Identification of Potential Drug Targets and Prediction of the Potential Impact of High Risk Non Synonymous Single Nucleotide Polymorphism in SARS-CoV-2 3C Like Proteinase (3CLpro): A Computational Approach

Sahar G. Elbager[1*], Abdelrahman H. Abdelmoneim[2], Afra M. Al Bkrye[3], Asia M. Elrashied[4], Entisar N.M. Ali[5], Hadeel A. Mohamed[6], Hazem A. Abubaker[7,] Israa A. Mohamed[8], Manal A. H. Goda[9], Mohammed Y. Basher[7], Naglla F.A. Gabir[4], Safinaz I. Khalil[10]

[1]Faculty of Medical Laboratory Sciences, University of Medical Sciences and Technology (UMST), Khartoum, Sudan

[2]Faculty of Medicine, Alneelain University, Khartoum, Sudan

[3]College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

[4]Faculty of Science, University of Khartoum, Khartoum, Sudan

[5]Faculty of Medical Laboratory, University of Kordofan, Kordofan, Sudan

[6]Faculty of Science and Technology. Omdurman Islamic University, Sudan

[7]Faculty of Veterinary Medicine, University of Khartoum, Khartoum, Sudan

[8]Faculty of Science and Technology, University of Bahri, Khartoum, Sudan

[9]Institute of Endemic Diseases, Khartoum, Sudan

[10]Faculty of Medicine, Al Fajr College for Science and Technology, Khartoum, Sudan

*Corresponding author: sahaelbager@gmail.com

Abstract

On January 2020, a new coronavirus (officially named SARS-CoV-2) was associated with alarming outbreak of a pneumonia-like illness, which was later named by the WHO as COVID-19, originating from Wuhan City, China. Although many clinical studies involving antiviral and immunomodulatory drug treatments for SARS-CoV-2 all without reported results, no approved drugs have been found to effectively inhibit the virus so far. Full genome sequencing of the virus was done, and uploaded to be freely available for the world scientists to explore. A promising target for SARS-CoV-2 drug design is a chymotrypsin-like cysteine protease (3CLpro), a main protease responsible for the replication and maturation of functional proteins in the life cycle of the SARS coronavirus. Here we aim to explore SARS-CoV-2 3CLpro as possible drug targets based on ligand- protein interactions. In addition, ADME properties of the ligands were also analyzed to predict their drug likeliness. The results revealed Out of 9 ligands, 8 ligands (JFM, X77, RZG, HWH, T8A, 0EN, PEPTIDE and DMS) showed best ADME properties. These findings suggest that these ligands can be used as potential molecules for developing potent inhibitors against SARS-CoV-2 3CLpro, which could be helpful in inhibiting the propagation of the COVID-19. Furthermore, 10 potential amino acids residues were recognized as potential drug binding site

(THR25, HIS41, GLY143, SER144, CYS145, MET165, GLU166, GLN189, ASP295 and ARG298). All those amino acid residues were subjected to missense SNP analysis were recognized to affect the structure and function of the protein. These characteristics provide them the promising to be target sites for the fresh generation inhibitors to work with and overcome drug resistance. These findings would be beneficial for the drug development for inhibiting SARS-CoV-2 3CLpro hence assisting the pharmacogenomics effort to manage the infection. of SARS-CoV-2.

Keywords: COVID-19 , coronaviruses; SARS-CoV-2; SARS-CoV; 3C like proteinase; 3CLpro

1. Introduction

The innovative COVID-19 in hominids, first revealed in Wuhan, China, in December 2019. The first cases were classified as "pneumonia of unknown etiology" as they were unable to identify the causative agent (1) The genotyping analysis and phylogenetic relationships of viruses isolates, displays that the viruses belong to genera Human Betacoronavirus (SARS-CoV, and MERS-CoV) have many similarities, but also have differences in their genomic and phenotypic structure that can influence their pathogenesis (2-4) . Subsequently, the International Committee on Taxonomy of Viruses (ICTV) termed COVID-19 (SARS-CoV-2 virus) and confirmed as the causative agent of disease (5). On 30 January the World Health Organisation has declared this a pandemic worldwid (6). However, there is a lack of specific antiviral treatment recommended for COVID-19, and no vaccine is currently available.

Human SARS-CoV-2 are enveloped viruses causes a severe acute respiratory tract infection with a high fatality rate in human. SARS-CoV-2 are enveloped viruses possesses a ~26.4–31.7 kb positive RNA genome associated with a nucleoprotein within a capsid comprised of matrix protein. The betacoronavirus genome encodes several structural proteins, including glycosylatedspike (S), envelope (E), membrane protein (M), and nucleoprotein (N). In addition, the viral genome also encodes tow overlapping polypeptides open reading frame (pp1a and pp1ab) that encode for the replicase (7, 8). Replication of COVID-19 is achieved by translated polypeptides open reading frame (pp1a and pp1ab), which further then cleaved by 3-chymotrypsin-like protease (3CLpro) into 16 functional polypeptides (non-structural proteins or nsp) that mediated replication and transcription of the viral genome(9, 10).

3CLpro polypeptide is 306 amino acids long with a molecular weight of 33,796.64 Da consists of three domains. Domain I (residue 8–101) Domain II (residue 102–184) and domain III (residue 201–303) involves of 5 α-helices (α5-α9), which are connected by a long loop (residues185–200) with domain II. Domain I and II are important functional domains that includes conserved His41 and Cys145 catalytic dyad, form a substrate binding region. Domain III controlling dimerization of the enzyme, as the protease is active only in dimeric conformation (11, 12).
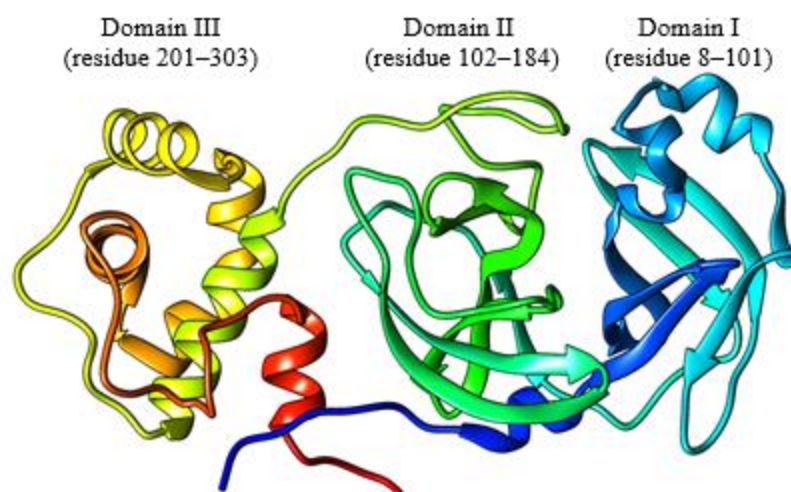
Fig. 1: Cartoon representation of 3D structure of SARS-CoV-2 3CL^{pro}

The unique function and structure of 3CLpro turns it into a promising drug target for the development of effective antiviral drugs against and other coronaviruses infections (11, 13). The aim of this study is to explore SARS-CoV-2 3CLpro as possible drug targets based on ligand- protein interactions and evolution of ligand drug likeliness using computational approaches. In addition, mapping of active sites on SARS-CoV-2 3CLpro beside predicting highly damaging missense SNPs on ligand sites.

2.Material and method

2.1 Protein sequence dataset

The protein sequence of SARS-CoV-2 (3CLpro) (accession ID: YP_009725301) was retrieved from the NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/) on 10 April 102020.

2.2. Protein conservation analysis

In order to identify similar sequences and key/conserved residues, Sequence comparison and alignment of SARS-CoV-2 3CLpro were performed using BLASTp (http://blast.ncbi.nim.nih.gov/ /Blast.cgi?PAGE=Proteins).

2. 3 Protein of 3D structure retrieval and verification

All the crystal structures of SARS-CoV-2 3CLpro were retrieved from Protein Data Bank Protein (PDB; http://wwpdb.org) Chimera used for the validation of 3D structure by plotting Ramachandran plot.

UCSF Chimera (http://www.cgl.ucsf.edu/chimera) is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, and conformational analysis Chimera (version 1.8)

The Ramachandran plot analysis provides a simple view of the conformation of a protein. The distinct regions in the Ramachandran plot based on φ-ψ angles cluster reflect a particular secondary structure. Residues are shown as blue dots, or when selected, as red dots colors in the plot indicate most favorable, allowed, generously allowed and disallowed regions. Probability contours based on a reference set of high-resolution proteins can be shown on the plot as green lines. Ideally, a model having more than 90% residues in favourable region is considered as a good-quality protein structure.

2.4 Identification of functional sites (PFSs)

Active site amino acid residues in SARS-CoV-2 3CLpro was identified using *fi*DPD. *fi*DPD (http://202.119.249.49/newfidpd.html) (14) is a sequence-based tool for the prediction of protein function-site and protein-ligand interaction.. The method is based on a functional site and physicochemical interaction-annotated domain profile database using protein domains found in the Protein Data Bank.

2.5. Ligand-protein interaction

PLIP (fully automated protein–ligand interaction profiler, http://projects.biotec.tu-dresden.de/plip-web) (15)is used to identify the residues involved in ligand interactions. The PLIP web service allows for comprehensive detection and visualization of protein–ligand interaction patterns from 3D structures, either directly from the PDB or in user-provided structures. Results for each binding site are provided as 3D interaction diagrams for manual inspection (online in

JSmol and offline with PyMOL) as well as XML and flat text files for further processing. The input options of this server generally consist of job name, Protein Data Bank ID (PDB ID) or file, and also PDB chain ID if proteins contain multiple subunits.

COACH (http://zhanglab.ccmb.med.umich.edu/COACH/) (16) ligand binding sites in SARS-CoV-2 3CLpro were analyzed. COACH is a meta-server approach for the prediction for ligand binding targets through two comparative methods TM-SITE and S-SITE. PDB file as an input. The overall results were given in the form of position of potential drug binding sites. The 3D-structure models were viewed using PyMOL. Only consensus binding sites predicted by these two server were taken.

2.6 Validation of the ligands as potential therapeutic agents

The molecular and physical features of compounds play a vital role in the selection of these agents as drug candidates. The molecular properties of the predicted ligands were analyzed using SwissADME (http://www.swissadme.ch) (17) online server to validate them as potential ligands against therapeutic targets. The ligands were then filtered through Lipinski's rule of five to predict their drug likeliness. Lipinski's rule of five (18) is a major standard to evaluate drug likeliness and if a particular chemical compound has physicochemical properties that would make it a likely orally active drug in humans. It states that an orally active drug should fall within the following criteria: a molecule with a molecular mass less than 500 Da, no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, and an octanol–water partition coefficient log P not greater than 5. Violating of more than one of these rules, predicts a molecule as do not fit into the criteria of drug likeliness and it is not considered in order to proceed with drug discovery

2.7 Retrieval of Missense Single-Nucleotide Polymorphisms

After ligand profiling, all amino acid residues showed potential binding activities, were subjected to missense SNP analysis.

2.8 Prediction of Functional Consequences of Missense Single-Nucleotide Polymorphisms

PROVEAN, PredictSNP1.0, SNPs & GO and Meta-SNP were used to assess the potential functional effect of the missense SNPs. SNPs were classified as deleterious by at least three tools.

2.8.1 PROVEAN (Protein Variation Effect Analyzer; http://provean.jcvi.org/index.php) is used to predict the possible impact of a substituted amino acid and indels on protein structure and biological function. It analyses the nsSNPs as deleterious or natural, if the final score was below the threshold score of −2.5 were considered deleterious; scores above this threshold were considered neutral (19). The input query is a protein FASTA sequence along with amino acid substitutions

2.8.2 SNPs & GO (https://snps-and-go.biocomp.unibo.it/snps-and-go/) web server was used to predict the human disease related single point protein mutations. This server was mainly based on

support vector machines which can corroborates all the information regarding variations from the existing databases. It annotates variations as deleterious based on information derived from Gene Ontology (GO) Predictor with overall accuracy of 82% (20) .protein FASTA sequence along with name and position of wild type and mutant amino acid was submitted as input for this server.

2.8.3 PMut (http://mmb.irbbarcelona.org/PMut/) a web-based tool for the annotation of pathological variants on proteins. PMUT method is based on the use of neural networks (NNs) trained with a large database of neutral mutations (NEMUs) and pathological mutations of mutational hot spots, which are obtained by alanine scanning, massive mutation, and genetically accessible mutations. The final output is displayed as a pathogenicity index ranging from 0 to 1, and the cutoff value is set to 0.5 (neutrual, 0 to 0.5; pathological, 0.5 to 1).

2.8.4 PredictSNP1.0 (http://loschmidt.chemi.muni.cz/predictsnp1/) (21)was used as the predictor of the disease related single point protein mutations. PredictSNP is a consensus classifier that integrates the results from nine in silico prediction tools: SIFT, PolyPhen-1, PolyPhen-2, MAPP, PhD-SNP, SNAP, PANTHER, PredictSNP, and nsSNPAnalyzer, thus resulting in significantly improve dprediction performance.The input query is a protein FASTA sequence along with amino acid substitutions

2.8.5 Meta-SNP(https://snps.biofold.org/meta-snp/)is a random forest-based binary classifier for Prediction of disease causing variants To reduce the bias of a single predictor, Meta-SNP was used since it integrates four existing methods: PANTHER, PhD-SNP, SIFT and SNAP. The input query is a protein FASTA sequence along with amino acid substitutions

Identification of mutant nsSNPs position in different domains

2.9Prediction of Change in Protein Stability Due to High Risk Missense Single-Nucleotide Polymorphisms

2.9.1 I-Mutant2.0 (http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi) (22) is a tool used for prediction of changes in protein stability due to single site mutations under different conditions. It is a web server based on support vector machine which worked on dataset derived from Protherm, a database of experimental records on protein mutations. It can predict the stability changes in protein with 80% accuracy based on its structure and with 77% of accuracy based on its sequence . The input can be submitted as either in the form of protein sequence or on a structure basis.

2.9.2 MUpro (http://mupro.proteomics.ics.uci.edu/) (23) , based on support vector machines and neural networks machine learning methods, which can be used to predict the effects of single-site amino acid mutations on protein stability. MUpro can predict protein stability changes merely using sequence information or combining that information with tertiary structure. A $\Delta\Delta G$ value

less than '0' indicates that the variant decreases the protein stability. On the contrary, a ΔΔG value greater than 0 indicates that the variant elevates the protein stability.

2.9.3 INPS-MD (Impact of Non-synonymous mutations on Protein Stability-Multi Dimension, https://inpsmd.biocomp.unibo.it) (24) is a method used to predict stability of protein variants from sequences and structures. The INPS-MD predictor using sequences is based on a simplified support vector (SVR) as implemented by the libsvm package, which was only tested by linear and radial basis function (RBF) kernels. INPS-MD predictions can be interpreted to identify stabilizing (ΔΔG > 0) and destabilizing (ΔΔG < 0) variations.

3. Results

3.1. Protein conservation analysis

Sequence comparison and alignment results revealed that SARS-CoV-2 3CLpro was conserved, with 100% identity among all SARS-CoV-2 genomes isolates till April 10, 2020. Then, the SARS-CoV-2 3CLpro protein sequence was aligned with that of SARS CoV. The results shown that SARS-CoV-2 3CLpro shares 96.08% sequence identity with SARS-CoV. There were 12 out of 306 residues different between SARS-CoV and SARS-CoV-2 (Val35Thr, Ser46Ala, Asn65Ser, Val86Leu, Lys88Arg, Ala94Ser, Phe134His, Asn180Lys, Val202Leu, Ser267Ala, Ala 285Thr and Leu286Iso).

| SARS-CoV-2 (3CLpro) | 1 | SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGH | 80 |
| SARS-CoV (3CLpro) | 1 | SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGH | 80 |

| SARS-CoV-2 (3CLpro) | 81 | SMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFC | 160 |
| SARS-CoV (3CLpro) | 81 | SMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFC | 160 |

| SARS-CoV-2 (3CLpro) | 161 | YMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE | 240 |
| SARS-CoV (3CLpro) | 161 | YMHHMELPTGVHAGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE | 240 |

| SARS-CoV-2 (3CLpro) | 241 | PLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ | 306 |
| SARS-CoV (3CLpro) | 241 | PLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTPFDVVRQCSGVTFQ | 306 |

Fig 2. Sequence alignment between SARS-CoV-2 3CL[pro] and SARS-CoV 3CL[pro]. Boxes are displaying mutations

3. 2 Protein of 3D structure retrieval and verification

The PDB contains more than 80 the 3D structure of the CoV-2 3CLpro. Ramachandran plot of SARS-CoV-2 3CLpro protein structure shows (PDB: 5R7Y, 6W63 , 5R80, 5R7Z and 5RFT) are the best structures identity with the query sequence since most of the residues are present in the core region (99.0%, 98.3%,98.0 %, 98.0% and 97.4%) respectively, and were chosen for further analysis. Fig. 2 Ramachandran plot analysis of (PDB: 5R7Y, 6W63)
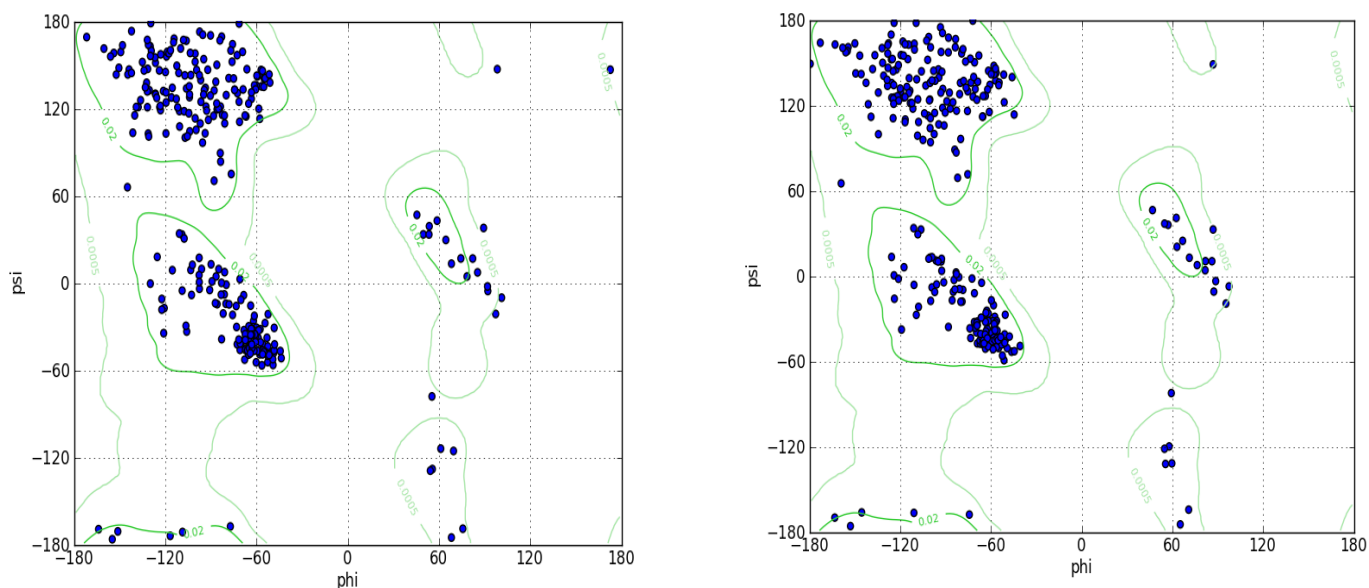
Fig. 3. Ramachandran plot analysis of SARS-CoV-2 3CLpro protein (Right: 5R7Y, Left: 6W63) showing most of the residues located at the allowed region (the blue dots represent amino acid residues; the green lines indicate the allowed region). (phi) ϕ and (psi) ψ are torsion angles. The torsion angle about the N—C bond is called ϕ and that about the C—C bond is ψ. This analysis is predicted by UCSF Chimera

3. 3 Protein functional sites identification

fiDPD server predicted 8 binding sites present in SARS-CoV-2 3CLpro (HIS41, PHE140, GLY143, CYS145, HIS163, HIS 164, GLU166 and GLN189). Considering the interactions types, fiDPD found very different types of interactions (i.e. covalent bond (COV), coordinate bond (COO), electrostatic interaction (ELE), H-bond donor (HBD), H-bond acceptor (HBA), π-stacking interactions (π-π.). Table 1

Table 1: Identification of functional sites within SARS-CoV-2 3CLpro using fiDPD

Residue	COV	COO	ELE	HBD	HBA	π-π
HIS41	T	T	0	T	T	T
PHE140	0	0	0	T	T	T
GLY143	0	0	0	0	T	T
CYS145	T	0	T	T	T	T
HIS163	0	T	T	T	T	T
HIS 164	0	0	T	T	T	T
GLU166	0	T	T	T	T	T
GLN189	0	T	0	T	T	T

"0" indicates the corresponding interaction is not present in

protein-ligand complex structure

3. 4 ligand binding sites predictions

DMS formed hydrogen bonds with ARG298, π-Cation Interactions with PHE8, Salt Bridges with ASP295 and GLU166. A π-Stacking interaction observed between JFM and HIS 41. Ligand 02J formed hydrophobic interactions with PRO168. Similarly, Ligand PJE formed hydrophobic interactions with THR25, THR26 and VAL3 along with hydrogen bondsGLY143. HIS163, HIS164, GLU166. Table 2 & Fig4.

By using COACH, ligand binding sites in SARS-CoV-2 was analyzed. ligands which have higher C-score (confidence score) indicate a more reliable prediction. The ligand AZP have higher C-score than rest of ligands. The name or synonyms of AZP is (5S,8S,14R)-ETHYL11-(3-AMINO-3-OXOPROPYL)-8-BENZYL-14-HYDROXY-5-ISOBUTYL-3,6,9,12-TETRAOXO-1-PHENYL-2-OXA-4,7,10,11-TETRAAZAPENTADECAN-15-OATE. Name of ligands and its possible binding sites are shown in Table 3.

The predicted ligands were further screened to evaluate their pharmacokinetics in the human body such as absorption, distribution, metabolism, and excretion (ADME), physicochemical properties and druglike nature. Moreover, the consensus binding sites predicted by these two tools were subjected to conservation and missense SNP analysis.

Table 2: Prediction of ligand binding sites within SARS-CoV-2 3CLpro protein using PLIP

PDB ID	Name of ligands	Interactions types	Residue number
5R7Y	JFM N-(2-phenylethyl methanesulfonamide)	π-Stacking interaction	HIS 41
6W63	X77 N-(4-tert-butylphenyl)-N-[(1R)-2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl] -1H-imidazole-4-carboxamide	Hydrophobic Interactions	HIS 41 MET165 GLU166 GLN189
		Hydrogen Bonds	HIS 41 GLY143 SER144 GLU166
5R80	RZG methyl 4-sulfamoylbenzoate	Hydrophobic Interactions	MET165 GLN189
		Hydrogen Bonds	GLU166 GLN189
		Water Bridges	GLU166
		Salt Bridges	HIS 41
5R7Z	HWH N-[2-(5-fluoranyl-1~ H -indol-3-yl)ethyl]ethanamide	Hydrophobic Interactions	MET165 GLN189
		Hydrogen Bonds	GLU166
		π-Stacking	HIS 41
5RFT	T8A	Hydrophobic Interactions	THR25 MET165
		Hydrogen Bonds	GLY143 SER144 CYS145
		π-Stacking	HIS 41
5R7Y 6W63 5R80 5R7Z 5RFT	DMS (Dimethyl Sulfoxide)	Hydrogen bonds	ARG298
		π-Cation interactions	PHE8
		Salt Bridges	GLU166 ASP295
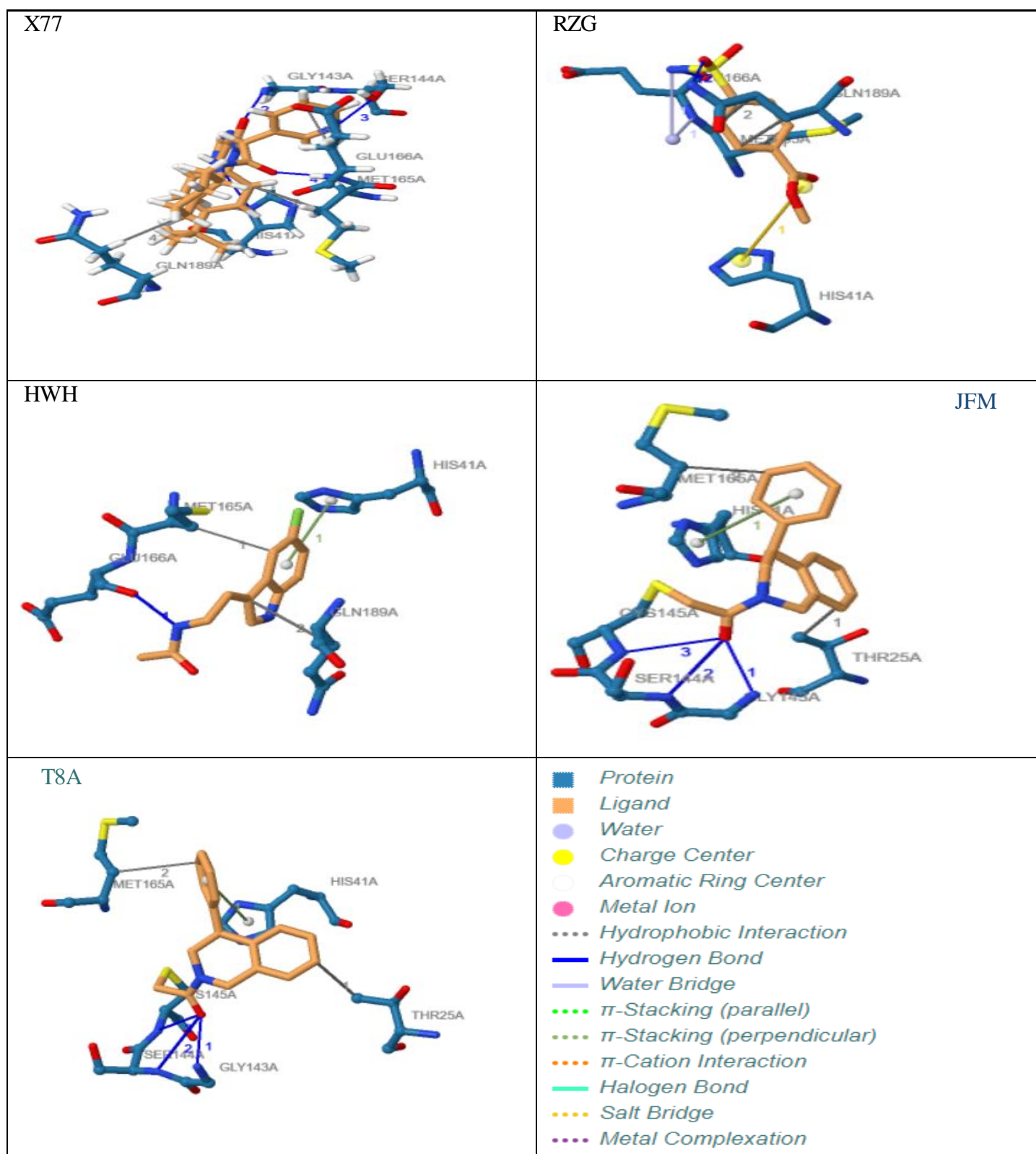
Fig 4: ligand- amino acid residues interactions
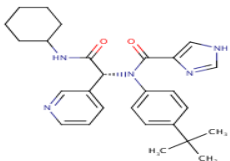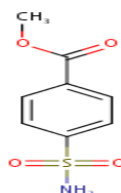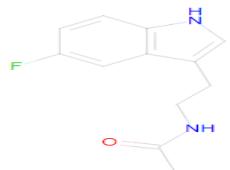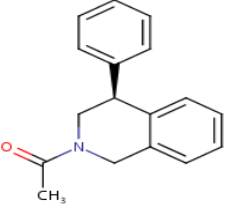
Table 3: Prediction of ligand binding sites within SARS-CoV-2 3CLpro protein using COACH.
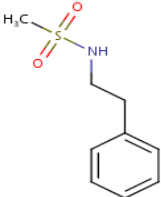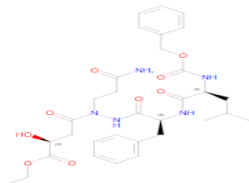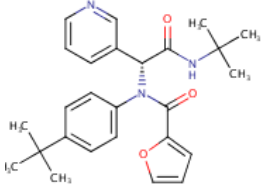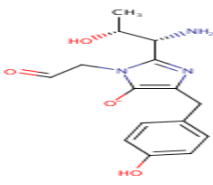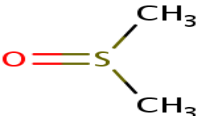
Rank	C-score	Cluster size	Ligand	Binding residues
1	1.00	213	AZP	25,41,49,54,140,141,142,143,144,145,163,164,165,166, 168,172,187,188,189,190,191,193
2	0.49	31	0EN	25,26,41,49,140,141,142,143,144,145,163,164,165,166, 187
3	0.13	11	PEPTIDE	140,141,142,143,144,145,163,164,165,166,187
4	0.01	2	HG	41,44,49,54
5	0.00	2	2Z3DA03	5,125,126,127

3. 5 Validation of the ligands as potential therapeutic agents

Swissadme server was used to analyze the molecular parameters of the ligands in the Lipinski's rule. The results include the molecular weight, no of hydrogen donors, acceptors, and lipophilicity of the ligand molecules (Log P). Out of the 9 predicted ligands, 9 ligands cleared Lipinski's rule of five, except AZP.

Table4: Lipinski properties of the ligands

Ligand	Formula	Structure	MW	(Lipinki's Rule of Five)				Drug likeliness
				#H-bond acceptors	#H-bond donors	Log P	violations	
X77	C27H33N5O2		459.58	4	2	3.83	0	Yes
RZG	C8H9NO4S		215.23	5	1	0.66	0	Yes
HWH	C12H13FN2O		220.24	2	2	2.25	0	Yes
T8A	C17H17NO		251.32	1	0	2.89	0	Yes

JFM	C9H13NO2S		199.27	3	1	1.33	0	Yes
AZP	C32H43N5O9		641.71	9	5	1.73	2	No
0EN	C26H31N3O3		433.54	4	1	4.19	0	Yes
PEPTIDE	C15H18N3O4		304.32	6	3	0.31	0	Yes
DMS	C2H6OS		78.13	1	0	0.05	0	Yes

3.6 Retrieval of Missense Single-Nucleotide Polymorphisms Datasets

Certain amino acid residues showed potential drug binding site, all such amino acid residues were subjected to missense SNP analysis. These amino acids include (THR25, HIS41, GLY143, SER144, CYS145, MET165, GLU166, GLN189, ASP295 and ARG298), were mutated to all possible missense mutation. As a result, a total of 191 mutations were subjected to further analysis. A variety of in silico SNP prediction tools were employed in order to determine the effect of a given missense mutation on the respective gene function.

3.7. Missense Single-Nucleotide Polymorphisms Analysis

To gather higher accuracy results, five in silico SNP prediction tools (PROVEAN, SNPs&GO, Pmut, PredictSNP and Meta-SNP) were employed to predict the high risk missense SNPs. We categorized SNPs as damaging if they were predicted to be damaging by four or more SNP prediction A total of 190 SNPs were subjected to analysis using theses algorithms. By using

PROVEAN tool there were 24 neutrals and 166 deleterious. In SNP&GO there were 146 neutral and 44 diseases causing. In PMUT there were 75 neutrals and 115 damaging. In Predict SNP were 26 neutrals and 164 damaging. MetaSNP were 71 neutrals and 119 disease causing. Out of a total of 190 SNPs, 97 SNPs were considered high risk and were subjected to further stability studies.

3.8 Prediction of Change in Protein Stability

The protein stability change was estimated using I-Mutant2.0, MUpro and INPS-MD. SNPs were considered as destabilizing if two or more algorithms showed a decrease in stability upon mutation. I -Mutant 3.0 showed that the 171 nsSNPs decreased stability ($\Delta\Delta G < 0$), whereas 19 nsSNPs increased stability ($\Delta\Delta G > 0$). Analysed by MUpro and INPS-MD, 177 and 159 nsSNPs were found to decrease protein stability, respectively. In total, 94 nsSNPs, were predicted to decrease the stability, 3 nsSNPs (H41L, E166I, E166M) were found to increase protein stability. H41L was found to increase protein stability by the three tools.

Table 5: Missense SNPs in 3CLpro predicted to be damaging/neutrall Destabilizing

	Damaging SNPs					Destabilizing SNPs		
Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP	I -Mutant	MUpro	INPS-MD
T25A	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
T25R	Neutral	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
T25N	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
T25D	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
T25C	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
T25Q	Neutral	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Decrease
T25E	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
T25G	Deleterious	Neutral	Damaging	Neutral	Neutral	Decrease	Decrease	Decrease
T25H	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Decrease
T25I	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Increase	Increase
T25L	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Increase	Increase
T25K	Neutral	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
T25M	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Increase	Increase
T25F	Neutral	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
T25P	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
T25S	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
T25W	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Increase
T25Y	Deleterious	Neutral	Damaging	Damaging	Neutral	Increase	Decrease	Increase
T25V	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Increase
H41A	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41R	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Increase
H41N	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41D	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41C	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
H41Q	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41E	Deleterious	Disease	Neutral	Damaging	Disease	Increase	Decrease	Decrease
H41G	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease

H41I	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Increase
H41L	Deleterious	Disease	Damaging	Damaging	Disease	Increase	Increase	Increase
H41K	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41M	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Increase
H41F	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Increase
H41P	Deleterious	Disease	Neutral	Damaging	Disease	Increase	Decrease	Decrease
H41S	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41T	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
H41W	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
H41Y	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Decrease
H41V	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Increase
G143A	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
G143R	Deleterious	Disease	Damaging	Disease	Disease	Decrease	Decrease	Decrease
G143N	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
G143D	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143C	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143Q	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143E	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143H	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143I	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143L	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Increase	Decrease
G143K	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Increase	Decrease
G143M	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143F	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143P	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
G143S	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
G143T	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
G143W	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143Y	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
G143V	Deleterious	Disease	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144A	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Increase
S144R	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144N	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
S144D	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
S144C	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144Q	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144E	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
S144G	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Decrease
S144H	Deleterious	Disease	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
S144I	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144L	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144K	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144M	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
S144F	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
S144P	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
S144T	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease

S144W	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
S144Y	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
S144V	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
C145A	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145R	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145N	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145D	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145Q	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145E	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145G	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145H	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145I	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145L	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145K	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145M	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145F	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145P	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145S	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145T	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
C145W	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145Y	Deleterious	Disease	Disease	Deleterious	Disease	Decrease	Decrease	Decrease
C145V	Deleterious	Disease	Neutral	Deleterious	Disease	Decrease	Decrease	Decrease
M165A	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
M165R	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165N	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165D	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165C	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165Q	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
M165E	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Decrease
M165G	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165H	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
M165I	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
M165L	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
M165K	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165F	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165P	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165S	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
M165T	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
M165W	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
M165Y	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
M165V	Neutral	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
E166A	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
E166R	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
E166N	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
E166D	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease

E166C	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
E166Q	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
E166G	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
E166H	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Increase
E166I	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Increase
E166L	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Decrease
E166K	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
E166M	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Increase
E166F	Deleterious	Disease	Damaging	Damaging	Disease	Increase	Decrease	Decrease
E166P	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Decrease
E166S	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
E166T	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
E166W	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
E166Y	Deleterious	Neutral	Damaging	Damaging	Disease	Increase	Decrease	Decrease
E166V	Deleterious	Disease	Damaging	Damaging	Disease	Increase	Decrease	Decrease
Q189A	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
Q189R	Neutral	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Decrease
Q189N	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
Q189D	Deleterious	Neutral	Damaging	Neutral	Disease	Decrease	Decrease	Decrease
Q189C	Deleterious	Neutral	Damaging	Neutral	Disease	Increase	Decrease	Decrease
Q189E	Neutral	Neutral	Neutral	Damaging	Neutral	Increase	Decrease	Decrease
Q189G	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
Q189H	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Increase
Q189I	Deleterious	Neutral	Damaging	Damaging	Neutral	Increase	Decrease	Increase
Q189L	Deleterious	Neutral	Damaging	Damaging	Neutral	Increase	Increase	Decrease
Q189K	Neutral	Neutral	Neutral	Neutral	Neutral	Increase	Decrease	Increase
Q189M	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
Q189F	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
Q189P	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
Q189S	Neutral	Neutral	Damaging	Neutral	Neutral	Decrease	Decrease	Decrease
Q189T	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
Q189W	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
Q189Y	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
Q189V	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Increase	Decrease
D295A	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Increase
D295R	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
D295N	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Decrease
D295C	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
D295Q	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
D295E	Neutral	Neutral	Neutral	Neutral	Neutral	Increase	Decrease	Decrease
D295G	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
D295H	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
D295I	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
D295L	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
D295K	Deleterious	Neutral	Neutral	Damaging	Disease	Decrease	Decrease	Increase
D295M	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Decrease

D295F	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
D295P	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
D295S	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
D295T	Deleterious	Neutral	Neutral	Damaging	Neutral	Decrease	Decrease	Decrease
D295W	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
D295Y	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Increase
D295V	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
R298A	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298N	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
R298D	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298C	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298Q	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
R298E	Neutral	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298G	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298H	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
R298I	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Increase	Increase
R298L	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Increase	Decrease
R298K	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
R298M	Deleterious	Neutral	Neutral	Neutral	Neutral	Decrease	Increase	Decrease
R298F	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Increase	Decrease
R298P	Deleterious	Neutral	Damaging	Damaging	Disease	Decrease	Decrease	Decrease
R298S	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298T	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298W	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Decrease
R298Y	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Decrease	Increase
R298V	Deleterious	Neutral	Damaging	Damaging	Neutral	Decrease	Increase	Increase

4. Discussion

The SARS-CoV-2 3CLpro play critical role in the maturation and posttranslational processing of the viral replicase protein and thus propagation of the COVID-19 infections. Therefore, great effort has been spent on studying this protein in order to identify therapeutics against the SARS-CoV-2 based on some of the previous progress of specific inhibitors development for the SARS-CoV enzyme because they share similar conserved regions, active sites and enzymatic mechanisms(25-28). Our study displayed that 3CLpro is conserved in all SARS-CoV-2. It is highly similar to SARS-CoV 3CLpro, with only 12 residues different. These differences may affect 3CLpro structure and function and might disrupt important hydrogen bonds and alter the receptor binding site, thereby affecting its ability to bind with the SARS-CoV inhibitor.

In the drug development process, the 3D structure of protein, protein active sites and ligand interactions of the is of central importance. Our analysis shows (PDB: 5R7Y, 6W63, 5R80, 5R7Z and 5RFT) are the best structures identity with the CoV-2 3CLpro sequence, and were chosen for further analysis. Additionally, results showed that HIS41, PHE140, GLY143, CYS145, HIS163, HIS 164, GLU166 and GLN189 are the amino acid residues participating in various physicochemical interaction at the active site of 3CLpro, particularly GLY145, HIS163, HIS 164 and GLU166 contribute significantly to strong electrostatic interactions in the active site. Moreover, nine ligands (JFM, X77, RZG, HWH, T8A, AZP, 0EN, PEPTIDE and DMS) were identified to have interaction with numerous 3CLpro residues. Furthermore, these ligands were subjected to drug likeliness evaluation.

The Pharmacokinetic parameters and physicochemical properties of compounds play a vital role in the selection of these agents as drug candidates. According to Lipinski's rule of five, out of 9 ligands, 8 were cleared ADME (Absorption, Distribution, Metabolism and Elimination) and exhibit good drug-likeness criteria, except AZP that tow violets. This results suggest that these ligands can be used for developing potent inhibitors against SARS-CoV-2 3CLpro.

Single-nucleotide polymorphisms (SNPs) in protein ligand sites may causing loss of protein–ligand interactions, disturb protein dynamics by changing protein stability, disrupting and blocking the active sit resulting in a loss of inhibitor efficiency. In the current study, 10 potential amino acids (THR25, HIS41, GLY143, SER144, CYS145, MET165, GLU166, GLN189, ASP295 and ARG298), showed potential drug binding site in SARS-CoV-2 3CLpro and were subjected to SNP analysis. Except for GLN189, all SNPs showed to be highly damaging.

The catalytic dyad of 3CLpro includes the conserved residues HIS41 and CYS145,which are essential for catalytic activity (29) . Mutational analysis of the 3CLpro established the importance of His41, Cys145, and Glu166 in the substrate-binding subsite for keeping the proteolytic activity(30). Complete loss of proteolytic activity of SARS-CoV 3CLpro were observed when

single amino acid mutations at His41 and Cys145 were generated using site-directed mutagenesis (29, 31) .The Glu166 located in the protein functional sites of 3CLpro, has an electrostatic interactions and hydrogen bonds other residues, being important for the substrate binding. Mutational analysis of Glu166 conformed the importance of Glu166 in the enzymatic function of the SARS-CoV 3CLpro. Results of the present study determined that HIS41, CYS145 and E166 are highly deleterious missense SNPs decreases protein stability except H41L, E166I and E166M SNPs, which increase the protein stability.

ASP295 and ARG298 of 3CLpro is located at the α-helix of domain 3 (aa 201 to 303), which has been shown to be critical for catalytic activity and enzyme dimerization. The substitution of A298 may disturbed the conformation of at the α-helix and this can have severe effects on the structure of the protein thereby impair catalysis (32, 33) . The current work determined ASP295 as a highly missense SNP decreases protein stability when mutated to (C, G, P, R, V, W, Y). Only R298P residue was found a highly missense SNP decreases protein stability. Prolines are known to be very rigid and therefore disturb a special backbone conformation which might be required at this position.

GLY143, SER144, MET165 residues are located at domain II, this domain is important for binding of substrate binding. This residue also involved in hydrophobic and electrostatic interactions. Mutation of this residue might disturb the interaction and as such disturbs the protein function. The current study determined that GLY143, SER144, MET165 are highly deleterious missense SNPs decreases protein stability.

The present study determined most of SNPs highly deleterious missense SNPs decreases protein stability. These characteristics provide them the promising to be target sites for the fresh generation inhibitors to work with and overcome drug resistance.

5. Conclusion

The computational approach can be fast and very useful tool to identify powerful inhibitors against the SARS-CoV-2 3CLpro. In this study, 8 ligands were identified as potential molecules for developing potent inhibitors against SARS-CoV-2 3CLpro and 10 potential amino acids residues were identified as candidates for drug targeting against SARS-CoV-2 3CLpro hence providing a valuable basis for pharmacogenomics practices. Additionally, the most deleterious mutation provides promising target sites for the fresh generation inhibitors to work with and overcome drug resistance.

Competing Interests

The authors declare that they have no competing interests.

References

1. Wuhan City Health Committee (WCHC). Wuhan Municipal Health and Health Commission's briefing on the current pneumonia epidemic situation in our city 2019 [updated 31 December 2019 J. Available from: http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989.

2. Mousavizadeh L, Ghasemi S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi. 2020.

3. Malik YS, Sircar S, Bhat S, Sharun K, Dhama K, Dadar M, et al. Emerging novel coronavirus (2019-nCoV)-current scenario, evolutionary perspective based on genome analysis and recent developments. The veterinary quarterly. 2020;40(1):68-76.

4. Zhou PY, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020; 579, 270–273.

5. Cascella M RM, Cuomo A, et al. Features, Evaluation and Treatment Coronavirus (COVID-19) [Updated 2020 Apr 6]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. . Available from: https://www.ncbi.nlm.nih.gov/books/NBK554776/.

6. 2020. WHONc-nSRJ. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4.

7. Woo PC, Huang Y, Lau SK, Yuen K-Y. Coronavirus genomics and bioinformatics analysis. viruses. 2010;2(8):1804-20.

8. Thiel V, Herold J, Schelle B, Siddell SG. Viral replicase gene products suffice for coronavirus discontinuous transcription. Journal of virology. 2001;75(14):6676-81.

9. Van Hemert MJ, van den Worm SH, Knoops K, Mommaas AM, Gorbalenya AE, Snijder EJ. SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. PLoS pathogens. 2008;4(5).

10. Ziebuhr J. Molecular biology of severe acute respiratory syndrome coronavirus. Current opinion in microbiology. 2004;7(4):412-9.

11. Anand K, Yang, H., Bartlam, M., Rao, Z., & R. Hilgenfeld. . Coronavirus main proteinase: Target for antiviral drug therapy. In A.Schmidt, O. Weber, & M. H. Wolff (Eds.), Coronaviruses with special emphasis on first insights concerning SARS. 2005(pp. 173–199).

12. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. Science. 2003;300(5626):1763-7.

13. Yang H, Bartlam, M., & Rao, Z. . Drug design targeting the main protease, the Achilles' Heel of coronaviruses. Current Pharmaceutical Design.2006;; 12(35), 4573–4590.

14. Han M, Song Y, Qian J, Ming D. Sequence-based prediction of physicochemical interactions at protein functional sites using a function-and-interaction-annotated domain profile database. BMC Bioinformatics. 2018;19(1):204-.

15. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015;43(W1):W443-W7.

16. Yang J, Roy A, Zhang Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics. 2013;29(20):2588-95.

17. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports. 2017;7(1):42717.

18. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2012;64:4-17.

19.   Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745-7.

20.   Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat. 2009;30(8):1237-44.

21.   Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol. 2014;10(1):e1003440-e.

22.   Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic acids research. 2005;33(Web Server issue):W306-W10.

23.   Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins. 2006;62(4):1125-32.

24.   Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics. 2016;32(16):2542-4.

25.   Chen YW, Yiu CB, Wong KY. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL (pro)) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. F1000Research. 2020;9:129.

26.   Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B. 2020.

27.   Jo S, Kim S, Shin DH, Kim MS. Inhibition of SARS-CoV 3CL protease by flavonoids. Journal of enzyme inhibition and medicinal chemistry. 2020;35(1):145-51.

28.   Xu X, Dang Z, Zhang L, Zhuang L, Jing W, Ji L, et al. Potential inhibitor for 2019-novel coronaviruses in drug development. Cancer Translational Medicine. 2020;6(1):17-20.

29.   Huang C, Wei P, Fan K, Liu Y, Lai L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry. 2004;43(15):4568-74.

30.   Lin CW, Tsai CH, Tsai FJ, Chen PJ, Lai CC, Wan L, et al. Characterization of trans- and cis-cleavage activity of the SARS coronavirus 3CLpro protease: basis for the in vitro screening of anti-SARS drugs. FEBS letters. 2004;574(1-3):131-7.

31.   Chen S, Chen L-l, Luo H-b, Sun T, Chen J, Ye F, et al. Enzymatic activity characterization of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer technique. Acta Pharmacol Sin. 2005;26(1):99-106.

32.   Xia B, Kang X. Activation and maturation of SARS-CoV main protease. Protein & cell. 2011;2(4):282-90.

33.   Kang X, Zhong N, Zou P, Zhang S, Jin C, Xia B. Foldon unfolding mediates the interconversion between Mpro-C monomer and 3D domain-swapped dimer. Proceedings of the National Academy of Sciences. 2012;109(37):14900-5.