

# High Risk functional nsSNP in SARS-CoV-2 (2019-nCoV) Main Peptidase as Potential Targets to Structure-Based Drug Design: A Computational Approach

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

<sup>1</sup>Faculty of Medical Laboratory Sciences, University of Medical Sciences and Technology (UMST), Khartoum, Sudan

<sup>2</sup>Faculty of Medicine, Alneelain University, Khartoum, Sudan

<sup>3</sup>College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

<sup>4</sup>Faculty of Science, University of Khartoum, Khartoum, Sudan

<sup>5</sup>Faculty of Medical Laboratory, University of Kordofan, Kordofan, Sudan

<sup>6</sup>Faculty of Science and Technology, Omdurman Islamic University, Sudan

<sup>7</sup>Faculty of Veterinary Medicine, University of Khartoum, Khartoum, Sudan

<sup>8</sup>Faculty of Science and Technology, University of Bahri, Khartoum, Sudan

<sup>9</sup>Institute of Endemic Diseases, Khartoum, Sudan

<sup>10</sup>Faculty of Medicine, Al Fajr College for Science and Technology, Khartoum, Sudan

**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 COVID-19, no approved drugs have been found to effectively inhibit the virus so far. A promising target for SARS-COV-2 drug design is a (Mpro), a main protease (M<sup>pro</sup>), responsible for the replication and maturation of functional proteins in the life cycle of the SARS coronavirus. Here we employed missense SNP analysis to all amino acid identified as functional residues and showed potential binding activities to predict the high risk missense mutations that are highly damaging to the structure and the function of the SARS-COV-2 M<sup>pro</sup> aiming to detect potential drug targets for inhibition of SARS-COV-2 M<sup>pro</sup>. Our results demonstrated essential roles of HIS41, GLY143, CYS145, HIS163, GLU166 as binding residues, structural and functional residues, participating in different physicochemical interaction for keeping the proteolytic activity and enzymatic function of the SARS-COV-2 M<sup>pro</sup>. Upon subjecting these residues to single-nucleotide polymorphism (SNP) analysis, out of 1527 nsSNP, 34 nsSNP recognized to alters the functionality and stability of SARS-COV-2 Mpro. Furthermore, most of these deleterious nsSNPs were predicted as ligand binding residues. We conclude that these residues can be target sites for the fresh generation inhibitors for of the SARS - CoV Mpro and overcome the current drugs ineffectiveness.

**Keywords** COVID-19, Coronaviruses, SARS-COV-2, Main Protease, Mpro

## 1. Introduction

The innovative SARS-COV-2 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 SARS-COV-2 (SARS-COV-2 virus) and confirmed as the causative agent of disease (5). On 30 January the World Health Organization has declared this a pandemic worldwide (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

\* Corresponding author:

saharelbager@gmail.com (Sahar G. Elbager)

Published online at <http://journal.sapub.org/ijvmb>

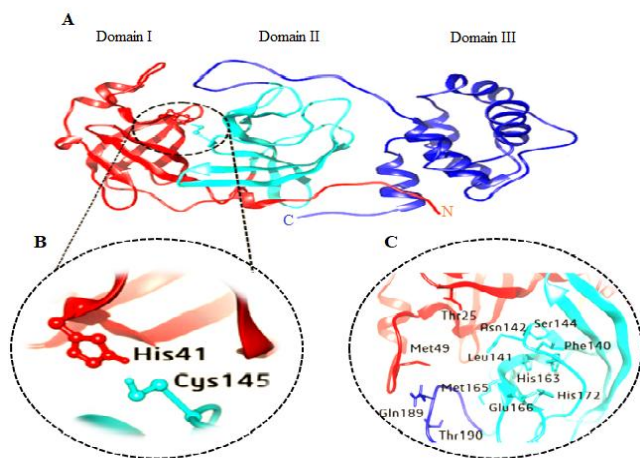
Copyright © 2020 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

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 glycosylated spike (S), envelope (E), membrane protein (M), and nucleoprotein (N). In addition, the viral genome also encodes two overlapping polypeptides open reading frame (pp1a and pp1ab) that encode for the replicase [7,8]. Replication of SARS-COV-2 is achieved by translated polypeptides open reading frame (pp1a and pp1ab), which further then cleaved by main peptidase (Mpro) into 16 functional polypeptides (non-structural proteins or nsp) that mediated replication and transcription of the viral genome [9,10].

The SARS-COV-2 Mpro, also known as 3C-like protease (3CLpro or Nsp5), like other coronaviruses, is 306 amino acids long with a molecular weight of 3797.0 Da. This enzyme naturally forms a dimer each monomer consists of three domains. Domain I (residue 8–101) Domain II (residue 102–184) and Domain III (residue 201–303) involves of 5  $\alpha$ -helices ( $\alpha$ 5– $\alpha$ 9), which are connected by a long loop (residues 185–200) 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 mediating the tight dimerization of the enzyme, as the protease is active only in dimeric conformation [11,12]. In addition to the catalytic centre there are five subsites (S1-S5). S1& S2 subsites are deeply buried, mainly involved in hydrophobic and electrostatic interactions while S3-S5 subsites, are superficial, involved in different functionalities. The S1 subsite consists of His163, Glu166, Cys145, Gly143, His172, and Phe140 while S2 consist of Cys145, His41, and Thr25 amino acid residues. The superficial subsites S3-S5 comprises Met49, His41, Met165, Glu166 and Gln189 amino acid residues [13-15]. (Figure 1).



**Figure 1.** A) Cartoon representation of 3D structure of one monomer of the dimeric SARS-COV-2 Mpro B) His41- Cys145 catalytic dyad. C) substrate-binding residues within S1-S5

The unique function and structure of Mpro turns it into

a promising drug target for the development of effective antiviral drugs against SARS-COV-2 viruses and other coronaviruses infections [11,16]. The aim of this study is to predict the high risk missense mutations that are highly damaging to the structure as well as the function of the SARS-COV-2 Mpro to identify possible drug targets for inhibition of SARS-COV-2 M<sup>pro</sup>.

## 2. Material and Method

### 2.1. Protein Sequence Analysis

The protein sequence of SARS-COV-2 (M<sup>pro</sup>) (accession ID: YP\_009725301) was retrieved from the NCBI Protein database (<https://www.ncbi.nlm.nih.gov/protein/>) on 10 April 2020. In order to identify similar sequences, Sequence comparison and alignment of SARS-COV-2 M<sup>pro</sup> were performed using BLASTp (<http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>).

### 2.2. Protein of 3D Structure Retrieval and Verification

All the crystal structures of SARS-COV-2 M<sup>pro</sup> 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 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) [17].

The Ramachandran plot analysis provides a simple view of the conformation of a protein. The distinct regions in the Ramachandran plot based on  $\phi$ - $\psi$  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.

### 2.3. Prediction of Ligand Binding Sites

Ligand binding sites predictions in SARS-COV-2 M<sup>pro</sup> were analyzed using COACH. COACH is a meta-server approach for the prediction for ligand binding targets through two comparative methods TM-SITE and S-SITE [18]. PDB file as an input. The overall results were given in the form of position of potential drug binding sites.

### 2.4. Identification of Functional Sites (PFSs)

Active site amino acid residues in SARS-COV-2 Mpro was identified using fiDPD. fiDPD is a sequence-based tool for the prediction of protein function-site and protein-ligand interaction [19]. The method is based on a functional site and physicochemical interaction-annotated domain profile

database using protein domains found in the Protein Data Bank. The residues predicted as active residues were subjected to conservation analysis and missense SNP analysis.

## 2.5. Protein Conservation Analysis

In order to identify key functional or structural /conserved residues, ConSurf web-server use to calculates the evolutionary conservation of amino acid substitution in proteins [20]. ConSurf gives the output in the form of score where score 9 represent the most conserved and 1 represent the highly variable amino acid.

## 2.6. Retrieval of Missense Single-Nucleotide Polymorphisms

All amino acid identified as functional residues and showed potential binding activities, were subjected to missense SNP analysis.

## 2.7. Prediction of Functional Consequences of nsSNPs

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

2.7.1. PROVEAN (Protein Variation Effect Analyzer) 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 Damaging or natural, if the final score was below the threshold score of  $-2.5$  were considered Damaging; scores above this threshold were considered neutral [21]. The input query is a protein FASTA sequence along with amino acid substitutions.

2.7.2. SNPs & 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 Damaging based on information derived from Gene Ontology (GO) Predictor with overall accuracy of 82% [22]. Protein FASTA sequence along with name and position of wild type and mutant amino acid was submitted as input for this server.

2.7.3. 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 [23]. The final output is displayed as a pathogenicity index ranging from 0 to 1, and the cut-off value is set to 0.5 (neutral, 0 to 0.5; pathological, 0.5 to 1).

2.7.4. PredictSNP1.0 was used as the predictor of the disease related single point protein mutations [24]. 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 prediction performance. The input query is a protein FASTA sequence along with amino acid substitutions.

2.7.5. 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 [25]. The input query is a protein FASTA sequence along with amino acid substitutions.

## 2.8. Prediction of Change in Protein Stability due to High Risk nsSNPs

2.8.1. I-Mutant2.0 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 [26]. 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.

2.8.2. MUpro 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 [27]. A  $\Delta\Delta G$  value less than '0' indicates that the variant decreases the protein stability. On the contrary, a  $\Delta\Delta G$  value greater than 0 indicates that the variant elevates the protein stability.

2.8.3. INPS-MD (Impact of Non-synonymous mutations on Protein Stability-Multi Dimension) 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 [28]. INPS-MD predictions can be interpreted to identify stabilizing ( $\Delta\Delta G > 0$ ) and destabilizing ( $\Delta\Delta G < 0$ ) variations.

## 2.9. Analyzing the Effect of nsSNPs on 3D Structure of the Proteins and Physicochemical Properties

Project Hope server was used to search protein 3D structures by collecting structural information from a series of sources, including calculations on the 3D coordinates of the protein, sequence annotations from the UniProt database, and predictions by DAS services [29]. Furthermore, it described the reaction and physicochemical properties of these candidates. Protein sequence and mutant variants were submitted to project hope server in order to analyze the structural and conformational variations that have resulted from single amino acid substitution.

### 3. Results

#### 3.1. Protein Sequence Analysis

Sequence comparison and alignment results revealed that SARS-COV-2 Mpro was conserved, with 100% identity among all SARS-COV-2 genomes isolates till April 10, 2020. Then, the SARS-COV-2 Mpro protein sequence was aligned

```

SARS-CoV-2 (3CLpro) 1  SGFRKMAFPSGKVEGCMVQVCTGTTTLNGLWLDLWYICPRHVICTSEDMLNPNVEDLLIRKSNHIFLVQAGNVQLRVIGH 80
SARS-CoV (3CLpro) 1  SGFRKMAFPSGKVEGCMVQVCTGTTTLNGLWLDLWYICPRHVICTAEDMLNPNVEDLLIRKSNHIFLVQAGNVQLRVIGH 80

SARS-CoV-2 (3CLpro) 81  SMQNCVLLKLVDTANPKTPKYKFKVRIQPGQTFSVLACYNQSPSGVYQCAMPRIHTIKGSFLNGSCGSGVFNIDYDCVSFC 160
SARS-CoV (3CLpro) 81  SMQNCVLLKLVDTANPKTPKYKFKVRIQPGQTFSVLACYNQSPSGVYQCAMPRIHTIKGSFLNGSCGSGVFNIDYDCVSFC 160

SARS-CoV-2 (3CLpro) 161  YMHHEMLPTGVHAGTDLEGIFYGPFFDRQTAQAAGDTTITIVNLAWLYAAVINGDRWFLNRFITTLNDFNLVAMKYNYE 240
SARS-CoV (3CLpro) 161  YMHHEMLPTGVHAGTDLEGIFYGPFFDRQTAQAAGDTTITIVNLAWLYAAVINGDRWFLNRFITTLNDFNLVAMKYNYE 240

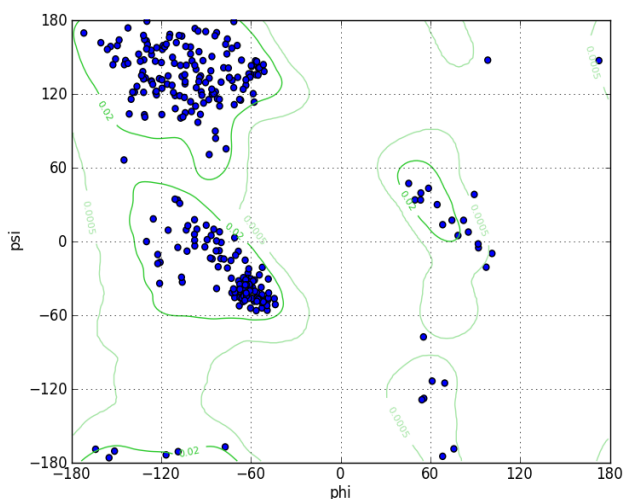
SARS-CoV-2 (3CLpro) 241  PLTQDHDVILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSLLEDEFTPFQVVRQCSGVTFQ 306
SARS-CoV (3CLpro) 241  PLTQDHDVILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTLLEDEFTPFQVVRQCSGVTFQ 306

```

**Figure 2.** Sequence alignment between SARS-COV-2 Mpro and SARS-CoV Mpro. 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 Mpro. Ramachandran plot of SARS-COV-2 Mpro protein structure shows (PDB: 5R7Y) is the best structures identity with the query sequence since most of the residues are present in the core region (99.0) and was chosen for further analysis. Fig. 3 Ramachandran plot analysis of (PDB: 5R7Y).



**Figure 3.** Ramachandran plot analysis of SARS-COV-2 Mpro protein (5R7Y) 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)  $\phi$  and (psi)  $\psi$  are torsion angles. The torsion angle about the N—C bond is called  $\phi$  and that about the C—C bond is  $\psi$ . This analysis is predicted by UCSF Chimera

with that of SARS CoV. The results shown that SARS-COV-2 Mpro 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).

#### 3.3. Prediction of Ligand Binding Sites

By using COACH, ligand binding sites in SARS-COV-2 Mpro 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 and its possible binding sites are 25, 26, 27, 41, 49, 140, 141, 142, 143, 144, 145, 163, 164, 165, 166, 168, 172, 187, 189, 190 and 192. Name of ligands and its possible binding sites are shown in Table 1.

**Table 1.** Prediction of ligand binding sites within SARS-COV-2 M<sup>pro</sup> protein using COACH

Rank	C-score	Cluster size	Ligand	Binding residues
1	1.00	213	AZP	25, 26, 27, 41, 49, 140, 141, 142, 143, 144, 145, 163, 164, 165, 166, 168, 172, 187, 189, 190, 192
2	0.49	31	OEN	25,27,41,145
3	0.13	11	PEPTIDE	25, 26, 41, 49, 140, 141, 142, 143, 144, 145, 163, 164, 165, 166, 187
4	0.01	2	HG	39, 41, 49, 145, 164, 165, 187, 188
5	0.01	2	2Z3DA03	145, 163, 164, 166
6	0.00	1	2Z3EA03	41, 44, 49, 54
7	0.00	1	3IH	62, 63, 64
8	0.00	1	N/A	10, 11, 14, 115, 122
9	0.00	1	N/A	73, 96, 97
10	0.00	1	N/A	134, 181, 182, 183

### 3.4. Identification of Functional Sites (PFSs)

fiDPD server identified 8 functional sites present in SARS-COV-2 Mpro (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),  $\pi$ -stacking interactions ( $\pi$ - $\pi$ )). Table 2

**Table 2.** Identification of functional sites within SARS-COV-2 Mpro using fiDPD

Residue	COV	COO	ELE	HBD	HBA	$\pi$ - $\pi$
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

### 3.5. Protein Conservation Analysis

ConSurf analysis predicted HIS41, PHE140, GLY143, CYS145 and HIS163 to be buried and conserved residue, i.e., a structural residue, which suggests that these positions are important for the SARS-COV-2 Mpro structure. GLU166, GLN189 were predicted to be exposed and conserved residue, i.e., a functional residue, which suggests that these positions are important for the SARS-COV-2 Mpro function. Table 3

**Table 3.** Conservation profile of amino acids in SARS-COV-2 Mpro

Residue	Consurf	Class assignment	Function
HIS41	Highly conserved	Buried	Structural
PHE140	Highly conserved	Buried	Structural
GLY143	Highly conserved	Buried	Structural
CYS145	Highly conserved	Buried	Structural
HIS163	Highly conserved	Buried	Structural
HIS 164	Conserved	Buried	-
GLU166	Highly conserved	Exposed	Functional
GLN189	Conserved	Exposed	Functional

### 3.6. Retrieval of Missense Single-Nucleotide Polymorphisms Datasets

All amino acid residues showed potential functional sites, were subjected to missense SNP analysis. These amino acids include (HIS41, PHE140, GLY143, CYS145, HIS163, HIS164, GLU166 and GLN189), were mutated to all possible missense mutation. SNPs identified theoretically by exchange of each 8 amino acid residues with all possible other nineteen amino acid ( $8 \times 19 = 152$  possible SNPs). Then a variety of computational tools were employed in order to determine the effect of a given missense mutation on the SARS-COV-2 Mpro function and structure.

### 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 152 SNPs were subjected to analysis using these algorithms. By using PROVEAN tool there were 5 neutrals and 147 Damaging. In SNP & GO there were 90 neutrals and 62 damaging. In PMUT there were 62 neutrals and 90 damaging. In Predict SNP were 9 neutrals and 143 damaging. MetaSNP were 17 neutrals and 135 disease causing. Out of a total of 152 SNPs, 34 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-Mutant 2.0, MUpro and INPS-MD. SNPs were considered as destabilizing if two or more algorithms showed a decrease in stability upon mutation. Out of 34 high risk nsSNPs I-Mutant 3.0 analysis showed that, 25 nsSNPs decreased stability ( $\Delta\Delta G < 0$ ), whereas 9 nsSNPs increased stability ( $\Delta\Delta G > 0$ ). MUpro and INPS-MD analysis showed that, 28 and 33 nsSNPs were found to decrease protein stability, respectively. In total Out of 34 high risk nsSNPs, 30 nsSNP were predicted to decrease the stability, 4 nsSNPs were found to increase protein stability. H41L was found to increase protein stability by the three tools.

### 3.9. Analyzing the Effect of nsSNPs on 3D Structure of the Proteins and Physicochemical Properties

The common physicochemical consequence of this nsSNPs were alteration in charge, size and hydrophobicity between wild-type and mutant residues which might lead to loss of interactions with other molecules and loss of hydrogen bonds in the core of the protein and as a result disturb correct folding. Table 6 & Fig 4-9.

Table 4. Missense SNPs in M<sup>pro</sup> predicted to be damaging/ neutral

Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP	Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP
H41A	Damaging	Neutral	Neutral	Damaging	Damaging	F140I	Damaging	Neutral	Neutral	Damaging	Damaging
H41C	Damaging	Neutral	Damaging	Damaging	Damaging	F140K	Damaging	Neutral	Damaging	Damaging	Damaging
H41D	Damaging	Damaging	Neutral	Damaging	Damaging	F140L	Damaging	Neutral	Neutral	Damaging	Damaging
H41E	Damaging	Damaging	Neutral	Damaging	Damaging	F140M	Damaging	Neutral	Neutral	Damaging	Damaging
H41F	Damaging	Neutral	Damaging	Damaging	Damaging	F140N	Damaging	Neutral	Damaging	Damaging	Damaging
H41G	Damaging	Damaging	Damaging	Damaging	Damaging	F140P	Damaging	Neutral	Damaging	Damaging	Damaging
H41I	Damaging	Neutral	Damaging	Damaging	Damaging	F140Q	Damaging	Neutral	Damaging	Damaging	Damaging
H41K	Damaging	Damaging	Neutral	Damaging	Damaging	F140R	Damaging	Neutral	Damaging	Damaging	Damaging
H41L	Damaging	Damaging	Damaging	Damaging	Damaging	F140S	Damaging	Neutral	Damaging	Damaging	Damaging
H41M	Damaging	Neutral	Neutral	Damaging	Damaging	F140T	Damaging	Neutral	Damaging	Damaging	Damaging
H41N	Damaging	Neutral	Neutral	Damaging	Damaging	F140V	Damaging	Neutral	Neutral	Damaging	Damaging
H41P	Damaging	Damaging	Neutral	Damaging	Damaging	F140W	Damaging	Neutral	Neutral	Damaging	Damaging
H41Q	Damaging	Damaging	Neutral	Damaging	Damaging	F140Y	Damaging	Neutral	Neutral	Damaging	Damaging
H41R	Damaging	Neutral	Neutral	Damaging	Damaging	G143A	Damaging	Neutral	Neutral	Damaging	Damaging
H41S	Damaging	Damaging	Neutral	Damaging	Damaging	G143C	Damaging	Damaging	Damaging	Damaging	Damaging
H41T	Damaging	Neutral	Neutral	Damaging	Damaging	G143D	Damaging	Damaging	Damaging	Damaging	Damaging
H41V	Damaging	Damaging	Neutral	Damaging	Damaging	G143E	Damaging	Damaging	Damaging	Damaging	Damaging
H41W	Damaging	Neutral	Damaging	Damaging	Damaging	G143F	Damaging	Damaging	Damaging	Damaging	Damaging
H41Y	Damaging	Neutral	Damaging	Damaging	Damaging	G143H	Damaging	Neutral	Damaging	Damaging	Damaging
F140A	Damaging	Neutral	Damaging	Damaging	Damaging	G143I	Damaging	Damaging	Damaging	Damaging	Damaging
F140C	Damaging	Neutral	Damaging	Damaging	Damaging	G143K	Damaging	Damaging	Damaging	Damaging	Damaging
F140D	Damaging	Neutral	Damaging	Damaging	Damaging	G143L	Damaging	Neutral	Damaging	Damaging	Damaging
F140E	Damaging	Neutral	Damaging	Damaging	Damaging	G143M	Damaging	Neutral	Damaging	Damaging	Damaging
F140G	Damaging	Neutral	Damaging	Damaging	Damaging	G143L	Damaging	Neutral	Damaging	Damaging	Damaging
F140H	Damaging	Neutral	Damaging	Damaging	Damaging	G143M	Damaging	Neutral	Damaging	Damaging	Damaging

Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP	Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	C145Y	Damaging	Damaging	Damaging	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163A	Damaging	Damaging	Neutral	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163C	Damaging	Damaging	Damaging	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163D	Damaging	Damaging	Neutral	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163E	Damaging	Damaging	Neutral	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163F	Damaging	Damaging	Damaging	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163G	Damaging	Damaging	Damaging	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163I	Damaging	Damaging	Damaging	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163K	Damaging	Damaging	Neutral	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163L	Damaging	Damaging	Damaging	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163M	Damaging	Damaging	Neutral	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163N	Damaging	Damaging	Neutral	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163P	Damaging	Damaging	Neutral	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163Q	Damaging	Damaging	Neutral	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163R	Damaging	Damaging	Neutral	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163S	Damaging	Damaging	Neutral	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163T	Damaging	Damaging	Neutral	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163V	Damaging	Damaging	Neutral	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H163W	Damaging	Damaging	Damaging	Damaging	Damaging
G143M	Damaging	Neutral	Damaging	Damaging	Damaging	H163Y	Damaging	Damaging	Damaging	Damaging	Damaging
G143L	Damaging	Neutral	Damaging	Damaging	Damaging	H164A	Damaging	Neutral	Damaging	Damaging	Damaging
C145S	Damaging	Damaging	Neutral	Damaging	Damaging	H164C	Damaging	Neutral	Damaging	Damaging	Damaging
C145T	Damaging	Damaging	Neutral	Damaging	Damaging	H164D	Damaging	Neutral	Damaging	Damaging	Damaging
C145V	Damaging	Damaging	Neutral	Damaging	Damaging	H164E	Damaging	Neutral	Neutral	Damaging	H164E
C145W	Damaging	Damaging	Damaging	Damaging	Damaging	H164F	Damaging	Neutral	Damaging	Damaging	H164F

Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP	Mutation	PROVEAN	SNP&GO	PMUT	PredictSNP	MetaSNP
H164G	Damaging	Neutral	Damaging	Damaging	Damaging	E166Q	Damaging	Neutral	Neutral	Damaging	Damaging
H164I	Damaging	Neutral	Damaging	Damaging	Damaging	E166R	Damaging	Neutral	Neutral	Damaging	Damaging
H164K	Damaging	Neutral	Damaging	Damaging	Damaging	E166S	Damaging	Neutral	Neutral	Damaging	Damaging
H164L	Damaging	Neutral	Damaging	Damaging	Damaging	E166T	Damaging	Neutral	Neutral	Damaging	Damaging
H164M	Damaging	Neutral	Damaging	Damaging	Damaging	E166V	Damaging	Damaging	Damaging	Damaging	Damaging
H164N	Damaging	Neutral	Damaging	Damaging	Damaging	E166W	Damaging	Neutral	Damaging	Damaging	Damaging
H164P	Damaging	Neutral	Damaging	Damaging	Damaging	E166Y	Damaging	Neutral	Damaging	Damaging	Damaging
H164Q	Damaging	Neutral	Neutral	Neutral	Damaging	Q189A	Damaging	Neutral	Neutral	Neutral	Neutral
H164R	Damaging	Neutral	Damaging	Damaging	Damaging	Q189C	Damaging	Neutral	Damaging	Neutral	Damaging
H164S	Damaging	Neutral	Damaging	Damaging	Damaging	Q189D	Damaging	Neutral	Damaging	Neutral	Damaging
H164T	Damaging	Neutral	Damaging	Damaging	Damaging	Q189E	Neutral	Neutral	Neutral	Damaging	Neutral
H164V	Damaging	Neutral	Damaging	Damaging	Neutral	Q189F	Damaging	Neutral	Damaging	Damaging	Neutral
H164W	Damaging	Neutral	Damaging	Damaging	Damaging	Q189G	Damaging	Neutral	Damaging	Damaging	Neutral
H164W	Damaging	Neutral	Damaging	Damaging	Damaging	Q189H	Damaging	Neutral	Neutral	Damaging	Neutral
E166A	Damaging	Neutral	Damaging	Damaging	Damaging	Q189I	Damaging	Neutral	Damaging	Damaging	Neutral
E166C	Damaging	Neutral	Damaging	Damaging	Damaging	Q189K	Neutral	Neutral	Neutral	Neutral	Neutral
E166D	Damaging	Neutral	Neutral	Damaging	Damaging	Q189L	Damaging	Neutral	Damaging	Damaging	Neutral
E166F	Damaging	Damaging	Damaging	Damaging	Damaging	Q189M	Damaging	Neutral	Damaging	Damaging	Neutral
E166G	Damaging	Neutral	Damaging	Damaging	Damaging	Q189N	Damaging	Neutral	Neutral	Neutral	Neutral
E166H	Damaging	Neutral	Neutral	Damaging	Damaging	Q189P	Neutral	Neutral	Neutral	Neutral	Neutral
E166I	Damaging	Neutral	Damaging	Damaging	Damaging	Q189R	Neutral	Neutral	Neutral	Damaging	Damaging
E166K	Damaging	Neutral	Neutral	Damaging	Damaging	Q189S	Neutral	Neutral	Damaging	Neutral	Neutral
E166L	Damaging	Neutral	Damaging	Damaging	Damaging	Q189T	Damaging	Neutral	Neutral	Neutral	Neutral
E166M	Damaging	Neutral	Damaging	Damaging	Damaging	Q189V	Damaging	Neutral	Neutral	Damaging	Neutral
E166N	Damaging	Neutral	Neutral	Damaging	Damaging	Q189W	Damaging	Neutral	Damaging	Damaging	Neutral
E166P	Damaging	Neutral	Damaging	Damaging	Damaging	Q189Y	Damaging	Neutral	Damaging	Damaging	Neutral

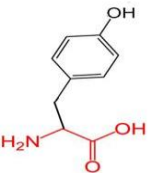
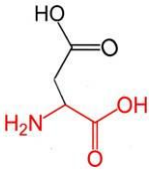
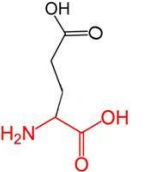
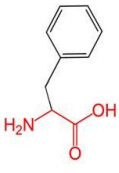
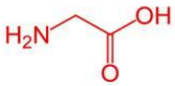
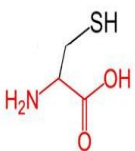
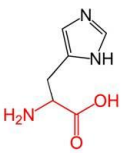
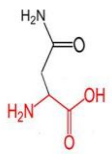
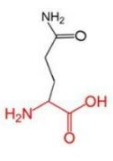
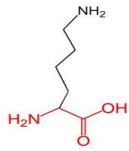
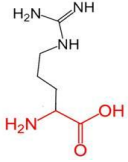
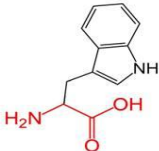
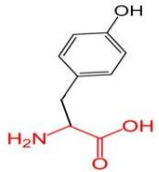


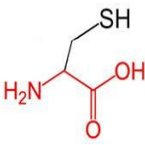
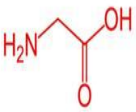
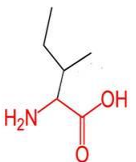
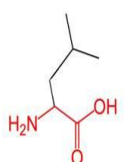
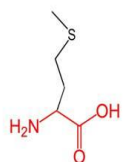
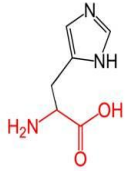
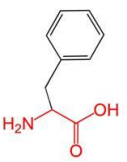
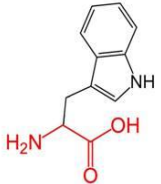
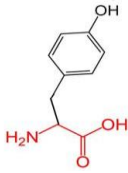
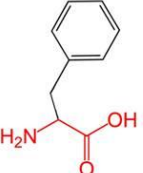
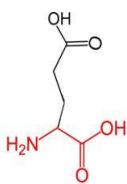
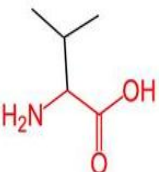
**Table 5.** Prediction result of Protein Stability

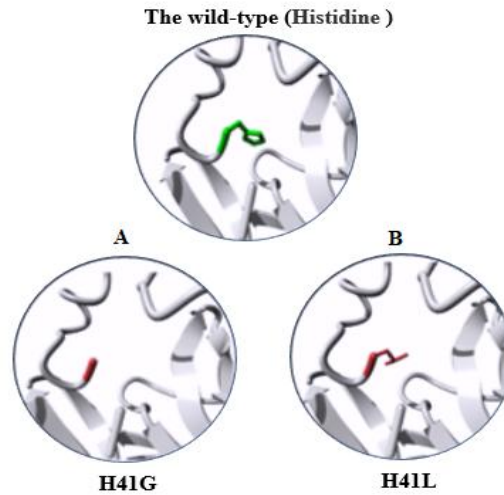
<b>Mutation</b>	<b>I-Mutant2</b>	<b>MUPRO</b>	<b>INPS-MD</b>
H41G	Decrease	Decrease	Decrease
H41L	Increase	Increase	Increase
G143C	Decrease	Decrease	Decrease
G143D	Decrease	Decrease	Decrease
G143E	Decrease	Decrease	Decrease
G143F	Decrease	Decrease	Decrease
G143I	Decrease	Decrease	Decrease
G143K	Decrease	Increase	Decrease
G143Q	Decrease	Decrease	Decrease
G143R	Decrease	Decrease	Decrease
G143V	Decrease	Decrease	Decrease
G143W	Decrease	Decrease	Decrease
G143Y	Decrease	Decrease	Decrease
C145D	Decrease	Decrease	Decrease
C145E	Decrease	Decrease	Decrease
C145F	Decrease	Decrease	Decrease
C145G	Decrease	Decrease	Decrease
C145H	Decrease	Decrease	Decrease
C145K	Decrease	Decrease	Decrease
C145N	Decrease	Decrease	Decrease
C145Q	Decrease	Decrease	Decrease
C145R	Decrease	Decrease	Decrease
C145W	Decrease	Decrease	Decrease
C145Y	Decrease	Decrease	Decrease
H163C	Increase	Decrease	Decrease
H163G	Decrease	Increase	Decrease
H163F	Increase	Decrease	Decrease
H163I	Increase	Increase	Decrease
H163L	Increase	Decrease	Decrease
H163M	Increase	Increase	Decrease
H163W	Decrease	Decrease	Decrease
H163Y	Increase	Increase	Decrease
E166F	Increase	Decrease	Decrease
E166V	Increase	Decrease	Decrease

**Table 6.** Schematic structures of the wild-type residue and mutant residue amino acid for each Mutation

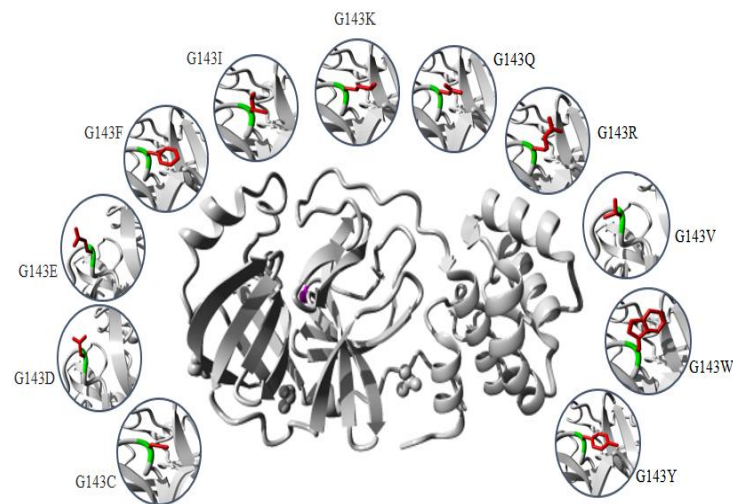
Residue	Wild-type Structure	Mutant residue Structure		Properties
H41G/ H41L	 His	 Gly	 Leu	<ul style="list-style-type: none"> <li>-This mutation resulted in a change of the wild-type histidine at position 41 to glycine and leucine respectively.</li> <li>-The mutant residues are smaller and more hydrophobic than the wild-type residue.</li> <li>-These mutations will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding.</li> </ul>
G143C/ G143F/ G143I /G143V/ G143W		 Cys	 Phe	<ul style="list-style-type: none"> <li>This mutation resulted in a change of the wild-type glycine at position 143 to cysteine, phenylalanine, isoleucine, valine and tryptophan respectively</li> <li>-The mutant residue is bigger and more hydrophobic than the wild-type residue.</li> </ul>
		 Iso	 Val	
G143D/ G143E	 Gly	 Asp	 Glu	<ul style="list-style-type: none"> <li>-This mutation resulted in a change of the wild-type glycine at position 143 to aspartic acid and glutamic acid respectively</li> <li>-The mutant residue is bigger and less hydrophobic than the wild-type residue.</li> <li>-The wild-type residue charge was neutral, the mutant residue charge is negative.</li> </ul>
G143K/ G143R		 Lys	 Arg	<ul style="list-style-type: none"> <li>-This mutation resulted in a change of the wild-type glycine at position 143 to lysine and arginine respectively.</li> <li>-The mutant residue is bigger and less hydrophobic than the wild-type residue.</li> <li>-The wild-type residue charge was neutral, the mutant residue charge is positive.</li> <li>- The mutation introduces a charge at this position, this can cause repulsion between the mutant residue and neighboring residues</li> </ul>
G143Q		 Gln		<ul style="list-style-type: none"> <li>-This mutation resulted in a change of the wild-type Glycine at position 143 to Glutamine.</li> <li>-The mutant residue is bigger and less hydrophobic than the wild-type residue.</li> </ul>

G143Y			-This mutation resulted in a change of the wild-type Glycine at position 143 to Tyrosine. -The mutant residue is bigger than the wild-type residue.		
C145D/ C145E		 Asp	 Glu	-This mutation resulted in a change of the wild-type Cysteine at position 145 to aspartic acid and glutamic acid respectively. -The mutant residue is bigger and less hydrophobic than the wild-type residue -The wild-type residue charge was neutral; the mutant residue charge is negative.	
C145F			-This mutation resulted in a change of the wild-type cysteine at position 145 to phenylalanine. -The mutant residue is bigger than the wild-type residue.		
C145G			-This mutation resulted in a change of the wild-type cysteine at position 145 to glycine. -The mutant residue is smaller and less hydrophobic than the wild-type residue.		
C145H/ C145N/ C145Q	 Cys	 His	 Asn	 Gln	-This mutation resulted in a change of the wild-type cysteine at position 145 to histidine, asparagine and glutamine. -The mutant residue is bigger and less hydrophobic than the mutant residue.
C145K/ C145R		 Lys	 Arg	-This mutation resulted in a change of the wild-type cysteine at position 145 to lysine and arginine respectively. -The mutant residue is bigger and less hydrophobic than the wild-type residue.	
C145W		 Try	-This mutation resulted in a change of the wild-type cysteine to tryptophan at position 145. -The mutant residue is bigger than the wild-type residue.		
C145Y		 Tyr	-This mutation resulted in a change of the wild-type cysteine to tyrosine at position 145. -The mutant residue is bigger than the wild-type residue. -The wild-type residue is more hydrophobic than the mutant residue.		

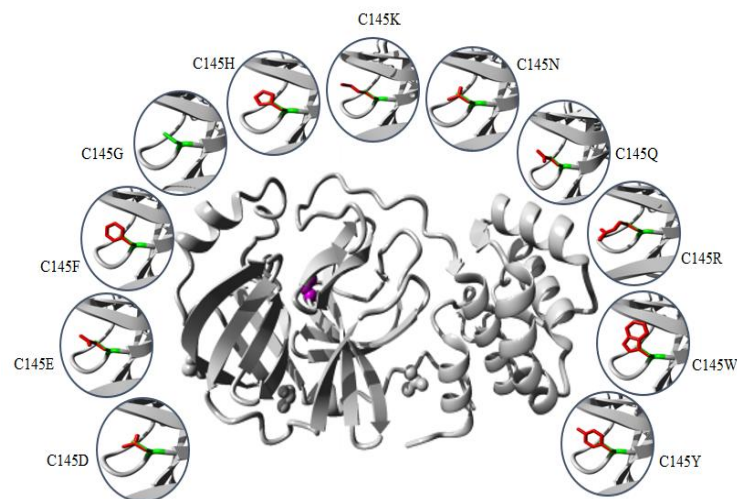
H163C/ H163G/ H163I/ H163L/ H163M		 Cys	 Gly	<p>-These mutations resulted in a change of the wild-type histidine at position 163 to cysteine, glycine, isoleucine, leucine and methionine respectively</p> <p>-The mutant residue is smaller and more hydrophobic than the wild-type residue.</p>	
		 Iso	 Leu	 Met	
H163F	 His	 Phe			<p>-This mutation resulted in a change of the wild-type histidine to phenylalanine at position 163.</p> <p>-The mutant residue is more hydrophobic than the wild-type residue.</p>
H163W/ H163Y		 Try	 Tyr		<p>-The mutant residue is bigger than the wild-type residue.</p> <p>-The mutant residue is more hydrophobic than the wild-type residue.</p>
E166F		 Phe			<p>This mutation resulted in a change of the wild-type Glutamic Acid at position 166 to Phenylalanine</p> <p>-The mutant residue is bigger and more hydrophobic than the wild-type residue</p> <p>The wild-type residue charge was negative; the mutant residue charge is neutral.</p> <p>- The wild-type residue forms a hydrogen bond with histidine at position 172.</p> <p>-The charge of the wild-type residue is lost by this mutation. This can cause loss of interactions with other molecules.</p>
E166V	 Glu	 Val			<p>-This mutation resulted in a change of the wild-type Glutamic Acid at position 166 to Valine.</p> <p>-The mutant residue is smaller and more hydrophobic than the wild-type residue.</p> <p>-The wild-type residue charge was negative; the mutant residue charge is neutral.</p> <p>- The wild-type residue forms a hydrogen bond with histidine at position 172.</p> <p>-The charge of the wild-type residue is lost by this mutation. This can cause loss of interactions with other molecules</p>



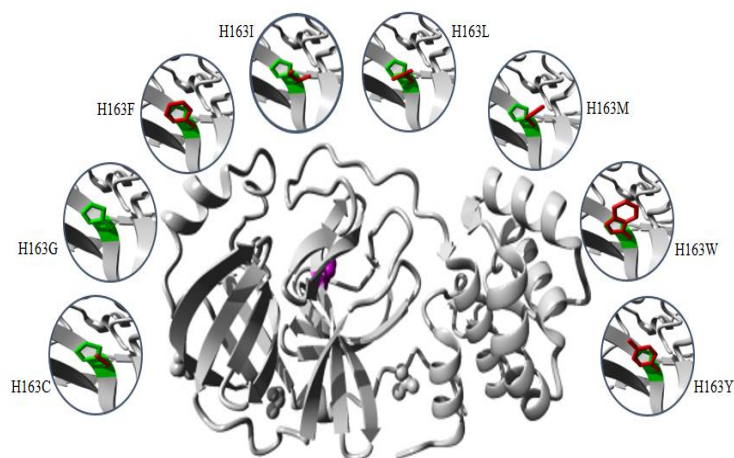
**Figure 4.** 3D structure of the wild-type Histidine substitution. A) the mutant residue H41G. B) the mutant residue H41L



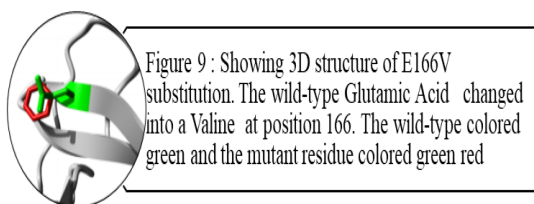
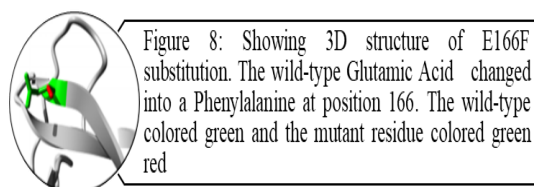
**Figure 5.** Structures of the protein (center) is colored grey, the side chain of the mutated residue is colored magenta and shown as small balls. 3D structures of the substitutions of wild-type Glycine at position 143. The wild-type colored green and the mutant residue colored green red



**Figure 6.** Structures of the protein (center) is colored grey, the side chain of the mutated residue is colored magenta and shown as small balls. 3D structures of the substitutions of wild-type Cysteine at position 145. The wild-type colored green and the mutant residue colored green red



**Figure 7.** Structures of the protein (center) is colored grey, the side chain of the mutated residue is colored magenta and shown as small balls. 3D structures of the substitutions of wild-type Histidine at position 163. The wild-type colored green and the mutant residue colored green red



**Figure 8&9**

## 4. Discussion

The SARS-COV-2 Mpro play critical role in the maturation and posttranslational processing of the viral replicase protein and thus propagation of the SARS-COV-2 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 [30-33]. Our study displayed that Mpro is conserved in all SARS-COV-2. It is highly similar to SARS-CoV Mpro, with only 12 residues different. These differences may affect Mpro 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 structure-based drug design process, identification of ligand-binding sites in the target protein is the first step. In the literature, several crystal structures of SARS-COV-2 Mpro in complex with different inhibitor are deposited in the protein data bank. In order to predict the ligand-binding in CoV-2 Mpro, firstly the Protein Data Bank (PDB)

structures were retrieved and verified. Our analysis shows (PDB: 5R7Y) is the best structures identity with the CoV-2 Mpro sequence and was chosen for ligand-binding sites identification. Ten potential binding sites were predicted using COACH. The ligand AZP have higher C-score possesses 21 binding residues (25, 26, 27, 41, 49, 140, 141, 142, 143, 144, 145, 163, 164, 165, 166, 168, 172, 187, 189, 190 and 192). 8 Out of this residue were identified as functional residues (HIS41, PHE140, GLY143, CYS145, HIS163, HIS 164, GLU166 and GLN189) participating in various physicochemical interaction at the active site of Mpro. particularly GLY145, HIS163, HIS 164 and GLU166 contribute significantly to strong electrostatic interactions in the active site. All these residues were subjected to conservation and missense SNP analysis. Single-nucleotide polymorphisms (SNPs) in protein functional 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.

The catalytic dyad of Mpro, located at the cleft of domain I and II, includes the conserved residues HIS41 and CYS145, which are essential for catalytic activity. GLY143,

HIS163, Glu166 residues are located at domain II and involved in the deeply buried subsite (S1), this domain is important for binding of substrate binding. Mutation of this residues might change the 3D conformation of the substrate-binding subsite and disturb the interaction and as such obstruct substrate binding. Except for PHE140, HIS164 and GLN189, all rest of residues showed to be highly risk damaging SNPs.

HIS41 involved in the catalytic centre of SARS-CoV Mpro, play a part in various physicochemical interaction. Based on conservation analysis, HIS41 is a highly conserved and buried in the core of domain I, thus mutation in the HIS41 residue is probably disrupting structure of this domain. The current study predicted that H41G and H41L are highly damaging SNPs destabilizing protein stability. H41 G and H41L resulted in a change of the wild-type histidine at position 41 to glycine and leucine respectively. The mutant residues are smaller and more hydrophobic than the histidine, these differences might disturb the conformation structure of this domain leading to loss of the proteolytic activity of the SARS-COV-2 Mpro. Table 6 & Fig. 4.

The GLY143 is located at the buried the oxyanion-binding loop involved in the deeply buried subsite (S1) of Domain II, has an electrostatic interactions and hydrogen bonds with other residues. ConSurf analysis showed that, GLY143 is an important residue for the SARS-COV-2 Mpro structure. Glycines is most flexible of all amino acids residues. This flexibility might be necessary for the protein's function. Any mutation in the GLY143 residue would disturb the required flexibility at the oxyanion-binding loop and therefore disturb a special backbone conformation which might be required at this position. Results of the current study predicted 11 nsSNPs are highly damaging result in the decrease of the protein stability. These mutations introduce some amino acids with different size, hydrophobicity and charge, thus these mutations are probably disrupting structure of this domain and decrease the enzymatic function of the SARS-COV-2 Mpro. Table 6 & Fig. 5.

The CYS145 involved in the catalytic center of SARS-CoV Mpro, participating in different physicochemical interaction. Based on conservation analysis, CYS145 is a highly conserved and buried in the core of domain II. Being important structural residue, any substitution in the CYS145 residue would disturb the local structure of domain II. Missense SNP analysis predicted 11 nsSNPs are highly damaging result in the decrease of the protein stability. These mutations introduce some amino acids with different size, hydrophobicity and charge, thus these mutations might cause loss of hydrophobic interactions with other molecules, disrupting structure of this domain and leading to loss or reduce of the proteolytic activity of the SARS-COV-2 Mpro. Table 6 & Fig. 6.

The HIS163 is buried in the core of a domain II, involved in several physicochemical interactions with other residues.

Conservation analysis showed that, HIS163 is an important residue for the SARS-COV-2 Mpro structure. Results of this study predicted 8 nsSNPs are highly damaging and destabilizing of the protein. The following project hope summary report shows the structural and physicochemical changes induced by these mutations, mainly as variation in the size, charge and hydrophobicity of the final protein. The wild-type histidine forms a hydrogen bond with: Tyrosine at position 161, thus these mutations might cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding of the SARS-COV-2 Mpro. Table 6 & Fig. 7.

The Glu166 located in the subsite (S1) of Domain II, involved in several physicochemical interactions with other residues. Conservation analysis showed that, Glu166 is an important functional residue; highly conserved and exposed in the surface of domain II, thus mutation in the Glu166 residue is probably disrupting function of this domain. Results of the current study predicted E166F and E166V are highly damaging SNPs decrease the protein stability. These mutations resulted in a change of the wild-type glutamic acid at position 166 to phenylalanine and valine respectively. Phenylalanine is bigger and more hydrophobic than glutamic acid, where valine is smaller and more hydrophobic than glutamic acid. Glutamic acid is negatively charged residue forms a hydrogen bond with histidine at position 172, where phenylalanine and valine residues neutrally charged. These differences in size, hydrophobicity and charge might affect hydrogen bond formation, lead to loss of interactions with other molecules, which can disturb the enzymatic function of the SARS-COV-2 Mpro. Table 6 & Fig. 8-9.

Previous mutagenesis study in other CoV Mpro established the importance of His41, Cys145, and Glu166 in the substrate - binding site for keeping the proteolytic activity and enzymatic function of the SARS - CoV Mpro [34,35]. Moreover site-directed mutagenesis studies in other CoV Mpro revealed that substitution of either His41, Cys145 or H163 residues resulted in complete loss of proteolytic activity [36,37]. Results of the present study determined that HIS41, CYS145, H163 and E166 are highly damaging missense SNPs decreases protein stability except H41L, H163I, H163M, and H163Y SNPs, which increase the protein stability.

Due to the urgent need for effective treatments, use of repurposed existing antiviral drugs approved and successes for treatment of other viral infections such as HIV, SARS-CoV-1 and MERS-CoV is somewhat promising. Based on literature search, (HIV-1) protease inhibitors (lopinavir/ritonavir) and new nucleoside analogues (remdesivir) are now most commonly protease inhibitor proposed for potential treatment of SARS-CoV-2 based on previously records of their therapeutic efficacy in SARS-COV and MERS-COV [38,39].

Recent computational studies showed that SARS-CoV-2 Mpro structure had four residues for lopinavir binding (M49,

M165, P168, and Q189), nine residues for ritonavir binding (L27, H41, M49, F140, N142, G143, H164, M165, and E166) and seven residues for remdesivir binding (Thr25, Thr26, His41, Phe140, Asn142, Cys145, His163) [40,41]. In contrast, our study identifies eight residues (HIS41, Phe140, GLY143, CYS145, HIS163, HIS164, GLU166 and Q189) as active functional sites showed potential binding activities. In addition to that, some nsSNPs related to these residues predicted to be highly damaging to the protein, which would make these residues suitable sites for drug targeting. Moreover, a newly in vitro experimental study shows powerful activity of remdesivir and lopinavir against SARS-CoV-2 [42,43]. Successful case reports describing the partially effective of remdesivir for the early confirmed patients of SARS-CoV-2 infection [44-46]. Despite these findings, a clinical randomized controlled trial performed on adult hospitalized patients with severe SARS-CoV-2, no benefits of lopinavir/ritonavir or Remdesivir treatment were observed [47-49]. This variability in effectiveness of HIV protease inhibitors might be relies on that HIV protease is belongs the aspartic protease enzymatic class, whereas SARS-CoV-2 Mpro belongs to the cysteine protease enzymatic class, do not contain a C2-symmetric pocket, which is the target of HIV protease inhibitors. Accordingly, on the basis of this evidences, the current antiviral drugs need to be modified in order to be effective against SARS-CoV-2 infection.

## 5. Conclusions

The computational approach can significantly reduce the financial cost and time for drug development against the SARS-COV-2 Mpro. In this study, five residues (HIS41, GLY143, CYS145, HIS163 and GLU166) predicted as active functional sites showed potential binding activities. Thirty four high-risk damaging nsSNPs of SARS-COV-2 Mpro were related to these residues resulting in stability and structural changes of proteins and changes in the interactions between domains, thus affecting the function of CoV-2 Mpro. These characteristics provide them the promising to be target sites for the fresh generation inhibitors to work with and overcome the current drugs ineffectiveness.

Combining methods such as site-directed mutagenesis assay and nuclear magnetic resonance potentially help scientists to better understanding the effect of this predicted SNPs on SARS-COV-2 Mpro structural stability and functional offering basis to develop specific drugs for inhibiting this virus.

## ACKNOWLEDGEMENTS

The authors acknowledge Dr. Ahmed Abdelbagi Hammad for language editing and proofreading of the manuscript.

## 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](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, Wadhvani P, Mesters JR, Hilgenfeld R. Coronavirus Main Proteinase (3CL<sup>pro</sup>) Structure: Basis for Design of Anti-SARS Drugs. *Science*. 2003; 300(5626): 1763-7.
- [13] Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*. 2020.
- [14] Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol*. 2005; 3(10): e324.



- [15] Ghosh AK, Xi K, Grum-Tokars V, Xu X, Ratia K, Fu W, et al. Structure-based design, synthesis, and biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors. *Bioorg Med Chem Lett*. 2007; 17(21): 5876-80.
- [16] 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.
- [17] Pettersen, EF, Goddard, TD, Huang, CC et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25: 1605-1612.
- [18] 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.
- [19] 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.
- [20] Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010; 38: W529–33.16.
- [21] 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.
- [22] Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human Damaging-related mutations in proteins. *Hum Mutat*. 2009; 30(8): 1237-44.
- [23] Orozco CF-CJLGLZIPXdlCM. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*. 2005; 21(14): 3176–8.
- [24] Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of Damaging-related mutations. *PLoS Comput Biol*. 2014; 10(1): e1003440-e.
- [25] Capriotti, E.; Altman, R.B.; Bromberg, Y. Collective judgment predicts Damaging-associated single nucleotide variants. *BMC Genom*. 2013, 14, S2.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable Damagings. An e-Science approach with life scientist friendly interfaces. *BMC bioinformatics*. 2010; 11: 548.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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 Mpro protease: basis for the in vitro screening of anti-SARS drugs. *FEBS letters*. 2004; 574(1-3): 131-7.
- [35] Hegyi A, Friebe A, Gorbalenya AE, Ziebuhr J. Mutational analysis of the active centre of coronavirus 3C-like proteases. *Journal of General Virology*. 2002; 83(3): 581-93.
- [36] 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.
- [37] 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.
- [38] Chan KS, Lai ST, Chu CM, Tsui E, Tam CY, Wong MM, et al. Treatment of severe acute respiratory syndrome with lopinavir/ritonavir: a multicentre retrospective matched cohort study. *Hong Kong medical journal = Xianggang yi xue za zhi*. 2003; 9(6): 399-406.
- [39] Chu CM, Cheng VC, Hung IF, Wong MM, Chan KH, Chan KS, et al. Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax*. 2004; 59(3): 252-6.
- [40] Khan SA, Zia K, Ashraf S, Uddin R, Ul-Haq Z. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *Journal of Biomolecular Structure and Dynamics*. 2020: 1-10.
- [41] Nutho B, Mahalapbutr P, Hengphasatporn K, Pattarangoon NC, Simanon N, Shigeta Y, et al. Why Are Lopinavir and Ritonavir Effective against the Newly Emerged Coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms. *Biochemistry*. 2020; 59(18): 1769-79.
- [42] Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research*. 2020; 30(3): 269-71.
- [43] Kang CK, Seong MW, Choi SJ, et al. In vitro activity of lopinavir/ritonavir and hydroxychloroquine against severe acute respiratory syndrome coronavirus 2 at concentrations achievable by usual doses [published online ahead of print, 2020 May 29]. *Korean J Intern Med*. 2020; 10. doi.org/10.3904/kjim.2020.157.

- [44] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med.* 2020; 382(10): 929-36.
- [45] Kujawski SA, Wong KK, Collins JP, Epstein L, Killerby ME, Midgley CM, et al. Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States. *Nature Medicine.* 2020.
- [46] Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *New England Journal of Medicine.* 2020; 382.
- [47] Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. *New England Journal of Medicine.* 2020; 382(19): 1787-99.
- [48] Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet.* 2020; 395(10236): 1569-78.
- [49] Li Y, Xie Z, Lin W, Cai W, Wen C, Guan Y, et al. An exploratory randomized controlled study on the efficacy and safety of lopinavir/ritonavir or arbidol treating adult patients hospitalized with mild/moderate COVID-19 (ELACOI). *medRxiv.* 2020:2020.03.19.20038984.