

# **Compilation of Potential Protein Targets for SARS-CoV-2: Preparation of Homology Model and Active Site Determination for Future Rational Antiviral Design**

Sourav Pal<sup>1,2</sup> and Arindam Talukdar<sup>1,2\*</sup>,

<sup>1</sup>Department of Organic and Medicinal Chemistry, CSIR-Indian Institute of Chemical Biology, 4 Raja S. C. Mullick Road, Kolkata 700032, WB (India).

<sup>2</sup>Academy of Innovative and Scientific Research, Ghaziabad-201002, India

## **ABSTRACT**

The recent pandemic due to the novel coronavirus SARS-CoV-2 (COVID-19) is causing significant mortality worldwide. However, there is a lack of specific drugs which can either prevent or treat the patient suffering from COVID-19. To understand the SARS-CoV-2 receptor recognition causing infectivity and pathogenesis, we have compiled a list of 20 probable drug targets on host and virus based on viral life cycle along with their PDB IDs for the rational development of future antivirals. We have prepared nine homology model for vital proteins for which no crystal structure is reported, which includes protein from host, viral membrane proteins and essential non-structural proteins (NSPs) of virus. The generated models were validated followed by Ramachandran plot along with their sequence and structural alignment. The active site residues of all the protein models are calculated by utilizing COACH meta-server and also cross verified with the CASTp webserver. All the active sites of the homology build proteins were evaluated after superimposition of the closely related X-ray crystallized structure bound with the co-crystal ligands. These information present in the manuscript can be used for the discovery effort towards new antivirals as well as repurposing FDA approved drugs against SARS-CoV-2.

**Result:** The compiled list of 20 probable drug targets on host and virus along with their role in pathogenesis will provide a snap shot for the drug discovery scientists. These targets present an attractive target for antiviral design. We have prepared homology model of TMPRSS2, Envelope

Protein (E), Membrane Protein (M), NSP1, NSP2, NSP4, NSP6, NSP7, NSP14, that have critical role in the virus recognition to human cell receptors, replication and transcription. The generated models were validated followed by Ramachandran plot along with their sequence and structural alignment. The manuscript also reports the mapping of the active sites and the amino acids residues involved for all these protein targets. The computational scientists and medicinal chemistry researchers can use these homology models and the active site for rational designing and docking studies for novel antivirals as well as can perform docking studying of the FDA approved drugs for repurposing against COVID-19.

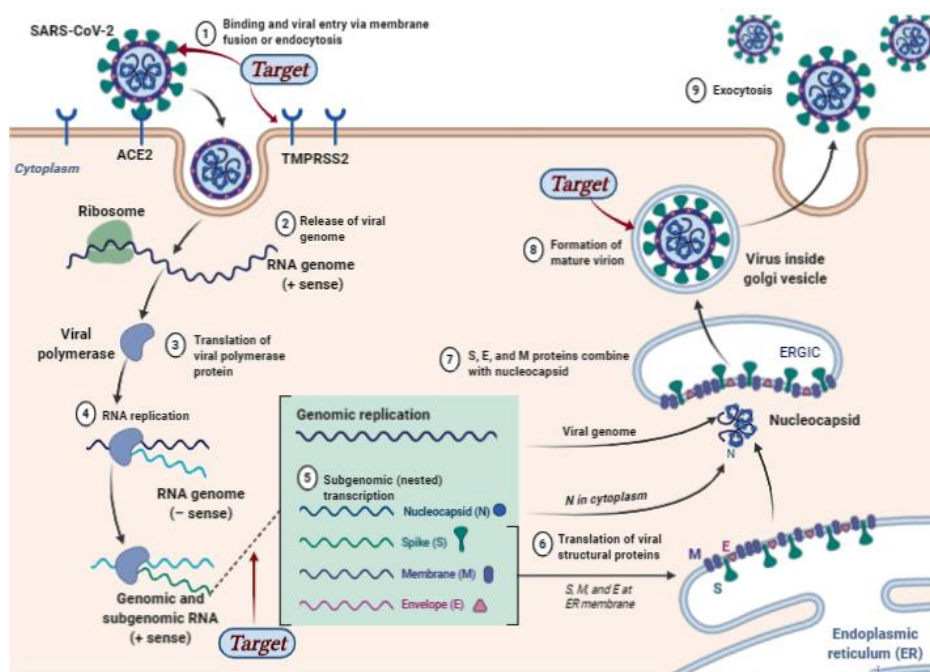
**Keywords:** COVID-19, homology model, active site, antivirals, drug design

## **INTRODUCTION**

The recent pandemic due to the novel coronavirus SARS-CoV-2 (COVID-19) is causing significant mortality worldwide (1). World Health Organization (WHO) named the recent viral pneumonia as 2019 novel coronavirus or "2019-nCoV" on January 12, 2020 (2,3). Because of profoundly homologous nature of 2019-nCoV with SARS-CoV, it is considered as a nearby relative of SARS-CoV. Thus, 2019-nCoV is grouped as a Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) on February 11, 2020 as per the ordered from The International Virus Classification Commission (ICTV). Simultaneously, WHO named the infection brought about by 2019-nCoV as COVID-19 (4). Coronaviruses (CoVs) belonging to the family Coronaviridae having a one of the longest single stranded positive sense RNA genome (~29.7 kb) protected by a enveloped structure (5). The shape of the overall nCoV-2 is either pleomorphic or round which has club-molded projections of glycoproteins on its surface having a diameter across 80–120 nm. Seven strains of human CoVs are available; significantly Middle East respiratory disorder (MERS)- CoV, extreme intense respiratory disorder (SARS)- CoV, and 2019-novel coronavirus (nCoV) are most vulnerable for causing the severe infection on both upper and lower respiratory tract (6).

Unfortunately until this point, no particular remedial medication or vaccine has been affirmed for the treatment of human coronavirus. The recent outbreak of novel coronavirus disease-

19 (COVID-19) calls for repurposing of the market drug as well as finding new antivirals based on rational design. Treatment of COVID-19 can follow approaches by blocking the virus binding to human cell receptors or inhibiting the virus's self-assembly process through acting on some structural proteins to prevent the synthesis of viral RNA through acting on the genetic material of the virus, inhibiting virus replication through acting on critical enzymes of virus (figure 1). The related viral target proteins include ACE2, TMPRSS2, Spike protein, Envelop protein, Membrane protein, non-structural proteins such as NSP1, NSP2, NSP4, RNA-dependent RNA polymerase (RdRp), helicase, Papain-Like Proteases (PLpro), main protease (3CLpro, also named 3-chymotrypsin-like protease) etc (7,8).



**Figure 1. Life cycle of novel coronavirus SARS-CoV-2 and possible drug targets.**


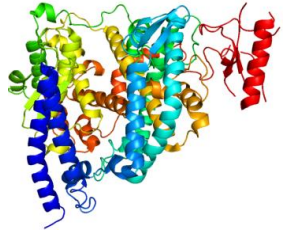
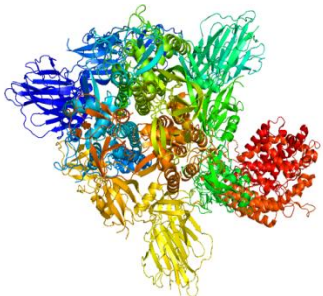
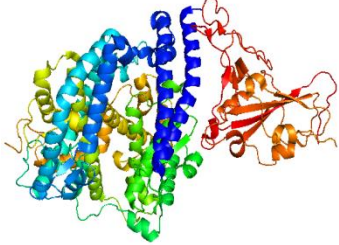
There is a lack of specific antiviral drugs for the treatment of COVID-19. Most treatment strategies are still not widely accepted and focus on symptomatic management and supportive therapy only. Antimalarial drug having lysosomotropic groups such as chloroquine (CQ) and hydroxychloroquine (HCQ) were found to be effective. Some therapeutic agents that require further experimental validation are ribavirin, remdesivir, ritonavir and lopinavir (8,9). In the present work, we have compiled a list of 20 probable drug targets on host and virus based on viral life cycle which can


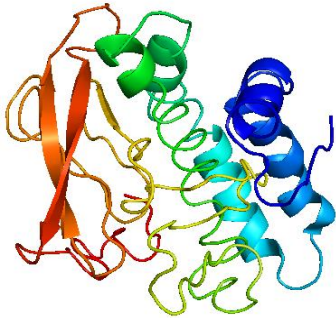
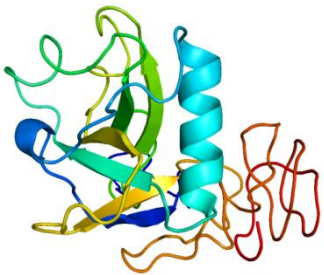
targeted for the rational development of future antivirals as well as computationally screening marketed drugs for repurposing. We prepared homology model for ten vital proteins for which no crystal structure is reported.

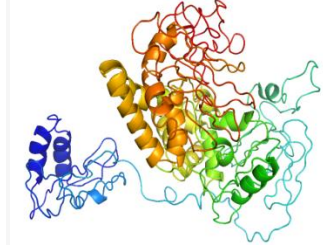
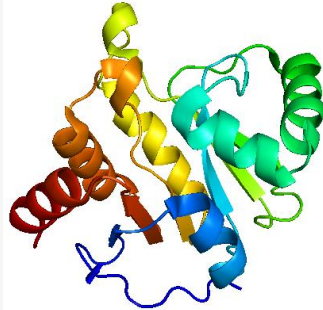

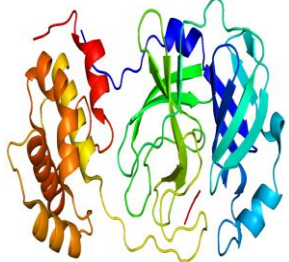
## **MATERIALS AND METHODS**

We have searched thoroughly the RCSB protein data bank of the possible drug target for the treatment against novel corona virus 2019. There are plenty of proteins associated with their life cycle. Among them few are already available in PDB. For the preparation of rest of the protein associated with SARS-n-Cov2, we followed up template based homology modeling. The most readily accessible full genome sequence for the recently emerged COVID-19 is recovered from the National Center for Biotechnology Information (NCBI) nucleotide database. The accession code for that sequence is NC\_045512.2. The required genome sequence of the respective structural target protein isolated as a query sequence. The Protein-Protein Blast (BLASTp) tools was performed against all known crystals structures deposited into protein data bank (PDB) in order to identify closely related homologous template structure for modeling (10). Based on the structure and sequence similarities with the template protein and E-value on the BLAST results, proper highest identical homologous template structure has been selected. Swiss model web-server was utilized to build the models (Table 1). Homology models for TMPRSS2, Envelope Protein (E), Membrane Protein (M), NSP1, NSP2, NSP4, NSP6, NSP7,, NSP14 protein targets were prepared for which no crystal structure is available. The generated models were validated followed by Ramachandran plot along with their sequence and structural alignment. The binding pocket residues of all the protein models are calculated by utilizing COACH meta-server (11) and also cross verified with the CASTp webservers. All the active sites of the homology build proteins were evaluated after superimposition of the closely related X-ray crystallized structure bound with the co-crystal ligands.

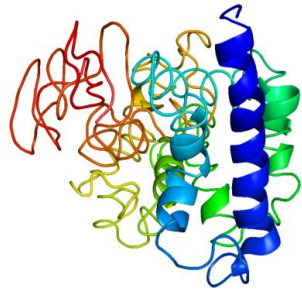
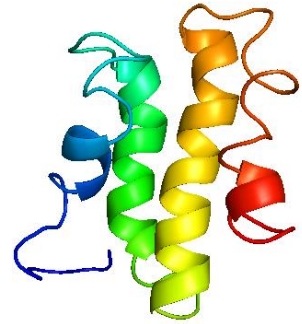
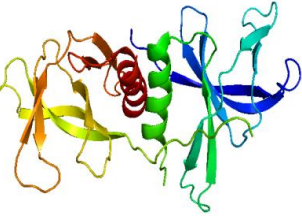
**Table 1. Compilation and preparation of homology model of possible protein targets for viral entry, replication and transcription**

Sl. No.	Protein Name	Template PDB	Final Structure
1.	Transmembrane protease, serine 2 (TMPRSS2)	<p><b>Template Used:</b> Crystal Structure of Serine protease Hepsin in complex with Inhibitor</p> <p>PDB ID: <b>5CE1</b></p> <p><b>Sequence Identity:</b> 33.82%</p>	
2.	Angiotensin Converting Enzyme 2 (ACE2)	<p>Crystal Structure Available.</p> <p>PDB ID: <b>1R42</b></p> <p>(<a href="https://www.rcsb.org/structure/1R42">https://www.rcsb.org/structure/1R42</a>)</p>	
3.	SARS Spike Glycoprotein - human ACE2 complex	<p>Crystal Structure Available.</p> <p>PDB ID: <b>6CS2</b></p> <p>(<a href="https://www.rcsb.org/structure/6CS2">https://www.rcsb.org/structure/6CS2</a>)</p>	
4.	Native Spike Protein (S)	<p>Crystal Structure Available.</p> <p><b>For SARS-COV</b></p> <p>PDB ID: <b>6CRV</b> (Resolution: 3.2 Å)</p> <p><a href="https://www.rcsb.org/structure/6CRV">https://www.rcsb.org/structure/6CRV</a></p> <p><b>SARS-CoV-2:</b></p> <p>PDB ID:</p> <p><b>6LZG</b> (Resolution: 2.5Å)</p> <p><a href="https://www.rcsb.org/structure/6lzg">https://www.rcsb.org/structure/6lzg</a></p> <p><b>6M0J</b> (resolution: 2.45 Å)</p> <p><a href="https://www.rcsb.org/structure/6M0J">https://www.rcsb.org/structure/6M0J</a></p> <p><b>6VW1</b> (resolution: 2.68 Å) :</p>	

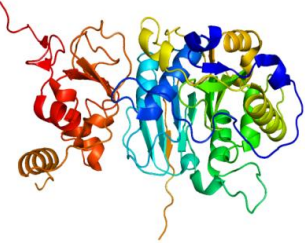
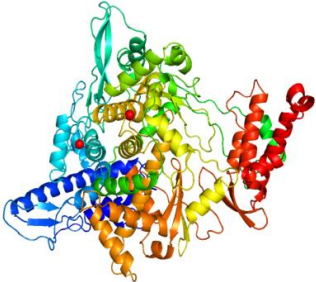
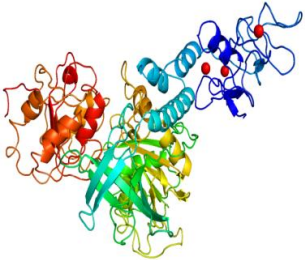
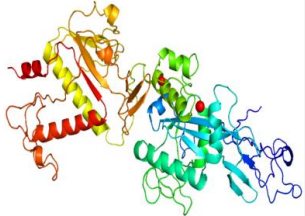
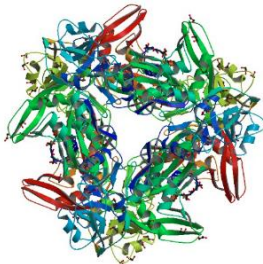
		<a href="https://www.rcsb.org/structure/6VW1">https://www.rcsb.org/structure/6VW1</a>	
5.	Envelope Protein (E)	<p>Crystal Structure Available.</p> <p><b><u>For SARS-COV</u></b></p> <p>PDB ID: <b>5X29</b></p> <p><a href="https://www.rcsb.org/structure/5X29">https://www.rcsb.org/structure/5X29</a></p> <p><b><u>For SARS-CoV-2</u></b></p> <p>Crystal Structure Not Available.</p> <p><b>Template Used:</b> NMR structure of the SARS Coronavirus E protein pentameric ion channel.</p> <p>PDB ID: <b>5X29</b></p> <p><b>Sequence Identity:</b> 91.38%</p>	
6.	Membrane Protein (M)	<p>Crystal Structure Available.</p> <p><b><u>For SARS-COV</u></b></p> <p>PDB ID: 3I6G (Resolution: 2.201 Å)</p> <p><a href="https://www.rcsb.org/structure/3I6G">https://www.rcsb.org/structure/3I6G</a></p> <p><b><u>SARS-CoV-2:</u></b></p> <p>Crystal Structure Not Available.</p> <p><b>Template Used:</b> Crystal structure of melB apo-protyrosinase from <i>Asperugillus oryzae</i></p> <p><b>PDB ID:</b> 3W6Q</p>	
7.	Non-structural protein 1 (NSP1)	<p>Crystal Structure Not Available.</p> <p><b>Template Used:</b> NMR Structure of the nonstructural protein 1 (nsp1) from the SARS coronavirus</p> <p>PDB ID: <b>2HSX</b></p> <p><b>Sequence Identity:</b> 86.09%</p>	

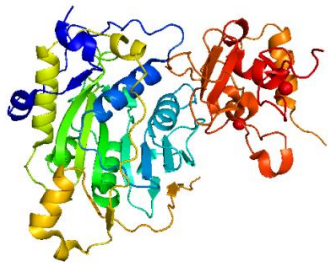
8.	Non-structural protein 2 (NSP2)	<p>Crystal Structure Not Available.</p> <p><b>Template Used:</b> Structure of the M1 Alanylaminopeptidase from malaria</p> <p>PDB ID: <b>3EBG</b></p>	
9.	Papain-Like Proteases (PLpro)-NSP3	<p>Crystal Structure Available.</p> <p><b>For SARS-COV</b></p> <p>PDB ID: <b>4OVZ</b> (Inhibitor: N-[(4-fluorophenyl)methyl]-1-[(1R)-1-naphthalen-1-ylethyl]piperidine-4-carboxamide) <a href="https://www.rcsb.org/structure/4OVZ">https://www.rcsb.org/structure/4OVZ</a></p> <p><b>SARS-CoV-2:</b></p> <p>PDB ID: <b>6VXS</b> (Resolution : 2.03 Å) <a href="https://www.rcsb.org/structure/6VXS">https://www.rcsb.org/structure/6VXS</a></p> <p><b>6W02</b> (Resolution : 1.5 Å) <a href="https://www.rcsb.org/structure/6w02">https://www.rcsb.org/structure/6w02</a></p> <p><b>6W6Y</b> (Resolution : 1.451 Å) <a href="https://www.rcsb.org/structure/6w6y">https://www.rcsb.org/structure/6w6y</a></p>	
10.	Non-structural protein 4 (NSP4)	<p>Crystal Structure Not Available.</p> <p><b>Template used:</b> C425S mutant of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis virus A59</p> <p>PDB ID: <b>3VCB</b></p> <p><b>Sequence Identity:</b> 61.36%</p>	
11.	Chymotrypsin-like cysteine protease/ main protease (Mpro)/ 3C-like protease (3CLpro) - NSP5	<p>Crystal Structure Available.</p> <p>PDB ID: <b>5R7Z</b> (Inhibitor: {N}-[2-(5-fluoranyl-1-{H}-indol-3-yl)ethyl]ethanamide) <a href="https://www.rcsb.org/structure/5r7z">https://www.rcsb.org/structure/5r7z</a></p> <p><b>5R7Y</b> (Inhibitor: N-(2-phenylethyl)methanesulfonamide)</p>	



		<p><a href="https://www.rcsb.org/structure/5r7y">https://www.rcsb.org/structure/5r7y</a></p> <p><b>5R80</b> (Inhibitor: methyl 4-sulfamoylbenzoate) <a href="https://www.rcsb.org/structure/5r80">https://www.rcsb.org/structure/5r80</a></p> <p><b>6Y84</b> (unliganded version) <a href="https://www.rcsb.org/structure/6y84">https://www.rcsb.org/structure/6y84</a></p> <p><b>6LU7</b> (Inhibitor: n-[(5-methylisoxazol-3-yl)carbonyl]alanyl-l-valyl-n~1~--((1r,2z)-4-(benzyloxy)-4-oxo-1-[(3r)-2-oxopyrrolidin-3-yl]methyl}but-2-enyl)-l-leucinamide) <a href="https://www.rcsb.org/structure/6LU7">https://www.rcsb.org/structure/6LU7</a></p>	
12.	Non-structural protein 6 (NSP6)	<p>Crystal Structure Not Available.</p> <p><b>Template used:</b> Crystal structure of <i>Thermotoga Maritima</i> sodium pumping membrane integral pyrophosphatase with metal</p> <p>PDB ID: <b>4AV3</b></p>	
13.	Non-structural protein 7 (NSP7)	<p>Crystal Structure Not Available.</p> <p><b>Template used:</b> NMR Structure of the nonstructural Protein 7 (nsP7) from the SARS CoronaVirus</p> <p>PDB ID: <b>1YSY</b></p> <p><b>Sequence Identity:</b> 98.80%</p>	
14.	Non-structural protein 9 (NSP9)	<p>Crystal Structure Available.</p> <p>PDB ID: <b>6W4B</b> <a href="https://www.rcsb.org/structure/6W4B">https://www.rcsb.org/structure/6W4B</a></p>	



15.	Non-structural protein 10 (NSP 10)	<p>Crystal Structure Available.</p> <p>PDB ID:  <b>6w4h</b> (Resolution : 1.80 Å)  <a href="https://www.rcsb.org/structure/6W4H">https://www.rcsb.org/structure/6W4H</a>  <b>6w75</b> (Resolution : 1.95 Å)  <a href="https://www.rcsb.org/structure/6W75">https://www.rcsb.org/structure/6W75</a>  <b>6w61</b> (Resolution : 2.00 Å)</p>	
16.	RNA dependent RNA polymerase (RdRp)—NSP12	<p>Crystal Structure Not Available.</p> <p><b>Template used:</b> SARS-Coronavirus NSP12 bound to NSP7 and NSP8 co-factors</p> <p>PDB ID: <b>6NUR</b></p> <p><b>Sequence Identity:</b> 96.35%</p>	
17.	RNA helicase (NSP13)	<p>Crystal Structure Not Available.</p> <p><b>Template Used:</b> Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis</p> <p>PDB ID: <b>6JYT</b></p> <p><b>Sequence Identity:</b> 99.83%</p>	
18.	Guanine-N7 methyltransferase (NSP14):	<p>Crystal Structure Not Available.</p> <p><b>Template Used:</b> Crystal structure of the SARS coronavirus nsp14-nsp10 complex with functional ligands SAH and GpppA</p> <p>PDB ID: <b>5C8S</b></p> <p><b>Sequence Identity:</b> 95.07%</p>	
19.	Uridylate-specific endoribonuclease (NendoU/NSP 15)	<p>Crystal Structure Available.</p> <p>PDB ID:  <b>6W01</b> (Resolution: 1.9 Å)  <a href="https://www.rcsb.org/structure/6W01">https://www.rcsb.org/structure/6W01</a>  <b>6VWW</b> (Resolution: 2.2 Å)  <a href="https://www.rcsb.org/structure/6VWW">https://www.rcsb.org/structure/6VWW</a></p>	

20.	2'-O-ribose methyltransferase (NSP16):	Crystal Structure Available. PDB ID: <b>6W4H</b> (Resolution: 1.8 Å) <a href="https://www.rcsb.org/structure/6W4H">https://www.rcsb.org/structure/6W4H</a> <b>6W75</b> (Resolution: 1.95 Å) <a href="https://www.rcsb.org/structure/6W75">https://www.rcsb.org/structure/6W75</a> <b>6W61</b> (Resolution: 2.00 Å) <a href="https://www.rcsb.org/structure/6W61">https://www.rcsb.org/structure/6W61</a>	
-----	--	---	---

## RRESULT AND DISCUSSION

The probable protein targets of our interest on host and virus along with their role in pathogenesis has been described below. (9, 12-17)

**ACE2 protein:** Angiotensin Converting Enzyme 2 (ACE2) is present in host and is involved during viral entry to host. The Spike (S) protein of the Covid-19 binds with ACE2 receptor for entry into host. Crystal structure of ACE2 is available (PDB ID: 1R42). Active site of the ACE2 is Pro426, Ile439, Leu440 and Gln 442 (Figure S17).

**Spike Protein (S):** S Protein is a viral surface protein responsible for binding to ACE2 receptor present in host. The binding of S protein with ACE2 is critical for entry of virus to host cell. There are various crystal structures available for S protein SAR-COV (PDB ID: 6CRV), SARS-n-CoV2 PDB ID: 6LZG, 6MOJ, 6VW1). Crystal structures of the bound Spike glycoprotein-ACE2 complex (PDB ID: 6CS2) provides the nature of binding of these two critical protein. Active site of the S protein is present on the surface of the protein so that it can capture host ACE2 protein receptors; names. Lys26, Asn90 and Phe32 are the active site residues (Figure S18).

**Transmembrane protease, serine 2 (TMPRSS2):** TMPRSS2 is a transmembrane protease 2 is that is responsible for proteolytic cleavage and processing of the viral spike glycoproteins. TMPRSS2 is a host protein which facilitates virus-cell membrane fusions. No crystal structure is available against this protein till now. Crystal Structure of Serine protease Hepsin (5CE1) was taken as a template by maintain the 33.82% sequence similarities (Figure S16). 99% residues are within the allowed regions according to the Ramachandran Plot (Figure S15). Active site residues are also calculated; named as His296, Ser441, Asp345 (Figure S19).

**Envelope (E) protein:** Envelope (E) protein is an integral membrane protein involved in the assembly and formation of envelop which is important for COVID-19 pathogenesis. Homology model of the membrane protein was builded by taking the SAS-CoV E protein (PDB ID: 5X29) (Table 2). Overall 98% residues are in the allowed regions (Figure S21). Active site of the protein is Ala32, Thr35, Tyr42 and Tyr57 which is present on the surface of the protein (Figure S20).

**Membrane (M) protein:** Membrane (M) protein is the central organizer that determines the shape of viral envelopes and interacts with the E protein in the budding compartment of the host cell. The crystal structure of SARS-COV (PDB ID: 3I6G) is present but not for COVID-19. Active Sites of M protein is Lys50, Leu51, Leu54, Leu93 and Ala98 (Figure S22).

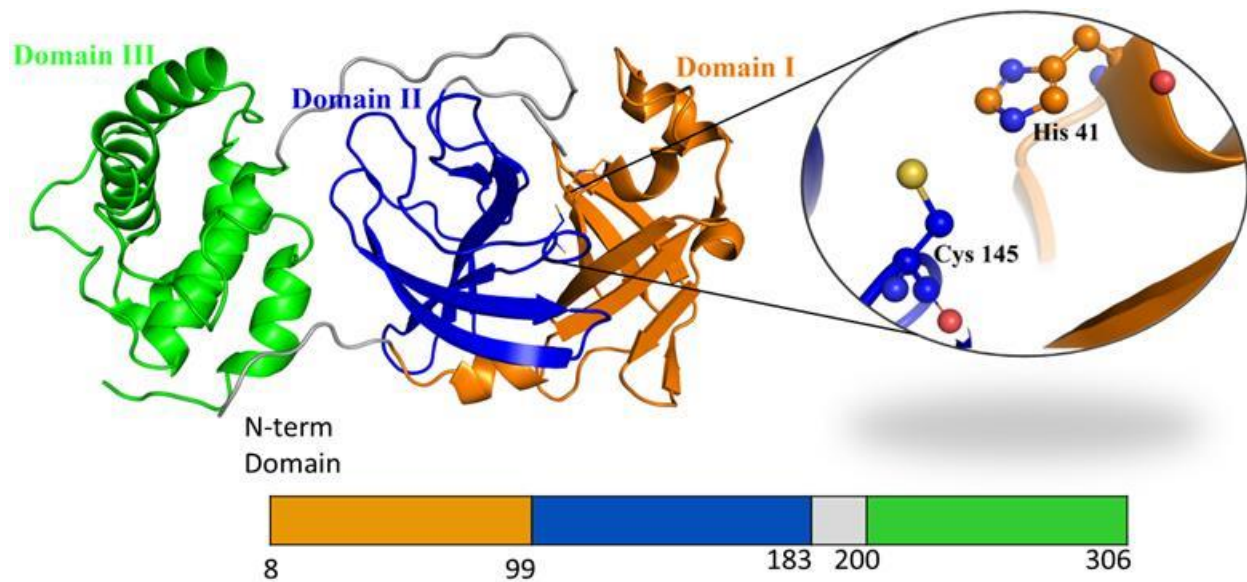
**Non-structural protein 1 (NSP1):** NSP1 is a membrane-associated protein which is responsible for anchoring the replication complex to the cellular membranes. In the infected cell through interaction with the 40S ribosomal subunit, NSP1 suppresses host gene expression which facilitates viral gene expression by avoiding host immune response. Crystal structure of Novel corona virus is still not discovered. So, homology model of NSP1 has been builded by taking SARS-CoV NSP1 (PDB ID: 2HSX) as a template structure. The Sequence identity is 86.09% (Figure S2). Overall 96% residues are in the allowed regions (Figure S1). The active sites region is built by Val 35, Glu36, Pro 67 and Tyr 68 (Figure S23).

**Non-structural protein 2 (NSP2):** NSP2 plays a role in the host cell survival signaling pathway. Homology model of this protein has been built with the M1 Alanyl aminopeptidase protein with 50.09% sequence similarities (Figure S4). 84% residues are on the allowed regions on Ramachandran plot [Figure S3]. Active site of NSP2 protein is depicted on supplementary figure 24.

**Papain-Like Proteases:** A viral Papain-Like Proteases (PLpro) is critical for virus replication and packaging within the host cells through cleavage of viral peptides into functional proteins. The PLpro cleaves polyproteins (PPs) to generate NSP1, NSP2 and NSP3. The crystal structures for SARS-COV (PDB ID: 4OVZ) and SARS-CoV-2 (PDB ID: 6VXS, 6WO2 AND 6W6Y) can be used to design future antivirals. Asp22, Ile23, Val49, Ile131 and Leu160 residues are present in the active site of PLpro (Figure S25).

**Non-structural protein 4 (NSP4):** NSP4 is necessary for viral replication. Due to unavailability of nCov2 NSP4 protein, homology structure has been built by maintain the 61.36% sequence identity with the template protein (3VCB) (Figure S6). Whole protein and along with the active sites are depicted on the supplementary figure 26.

**3-Chymotrypsin-like cysteine protease (3-CLpro)/main protease (Mpro):** 3CLpro cleaves viral peptides into functional proteins for virus replication and packaging within the host cells. It can cleave PPs to release 16 NSPs and produce mature protein that anchors the replication/transcription. There are many crystal structures (PDB ID: 5R7Z, 5R7Y, 5R80, 6Y84, 6LU7) for Mpro protein is available. In their crystal structure three domains are present. Residues 8-99 are in domain I, 100-183 are in the domain II and 183 to 200 are the loop regions and lastly 200 to 306 are the domain III. His41 and Cys 145 are the active residues (Figure 2).



**Figure 2: Structural alignment of NSP4 protein. Active site residues are marked with blue and orange elemental colored ball and sticks representations**

---

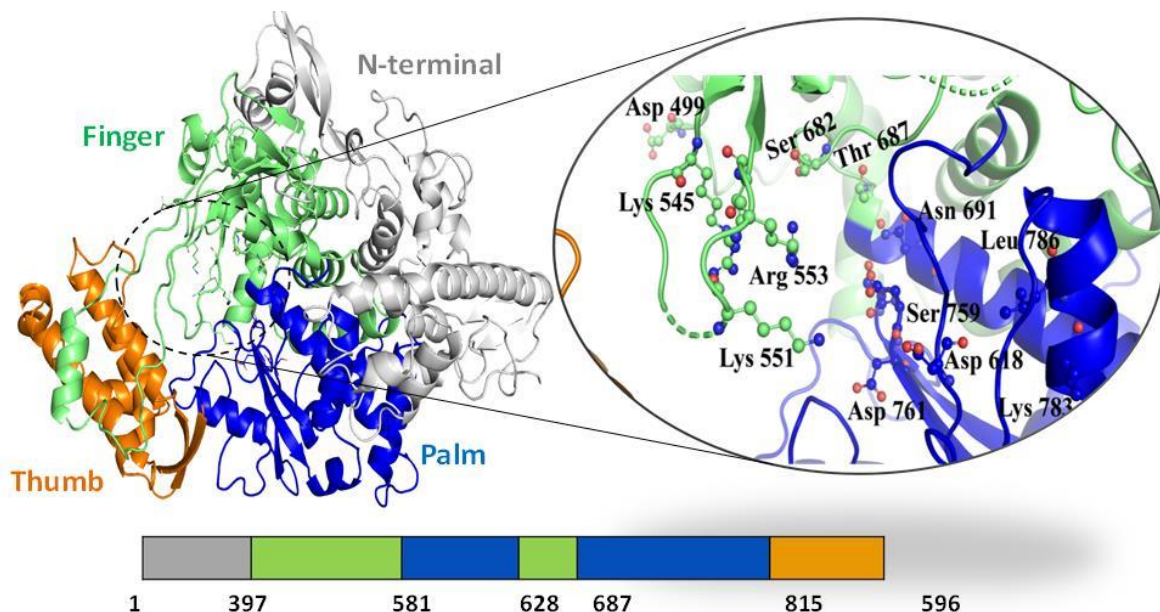
**Non-structural protein 6 (NSP6):** NSP6 is a viral replicase protein capable of inducing autophagy. Homology structure has been made for this protein. Ramachandran plot and structure and sequence

alignment has incorporated on supplementary figure S7 and S8 respectively. His11, Leu15, Ile 18 and Phe70 are the active sites residues (Figure S27).

**Non-structural protein 7 (NSP7):** NSP7 is a primase may participate in viral replication. Due to unavailability of the crystal structure it is built by homology modeling with SARS NSP7 template protein (PDB ID: 1YSY). Sequence alignment and Ramachandran plot is on the figure S9 and figure S10. Val16, Cys32 and Leu55 are active site residues (Figure S28).

NSP7 and NSP8 can form hexadecamer. NSP9 is a RNA binding protein of COVID-19 and interacts with NSP8 for its functions. NSP9 is a replicase protein that participates in viral replication. The crystal structure (PDB ID: 6W4B) of NSP9 is available. Active site of this protein is depicted on figure S29.

**Non-structural protein 10 (NSP10):** NSP10 is a cofactor for the activation of the replicative enzyme and plays important role in viral transcription. The available PDB IDs for NSP10 are 6W4H, 6W75 and 6W61 (Figure S30).

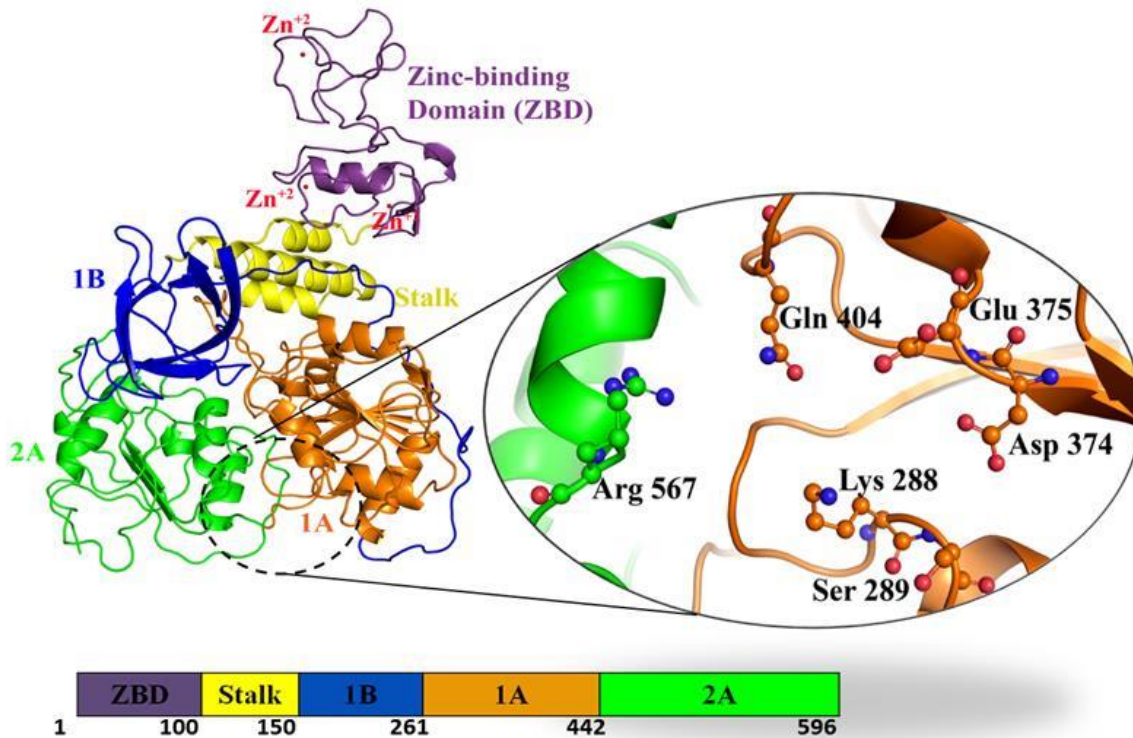


**Figure 3: Structural alignment of RdRp protein. Active site residues are marked with blue and green elemental colored ball and sticks representations.**



**RNA dependent RNA polymerase (RdRp):** RNA dependent RNA polymerase (RdRp)/NSP12 is a RNA dependent RNA polymerase. RdRp is the main protein for RNA replication and transcription of viral RNA genome. This protein is a cluster of several domains. 1-397 is N-terminal domain, (397-581) and (629-687) is finger domain, (582-628) and (688-815) is the palm region and (816-596) is the thumb region (Figure 3).

**Non-structural protein 13 (NSP13):** NSP13 is a helicase that separates double-stranded RNA (dsRNA) or DNA (dsDNA). NSP14 shows exoribonuclease activity acting on both ssRNA and dsRNA. It also has guanine methyltransferase activity. Active sites and several structural domains are depicted on figure 4.



**Figure 4: Structural alignment of NSP13 protein. Active site residues are marked with blue and green elemental colored ball and sticks representations. Metal binding domain is represented by red mesh color map.**

**Uridylate-specific endoribonuclease (NendoU/NSP15):** NSP15 shows endoribonuclease activity. The crystal structure (PDB ID: 6W01, 6VWW) for NSP15 is available His235, Gln245, Gly248, Gln294 Thr341 are the active sites residues (Figure S32).

**2'-O-Ribose methyltransferase (NSP16):** NSP16 has methyltransferase activity plays vital role in viral mRNAs cap methylation which is important to evade immune system. Structural domain and active sites are depicted on figure S33. All these target proteins mentioned above have critical role in the virus binding to human cell receptors, replication and transcription.

**Table 2: Protein Analysis Table**

<b>Protein Name</b>	<b>Ramachandran Plot Statistics</b>	<b>Sequence alignment</b>	<b>Ramachandran Plot Figure Number</b>	<b>Sequence alignment Fig. Number</b>
<b>Non-structural protein 1 (NSP1)</b>	Residues in most favored regions: 77.1% Residues in additional allowed regions : 18.8% Residues in generously allowed regions: 3.1% Residues in disallowed regions: 1.0%	Sequence Identity: 86.09%	Figure S1	Figure S2
<b>Non-structural protein 2 (NSP2)</b>	Residues in most favoured regions: 62.5% Residues in additional allowed regions: 25.4% Residues in generously allowed regions: 7.2% Residues in disallowed regions: 4.8%	Sequence Identity: 50.09%	Figure S3	Figure S4
<b>Non-structural protein 4 (NSP4)</b>	Residues in most favoured regions: 94.5% Residues in additional allowed regions : 4.9% Residues in generously allowed regions : 0.6% Residues in disallowed regions: 0.0%	Sequence Identity: 61.36%	Figure S5	Figure S6
<b>Non-structural protein 6 (NSP6)</b>	Residues in most favoured regions: 33.5% Residues in additional allowed regions: 35.0% Residues in generously allowed regions: 18.4% Residues in disallowed regions: 13.2%	Sequence Identity: 66%	Figure S7	Figure S8
<b>Non-structural protein 7 (NSP7)</b>	Residues in most favoured regions: 72.5% Residues in additional allowed regions: 18.22.5% Residues in generously allowed regions: 3.8% Residues in disallowed regions: 1.2%	Sequence Identity: 98.80%	Figure S9	Figure S10
<b>RNA helicase</b>	Residues in most favoured regions: 80.3% Residues in additional allowed regions: 79	Sequence Identity:	Figure S11	Figure S12



<b>(NSP13)</b>	14.8% Residues in generously allowed regions: 3.6% Residues in disallowed regions: 1.3%	99.83%		
<b>Guanine-N7 methyltransferase (NSP14):</b>	Residues in most favoured regions: 91.4% Residues in additional allowed regions: 8.1% Residues in generously allowed regions: 0.4% Residues in disallowed regions: 0.0%	Sequence Identity: 95.07%	Figure S13	Figure S14
<b>Transmembrane protease, serine 2 (TMPRSS2)</b>	Residues in most favoured regions 84.6% Residues in additional allowed regions: 14.7% Residues in generously allowed regions: 0.3% Residues in disallowed regions:0.3%	Sequence Identity: 33.82%	Figure S15	Figure S16
<b>Envelope (E) protein</b>	Residues in most favoured regions: 84.4% Residues in additional allowed regions: 14.1% Residues in generously allowed regions: 1.5% Residues in disallowed regions 0 0.0%	Sequence Identity: 91.38%	Figure S20	

In summary, the compiled list of 20 probable drug targets on host and virus along with their role in pathogenesis will provide a snap shot for the drug discovery scientists. They present an attractive target for new antiviral design. We have prepared homology model of TMPRSS2, Envelope Protein (E), Membrane Protein (M), NSP1, NSP2, NSP4, NSP6, NSP7, NSP14, that have critical role in the virus recognition to human cell receptors, replication and transcription. The generated models were validated followed by Ramachandran plot along with their sequence and structural alignment (table 2). The residues present in the active site of all the protein models were calculated by utilizing COACH meta-server and subsequently cross verified with the CASTp web servers. All the active sites of the homology build proteins were evaluated after superimposition of the closely related X-ray crystallized structure bound with the co-crystal ligands. The computational scientists and medicinal chemistry researchers can use these homology models and the active site analysis described in the manuscript for rational designing and docking studies for novel antivirals as well as can perform docking studying of the FDA approved drugs for repurposing against COVID-19.

## Author Contribution

SP and AT have contributed to the concept of the study, preparation of model and writing of the manuscript. Both the authors have approved the final version of the article.

## Acknowledgments

SP has received Senior Research Fellowship from Indian Council of Medical Research

## Disclosure Summary

The authors declare no conflict of interest

## References

1. World Health Organization. Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> Date of accessed: March 31, 2020.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-733. doi: 10.1056/NEJMoa2001017.
3. Chen Y, Liu Q, Guo D, Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol.* 2020;92(4):418-423. doi: 10.1002/jmv.25681.
4. Canrong Wu, Yang Liu, Yueying Yang, Peng Zhang, Wu Zhong, Yali Wang, Qiqi Wang, Yang Xu, Mingxue Li, Xingzhou Li, Mengzhu Zheng, Lixia Chen, Hua Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B*, 2020; 2211-3835. doi.org/10.1016/j.apsb.2020.02.008.
5. Yang H, Bartlam M, Rao Z. Drug design targeting the main protease, the Achilles' heel of coronaviruses. *Curr Pharm Des.* 2006;12(35):4573-90. doi: 10.2174/138161206779010369.
6. Paules CI, Marston HD, Fauci AS. Coronavirus Infections-More than Just the Common Cold. *JAMA.* 2020. doi: 10.1001/jama.2020.0757.
7. Canrong Wu, Yang Liu, Yueying Yang, Peng Zhang, Wu Zhong, Yali Wang, Qiqi Wang, Yang Xu, Mingxue Li, Xingzhou Li, Mengzhu Zheng, Lixia Chen, Hua Li Analysis

- of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods *Acta Pharmaceutica Sinica B*. doi: <https://doi.org/10.1016/j.apsb.2020.02.008>.
8. Senanayake, SL. Drug repurposing strategies for COVID-19. *Future Drug. Discov.* eISSN 2631-3316. doi: 10.4155/fdd-2020-0010.
  9. Prajapat M, Sarma P, Shekhar N, Avti P, Sinha S, Kaur H, Kumar S, Bhattacharyya A, Kumar H, Bansal S, Medhi B. Drug targets for corona virus: A systematic review. *Indian J Pharmacol.* 2020;52(1):56–65. doi: 10.4103/ijp.IJP\_115\_20.
  10. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server issue):W5–W9. doi: 10.1093/nar/gkn201
  11. Yang J, Roy, A, Zhang Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. 2013;29(20): 2588–2595.
  12. te Velhuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Res.* 2012;40:1737–47.
  13. Stobart CC, Sexton NR, Munjal H, Lu X, Molland KL, Tomar S, et al. Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *J Virol.* 2013;87:12611–8.
  14. Wang H, Xue S, Yang H, Chen C. Recent progress in the discovery of inhibitors targeting coronavirus proteases. *Virol Sin.* 2016;31:24–30.
  15. Egloff MP, Ferron F, Campanacci V, Longhi S, Rancurel C, Dutartre H, et al. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc Natl Acad Sci U S A.* 2004;101:3792–6.
  16. Hu T, Chen C, Li H, Dou Y, Zhou M, Lu D, et al. Structural basis for dimerization and RNA binding of avian infectious bronchitis virus nsp9. *Protein Sci.* 2017;26:1037–48
  17. Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, et al. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem.* 2014;289:25783–96.