and Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking
673 with a new scoring function, efficient optimization, and multithreading. Journal of
674 computational chemistry, 31(2), pp.455-461.

675 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T.,
676 de Beer, T.A.P., Rempfer, C., Bordoli, L. and Lepore, R., 2018. SWISS-MODEL:
677 homology modelling of protein structures and complexes. Nucleic Acids Research,
678 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.

Supplementary Fig.1 Multiple alignment of the Polyproteins of Severe acute respiratory syndrome coronavirus 2 isolate virus isolates

702

703

704 Supplementary Table 1 List of Severe Acute Respiratory Syndrome coronavirus 2 isolate sequences taken for bioinformatic analysis

Genbank Accession Number	Title	Description
MN988713.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV/USA-IL1/2020	Complete genome
MN938384.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-002a_2020	Complete genome
MN975262.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-005b_2020	Complete genome
MN985325.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV/USA-WA1/2020	Complete genome
NC_045512.2	Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1	Complete genome
MN908947.3	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1	Complete genome
MN938385.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-001_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN938386.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-004_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN975263.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-007a_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN975264.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-007b_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN975265.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-007c_2020 ORF1ab polyprotein, RdRp region, (orf1ab) gene, partial cds	Polyprotein, RdRp region

MN970003.1	Severe acute respiratory syndrome coronavirus 2 isolate SI200040-SP orf1ab polyprotein, RdRP region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN970004.1	Severe acute respiratory syndrome coronavirus 2 isolate SI200121-SP orf1ab polyprotein, RdRP region, (orf1ab) gene, partial cds	Polyprotein, RdRp region
MN938387.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-001_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN938388.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-002b_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN938389.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-004_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN938390.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-005_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN975266.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-007a_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN975267.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-007b_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein
MN975268.1	Severe acute respiratory syndrome coronavirus 2 isolate 2019-nCoV_HKU-SZ-007c_2020 surface glycoprotein (S) gene, partial cds	Glycoprotein

705

706

Supplementary Table 2 Comparison of binding properties of Novel Coronavirus protein from region 3268-3573 (2a5i _SARS-CoV-2) and 2a5i template to ligand AZP

Hydrophobic Interactions

Index	Residue		AA		Distance		Ligand Atom		Protein Atom	
	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i
1	3308A	41A	HIS	HIS	3.64	3.70	2393	2461	307	308
2	3316A	49A	MET	MET	3.81	3.86	2395	2463	368	368
3	3435A	168A	PRO	PRO	3.42	3.73	2376	2443	1303	1347
4	3456A	189A	GLN	GLN	3.84	3.93	2396	2464	1462	1507

Hydrogen Bonds

Index	Residue		AA		Distance H-A		Distance D-A		Donor Angle		Protein donor?		Sidechain		Donor Atom		Acceptor Atom	
	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i
1	3407A	140A	PHE	PHE	2.61	2.46	3.47	3.33	146.60	147.28	✖	✖	✖	✖	2404 [Nam]	2472 [Nam]	1081 [O2]	1112 [O2]

2	3409A	142A	ASN	ASN	2.51	2.52	2.87	2.87	102.18	100.93	✗	✗	✓	✓	2410 [O3]	2478 [O3]	1103 [O2]	1134 [O2]
3	3410A	143A	GLY	GLY	1.94	1.83	2.78	2.73	142.36	150.57	✓	✓	✗	✗	1105 [Nam]	1136 [Nam]	2407 [O2]	2475 [O2]
4	3411A	144A	SER	SER	3.37	3.44	3.76	3.80	106.27	104.79	✓	✓	✓	✓	1114 [O3]	1145 [O3]	2405 [O2]	2473 [O2]
5	3411A	144A	SER	SER	2.60	2.61	3.24	3.20	122.58	118.23	✓	✓	✗	✗	1109 [Nam]	1140 [Nam]	2407 [O2]	2475 [O2]
6	3412A	145A	CYS	CYS	2.50	2.57	3.39	3.37	150.10	137.96	✓	✓	✗	✗	1115 [Nam]	1146 [Nam]	2407 [O2]	2475 [O2]
7	3431A	164A	HIS	HIS	1.70	1.73	2.63	2.67	156.72	157.31	✗	✗	✗	✗	2399 [Nam]	2467 [Nam]	1266 [O2]	1307 [O2]
8	3433A	166A	GLU	GLU	2.08	2.01	3.04	2.97	165.52	163.22	✓	✓	✗	✗	1281 [Nam]	1325 [Nam]	2387 [O2]	2455 [O2]
9	3433A	166A	GLU	GLU	2.06	2.13	2.88	2.92	140.23	135.67	✗	✗	✗	✗	2370 [Nam]	2438 [Nam]	1284 [O2]	1328 [O2]
10	3456A	189A	GLN	GLN	1.90	1.82	2.84	2.77	158.82	161.93	✗	✗	✓	✓	2388 [Nam]	2456 [Nam]	1464 [O2]	1509 [O2]

Salt Bridges

Index	Residue		AA		Distance		Protein positive?		Ligand Group		Ligand Atoms	
	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i
1	3308A	41A	HIS	HIS	5.10	5.08	✔	✔	Carboxylate	Carboxylate	2412, 2413	2481, 2480

Water Bridges

Index	Residue		AA		Dist. A-W		Dist. D-W		Donor Angle		Water Angle		Protein donor?		Donor Atom		Acceptor Atom		Water Atom	
	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i	2a5i _SARS-CoV-2	2a5i
1	189A	-	GLN	-	4.07	-	3.93	-	115.60	-	91.13	-	✔	-	1510 [Nam]	-	2456 [Nam]	-	2543	-

Supplementary Table 3 Comparison of binding properties of Novel Coronavirus protein from region 1568-1882 (3e9s_SARS-CoV-2) and 3e9s template to ligand Small molecule Noncovalent Lead Inhibitor

Hydrophobic Interactions

Index	Residue		AA		Distance		Ligand Atom		Protein Atom	
	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s
1	1731A	165A	ASP	ASP	3.82	3.83	2502	2504	1308	1320
2	1814A	248A	PRO	PRO	3.79	3.80	2501	2503	1955	1952
3	1815A	249A	PRO	PRO	3.52	3.75	2503	2505	1963	1960
4	1831A	265A	TYR	TYR	3.50	3.57	2503	2505	2090	2089
5	1831A	265A	TYR	TYR	3.63	3.67	2504	2506	2091	2090
6	1835A	-	TYR	-	3.62	-	2502	-	2121	-
7	1836A	270A	GLN	GLN	3.60	3.59	2502	2504	2130	2129
8	1836A	270A	GLN	GLN	3.58	3.62	2507	2509	2130	2129
9	1868A	302A	THR	THR	3.37	3.54	2514	2516	2381	2385

Hydrogen Bonds

Index	Residue		AA		Distance H-A		Distance D-A		Donor Angle		Protein donor?		Sidechain		Donor Atom		Acceptor Atom	
	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s
1	1731A	165A	ASP	ASP	2.12	2.19	2.98	3.03	151.51	149.71	✓	✓	✓	✓	1311 [O3]	1323 [O3]	2512 [Nam]	2514 [Nam]
2	1731A	165A	ASP	ASP	2.00	2.05	2.98	3.03	173.11	175.80	✗	✗	✓	✓	2512 [Nam]	2514 [Nam]	1311 [O3]	1323 [O3]
3	1835A	269A	TYR	TYR	2.78	2.81	3.61	3.64	143.08	142.33	✗	✗	✓	✓	2511 [Npl]	2513 [Npl]	2124 [O3]	2123 [O3]
4	1836A	270A	GLN	GLN	1.87	1.77	2.83	2.75	164.21	174.65	✓	✓	✗	✗	2125 [Nam]	2124 [Nam]	2509 [O2]	2511 [O2]

π-Stacking

Index	Residue		AA		Distance		Angle		Offset		Type		Ligand Atoms	
	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s	3e9s_SARS-CoV-2	3e9s
1	1835A	269A	TYR	TYR	5.30	5.31	84.17	85.46	1.70	1.70	T	T	2497, 2500, 2503, 2506, 2508, 2516	2499, 2502, 2505, 2508, 2510, 2518
2	1835A	269A	TYR	TYR	5.09	5.10	84.10	85.40	0.83	0.85	T	T	2497, 2498, 2500, 2501, 2515, 2517	2499, 2500, 2502, 2503, 2517, 2519
3	-	269A	-	TYR	-	5.18	-	73.97	-	1.96	-	T	-	2498, 2501, 2504, 2507, 2509, 2512

761

762

763