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

Sourav Pal, Dr. Arindam Talukdar

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 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. 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.

File list (2)

Manuscript_Compilation of Potential Protein Targets for S... (1.15 MiB)	view on ChemRxiv • download file
Supplementary_Compilation of Potential Protein Targets fo... (2.42 MiB)	view on ChemRxiv • download file

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

Sourav Pal[1,2] and Arindam Talukdar[1,2*,]

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

[2] 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 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. 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, TMPRSSS2, 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).
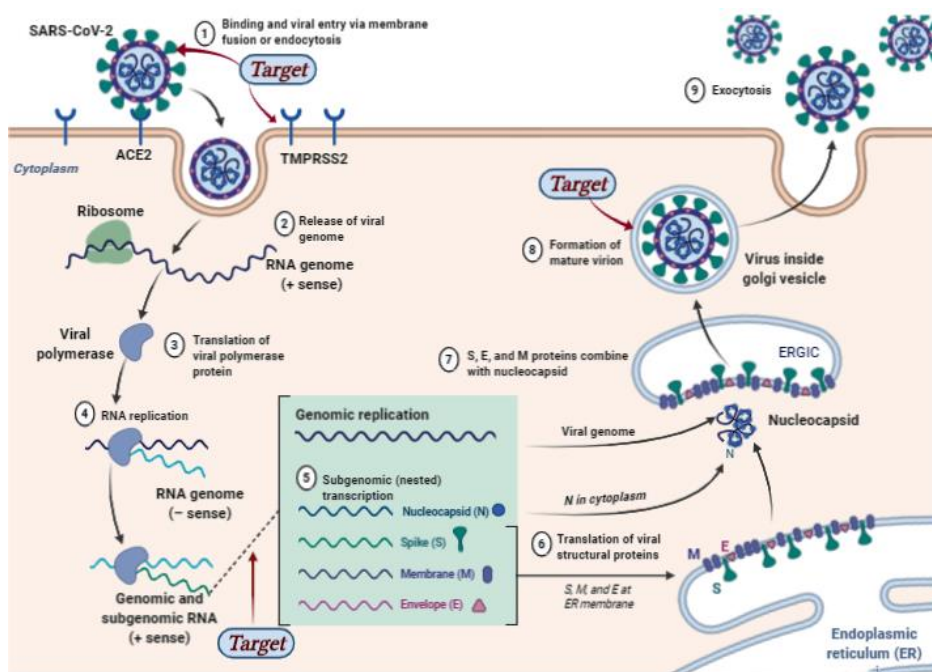
Figure 1. Life cycle of novel coronavirus SARS-CoV-2 and possible drug targets.
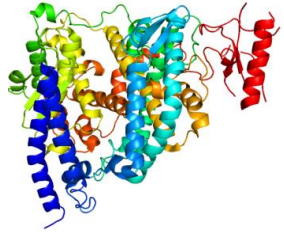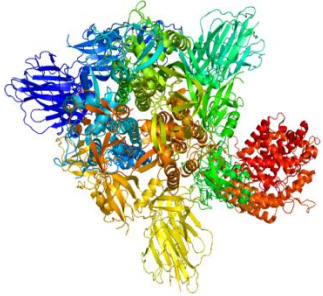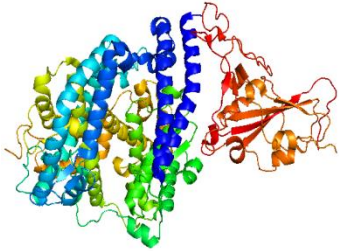
There is a lack of specific antiviral drugs for the treatment of COVID-19. Most treatment strategies are still not widely accepted and focus on symptomatic management and supportive therapy only. Antimalarial drug having lysosomotropic groups such as chloroquine (CQ) and hydroxychloroquine (HCQ) were 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
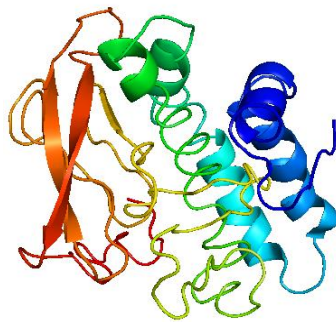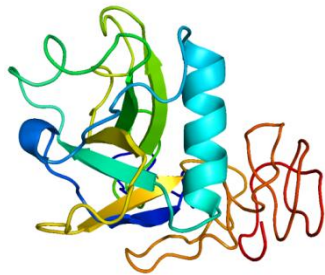
targeted for the rational development of future antivirals as well as computationally screening marketed drugs for repurposing. We prepared homology model for ten vital proteins for which no crystal structure is reported.

MATERIALS AND METHODS
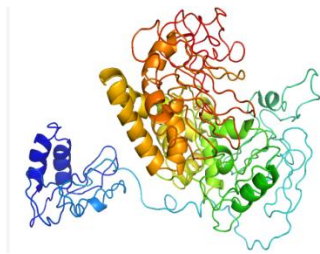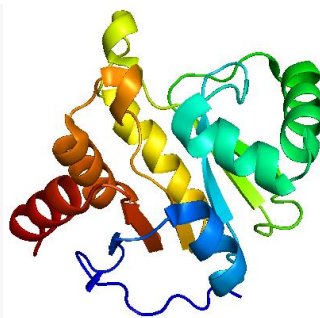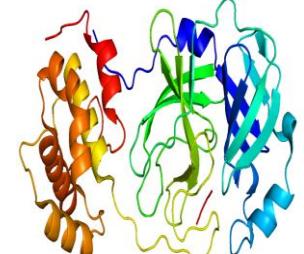
We have searched thoroughly the RCSB protein data bank of the possible drug target for the treatment against novel corona virus 2019. There are plenty of proteins associated with their life cycle. Among them few are already available in PDB. For the preparation of rest of the protein associated with 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 highes t 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
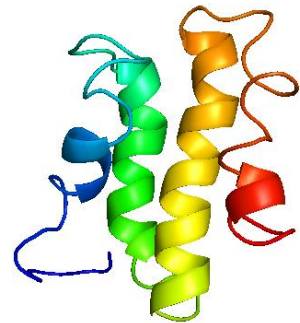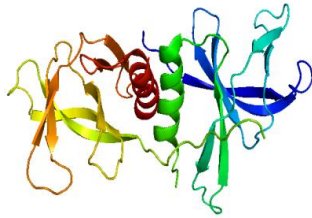
Sl. No.	Protein Name	Template PDB	Final Structure
1.	Transmembrane protease, serine 2 (TMPRSS2)	**Template Used:** Crystal Structure of Serine protease Hepsin in complex with Inhibitor PDB ID: **5CE1** **Sequence Identity:** 33.82%	
2.	Angiotensin Converting Enzyme 2 (ACE2)	Crystal Structure Available. PDB ID: **1R42** (https://www.rcsb.org/structure/1R42)	
3.	SARS Spike Glycoprotein - human ACE2 complex	Crystal Structure Available. PDB ID: **6CS2** (https://www.rcsb.org/structure/6CS2)	
4.	Native Spike Protein (S)	Crystal Structure Available. **For SARS-COV** PDB ID: **6CRV** (Resolution: 3.2 Å) https://www.rcsb.org/structure/6CRV **SARS-CoV-2:** PDB ID: **6LZG (**Resolution: 2.5Å) https://www.rcsb.org/structure/6lzg **6M0J** (resolution: 2.45 Å) :https://www.rcsb.org/structure/6M0J **6VW1** (resolution: 2.68 Å) :	
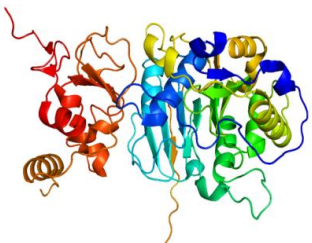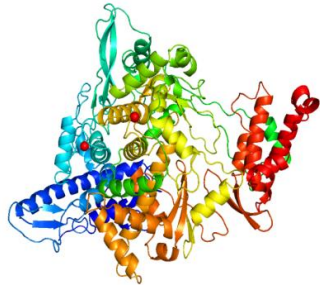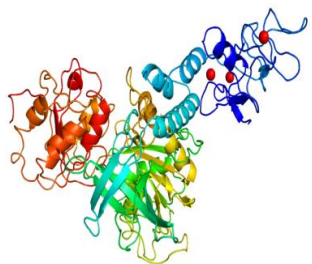
		https://www.rcsb.org/structure/6VW1	
5.	Envelope Protein (E)	Crystal Structure Available. **For SARS-COV** PDB ID: **5X29** https://www.rcsb.org/structure/5X29 **For SARS-CoV-2** Crystal Structure Not Available. **Template Used:** NMR structure of the SARS Coronavirus E protein pentameric ion channel. PDB ID: **5X29** **Sequence Identity:** 91.38%	
6.	Membrane Protein (M)	Crystal Structure Available. **For SARS-COV** PDB ID: 3I6G (Resolution: 2.201 Å) https://www.rcsb.org/structure/3I6G **SARS-CoV-2:** Crystal Structure Not Available. **Template Used:** Crystal structure of melB apo-protyrosinase from *Asperugillus oryzae* **PDB ID:** 3W6Q	
7.	Non-structural protein 1 (NSP1)	Crystal Structure Not Available. **Template Used**: NMR Structure of the nonstructural protein 1 (nsp1) from the SARS coronavirus PDB ID: **2HSX** **Sequence Identity:** 86.09%	

8.	Non-structural protein 2 (NSP2)	Crystal Structure Not Available. **Template Used:** Structure of the M1 Alanylaminopeptidase from malaria PDB ID: **3EBG**	
9.	Papain-Like Proteases (PLpro)- NSP3	Crystal Structure Available. **For SARS-COV** PDB ID: **4OVZ (Inhibitor**: N-[(4-fluorophenyl)methyl]-1-[(1R)-1-naphthalen-1-ylethyl]piperidine-4-carboxamide) https://www.rcsb.org/structure/4OVZ **SARS-CoV-2:** PDB ID: **6VXS** (Resolution : 2.03 Å) https://www.rcsb.org/structure/6VXS **6W02** (Resolution : 1.5 Å) https://www.rcsb.org/structure/6w02 **6W6Y** (Resolution : 1.451 Å) https://www.rcsb.org/structure/6w6y	
10.	Non-structural protein 4 (NSP4)	Crystal Structure Not Available. **Template used:** C425S mutant of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis virus A59 PDB ID: **3VCB** **Sequence Identity:** 61.36%	
11.	Chymotrypsin-like cysteine protease/ main protease (Mpro)/ 3C-like protease (3CLpro) - NSP5	Crystal Structure Available. PDB ID: **5R7Z** (Inhibitor: {N}-[2-(5-fluoranyl-1~{H}-indol-3-yl)ethyl]ethanamide) *https://www.rcsb.org/structure/5r7Z* **5R7Y** (Inhibitor: N-(2-phenylethyl)methanesulfonamide)	

		https://www.rcsb.org/structure/5r7y **5R80** (Inhibitor: methyl 4-sulfamoylbenzoate) *https://www.rcsb.org/structure/5r80* **6Y84** (unliganded version) *https://www.rcsb.org/structure/6y84* **6LU7** (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) *https://www.rcsb.org/structure/6LU7*	
12.	Non-structural protein 6 (NSP6)	Crystal Structure Not Available. **Template used:** Crystal structure of *Thermotoga Maritima* sodium pumping membrane integral pyrophosphatase with metal PDB ID: **4AV3**	
13.	Non-structural protein 7 (NSP7)	Crystal Structure Not Available. **Template used:** NMR Structure of the nonstructural Protein 7 (nsP7) from the SARS CoronaVirus PDB ID: **1YSY** **Sequence Identity:** 98.80%	
14.	Non-structural protein 9 (NSP 9)	Crystal Structure Available. PDB ID: **6W4B** (https://www.rcsb.org/structure/6W4B)	

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

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

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 sturucture 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 envelops 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).

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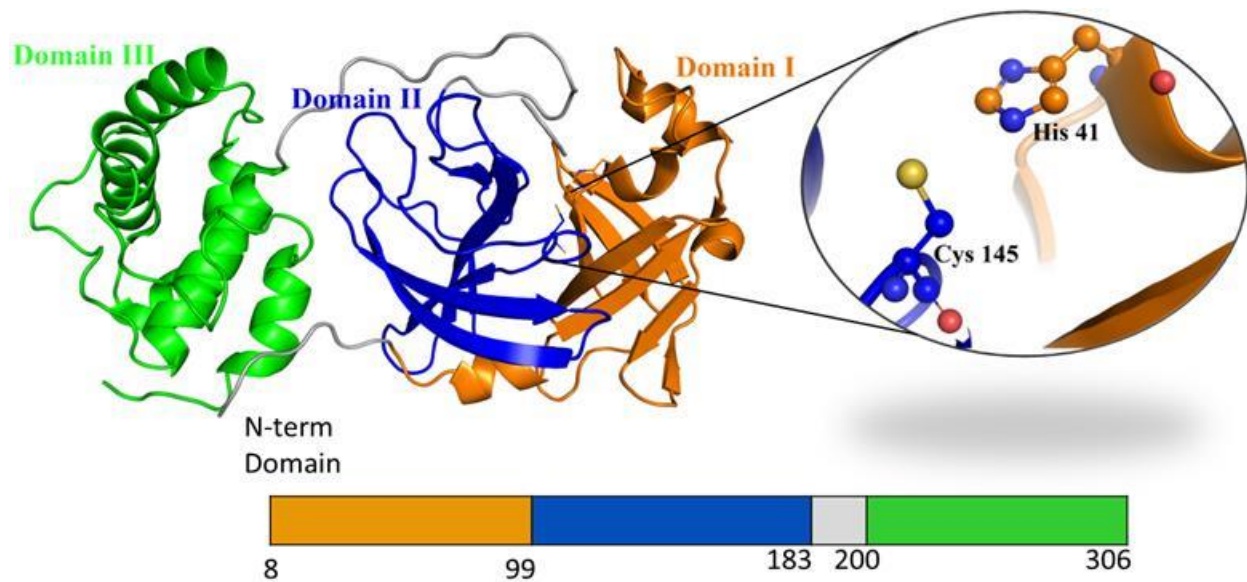
Figure 2: Structural alignment of NSP4 protein. Active site residues are marked with blue and orange elemental colored ball and sticks representations

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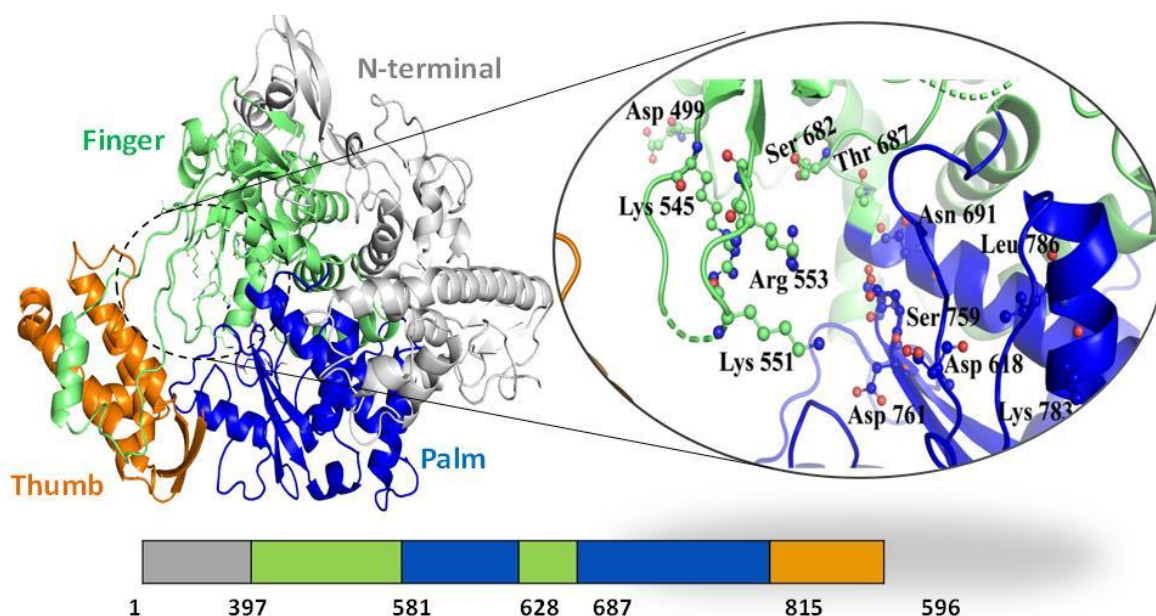
Figure 3: Structural alignment of RdRp protein. Active site residues are marked with blue and green elemental colored ball and sticks representations.

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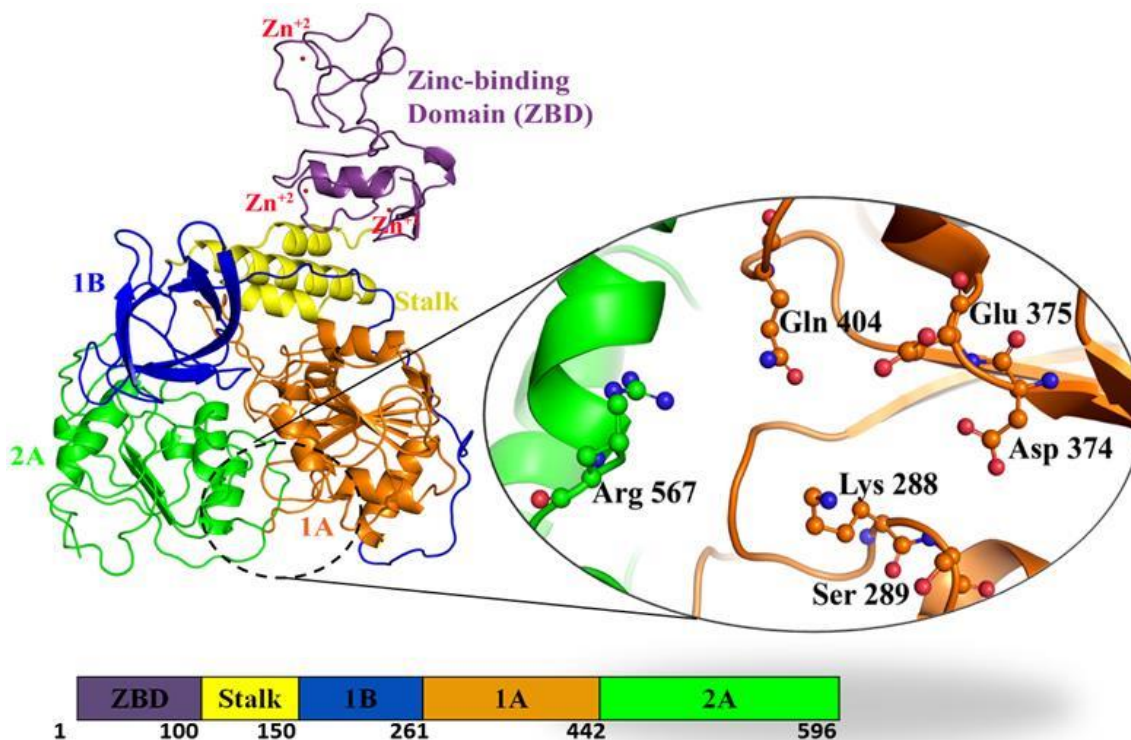
Figure 4: Structural alignment of NSP13 protein. Active site residues are marked with blue and green elemental colored ball and sticks representations. Metal binding domain is represented by red mesh color map.

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 toevade immune system. Structural domain and active sites are depicted on figure S33. All these target proteinsmentioned above have critical role in the virus binding to human cell receptors, replication and transcription.

Table 2: Protein Analysis Table

Protein Name	Ramachandran Plot Statistics	Sequence alignment	Ramachandran Plot Figure Number	Sequence alignment Fig. Number
Non-structural protein 1 (NSP1)	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
Non-structural protein 2 (NSP2)	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
Non-structural protein 4 (NSP4)	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
Non-structural protein 6 (NSP6)	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
Non-structural protein 7 (NSP7)	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
RNA helicase	Residues in most favoured regions: 80.3% Residues in additional allowed regions: 79	Sequence Identity:	Figure S11	Figure S12

(NSP13)	14.8% Residues in generously allowed regions: 3.6% Residues in disallowed regions: 1.3%	99.83%		
Guanine-N7 methyltransferase (NSP14):	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
Transmembrane protease, serine 2 (TMPRSS2)	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
Envelope (E) protein	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 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. 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

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 Velthuis 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.

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

Sourav Pal[1,2] and Arindam Talukdar[1,2*]

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

[2] Academy of Innovative and Scientific Research, Ghaziabad-201002, India

Supplementary Information:

PROCHECK

Ramachandran Plot
x802

Plot statistics

Residues in most favoured regions [A,B,L]	74	77.1%
Residues in additional allowed regions [a,b,l,p]	18	18.8%
Residues in generously allowed regions [~a,~b,~l,~p]	3	3.1%
Residues in disallowed regions	1	1.0%
	----	------
Number of non-glycine and non-proline residues	96	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	12	
Number of proline residues	6	
Total number of residues	115	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

x802_01.ps

Figure S1: Ramachandran plot of NSP1

A

```
                  A              ββββγβ       β        H1                A            β
nCov2-NSP1   MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGV        60
2hsx         ------------HVQLSLPVLQVRDVLVRGFGDSVEEALSEAREHLKNGTCGLVELEKGV        48
                         ******************************* *****:***:********:****

                 H2       γ  A      γ   β        γ A         β      β  ββγ   β        A       β       A
nCov2-NSP1   LPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETLGVLVPHVGEIPVAYRK
                120
2hsx         LPQLEQPYVFIKRSDALSTNHGHKVVELVAEMDGIQYGRSGITLGVLVPHVGETPIAYRN        108
             ****************  :: ***  :*******::******** *********** *:***:.

nCov2-NSP1   VLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGVTRELMRELNGG
                180
2hsx         VLLRKNG-----------------------------------------------        115
             *******
```

B

Figure S2: A. Sequence alignment of homology modeled NSP1 protein with the template SARS-Cov protein having 86.09% of sequence identity. The entire identical region is marked with a yellow highlighted portion. The secondary structure are also predicted and attached herewith. B. Structure alignment of NSP1 protein with the template SARS-Cov protein (2HSX) with a RMSD value of 0.077Å. Here homology model is represented by blue color and template model are marked with a orange color.

Figure S3: Ramachandran plot of NSP2

A

```
                         ββ      H1              γ   β      H2                β         β            γ
nCov2-NSP2   AYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQLDFIDTKRGVYCCREHEHEIAW                    60
3ebg         AYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQLDFIDTKRGVYCCREHEHEIAW                    60
             ************************************************************

                 γβ    ββ  γββββγβ    β   H3       β        γ    β   β        β          βββ      β
nCov2-NSP2   YTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFPLNSIIKTIQPRVEKKKLDGFMGRI                    120
3ebg         YTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFPLNSIIKTIQPRVEKKKLDGFMGRI                    120
             ************************************************************

                 β  β        γγβ           β  γ                   β           ββ  β
nCov2-NSP2   RSVYPVASPNECNQMCLSTLMKCDHCGETSWQTGDFVKATCEFCGTENLTKEGATTCGYL
                   180
3ebg         RSVYPVASPNECNQMCLSTLMKCDHCGETSWQTGDFVKATCEFCGTENLTKEGATTCGYL  180
             ************************************************************

                  β        β       β          β   H4              ββ
nCov2-NSP2   PQNAVVKIYCPACHNSEVGPEHSLAEYHNESGLKTILRKGGRTIAFGGCVFSYVGCHNKC
                   240
3ebg         PQNAVVKIYCPACHNSEVGPEHSLAEYHNESGLKTILRKGGRTIAFGGCVFSYVGCHNKC  240
             ************************************************************

                      β      γ         βγ  H5             β  β   βγ  H6       β
nCov2-NSP2   AYWVPRASANIGCNHTGVVGEGSEGLNDNLLEILQKEKVNINIVGDFKLNEEIAIILASF    300
3ebg         AYWVPRASANIGCNHTGVVGEGSEGLNDNLLEILQKEKVNINIVGDFKLNEEIAIILASF    300
             ************************************************************

                  β    H7        H8         β           β        βγ      H9    H10
nCov2-NSP2   SASTSAFVETVKGLDYKAFKQIVESCGNFKVTKGKAKGAWNIGEQKSILSPLYAFASEA
                   360
3ebg         SASTSAFVETVKGLDYKAFKQIVESCGNFKVTKGKAKGAWNIGEQKSILSPLYAFASEA  360
             ************************************************************

                  H11           H12            βββ   H13          β       γ    βββ
nCov2-NSP2   ARVVRSIFSRTLETAQNSVRVLQKAAITILDGISQYSLRLIDAMMFTSDLATNNLVVMAY 420
3ebg         ARVVRSIFSRTLETAQNSVRVLQKAAITILDGISQYSLRLIDAMMFTSDLATNNLVVMAY 420
             ************************************************************

                 βββ  H14        H15            H16         βββ       β
nCov2-NSP2   ITGGVVQLTSQWLTNIFGTVYEKLKPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIV 480
3ebg         ITGGVVQLTSQWLTNIFGTVYEKLKPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIV 480
             ************************************************************

                ββ       β  H17        H18    βββ   H19    βββγγ    ββ      β    β
nCov2-NSP2   GGQIVTCAKEIKESVQTFFKLVNKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKC 540
3ebg         GGQIVTCAKEIKESVQTFFKLVNKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKC 540
             ************************************************************

                H20   β  β    β   β         β  β    β        β    ββ     βββ    β
nCov2-NSP2   VKSREETGLLMPLKAPKEIIFLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVG 600
3ebg         VKSREETGLLMPLKAPKEIIFLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVG 600
             ************************************************************

                 β   βββ   γ    β   β     β  ββ
nCov2-NSP2   TPVCINGLMLLEIKDTEKYCALAPNMMVTNNTFTLKGG----------------------    638
3ebg         TPVCINGLMLLEIKDTEKYCALAPNMMVTNNTFTLKGGPKIHYRKDYKPSGFIINQVTLN  660
             *************************************

nCov2-NSP2   -----------------------------------------------------------    638
```

3ebg   INIHDQETIVRSVLDMDISKHNVGEDLVFDGVGLKINEISINNKKLVEGEEYTYDNEFLT 720

nCov2-NSP2 --- 638
3ebg   IFSKFVPKSKFAFSSEVIIHPETNYALTGLYKSKNIIVSQCEATGFRRITFFIDRPDMMA 780

nCov2-NSP2 --- 638
3ebg   KYDVTVTADKEKYPVLLSNGDKVNEFEIPGGRHGARFNDPPLKPCYLFAVVAGDLKHLSA 840

nCov2-NSP2 --- 638
3ebg   TYITKYTKKKVELYVFSEEKYVSKLQWALECLKKSMAFDEDYFGLEYDLSRLNLVAVSDF 900

nCov2-NSP2 --- 638
3ebg   NVGAMENKGLNIFNANSLLASKKNSIDFSYARILTVVGHEYFHQYTGNRVTLRDWFQLTL 960

nCov2-NSP2 --- 638
3ebg   KEGLTVHRENLFSEEMTKTVTTRLSHVDLLRSVQFLEDSSPLSHPIRPESYVSMENFYTT 1020

nCov2-NSP2 --- 638
3ebg   TVYDKGSEVMRMYLTILGEEYYKKGFDIYIKKNDGNTATCEDFNYAMEQAYKMKKADNSA 1080

nCov2-NSP2 --- 638
3ebg   NLNQYLLWFSQSGTPHVSFKYNYDAEKKQYSIHVNQYTKPDENQKEKKPLFIPISVGLIN 1140

nCov2-NSP2 --- 638
3ebg   PENGKEMISQTTLELTKESDTFVFNNIAVKPIPSLFRGFSAPVYIEDQLTDEERILLLKY 1200

nCov2-NSP2 --- 638
3ebg   DSDAFVRYNSCTNIYMKQILMNYNEFLKAKNEKLESFQLTPVNAQFIDAIKYLLEDPHAD 1260

nCov2-NSP2 --- 638
3ebg   AGFKSYIVSLPQDRYIINFVSNLDTDVLADTKEYIYKQIGDKLNDVYYKMFKSLEAKADD 1320

nCov2-NSP2 --- 638
3ebg   LTYFNDESHVDFDQMNMRTLRNTLLSLLSKAQYPNILNEIIEHSKSPYPSNWLTSLSVSA 1380

nCov2-NSP2 --- 638
3ebg   YFDKYFELYDKTYKLSKDDELLLQEWLKTVSRSDRKDIYEILKKLENEVLKDSKNPNDIR 1440

nCov2-NSP2 --- 638
3ebg   AVYLPFTNNLRRFHDISGKGYKLIAEVITKTDKFNPMVATQLCEPFKLWNKLDTKRQELM 1500

nCov2-NSP2 --------------------------- 638
3ebg   LNEMNTMLQEPQISNNLKEYLLRLTNK   1527

B
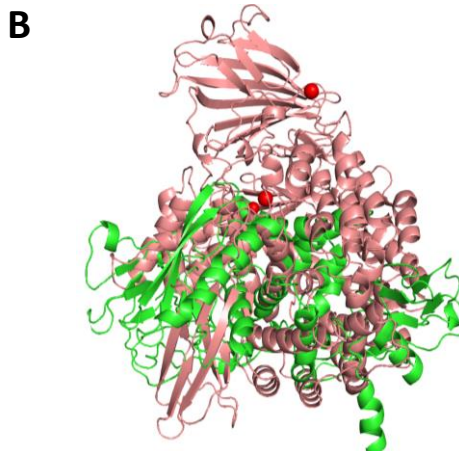
Figure S4: A. Sequence alignment of homology modeled NSP2 protein with the template SARS-CoV protein (3EBG) having 50.09% of sequence identity. The entire identical region is marked with a yellow highlighted portion. The secondary structure are also predicted and attached herewith. **B.** Structure alignment of NSP2 protein with the template SARS-CoV protein (3EBG) with a RMSD value of 23.606 Å. Here homology model is represented by green color and template model are marked with red color.

Figure S5: Ramachandran plot of NSP4

A

```
nCov2-NSP4:A    KIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAST  60
nCov2-NSP4:B    KIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAST  60
3vcb.1.A        ------------------------------------------------------------   0


Model_01:A    DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLH 120
Model_01:B    DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLH 120
3vcb          ------------------------------------------------------------  0


Model_01:A    FLPRVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSV  180
```

```
Model_01:B    FLPRVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSV  180
3vcb          ------------------------------------------------------------   0


Model_01:A    AYESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGR  240
Model_01:B    AYESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGR  240
3vcb          ------------------------------------------------------------   0


Model_01:A    WVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYY  300
Model_01:B    WVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYY  300
3vcb.1.A      -----------------------------------------------------------    0


Model_01:A    FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL  360
Model_01:B    FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL  360
3vcb          ------------------------------------------------------------   0

                                                                       H1

Model_01:A    AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
Model_01:B    AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
3vcb          --------------------------------------------------TFEEMALTTF  10
                                                                 **** ** **

                       A    H2               β     H3              β    B     H4
Model_01:A    LLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFS-
              479
Model_01:B    LLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFS-
              479
3vcb          MITKESYSKLKNSV--SDVAFNRYLSLYNKYRYFSGKMDTAAYREAACSQLAKAMETFNH  68
              :::.** * **::.*    .:****:*****:**** ***::******.:****:: *.

                    ββ           A                B
Model_01:A    NSGSDVLYQPPQTSITSAVLQ  500
Model_01:B    NSGSDVLYQPPQTSITSAVLQ  500
3vcb          NNGNDVLYQPPTASVTTSFLQ   89
              * *.:******* :*.*::. **
```

B

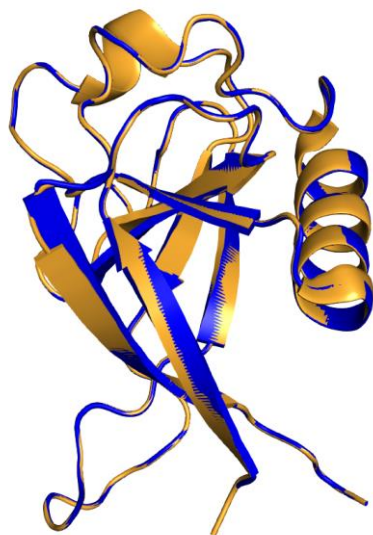**Figure S6: A. Sequence alignment of homology modeled NSP4 protein with the template SARS-Cov protein (3VCB) having 61.36% of sequence identity. The entire identical region is marked with a yellow highlighted portion. The secondary structure are also predicted and attached herewith. B. Structure alignment of NSP4 protein with the template SARS-Cov protein (3VCB) with a RMSD value of 0.115Å. Here homology model is represented by blue color and template model are marked with a yellow color.**
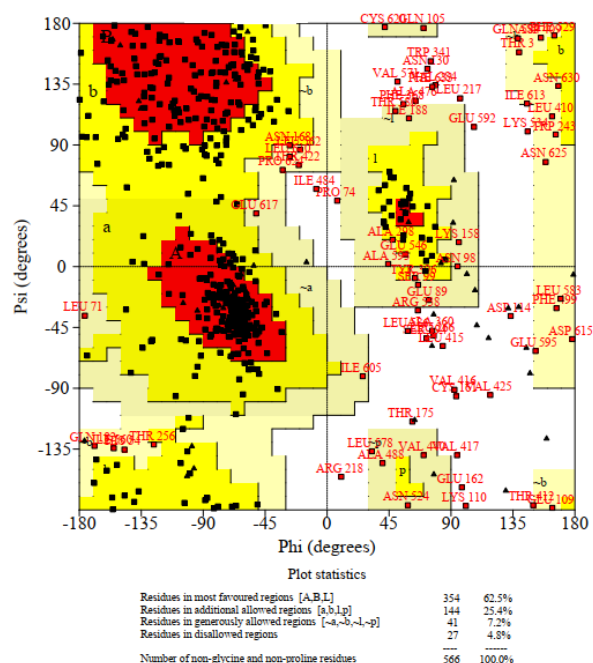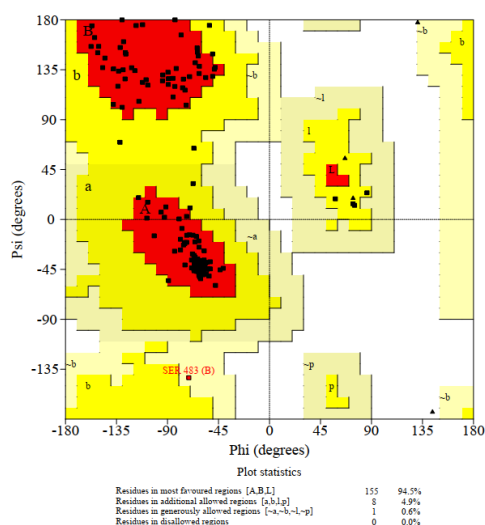
Figure S7: Ramachandran plot of NSP6

A

```
                    βββ  H1      H2              H3    β  βγ   βββ  H4      H5          β
nCov2-NSP6  SAVKRTIKGTHHWLLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFV 60
4aV3        SAVKRTIKGTHHWLLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFV 60
            ************************************************************

            β  ββββββββββββββββββββ  β  ββββββββββββββββ    γ  βββββββββββ
nCov2-NSP6  KHKHAFLCLFLLPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAV 120
4aV3        KHKHAFLCLFLLPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAV 120
            ************************************************************

            ββ  H6    βββγββ  ββββββββ   H7   ββ  ββ   βββ  ββββββββ  ββγ  ββγ  ββββββ
nCov2-NSP6  VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT 180
4aV3        VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT 180
            ************************************************************

            ββ      ββββββ      β  β  ββ  γββββββββ  βββββ  ββββββββ  βββ  β
nCov2-NSP6  TVMFLARGIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLG 240
4aV3        TVMFLARGIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLG 240
            ************************************************************

            β    ββ      ββ  β    β    βββ    ββ
nCov2-NSP6  VYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQ---------- 290
4aV3        VYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQYVAALFFLIP 300
            *************************************************

nCov2-NSP6  ------------------------------------------------------------ 290
4aV3        LVALGFAAANFAAVVRKPETERMKEISSYIRSGADSFLAHETKAIFKVAIVIAILLMIFT 360


nCov2-NSP6  ------------------------------------------------------------ 290
```

```
4aV3        TWQTGVAFLLGAVMSASAGIVGMKMATRANVRVAEAARTTKKIGPALKVAYQGGSVMGLS
            420


nCov2-NSP6  ------------------------------------------------------        290
4aV3        VGGFALLGLVLVYLIFGKWMGQVDNLNIYTNWLGINFVPFAMTVSGYALGCSIIAMFDRV  480


nCov2-NSP6  ------------------------------------------------------        290
4aV3        GGGVYTKAADMAADLVGKNPATIADNVGDNVGDVAGLGADLLESFVGAIVSSIILASYMF  540


nCov2-NSP6  ------------------------------------------------------        290
4aV3        PIYVQKIGENLVHQVPKETIQALISYPIFFALVGLGCSMLGILYVIVKKPSDNPQRELNI      600


nCov2-NSP6  ------------------------------------------------------        290
4aV3        SLWTSALLTVVLTAFLTYFYLKDLQGLDVLGFRFGAISPWFSAIIGIFSGILIGFWAEYY      660


nCov2-NSP6  ------------------------------------------------------        290
4aV3        TSYRYKPTQFLGKSSIEGTGMVISNGLSLGMKSVFPPTLTLVLGILFADYFAGLYGVAIA      720


nCov2-NSP6  ------------------------------------------------------        290
4aV3        ALGMLSFVATSVSVDSYGPIADNAGGISEMCELDPEVRKITDHLDAVGNTTAAIGKGFAI      780


nCov2-NSP6  ------------------------------------------------------        290
4aV3        GSAIFAALSLFASYMFSQISPSDIGKPPSLVLLLNMLDARVIAGALLGAAITYYFSGYLI      840


nCov2-NSP6  ------------------------------------------------------        290
4aV3        SAVTKAAMKMVDEYNRCIEITSDNALKQMGYPAFIAILTPLVTGFLLGAEFVGGVLIGTV      900


nCov2-NSP6  ------------------------------------------------------        290
4aV3        LSGAMLAILTANSGGAWDNAKKYLEAGNLEGYGKGSEPHKALVIGDTVGDPLKDTVGPSL  960


nCov2-NSP6  ----------------------- 290
4aV3        DILIKIMSVVSVIAVSIFKHVHLF  984
```

B
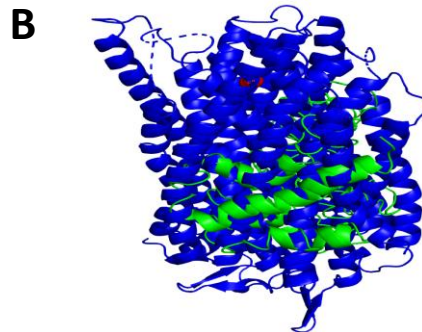
Figure S8: A. Sequence alignment of homology modeled NSP6 protein with the template SARS-Cov protein (4AV3) having 66% of sequence identity. The entire identical region is marked with a yellow highlighted

portion. The secondary structure are also predicted and attached herewith. B. Structure alignment of NSP6 protein with the template SARS-Cov protein (3VCB) with a RMSD value of 16.675Å. Here homology model is represented by green color and template model are marked with a blue color.

Figure S9: Ramachandran plot of NSP7

A

nCoV-NSP7   SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLL   60
1ysy   SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLL 60
             ****

nCoV-NSP7   SMQGAVDINKLCEEMLDNRATLQ   83
1ysy   SMQGAVDINRLCEEMLDNRATLQ 83
             *********·**************

B

Figure S10: A. Sequence alignment of homology modeled NSP7 protein with the template SARS-Cov protein (1YSY) having 98.80% of sequence identity. The entire identical region is marked with a yellow highlighted

portion. The secondary structure are also predicted and attached herewith. B. Structure alignment of NSP7 protein with the template SARS-Cov protein (1YSY) with a RMSD value of 0.455Å. Here homology model is represented by orange color and template model are marked with a blue color.
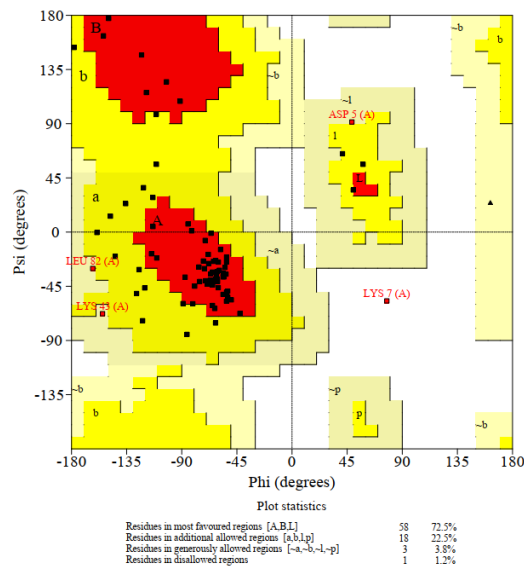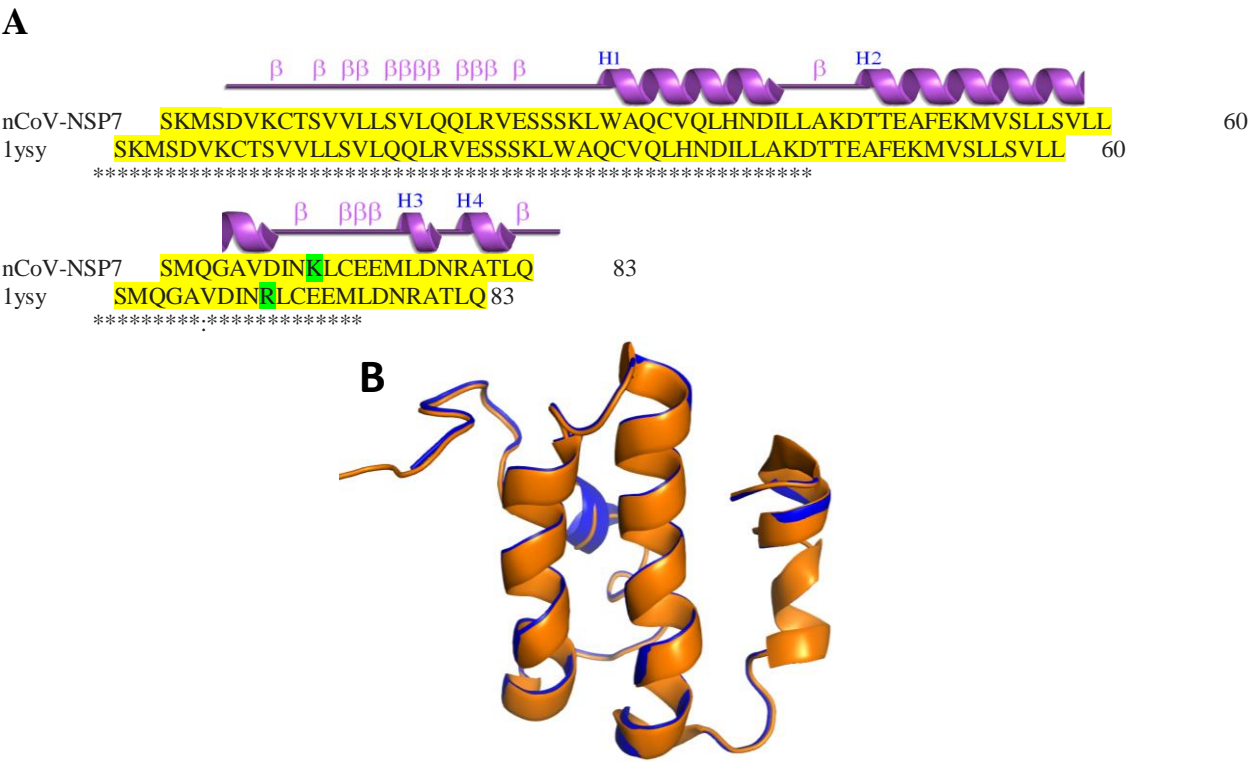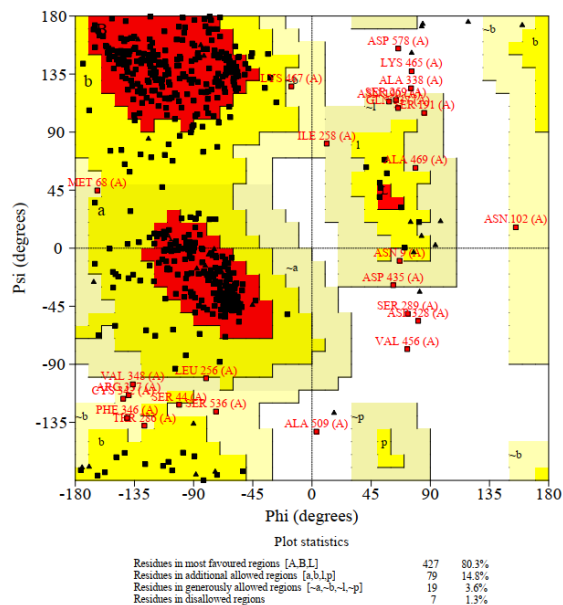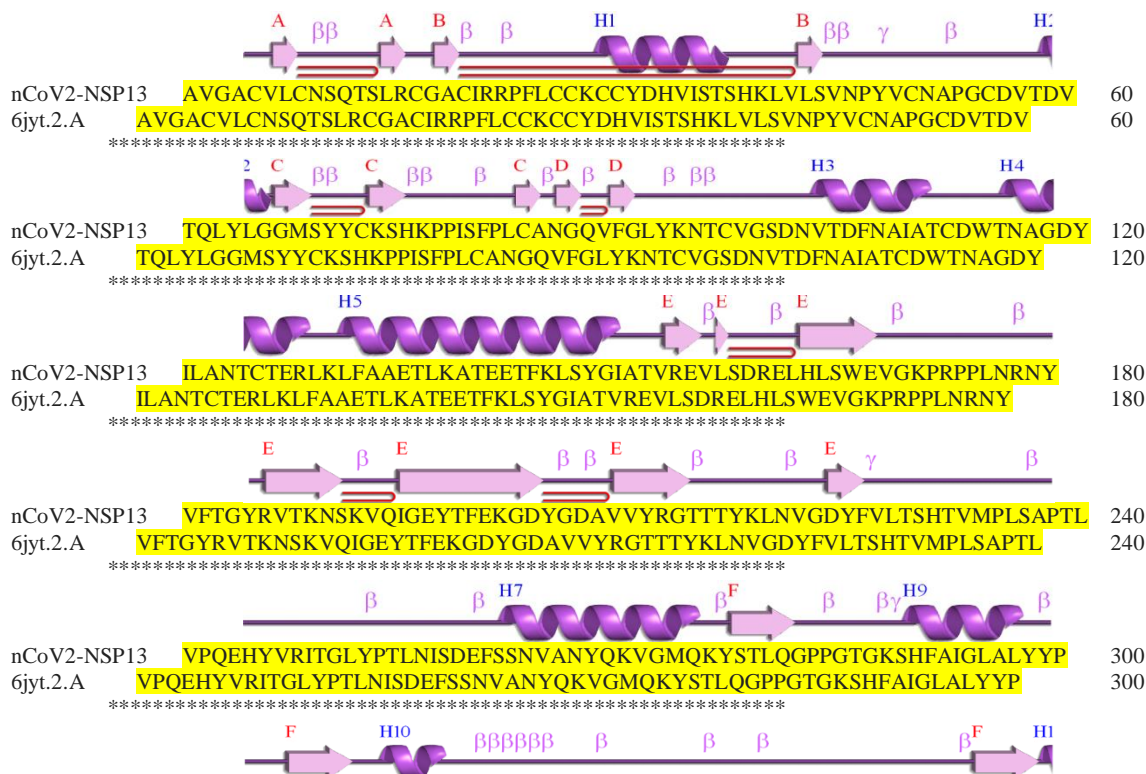

Figure S11: Ramachandran plot of NSP13

A


```
nCoV2-NSP13   AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV   60
6jyt.2.A      AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV   60
              ***********************************************************

nCoV2-NSP13   TQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDY   120
6jyt.2.A      TQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDY   120
              ***********************************************************

nCoV2-NSP13   ILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNY   180
6jyt.2.A      ILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNY   180
              ***********************************************************

nCoV2-NSP13   VFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMPLSAPTL   240
6jyt.2.A      VFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMPLSAPTL   240
              ***********************************************************

nCoV2-NSP13   VPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYP   300
6jyt.2.A      VPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYP   300
              ***********************************************************
```

```
nCoV2-NSP13   SARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTV   360
6jyt.2.A      SARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTV   360
              ***********************************************************

nCoV2-NSP13   NALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQLPAPRTLLTKGTLEPE   420
6jyt.2.A      NALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQLPAPRTLLTKGTLEPE   420
              ***********************************************************

nCoV2-NSP13   YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI   480
6jyt.2.A      YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI   480
              ***********************************************************

nCoV2-NSP13   THDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSE   540
6jyt.2.A      THDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSE   540
              ***********************************************************

nCoV2-NSP13   YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL   600
6jyt.2.A      YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEIPRRNVATL   600
              *****************************.*****************************

nCoV2-NSP13   Q   601
6jyt.2.A      Q   601
              *
```

Figure S12: A. Sequence alignment of homology modeled NSP13 protein with the template SARS-Cov protein (6JYT) having 99.83% of sequence identity. The entire identical region is marked with a yellow highlighted portion. The secondary structure are also predicted and attached herewith. **B.** Structure alignment of NSP13 protein with the template SARS-Cov protein (6JYT) with a RMSD value of 0.125Å. Here homology model is represented by pink color and template model are marked with a blue color.
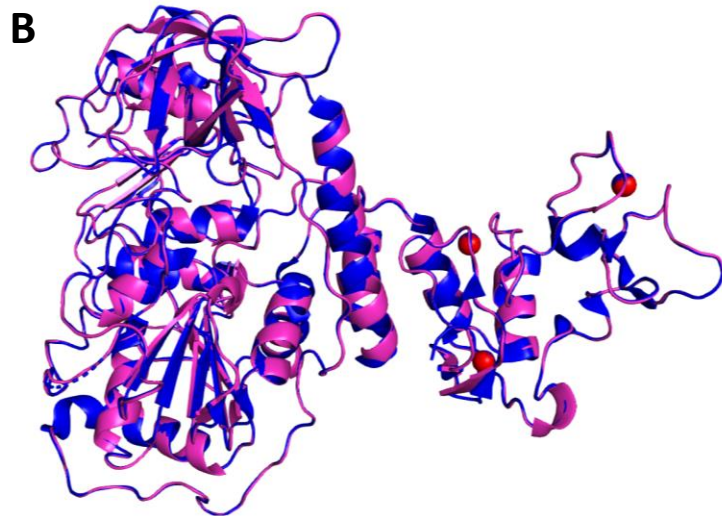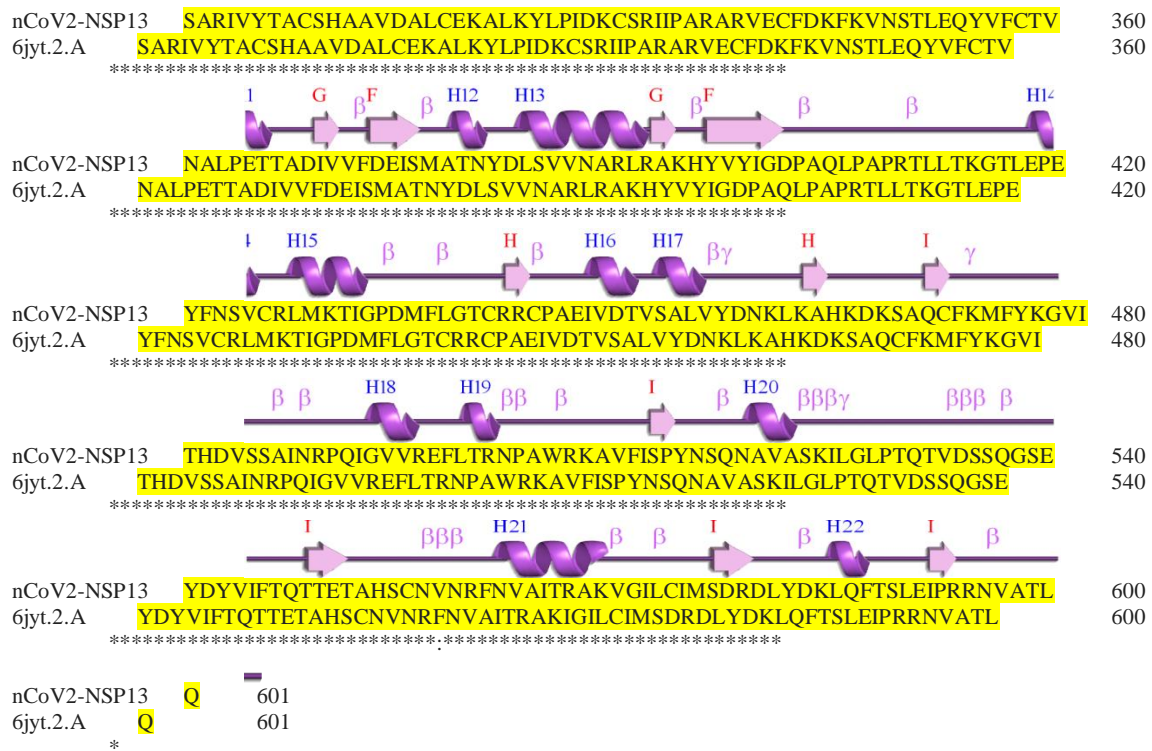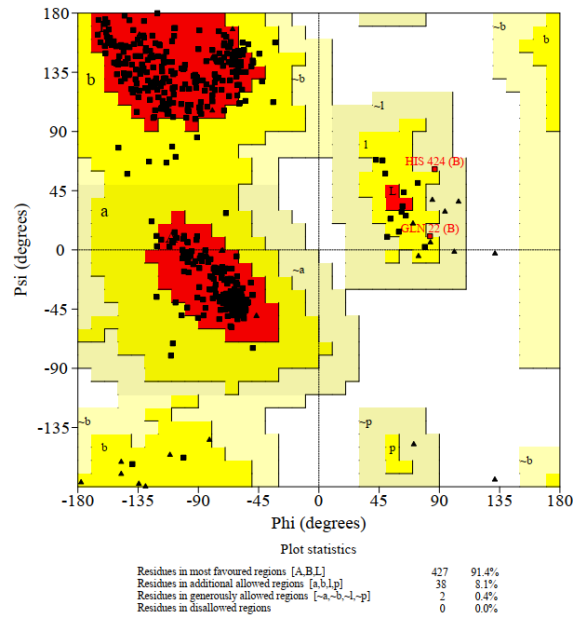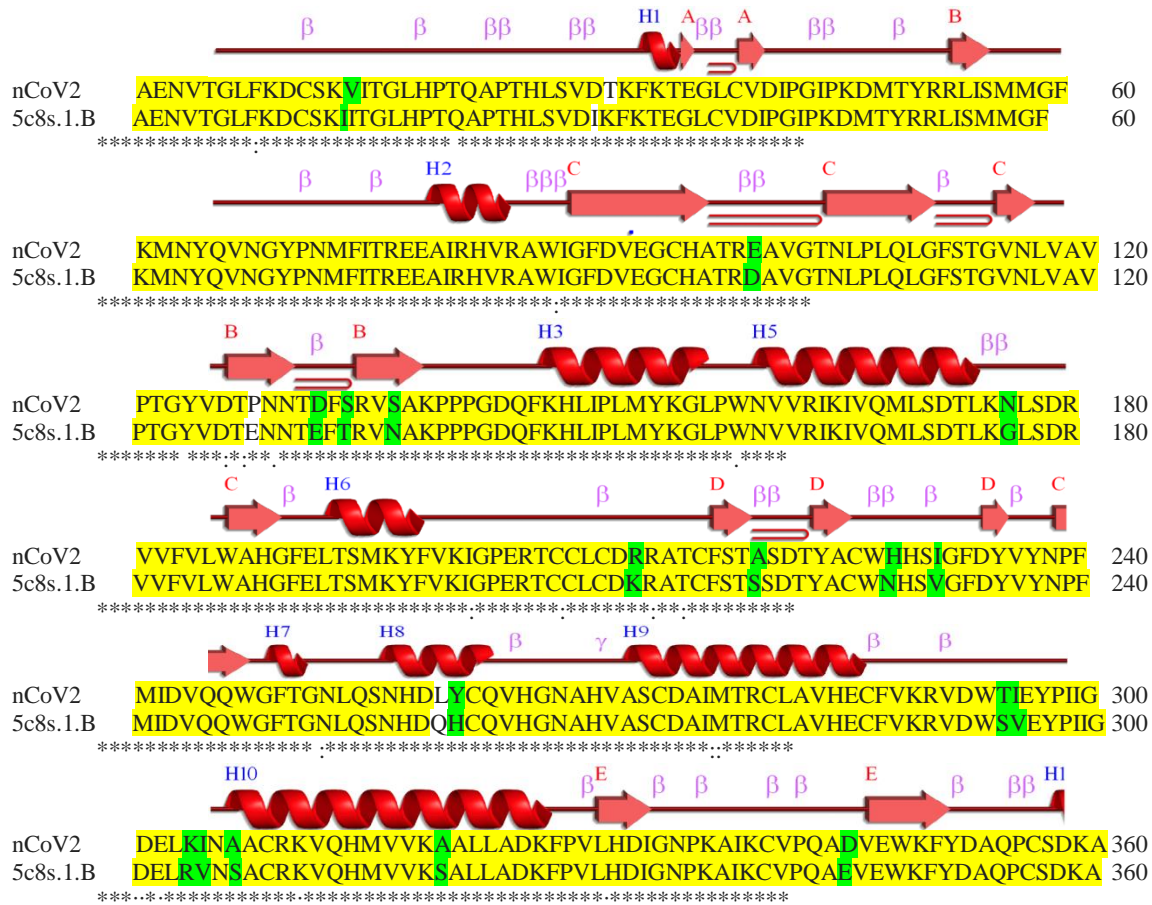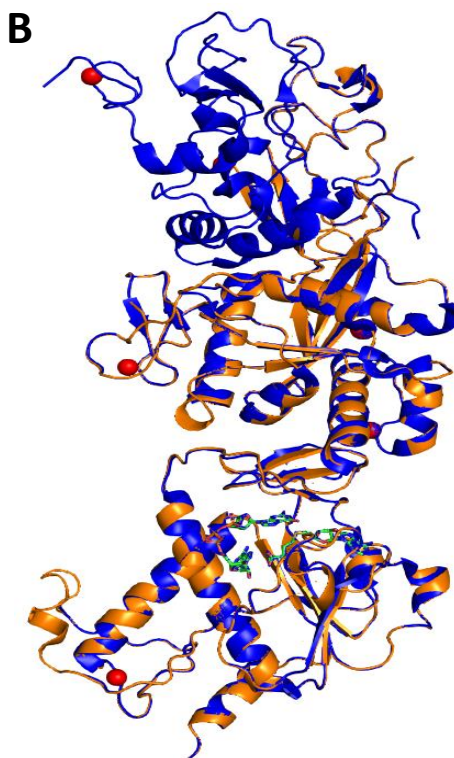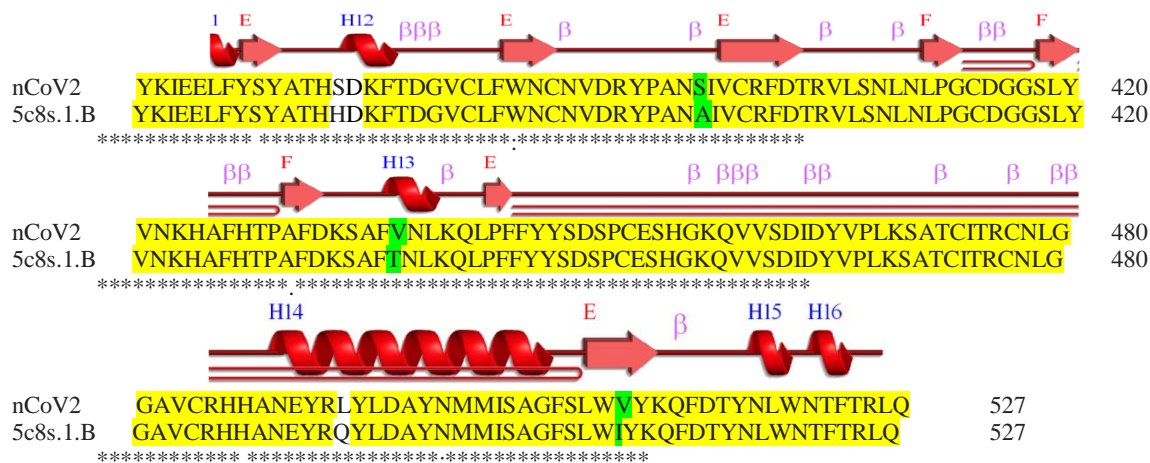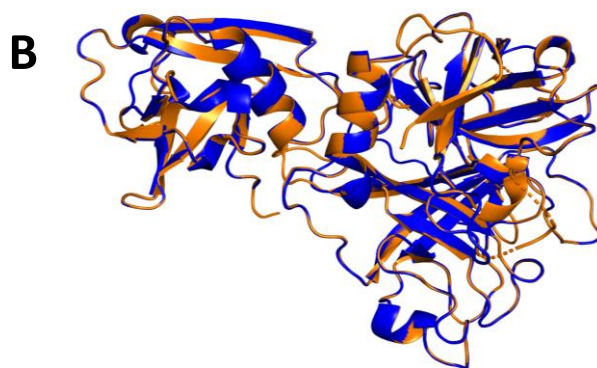
Residues in most favoured regions [A,B,L]	427	91.4%
Residues in additional allowed regions [a,b,l,p]	38	8.1%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.4%
Residues in disallowed regions	0	0.0%

Figure S13: Ramachandran plot of NSP14

nCoV2   AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF 60
5c8s.1.B AENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIPKDMTYRRLISMMGF 60
        *************.*************** ******************************

nCoV2   KMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATREAVGTNLPLQLGFSTGVNLVAV 120
5c8s.1.B KMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATRDAVGTNLPLQLGFSTGVNLVAV 120
        *************************************:*********************

nCoV2   PTGYVDTPNNTDFSRVSAKPPPGDQFKHLIPLMYKGLPWNVVRIKIVQMLSDTLKNLSDR 180
5c8s.1.B PTGYVDTENNTEFTRVNAKPPPGDQFKHLIPLMYKGLPWNVVRIKIVQMLSDTLKGLSDR 180
        ******* ***.*:.* ********************************** ****

nCoV2   VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF 240
5c8s.1.B VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWNHSVGFDYVYNPF 240
        *******************************:*******:*******:***:*******

nCoV2   MIDVQQWGFTGNLQSNHDLYCQVHGNAHVASCDAIMTRCLAVHECFVKRVDWTIEYPIIG 300
5c8s.1.B MIDVQQWGFTGNLQSNHDQHCQVHGNAHVASCDAIMTRCLAVHECFVKRVDWSVEYPIIG 300
        *****************: ********************************::******

nCoV2   DELKINAACRKVQHMVVKAALLADKFPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKA 360
5c8s.1.B DELRVNSACRKVQHMVVKSALLADKFPVLHDIGNPKAIKCVPQAEVEWKFYDAQPCSDKA 360
        ::.:********:*************************.*************

```
           1  E              H12      ββββ      E      β        β   E        β      β   β  F    ββ    F
nCoV2      YKIEELFYSYATHSDKFTDGVCLFWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLY      420
5c8s.1.B   YKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSLY      420
           ***************** **************************:*********************

           ββ    F          H13   β    E                   β   βββ    ββ         β       β   ββ
nCoV2      VNKHAFHTPAFDKSAFVNLKQLPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLG      480
5c8s.1.B   VNKHAFHTPAFDKSAFTNLKQLPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLG      480
           *************** ********************************************

                   H14                        E       β         H15   H16
nCoV2      GAVCRHHANEYRLYLDAYNMMISAGFSLWVYKQFDTYNLWNTFTRLQ      527
5c8s.1.B   GAVCRHHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQ      527
           ************ *****************:***************
```

B

Figure S14: A. Sequence alignment of homology modeled NSP14 protein with the template SARS-Cov protein (5C8S) having 95.07% of sequence identity. The entire identical region is marked with a yellow highlighted portion. The secondary structure are also predicted and attached herewith. B. Structure alignment of NSP14 protein with the template SARS-Cov protein (5C8S) with a RMSD value of 0.125Å. Here homology model is represented by blue color and template model are marked with a orange color.

Figure S15: Ramachandran plot of Transmembrane protease, serine 2 (TMPRSS2)

A

```
MODEL      NRCVRLYGPNFILQVYSSQRKSWHPVCQDDWNENYGRAACRDMGYKNNFYSSQGIVDDSG        60
5ce1.1.A   LYPVQVSSADARLMVFDKTEGTWRLLCSSRSNARVAGLSCEEMGFLRALTHSELDVRTAG  60
           *::. : **::...:*::*:  * . :*:**.:. *: * :*

MODEL      ---STSFMKLNT-SAGNVDIYKKLYHSDACSSKAVVSLRCIACGVNLNSSRQSRIVGGES       116
5ce1.1.A   ANGTSGFFCVDEGRLPHTQRLLEVISVCDCPRGRFLAAICQDCGRRKLP--VDRIVGGRD  118
           :..*:::  ::. ::   *  .:: * **.  .*****..

MODEL      ALPGAWPWQVSLHVQNVHVCGGSIITPEWIVTAAHCVEKPLNNPWHWTAFAGILRQSFMF       176
5ce1.1.A   TSLGRWPWQVSLRYDGAHLCGGSLLSGDWVLTAAHCFPERNRVLSRWRVFAGAVAQASP-  177
           : * *******:..*:*****::. *:.:*****. :. :* *** :*:

MODEL      YGAGYQVEKVISHPNYDS------KTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQP       230
5ce1.1.A   HGLQLGVQAVVYHGGYLPFRDPNSEENSNDIALVHLSSPLPLTEYIQPVCLPAAGQALVD  237
           .:*  *:*:*.*      .::*****:* **.::.:***** * *

MODEL      EQLCWISGWGATEEKGKTSEVLNAAKVLLIETQRCNSRYVYDNLITPAMICAGFLQGNVD       290
5ce1.1.A   GKICTVTGWGNTQYYGQQAGVLQEARVPIISNDVCNGADFYGNQIKPKMFCAGYPEGGID  297
           .::* ::****.*: *:.:  **. :*  **.  *.***.:** :*.:*

MODEL      SCQGDSGGPLVTSK----NNIWWLIGDTSWGSGCAKAYRPGVYGNVMVFTDWIYRQMRAD 346
5ce1.1.A   ACQGDSGGPFVCEDSISRTPRWRLCGIVSWGTGCALAQKPGVYTKVSDFREWIFQAIKTH  357
           .*********:*  ..   * * *.***:*** *.**** :* *.:**::. :::.
```

B

Figure S16: **A. Sequence alignment of homology modeled TMPRSS2 protein with the template SARS-Cov protein (5CE1) having 95.07% of sequence identity. The entire identical region is marked with a yellow highlighted portion. The secondary structure are also predicted and attached herewith. B. Structure alignment of TMPRSS2 protein with the template SARS-Cov protein (5CE1) with a RMSD value of 0.132Å. Here homology model is represented by blue color and template model are marked with a orange color.**
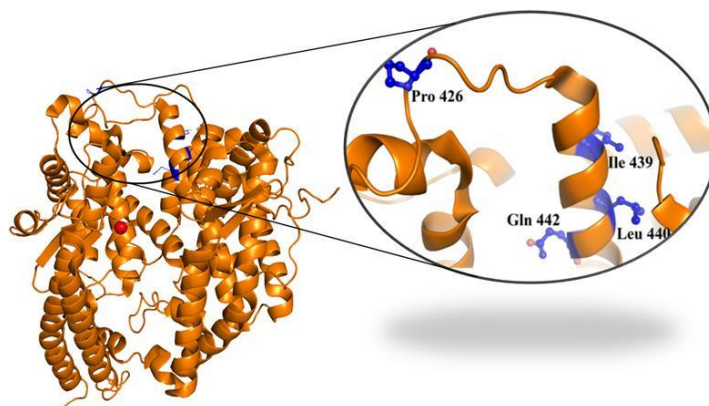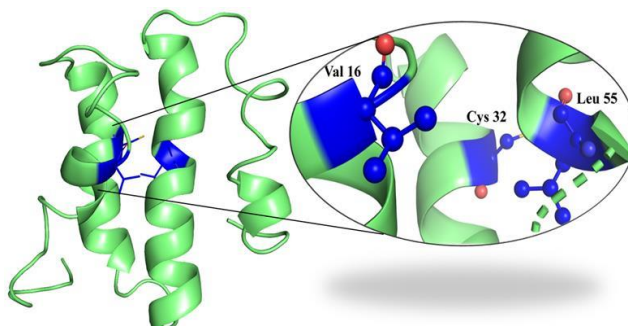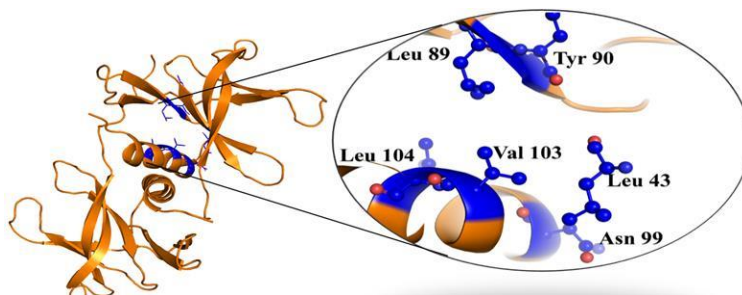

Figure S17: **Active site of ACE2 protein. Active Site residues are marked with blue elemental colored ball and sticks representations.**
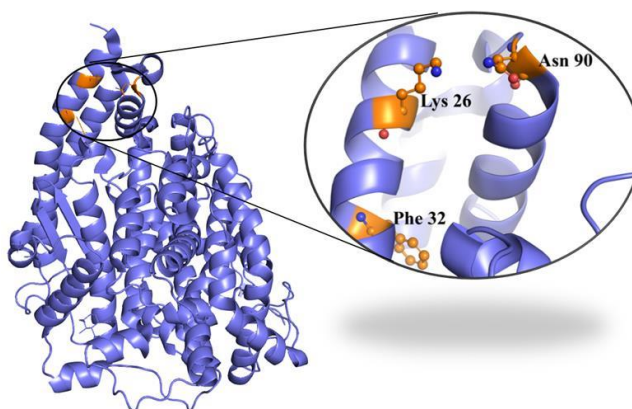
Figure S18: **Active site of spike protein. Active site residues are marked with orange elemental colored ball and sticks representations.**
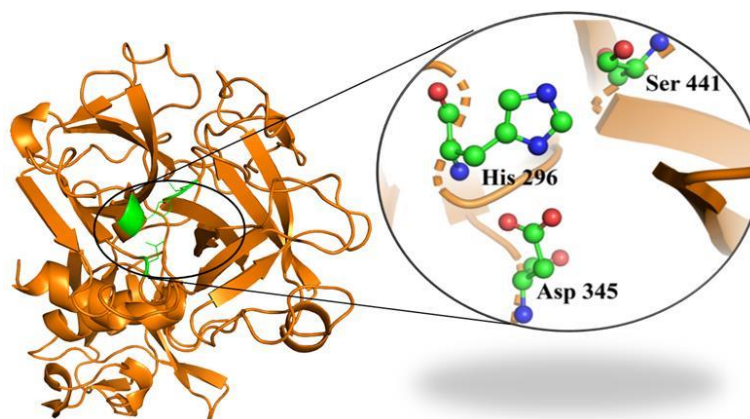
Figure S19: Active site of TMPRSS2 protein. Active site residues are marked with green elemental colored ball and sticks representations.
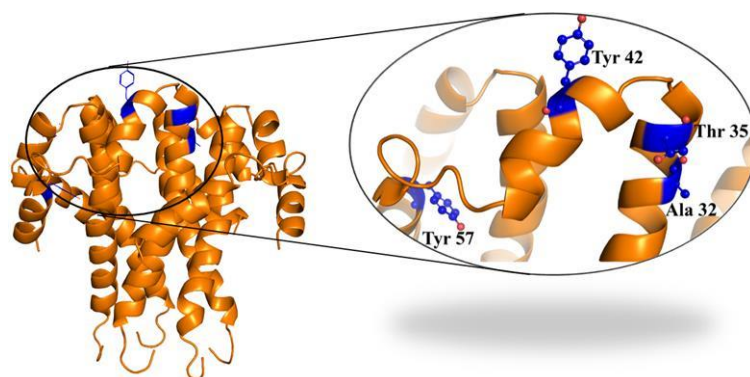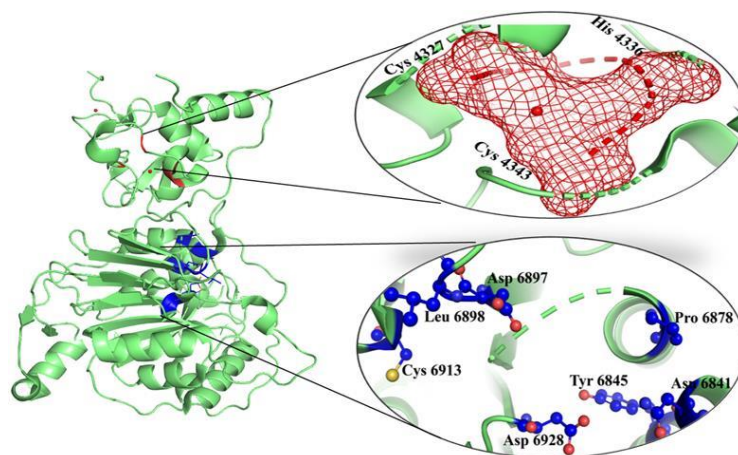

Figure S20: Active site of envelope (E) protein. Active site residues are marked with blue elemental colored ball and sticks representations.
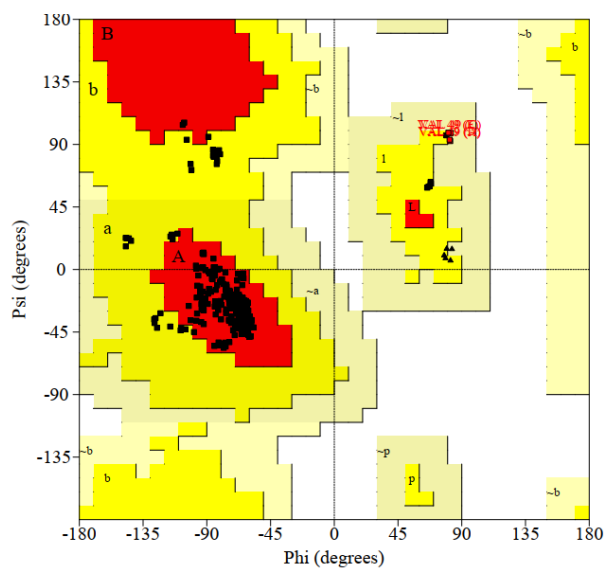
Figure S21: Ramachandran Plot of Envelop protein.
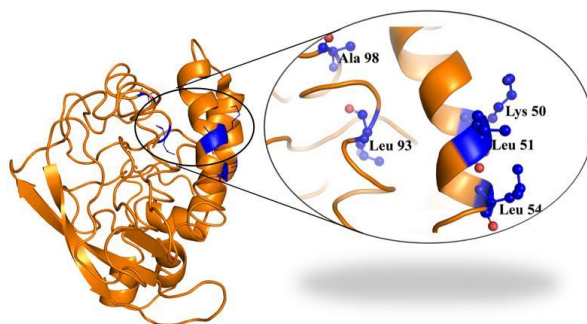
Figure S22: Active site of membrane (M) protein. Active site residues are marked with blue elemental colored ball and sticks representations.
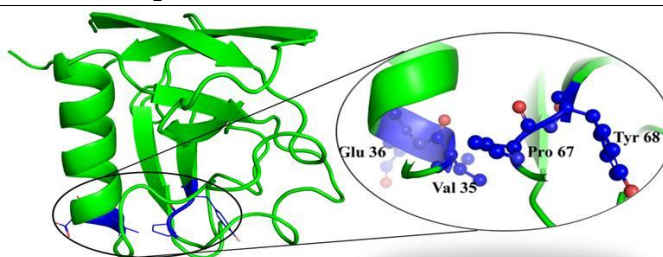
Figure S23: Active site of NSP1 protein. Active site residues are marked with blue elemental colored ball and sticks representations.
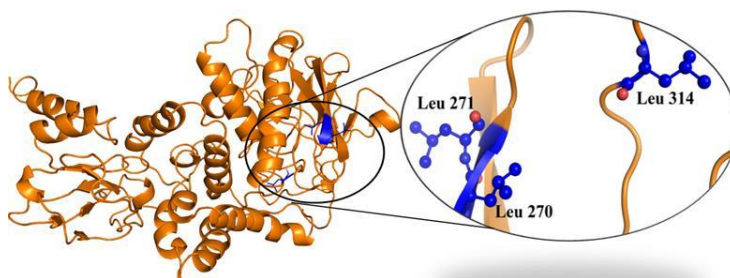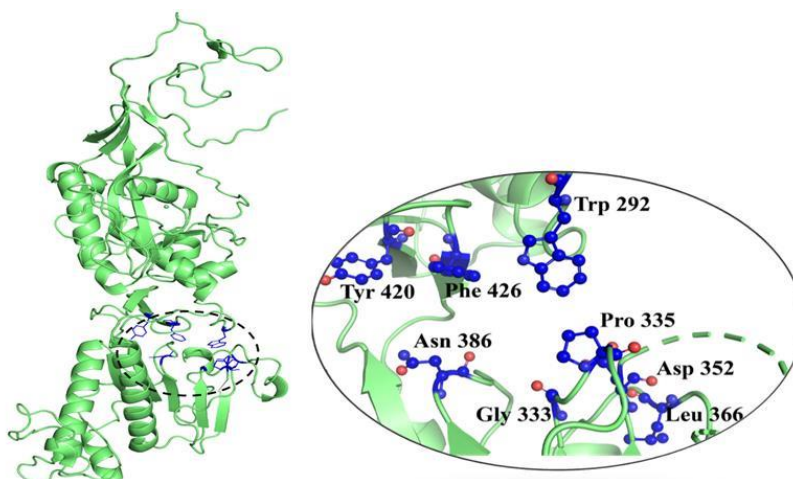

Figure S24: Active site of NSP2 protein. Active site residues are marked with blue elemental colored ball and sticks representations.
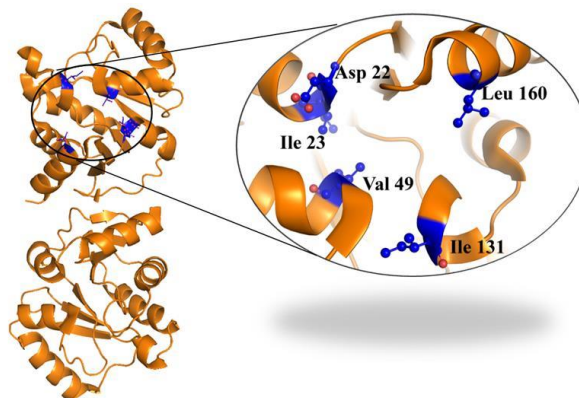
Figure S25: Active site of Papain-Like Proteases (PLpro)-NSP3 protein. Active site residues are marked with blue elemental colored ball and sticks representations.
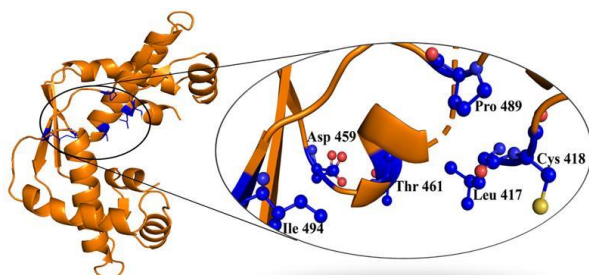
Figure S26: Active site of NSP4 protein. Active site residues are marked with blue elemental colored ball and sticks representations.
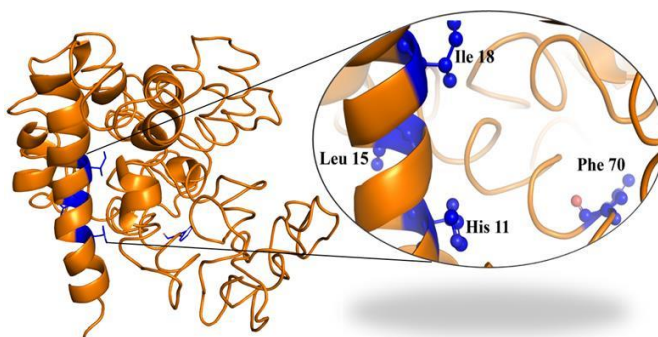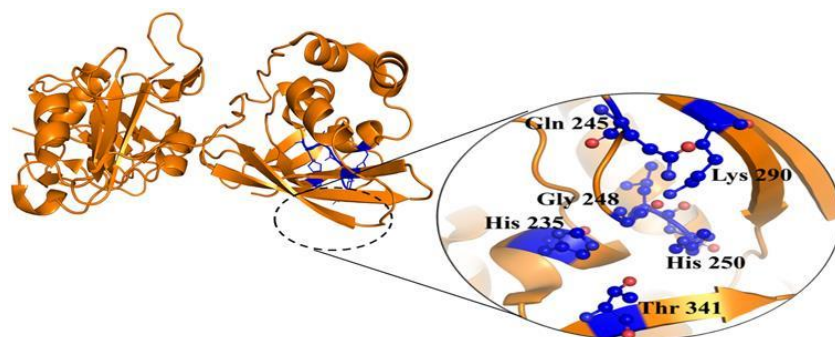

Figure S27: Active site of NSP6 protein. Active site residues are marked with blue elemental colored ball and sticks representations.

Figure S28: Active site of NSP7 protein. Active site residues are marked with blue elemental colored ball and sticks representations.

Figure S29: Active site of NSP9 protein. Active site residues are marked with blue elemental colored ball and sticks representations.

Figure S30: Active site of NSP10 protein. Active site residues are marked with blue elemental colored ball and sticks representations. Metal binding domain is represented by red mesh color map.
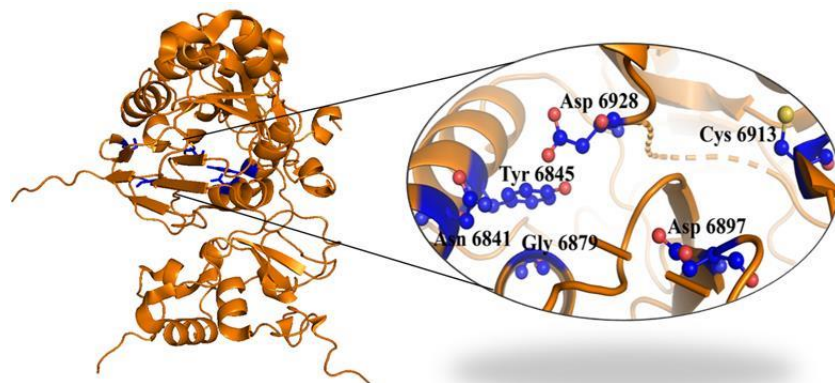

Figure S31: Active site of NSP14 protein. Active site residues are marked with blue elemental colored ball and sticks representations.

Figure S32: Active site of NSP15 protein. Active site residues are marked with blue elemental colored ball and sticks representations.

Figure S33: Active site of NSP16 protein. Active site residues are marked with blue elemental colored ball and sticks representations.