IJMPR 2023, 7(2), 51-63

International Journal of Modern Pharmaceutical Research
www.ijmpronline.com

ISSN: 2319-5878
IJMPR
Research Article
SJIF Impact Factor: 5.273

IN SILICO DRUG REPURPOSING METHODS TO COMBAT SARS-CoV-2: DISCOVERY OF POTENTIAL DRUG CANDIDATES TARGETING THE MAIN PROTEASE M^{pro}

Prakriti Seth, Namrata Chatterjee, Ushashi Chakraborty and Prof. Dr. Nandan Kumar Jana*

Department of Biotechnology, Heritage Institute of Technology Kolkata- 700107, West Bengal, India.

Received on: 06/12/2022
Revised on: 26/12/2022
Accepted on: 16/01/2023

***Corresponding Author**
Prof. Dr. Nandan Kumar Jana
Department of Biotechnology, Heritage Institute of Technology Kolkata- 700107, West Bengal, India.

ABSTRACT

Over the last 3 years, Coronavirus or COVID 19 has been a biological entity which led the entire world to a standstill. With over 630 million affected cases and 6.59 million deaths (recorded on 31stOctober 22), the world faced a dangerous pandemic situation and has foreseen millions of casualties in this period. Science and time came to rescue and vaccines were introduced in the market. Even though over time the fatal COVID-19 became a "flu-like" phenomenon its severity still remains unhindered. SARS-CoV-2 is still mutating every day and research is still going on in almost every nook and corner of the world to resist and more importantly, cure this disease. In this experiment, we performed *in-silico* drug-repurposing, i.e., virtually screening priorly approved drugs that can be also used for combating SARS-CoV-2. For this, we obtained the phylogenetic tree by comparing the whole genome sequence of SARS-CoV-2 with other closely-related viruses as noted in various scientific research. Following this, the target sequence of the main 3C-like protease ($3CL^{pro}$) of SARS-CoV-2 extracted from PDB and similar template protein from the closest virus SARS-CoV were checked by homology modelling which yielded a 3-dimensional structure with 96% sequence identity. Virtual screening of the refined target with specified grid coordinates against diverse drug libraries revealed about 100 suitable, potential drug candidates. These drugs were then made to undergo molecular docking and those drugs with better binding energy and higher affinity to attach to the active sites were selected. These were then further shortlisted on the basis of site-specific individual docking scores and the number of active site attachments. Lastly, the top three drugs were filtered from the rest in accordance with appropriate ADMET profiles and carcinogenicity parameters. Evatanepag, Telcagepant and Tasosartan were found to be safe and non-toxic drugs capable of effectively interfering with the replication of the virus.

KEYWORDS: SARS-CoV-2, COVID-19, 3C-like protease, Homology modelling, Virtual screening, Molecular docking, Druggability, Evatanepag, Telcagepant, Tasosartan.

1. INTRODUCTION

COVID 19 or Coronavirus disease 2019 is an infectious disease leading to severe respiratory problems and has led the world into a grave pandemic situation, the first confirmed case being traced back to November 2019 in Wuhan. The disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Coronaviruses are a group of related RNA viruses that cause diseases in mammals. The genus coronavirus [International Committee of Virus Database, virus number - 03.019.0.1] belongs to the sub family *Orthocoronavirinae*, in the family *Coronaviridae*, order *Nidovirales*, and realm *Riboviria*.[1] Structurally coronaviruses are roughly spherical and pleomorphic enveloped viruses with a positive sense single stranded RNA and a nucleocapsid of helical symmetry, the capsid is formed of matrix protein. The envelope bears club shaped glycoprotein projections which look like spikes protruding from the surface that are about 12 to 24 nm long.[2,3] The genome size of coronavirus ranges from 26 to 32 Kbp, one of the largest amongst RNA with the G+C content varying from 32% to 43%.[4-6] The coronaviral genes encode trimeric structural spike protein, a homodimeric cysteine proteinase, an RNA polymerase and several non-structural proteins amongst others. Some coronavirus contains a hemagglutinin-esterase protein.[7]

Genes for major structural protein are coded in 5'- 3' manner. There are six ORFs in a typical SARS-CoV genome. ORF1a and ORF1b contain a frameshift producing two polypeptides: pp1a and pp1ab. These polypeptides are processed by virally encoded chymotrypsin-like protease ($3CL^{pro}$) or main protease (M^{pro}) and one or two papain-like proteases (PL^{pro}) into

16 nsps (non-structural proteins). The four main structural proteins are spike protein(S), membrane protein (M), envelope protein (E), and nucleocapsid protein (N) which are encoded by ORFs 10, 11 near the 3′-terminus.[8-10] Besides these four main structural proteins, different CoVs encode special structural and accessory proteins, such as HE protein, 3a/b protein, and 4a/b protein. (Fig 1) All the mature structural and non-structural proteins are responsible for several important functions in genome maintenance and virus replication.[5,11-13]

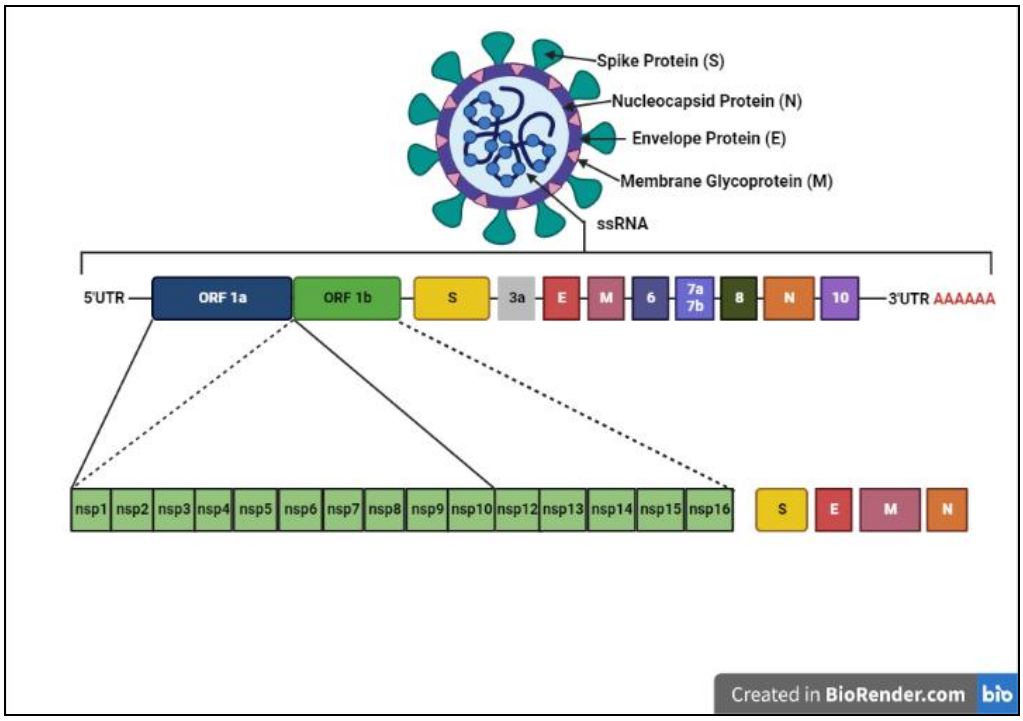

Fig. 1: Schematic presentation of the SARS-CoV-2 genome. It has a spherical structure showing an outer lipid envelope, covered with spike glycoprotein. The full-length RNA genome is made up of approximately 29,903 nucleotides and has a replicase complex composed of ORF1a and ORF1b at the 5′UTR end. The ORF1a encodes for nsp1–nsp10, while ORF1b encodes for nsp1–nsp16. Four genes that encode for the Structural proteins: Spike gene, Envelope gene, Membrane gene, Nucleocapsid gene are situated at the 3′UTR region.

The symptoms of this deadly virus as reported vary over a wide range. Most individuals affected experience mild to moderate symptoms. However older adults and individuals with pertaining medical conditions like cancer, diabetes etc are at higher risk of developing serious complications due to this. The symptoms generally develop within 2-14 days of infection. The most common symptoms are seen to be fever, dry cough and tiredness. However along with these aches and pain, sore throat, diarrhoea, loss of taste and smell are also prevalent in most cases. The serious complications lead to difficulties in breathing and shortness of breath, chest pain or pressure, loss of speech and movement. In this case immediate medical action should be taken.[6, 14-16]

SARS CoV-2 binds to ACE2 receptor (Angiotensin Converting Enzyme 2) by its spikes. After the initial entry the spike protein has to be primed by a protease. SARS CoV2 uses the same protease as SARS CoV which is TMPRSS2. In order to attach the viral receptor (spike protein) to its cellular ligand ACE2 activation by TMPRSS2 is necessary. The SARS-CoV-2 infection process starts with the viral entry mediated by the interaction of the spike (S) glycoprotein with the host angiotensin-converting enzyme 2 (ACE2) receptor and cleavage of the S protein by the host transmembrane serine protease 2 (TMPRSS2) prior to the fusion to the host cell membrane.[17-20]

There are two theories pertaining to the mechanism of entry of SARS-CoV-2. One non-endosomal pathway was initially thought to be the CoVs mechanism to enter the host cell. In 2004, it was shown that SARS-CoV fused with the cellular surface after attaching the host cell membrane. The nucleocapsids were then blurred after the virions lost their envelopes, and no endocytic-related events were described.[21] However, recent evidence points to the endosomal pathway as the main entry route for CoVs to infect the cells. The endocytic pathway has also been established as an alternative entry pathway apart from direct fusion with the plasma membrane based on their observations of SARS-CoV.[22] Here, the virus enters the cell via a pH- and receptor-mediated endocytosis-dependent manner. After entry the S protein is cleaved by different proteases. Protease activation of the spike protein is necessary for further infection by the virus. Once the virus enters the host cell, it gets disassembled to release the nucleocapsid and the viral

genome. Host ribosomes translate the open reading frame (ORF) 1a/b into two polyproteins (pp1a and pp1ab) that encode 16 nsps. Both the proteases, the main protease (3CLpro, nsp5) and the papain-like protease (PL^{pro}, nsp3) play the role in the cleavage of the polyproteins, to produce nsp2–16. Some of those are the RNA-dependent RNA polymerase (RdRp, nsp12) and helicase (nsp13).In coronavirus, this process is followed by assembly of the virion components into the endoplasmic reticulum Golgi intermediate compartment complex and release from the infected cells by exocytosis.[23-26] (Fig 2).

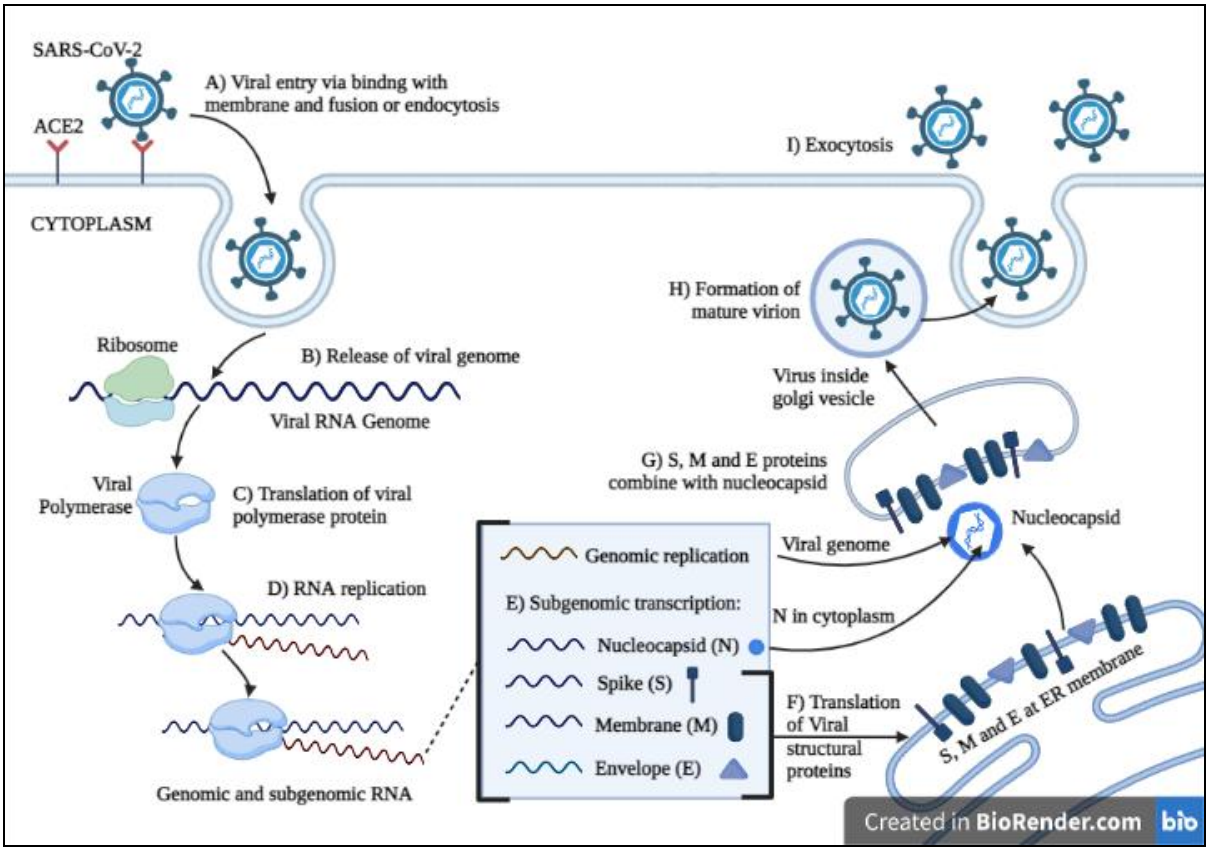

Fig. 2: SARS-CoV-2 infection cycle. (A) Entry of SARS-CoV-2 in target cell expressing ACE2 or CD147. (B) Uncoating and releasing SARS-CoV-2 single stranded RNA genome. (C) Translation of the replicase–transcriptase complex from RNA genome. (D) RNA genome replication via negative template. (E) Nested production of subgenomic RNA encoding for structural proteins. (F) Translation of viral S, E and M that gets inserted in endoplasmic reticulum. (G) Nucleocapsid coupled to the genome, forming nucleoprotein, combining to S, E and M, (H) forming a mature virion. (I) Exocytosis of the mature SARS-CoV-2.

Amongst the structural proteins that form the viral particle Spike(S) Glycoprotein and nucleocapsid(N) protein are the possible drug targets for therapeutic intervention.[27] Spike (S) protein has an important role in virus pathogenesis and organ tropism, being responsible for the viral entry through receptor recognition and membrane fusion while the nucleocapsid (N) protein is a multifunctional protein that packages the viral RNA genome into a ribonucleoprotein complex called nucleocapsid to protect the genome.[27,28,31] Inhibition of these targets will nip the virus in the early stages of their life cycle and prevent further infection. Various host based druggable targets can be exploited by using a possible drug that will prevent the binding of the virus. Such targets are Angiotensin I Converting Enzyme 2 (ACE2) and Transmembrane Serine Protease 2 (TMPRSS2) as ACE2 is the main host receptor of the virus, while TMPRSS2 is required for proteolytic activation of the S protein which is necessary for binding to ACE2 receptor.[29-32]

In this study we exploit 3 chymotrypsin like protease (3CLpro) also known as Main Protease (Mpro) as a potent target for inhibition using virtual screening and docking as an approach to find feasible drugs. It can be considered as an attractive drug target as it has an essential role in processing polyproteins translated from viral DNA.[33] 3CLpro forms a dimer, and each monomer contains two regions, the N-terminal catalytic region and the C-terminal region. The Mpro operates at 11 cleavage sites on the polyprotein PP1AB thus facilitating viral replication. Inhibition of this enzymatic activity will block viral replication.[34,35]The sequences of 3CLpro in SARS-CoV and SARSCoV-2 share 96% identity, and the minimal differences between the two enzymes are at the surface of the proteins. Therefore, inhibitors against SARS-CoV 3CLpro are expected to inhibit SARS-CoV-

2 3CLpro. As no human protease has the same cleavage site the drugs are unlikely to be toxic to the individuals.[36-38]Given the high rate of transmission of this virus between humans and its pandemics, it is important to identify the basis of its replication, structure, and pathogenicity for discovering a way to the special treatment or the prevention. Also, identifying the specific molecular details of the virus is helpful in achieving treatment goals.

2. MATERIALS AND METHODS

2.1. Genome sequence retrieval and generation of Phylogenetic Tree

NCBI (National Center for Biotechnology Information) provides a large suite of online resources for biological information and data.[39] Similarly PDB is a database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids. The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.[40.41] Clustal Omega on the other hand is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences.[42,43]

The fully annotated genome sequence of 2019-nCoV or SARS-CoV-2 was retrieved from NCBI (GenBank: MN908947.3) and subsequent Multiple Sequence Alignment was done using user based templates of the genome sequences of Bat-SL ZXC 21 (GenBank: MG772934.1), SARS-CoV (GenBank: MT240479.1), MERS-CoV (GenBank: MG923480.1), HKU-1 (GenBank: KF686344.1) and MHV-A59 (GenBank: MF618252.1) and a phylogenetic tree was generated. The closest resembling species with SARS-CoV-2 was found to be SARS-CoV. Then the main protease amino acid sequence of SARS-CoV-2 was retrieved from PDB (PDB ID: 6M2Q) and along the main protease amino acid sequence of SARS-CoV was retrieved (PDB ID: 2C3S).

2.2. Homology Modelling using SARS CoV-2 M^{pro} as target

Swiss Model is a structural bioinformatics web-server dedicated to homology modeling of 3D protein structures.[44] We used Swiss Model to generate a 3D structure of our target protein, the main protease of 2019-nCoV (6M2Q) using the main protease of SARS-CoV (2C3S) as the template as it was the closest related species of the novel Coronavirus. The 3D structure obtained had 96.08% structural similarity along with a Qmean score of 0.2. The Q-Mean Qualitative Model Energy ANalysis, is a composite scoring function assessing the major geometrical aspects of protein structures. It is able to derive both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates on the basis of one single model.[45] A Q-Mean score of 0.2 signified the generation of a pretty good 3D structure of our target protein.

2.3. Refinement of obtained 3D structure and Visualisation of Alignment

ModRefiner algorithm for atomic level protein structure refinement. It brings the generated protein model nearer to its native state in terms of hydrogen bonds, backbone topology and side chain positioning.[46] The PDB sequence of the generated 3D structure of the target protein (Main protease of SARS-CoV-2) was given as input and structure refinement was done using ModRefiner. PyMol is one of the few open-source visualization tools available for use in structural biology.[47] The refined structure was visualised in PyMol and alignment was done against the already available crystal structure of the main protease of SARS-CoV-2.

2.4. Virtual Screening using software curated Drug Library

Virtual screening can be defined as a set of computational methods that analyzes large databases or collections of compounds in order to identify potential hit candidates. MTiOpenScreenis an online server that performs virtual screening against some software curated databases to screen for potential drug candidates.[48] The refined structure of the main protease was given as input was used for screening against its library of 7173 purchasable drugs (Drugs-lib), with 4574 unique compounds and their stereoisomers. Each library entry is identified with the name of the compound as well as a PubChem ID.

The virtual screening was conducted in two phases, once the screening 'Mode' was selected as ‘List of Residues', specified as: H41, M49, G143, S144, C145, H163, H164, M165, E166, L167, D187, R188, Q189, T190, A191 and Q192. The next time the 'Mode' was selected as 'Grid Coordinates' and the grid centre points were set at X = -24.9165, Y = 12.3378, Z = 57.4364 and box dimensions were set as 48.30 Å × 69.97 Å × 60.87 Å with exhaustiveness 8.[49,50] Each screening yielded 1500 compounds from which top 100 compounds (in total 200 compounds) were chosen based on affinity energy.

2.5 Active site generation and Re-Screening of top 100 candidates

Before re-screening, the top 100 compounds (sets of two making a total of 200 compounds), active sites of the target protein were generated. Active sites available from literature review[49]were used as well as a software was used to predict the same. 3DLigandSite is an automated method for the prediction of ligand binding sites.[51] Either the .pdb file of the concerned protein or FASTA sequence can be given as input to generate active ligand sites. Once the active sites were generated which took about an hour or so, the top 200 compounds (100x2) were divided into 3 groups of 70, 70 and 60 compounds. The acute toxic and hazardous compounds were screened out and the remaining elements were downloaded from PubChem.[52] 3D structures of the compounds were downloaded but some of the compounds yielded only 2D

structures. For them, the 2D structures were converted into respective 3D structures with the help of Open Babel software.[53,54]

The 3D structures of the compounds were loaded in PyRx for site-specific docking. PyRx is a virtual screening software for Computational Drug Discovery that can be used to screen potential compounds against potential drug targets.[55] Here, the protein target was the generated 3D model of the main protease of SARS-Cov-2 and the drug compounds were the 200 top compounds obtained from the two virtual screenings performed previously in MTiOpenScreen. Blind docking in PyRx was done in three phases with 70, 70 and 60 compounds respectively. The active sites generated from 3DLigandSite were set and a customised docking grid was generated. Energy minimization of the drugs was done by conjugate-gradient method.[56] Finally docking was performed.

2.6. Selection of top 10 candidates and visualisation of active site binding

After docking was completed in PyRx, the ligand-protein interactions which gave the highest binding affinity were visualised in PyMol to determine the site of binding. The binding sites were noted and the active sites out of those binding sites were highlighted. The compounds with the highest number of active site binding with the target were screened and the top 10 compounds were determined.

2.7. Selection of top 5 candidates and Site-Specific Docking

The top 10 compounds were individually docked once again in PyRx by the method of site specific docking mentioning all the active sites and generation of customised grid. After individual site-specific docking the binding energy alterations were noted and accordingly top 5 candidates were chosen for toxicity analysis.

2.8. Analysis of Druggability and Toxicity

The top 5 compounds were analysed for toxicity and druggability to render the best possible and safest candidates. PkCSM and Swiss ADME are two independent softwares that test the toxicity and druggability of the candidate compounds.[57,58] Druggability map was generated with the help of Swiss ADME and on the other hand ADMET analysis (Absorption, Distribution, Metabolism, Excretion, Toxicity) was in done with the help of PkCSM.[59]

Finally the safest 3 potential candidates were chosen as probable drug targets against the concerned protein target.

3. RESULTS AND DISCUSSION

3.1. Genome sequence retrieval and generation of phylogenetic tree

3.1.1. Genome Retrieval

The fully annotated genome sequence of SARS-CoV-2 viral strain, 305 amino acid long(given below) was retrieved from PDB. The PDB ID of the same is 6M2Q (Fig. 3) and has a weight of 33.83 KDa.

Fig. 3: Crystal structure of target protease of SARS-CoV-2, retrieved from PDB (PDB ID: 6M2Q).

>protein sequence
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDV
VYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGN
VQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQP
GQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGN
FYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDR
WFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLS
AQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEF
TPFDVVRQCSGVTFQ

The experimental data of the protein model procured from X-ray diffraction methods shows a resolution of 1.70 A and a R-value free of 0.204. The R-value measures how well the simulated diffraction patterns match the experimental diffraction pattern.[60] The typical R-value free is around 0.2 for a good model, thus the model considered is a good model.

3.1.2. Generation of Phylogenetic Tree

Phylogeny represents the evolutionary relationships between a set of organisms or a group of organisms.[61] Since SARS-CoV-2 is a beta coronavirus it has similarities between the other coronavirus strains of the same group like SARS CoV, MERS etc, with the help multiple sequence alignment using CLUSTAL OMEGA, we obtained a evolutionary tree.[62]

From the results obtained SARS-CoV-2 has the closest relationship with SARS-CoV. (Fig. 4) Thus it can be considered that these two strains of coronavirus share evolutionary relations with each other and our required strain shares a sequence identity with SARS-CoV. Thus we have used SARS-CoV as a template for the further steps of homology modelling.

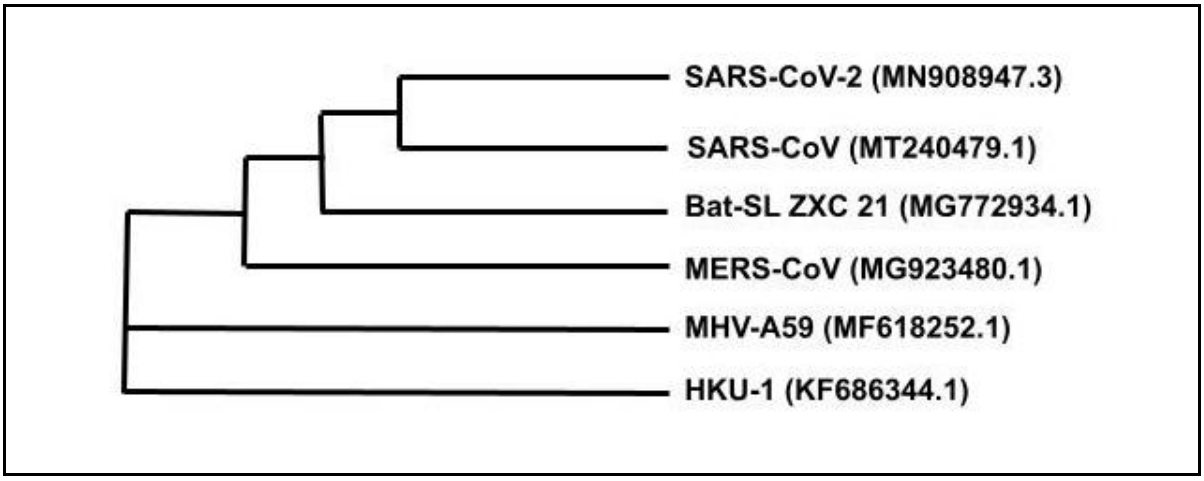

Fig. 4: Phylogenetic tree of SARS-CoV-2 along with other strains of Beta Coronavirus showing closest evolutionary relationship between SARS-CoV-2 and that of SARS-CoV; respective GenBank ID of the genome sequences as retrieved from NCBI are mentioned in brackets.

3.2. Homology Modelling of the Main Protease of SARS-CoV-2

The FASTA sequence of the target structure of SARS-CoV-2 3CL main protease was extracted from PDB (PDB ID: 62MQ). The raw sequences were given in FASTA format while the tertiary structure was retrieved in PDB format. Homology modelling of the target against the related homologous protein template of SARS-CoV (PDB ID: 2C3S) revealed a 3-dimensional structure with 96.08% sequence identity with the template protease. The obtained model, existing in the oligo-state of a monomer and with no attached ligand, was found to have a Q-MEAN score of 0.2. (Fig. 5).

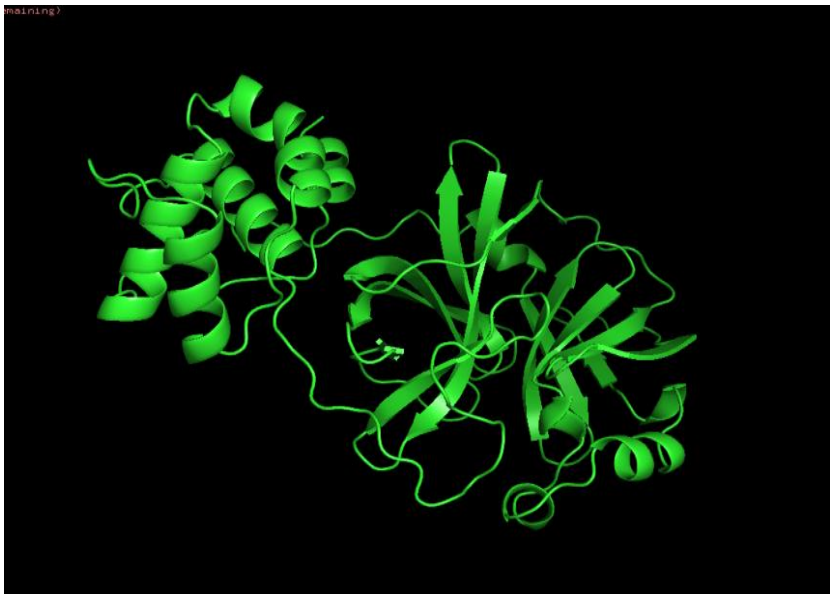

Fig. 5: Tertiary Structure of Main Protease of SARS-CoV-2 as predicted by SWISS MODEL against the closest homologous protein template of SARS-CoV (PDB ID: 2C3S), showing a 3D structure with 96.08% sequence identity with the template protease and having a Q-MEAN score of 0.2.

3.3. Refinement of obtained 3D structure and Visualisation of Alignment

ModRefiner structurally refined the obtained 3D structure and the alignment of the generated 3D model with the template protein structure (PDB ID:2C3S) was visualised in PyMol. Almost all the major grooves and helices were found out to be perfectly aligned with the target protein structure generated. RMSD score i.e the Root Mean Score Deviation is the measure of the average distance between the atoms of the backbone

structure and that of the imposed atoms.[63] An RMSD score between 3 and 4 is considered to be good. The RMSD score was 0.520 in this case which proves it to be an excellent alignment. (Fig 6).

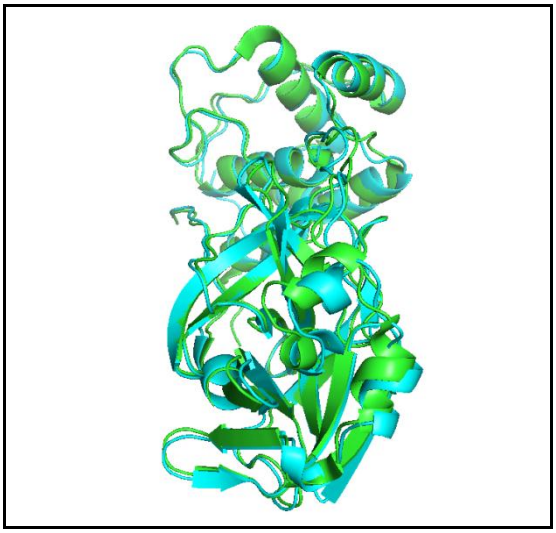

Fig. 6: Alignment of target protein of SARS-CoV-2 almost overlapping with the template protein of SARS-CoV showing almost all the major grooves and helices in perfect alignment having a RMSD score of 0.520, predicted by PyMOL (target protein PDB ID: 6M2Q in blue and template protein PDB ID: 2C3S in green).

3.4 Virtual Screening using MTi Open Screen

The library of purchasable drug inhibitors (DrugLib) against the main protease or 3CL protease of SARS-CoV-2 was screened with the help of MTi Open Screen. The screening procedure was done twice using two different sets of parameters, grid coordinates (mentioned earlier) and active site residues-H41, M49, G143, S144, C145, H163, H164, M165, E166, L167, D187, R188, Q189, T190, A191 and Q192, yielding about 1500 drugs in each case. Out of each 1500 compounds top 100 was selected from both the set of inhibitory drugs obtained giving 200 top drugs in total with desired affinities towards the residues. The top 200 drugs were then divided into sets of 70,70,60 drugs for ease of handling and were further docked in PyRx.

3.5. Active Site Generation and Rescreening using PyRx

3.5.1. Active Site Generation

The active site residues of SARS-CoV-2 was determined using 3D Ligand Site server. Besides the residues obtained from the server, the ones obtained from various literature papers were looked into and all the residues were considered for docking and visualization of binding of the drugs to the active sites of the modelled SARS-CoV-2 protease. The sites obtained from 3D Ligand site are: **ARG4, LYS5, MET6, ALA7, PHE8, PRO9, SER10, ASP34, HIS41, LYS90, THR93, PHE112, SER113, VAL125, GLN127, PHE150, ASN151, ILU152, HIS164.** The sites compiled from various literature sources are**: THR24, THR25, THR26, LEU27, CYS44, MET49, TYR54, HIS64, LEU67, ASN72, GLN74, ARG76, ASP92, PHE140, LEU141, ASN142, GLY143, SER144, CYS145, VAL157, HIS163, MET165, GLU166, LEU167, PRO168, HIS172, PHE185, ASP187, GLN189, THR190, ALA191**. All these active sites were considered in further steps.

3.5.2. Rescreening Using Pyrx

The top 200 drugs divided into three sets of 70,70 and 60 were individually screened and docked using PyRx. The elements were downloaded from PubChem server, eliminating the ones that are classified as health hazard or acute toxic. Energy minimization of the drugs was done by conjugate-gradient method.[56] After the software was run we obtained top 10 docked drug targets from each set on the basis of binding affinity. Higher the binding affinity greater is the attractive force between the drug and the target.[64] The top 30 drugs were considered within a binding affinity range of -7.6 to -8.5 kJ/mol.

3.6 Visualisation in Pymol and Selection of 10 drug compounds

Each of the 30 protein-ligand interactions were viewed in Pymol individually and the active sites of the target to which they bind were observed. All the polar and non-polar bonds were considered. On the basis of number of active sites binding in case of a particular drug compound and binding affinity top 10 drugs were chosen. (Table 1).

Table 1: Top 10 compounds obtained after site-specific virtual screening in PyRx and visualisation in PyMol on the basis of binding affinity and higher number of active-site interaction.

Sr. No.	Compound name	PubChem C-ID	Binding affinity	Active sites bound with
01.	Telcagepant	11319053	-8.7	GLU166
02.	Hesperidin	93473245	-8.5	LYS5, ALA7,GLN127
03.	Enmd-2076	16041424	-8.3	LYS5
04.	Tasosartan	60919	-7.8	ARG4, LYS5
05.	Etopside Phosphate	6818092	-7.8	THR26, CYS145, LEU167
06.	Azaftozine	170360	-7.7	THR24, THR25,THR26, HIS41, CYS44, GLU143, SER144, GLU166, MET165
07.	Evatanepag	9890801	-7.6	GLU166, ASN142
08.	Raloxifene	5035	-7.3	THR25,CYS144, HIS163
09.	Polydatin	5281718	-7.3	HIS41,CYS144,GLU166
10.	Peliglitazar	6451147	-5.9	GLN127,ALA7, ARG4, TRY126

3.7. Selection of top 5 compounds

The top 10 compounds are subjected to individual docking in PyRx which led to alterations in energy and binding affinity in some compounds. Againconsidering the parameter of number of active binding sites and binding affinity top 5 compounds are finally considered. (Table 2) (Fig 7).

Table 2: Top 5 compounds obtained after individual site-specific docking in PyRx and visualisation in PyMol on the basis of binding affinity and higher number of active-site interaction.

Sr.No.	Compound name	PubChem C-ID	Binding affinity	Active sites bound with
01.	Telcagepant	11319053	-8.7	GLU166
02.	Hesperidin	93473245	-8.5	LYS5, ALA7,GLN127
03.	Tasosartan	60919	-7.8	ARG4, LYS5
04.	Azaftozine	170360	-7.7	THR24, THR25,THR26, HIS41, CYS44, GLU143, SER144, GLU166, MET165
05.	Evantanepag	9890801	-7.2	GLU166, ASN142

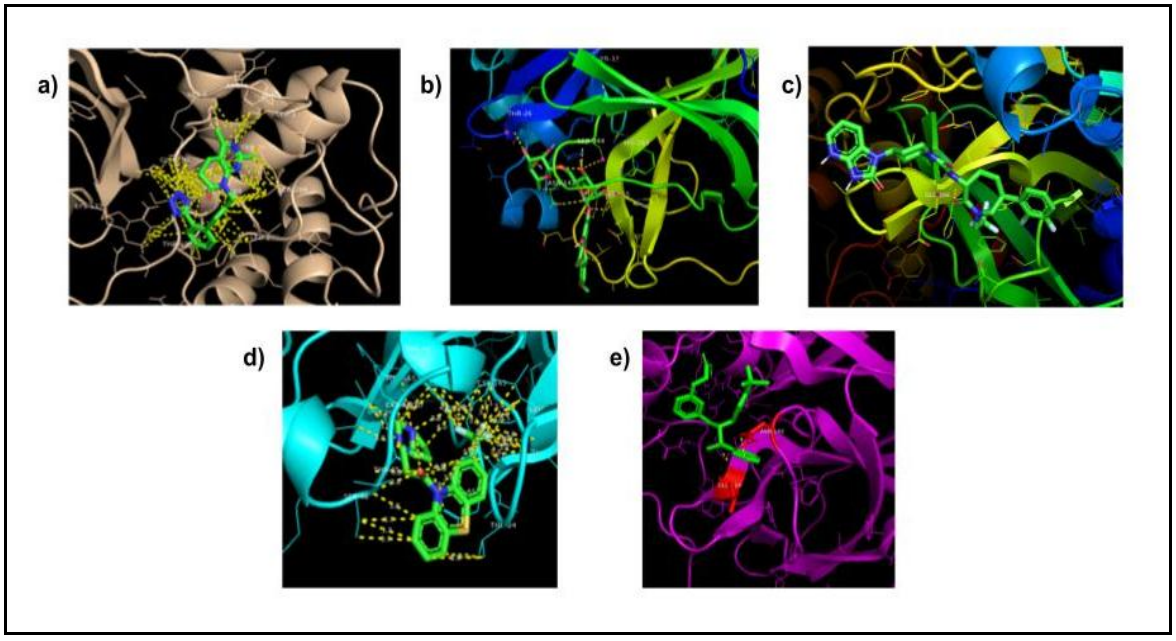

Fig. 7: Interaction between top 5 drug candidates namely a) Tasosartan, b) Hesperidin, c) Telcagepant, d) Azaftozine, e) Evatanepag with target protein after performing site-specific docking in PyRx and visualisation in PyMOL, showing the bonds formed between the drug candidates and the amino acid residues of the protease.

3.8 Analysis of druggability and toxicity

The top 5 compounds obtained after individual docking and energy and affinity alterations were subjected to druggability and toxicity tests using Swiss ADME and pKCSM.Amongst various parameters obtained from both the softwares, the final selection was made based on the following criterias- Bioavailability, LD50, AMES toxicity, GI absorption, P-glycoprotein substrate, CaCO2 permeability, Lipinski Rule of 5.

Lipinski's rule of 5(Lo5) evaluates drug likeness based on 5 criteria- Molecular mass less than 500 Dalton, octanol-water partition coefficient (expressed as LogP less than 5), less than 5 hydrogen bond donors (HBD), less than 10 hydrogen bond acceptors (HBA), molar refractivity (MR) should be between 40-130.[65] Ames Toxicity evaluates whether the drug is mutagenic thus can be carcinogenic to the human body.[66] LD50 or Lethal dose is a measure of lethal toxicity of a given substance.[67] GI absorption gives the measure of the amount of substance absorbed from the gastrointestinal tract to exert a toxic effect throughout the body, thus for a drug to be a good drug its GI absorption should be low.[68,69] In pharmacology, bioavailability is the fraction of administered drug that reaches systemic circulation, greater the bioavailability more favourable the drug is.[70] CaCO2 permeability assay is an established method that measures the rate of flux of a compound across polarised CaCO2 cell monolayers and can predict in vivo absorption of drugs.[71] Efflux transporters such as P-glycoprotein play an important role in drug transport in many organs.Drugs which induce P-glycoprotein can reduce the bioavailability of some other drugs. Inhibitors of P-glycoprotein increase the bioavailability of susceptible drugs.[72,73]

Hesperidin and Azaftozine were rejected since they were found to violate the required criteria. It was noted that Hesperidin showed 3 Lipinski violations while Azaftozinepotrayed Ames toxicity i.e it can be carcinogenic. According to the given parameters the other three drugs- Evatanepag, Tasosartan and Telcagepant were more or less found to have suitable values of the aforementioned parameters. The analysis of the drugs given in Table 3 showed all of the three drugs were non carcinogenic with moderate GI absorption and no Lipinski violation in case of Evantepag and Tasosartan and only 1 violation (MW>500) for Telcagepant. The moderate P-glycoprotein substrate propensity for Ligand 6 suggested that the compounds possess adequate chances of having efficient drug effluxes. The table also revealed that all the three hits have a good bioavailability score as well as good to moderate LD50 and CaCO2 permeability, thus rendering the drugs as possible targets to bind to the active sites of 3CL protease or Main Protease of SARS-CoV-2. Evantepag, Tasosartan and Telcagepant were observed to be novel finds of the team members; with therapeutic and binding properties, capable of qualifying as potential anti-COVID drugs in clinical trials. (Fig. 8a, Fig. 8b).

Table 3: Druggability and toxicity analysis mentioning the parameters that were considered for screening the top 5 candidates.

Sr. No.	Compound name	Bio-availability	LD50 (mol/kg)	AMES	Poly-glycoprotein substrate	GI-absorption (% absorbed)	Lo5	CaCO2 permeability (logP in 10^{-6} cm/s)
01.	Telcagepant	0.55	2.558	NO	YES	75.425	NO(1)	0.036
02.	Hesperidin	0.17	2.506	NO	YES	31.481	NO(3)	0.505
03.	Tasosartan	0.56	2.791	NO	YES	81.38	YES	0.166
04.	Azaftozine	0.55	3.063	YES	YES	91.524	YES	1.028
05.	Evantanepag	0.56	3.089	NO	NO	70.373	YES	0.6

Note: The no of Lipinski's rules that have been violated are written in brackets.

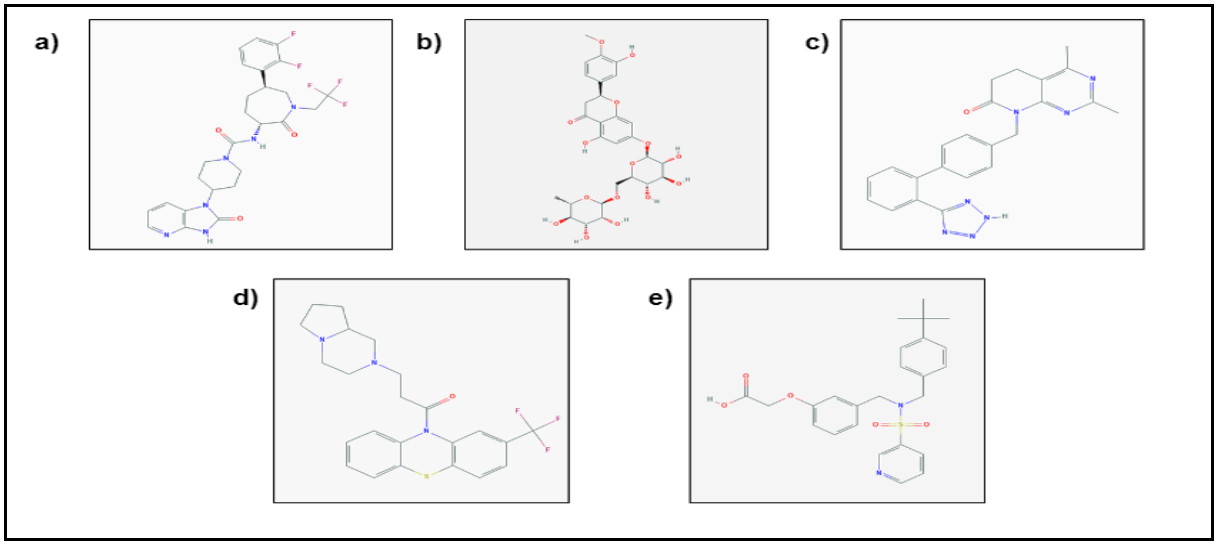

Fig. 8a: 2D chemical structures of the top 5 drug candidates namely a) Telcagepant, b) Hesperidin, c) Tasosartan, d) Azaftozine, e) Evatanepag obtained from PubChem.

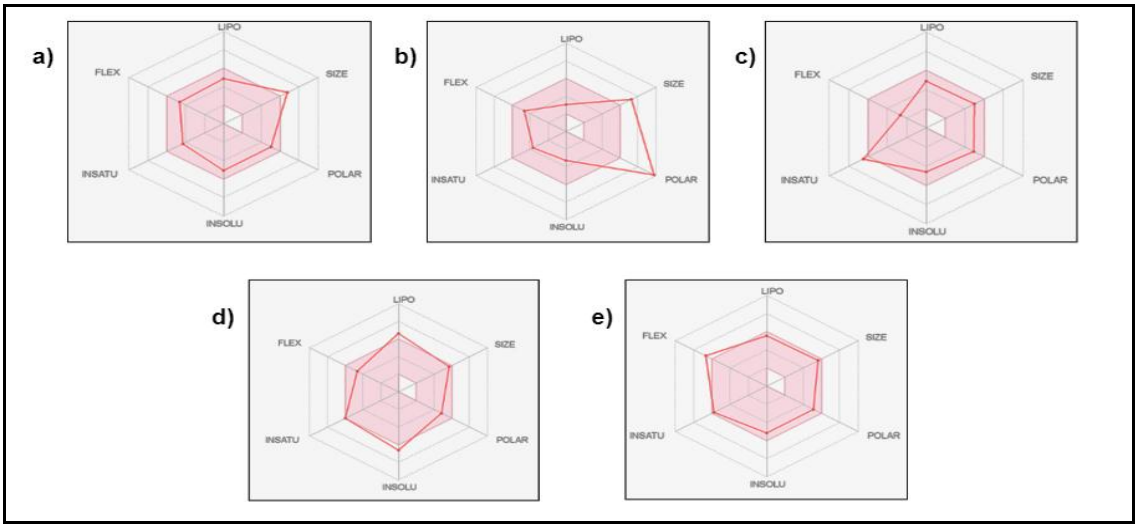

Fig. 8b: Druggability map of the top 5 drug candidates namely a) Telcagepant, b) Hesperidin, c) Tasosartan, d) Azaftozine, e) Evatanepagobtained from Swiss-ADME, explained in section 3.8.

4. CONCLUSION

In brief, to predict the 3-dimensional tertiary structural model of our target sequence of SARS-CoV-2 $3CL^{pro}$, homology modelling was performed against the template of related proteinase of SARS-CoV. Then the sequence of the obtained model was made to undergo refinement in ModRefiner and the alignment of the target against the template structure was visualised in PyMol. The sequence of the target was then loaded into the open virtual screening software MTi Open Screen and virtual screening was performed against purchasable drug library (DrugLib), specifying the active sites and grid dimensions in two phases. The screening results yielded top 200 compounds which were further divided into groups of 70, 70 and 60 compounds which were virtually screened in PyRx. Apart from active sites obtained from literature, specific active sites were generated in 3DLigandSite. Each operation in PyRx yielded top 10 compounds based on affinity energy. The top 30 compounds were visualised in PyMol and according to the number of bound active sites the top 10 compounds were selected. The top 10 compounds were then individually docked in PyRx against our target protein. This resulted in a few alterations in the binding energy which were noted and re-visualised in PyMol. After analysis the best 5 compounds were selected. The top 5 drug candidates were subjected to druggability and toxicity analysis in Swiss ADME and PkCSM respectively. The hazardous and carcinogenic compounds were screened off and the best 3 compounds were ultimately sorted out.

Finally it can be concluded that more than one month of work resulted in the determination of three FDA approved drugs namely Evatanepag, Tasosartan and Telcagepant as potential drug candidates that can be repurposed further for extensive use and ultimately give fruitful results in treatment against the ever spreading COVID19.

5. ACKNOWLEDGEMENT

The authors acknowledge the infrastructural facilities provided by Heritage Institute of Technology, Kolkata.

6. CONFLICT OF INTEREST

The authors declare no conflict of interest.

7. REFERENCES

1. Pal M, Berhanu G, Desalegn C, Kandi V. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. Cureus, 2020; 12(3): e7423.
2. Liu J, Liao X, Qian S, et al. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, Emerg Infect Dis., 2020; 26(6).
3. Malik YA. Properties of Coronavirus and SARS-CoV-2. Malays J Pathol, 2020; 42(1): 3-11.
4. Neuman BW, Adair BD, Yoshioka C, et al. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. J. Virol, 2006; 80(16): 7918-28.
5. Mousavizadeh L, Ghasemi S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. J Microbiol Immunol Infect, 2021; 54(2): 159-163.
6. Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, Atif SM, Hariprasad G, Hasan GM, Hassan MI. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. BiochimBiophys Acta Mol Basis Dis., 2020; 1866(10): 165878.
7. Fung TS, Liu DX. Post-translational modifications of coronavirus proteins: roles and function. Future Virol, 2018; 13(6): 405-430.
8. Wu Cr, Yin Wc, Jiang Y, et al. Structure genomics of SARS-CoV-2 and its Omicron variant: drug design templates for COVID-19. Acta Pharmacol Sin, 2022; 43: 3021–3033.
9. Rastogi M, Pandey N, Shukla A, et al. SARS coronavirus 2: from genome to infectome. Respir Res. 2020; 21: 318.
10. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel Coronavirus-infected pneumonia. N Engl J Med, 2020; 382: 1199–207.
11. Liu DX, Fung TS, Chong KK, Shukla A, Hilgenfeld R. Accessory proteins of SARS-CoV and other coronaviruses. Antiviral Res., 2014; 109: 97-109.
12. Natalia R, Sara ZL, Juan GJ, Maria M. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. Frontiers in Immunology, 2021; 12: 1664-3224.
13. Giuseppina M, Rebecca FJ, Shamar LFLM, Julien BRC. Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. Frontiers in Molecular Biosciences, 2020; 7: 2296-889X.
14. Mason RJ. Pathogenesis of COVID-19 from a cell biology perspective. European Respiratory Journal, 2020; 55(4): 2000607.
15. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. Journal of Infection, 2020; 80(6): 607-613.
16. Hussin RA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of Autoimmunity, 2020; 109: 102433.
17. Ni W, Yang X, Yang D, et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Crit Care, 2020; 24: 422.
18. Gheblawi M, Wang K, Viveiros A, Nguyen Q, Zhong JC, Turner AJ, Raizada MK, Grant MB, Oudit GY. Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System: Celebrating the 20th

Anniversary of the Discovery of ACE2. Circ Res., 2020; 126(10): 1456-1474.
19. Yang J, Petitjean SJL, Koehler M, et al. Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. Nat Commun, 2020; 11: 4541.
20. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med, 2020; 46(4): 586-590.
21. Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. Antiviral Res., 2020; 178: 104792.
22. Hofmann H, Pöhlmann S. Cellular entry of the SARS coronavirus. Trends Microbiol, 2004; 12(10): 466-72.
23. V'kovski P, Kratzel A, Steiner S, et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol, 2021; 19: 155–170.
24. Wong NA, Saier MH Jr. The SARS-Coronavirus Infection Cycle: A Survey of Viral Membrane Proteins, Their Functional Interactions and Pathogenesis. International Journal of Molecular Sciences, 2021; 22(3): 1308.
25. Jain J, Gaur S, Chaudhary Y, et al. The molecular biology of intracellular events during Coronavirus infection cycle. Virus Dis, 2020; 31: 75–79.
26. Mingjun S, Yaping C, Shanshan Q, Da S, Li F, Dongbo S. A Mini-Review on Cell Cycle Regulation of Coronavirus Infection. Frontiers in Veterinary Science, 2020; 7: 586826.
27. 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.
28. Papageorgiou AC, Mohsin I. The SARS-CoV-2 Spike Glycoprotein as a Drug and Vaccine Target: Structural Insights into Its Complexes with ACE2 and Antibodies. Cells, 2020; 9(11): 2343.
29. Dai L, Gao GF. Viral targets for vaccines against COVID-19. Nat Rev Immunol, 2021; 21: 73-82.
30. Yan W, Zheng Y, Zeng X, et al. Structural biology of SARS-CoV-2: open the door for novel therapies. Sig Transduct Target Ther, 2022; 7: 26.
31. Noori R, Sardar M. An outlook on potential protein targets of COVID-19 as a druggable site. Mol Biol Rep. 2022; 49: 10729-10748.
32. Gil C, Ginex T, Maestro I, Nozal V, Gil LB, Geijo MAC, Urquiza J, Ramírez D, Alonso C, Campillo NE, Martinez A. COVID-19: Drug Targets and Potential Treatments. Journal of Medicinal Chemistry. 2020; 63(21): 12359-12386.
33. Jo S, Kim HY, Shin DH, Kim MS. Dimerization Tendency of 3CLpros of Human Coronaviruses Based on the X-ray Crystal Structure of the Catalytic Domain of SARS-CoV-2 3CLpro. Int J Mol Sci. 2022;23(9):5268.
34. Zhu J, Zhang H, Lin Q, Lyu J, Lu L, Chen H, Zhang X, Zhang Y, Chen K. Progress on SARS-CoV-2 3CLpro Inhibitors: Inspiration from SARS-CoV 3CLpro Peptidomimetics and Small-Molecule Anti-Inflammatory Compounds. Drug Des DevelTher. 2022; 16: 1067-1082.
35. Hameedi MA, Prates TE, Garvin MR, et al. Structural and functional characterization of NEMO cleavage by SARS-CoV-2 3CLpro. Nat Commun. 2022; 13: 5285.
36. Qamar MT, Alqahtani SM, Alamri MA, Chen LL. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. Journal of Pharmaceutical Analysis. 2020; 10(4): 313-319.
37. Mody V, Ho J, Wills S. Identification of 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents. Commun Biol, 2021; 93(4).
38. Ivanov J, Polshakov D, Kato-Weinstein J, Zhou Q, Li Y, Granet R, Garner L, Deng Y, Liu C, Albaiu D, Wilson J, Aultman C. Quantitative Structure-Activity Relationship Machine Learning Models and their Applications for Identifying Viral 3CLpro- and RdRp-Targeting Compounds as Potential Therapeutics for COVID-19 and Related Viral Infections. ACS Omega, 2020; 5(42): 27344-27358.
39. National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/
40. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE.The Protein Data Bank. Nucleic Acids Research, 2000; 28: 235-242. rcsb.org
41. Berman HM, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nature Structural Biology. 2003; 10(12): 980.www.wwpdb.org
42. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins D. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular SystemsBiology, 2011; 7(539). https://www.ebi.ac.uk/Tools/msa/clustalo/
43. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic acids research, 2010; 38. https://www.ebi.ac.uk/Tools/msa/clustalo/
44. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res., 2018; 46: 296-303. https://swissmodel.expasy.org/.
45. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 2011; 27: 343-350.

46. Xu D, Zhang Y. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. Biophysical Journal. 2011; 101: 2525-2534. https://seq2fun.dcmb.med.umich.edu//ModRefiner/
47. The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC.https://pymol.org/2/
48. Labbé CM, Rey J, Lagorce D, Vavruša M, Becot J, Sperandio O, Villoutreix BO, Tufféry P, Miteva MA. MTiOpenScreen: a web server for structure-based virtual screening. Nucleic Acids Res, 2015; 43(W1): W448-54. https://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/
49. Das S, Sarmah S, Lyndem S, Singha Roy A. An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. J Biomol Struct Dyn, 2021; 39(9): 3347-3357.
50. Pham QM, Trần TQ, Thuy TTT, Cuong NM. Initial study on SARS-CoV-2 main protease inhibition mechanism of some potential drugs using molecular docking simulation. Vietnam Journal of Science and Technology, 2020; 58(665).
51. Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. NAR, 2010; 38: 469-473. http://www.sbg.bio.ic.ac.uk/3dligandsite/advanced.cgi
52. Kim S, Chen J, Cheng T, et al. PubChem in new data content and improved web interfaces. Nucleic Acids Res., 2021; 49(D1): D1388–D1395. https://pubchem.ncbi.nlm.nih.gov/
53. Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J. Cheminf, 2011; 33(3).
54. The Open Babel Package, version 2.4.1 http: //openbabel.org.
55. Dallakyan S, Olson AJ. Small-Molecule Library Screening by Docking with PyRx.Methods Mol Biol, 2015; 1263: 243-50.https://pyrx.sourceforge.io/
56. Watowich SJ, Meyer ES, Hagstrom R, Josephs R. A stable, rapidly converging conjugate gradient method for energy minimization. Journal of Computational Chemistry. 1988; 9: 650-661.
57. Pires DEV, Blundell TL, Ascher DB. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. Journal of Medicinal Chemistry, 2015; 58(9): 4066-4072. https://biosig.lab.uq.edu.au/pkcsm/prediction.
58. Daina A, et al.SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci. Rep., 2017; 7: 42717. http://www.swissadme.ch/
59. Guan L, Yang H, Cai Y, Sun L, Di P, Li W, Liu G, Tang Y. ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness. Medchemcomm, 2018; 10(1): 148-157.
60. Wang J. Estimation of the quality of refined protein crystal structures. Protein Sci., 2015; 24(5): 661-9.
61. McLennan DA. How to Read a Phylogenetic Tree. Evo Edu Outreach, 2010; 3: 506–519.
62. Zeinab A, Mengyuan L, Xiaosheng W. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses.Frontiers in Immunology, 2020; 11.
63. Kufareva I, Abagyan R. Methods of protein structure comparison. Methods Mol Biol., 2012; 857: 231-57.
64. Pantsar T, Poso A. Binding Affinity via Docking: Fact and Fiction. Molecules., 2018; 23(8): 1899.
65. Benet LZ, Hosey CM, Ursu O, Oprea TI. BDDCS, the Rule of 5 and drugability. Adv Drug Deliv Rev., 2016; 101: 89-98.
66. Vijay U, Gupta S, Mathur P, Suravajhala P, Bhatnagar P. Microbial Mutagenicity Assay: Ames Test. Bio Protoc. 2018; 8(6): e2763.
67. Raj J, Chandra M, Dogra TD, Pahuja M, Raina A. Determination of median lethal dose of combination of endosulfan and cypermethrin in wistar rat. Toxicol Int., 2013; 20(1): 1-5.
68. Kararli TT. Gastrointestinal absorption of drugs. Crit Rev Ther Drug Carrier Syst. 1989;6(1):39-86
69. Daina A, Zoete V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. ChemMedChem. 2016; 11: 1117.
70. Chow SC. Bioavailability and Bioequivalence in Drug Development. Wiley Interdiscip Rev Comput Stat. 2014; 6(4): 304-312.
71. Van Breemen RB, Li Y. Caco-2 cell permeability assays to measure drug absorption. Expert Opin Drug MetabToxicol, 2005; 1(2): 175-85.
72. Lin JH, Yamazaki M. Role of P-glycoprotein in pharmacokinetics: clinical implications. Clin Pharmacokinet, 2003; 42(1): 59-98.