

Binding site identification of COVID-19 main protease 3D structure by homology modeling

Marion Adebisi, Oludayo O. Olugbara

ICT and Society Research Group, South Africa Luban Workshop, Durban University of Technology, Durban, South Africa

Article Info

Article history:

Received May 8, 2020

Revised Aug 21, 2020

Accepted Oct 14, 2020

Keywords:

3D structure

Binding site

COVID-19 protease

Homology model

Main protease

ABSTRACT

The influx of coronavirus in 2019 (COVID-19) has recorded millions of infection cases with several deaths worldwide. There is no effective treatment, but recent studies have shown that its enzymes maybe considered as potential drug target. The purpose of this work was to identify the binding site in-silico and present the 3D structure of COVID-19 main-protease (Mpro) by homology modeling through multiple alignment followed by optimization and validation. The modeling was done by Swiss-Model template library. The obtained homotrimer oligo-state model was verified for reliability using PROCHECK, Verify3D, MolProbity and QMEAN. HHBlits software was used to determine structures that matched the target sequence by evolution. Structure quality verification through Ramachandran plot showed an abundance of 99.3% of amino acid residues in allowed regions while 0.1% in disallowed region. The Verify3D rated the structure a 90.87% PASS of residues having an average 3D-1D score of at least 0.2, which validates a good environment profile for the Mpro model. The features of the secondary structure indicated that the structure contains 32.05% α -helix and 37.17% random coil with 25.92 extended strand. The result of this study suggests that blocking expression of this protein may constitute an efficient approach for infection transmission blockage.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Marion Adebisi

ICT and Society Research Group, South Africa Luban Workshop

Durban University of Technology

P.O Box 1334, Durban 4000, South Africa

Email: mariona@dut.ac.za

1. INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is often called COVID-19 is a novel strain of coronavirus that was first discovered in late 2019 at Wuhan in China as confirmed by the Chinese authority on 7th January 2020. It is responsible for a viral outbreak of respiratory illness known as COVID-19 as christened by the world health organization (WHO). The virus has spread rapidly to every continent of the world and as of 14th April 2020, about 1,918, 855 cases with 119, 588 deaths across 185 nations had been confirmed globally by the center for systems science and engineering (CSSE) at Johns Hopkins University (JHU) [1]. There is still neither known effective drug nor vaccine in use or targeted therapies for its cure besides public health and social measures such as observing regular hygiene, social distancing, wearing of masks and avoiding large crowd that have been prescribed to hopefully halt its persistent spread. However, there is the need for greater understanding of its origin, epidemiology, evolution, transmission, pathogenesis and diagnosis. The discovery of vaccines, antivirals, laboratory biosafety and issues such as how immune evasion occurs in its hosts cannot be over emphasized. The transmission rate of

the virus is in exponential proportion hourly while death rate is fizzy and alarming [2]. More transmissions and symptoms are revealed daily as the virus evolves with depth of unexplainable mutation. Research scientists have clearly situated that successful diagnosis, treatment and prevention of COVID-19 are highly dependent on a deep understanding of the role of complex structures of glycosylation [3]. Moreover, isolation of protease binding site in this viroid can be the breaking point for understanding and blocking mutations responsible for fast and easy transmission of the immunity of hosts. A binding site is a region on a protein, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) to which other specific molecules and ions bind and it exhibits chemical specificity and affinity. Specificity is a measure of the types of ligand that will bind while affinity is a measure of the strength of a chemical bond [4]. Hence, an understanding of the complex structures of COVID-19 enzymes, proteases and exhuming their active site can give a useful insight to the mechanisms deployed for brisk cross-species, host-to-host transmission and more interestingly an understanding of the coverage area of amino acids involved in cavity building. This would help to establish the active sites of such proteins that may assist in speedy vaccine discovery and drug development.

The only public domain of experimentally generated atom file from this specific coronavirus at present is the COVID-19 coronavirus 3CL hydrolase (Mpro), which is also known as COVID-19 main protease (Mpro) with evidence of experimental analysis [5]. The first 3D structures were an outcome of coronavirus related outbreak of 2003 that has culminated into the existence of over 200 SARS protein database (PDB) experimental structures [6]. These related proteins make conservation of vital structural information that may further the understanding of the current strain of coronavirus. This may lead to new treatment, discovery of vaccines and drug development that could facilitate the containment of the virus outbreak. However, coronavirus 3CL hydrolase (Mpro) enzyme is essential for proteolytic maturation of the virus. Investigating this protein may expose a promising target for small molecule drug discovery that is capable of inhibiting cleavage of the viral polyprotein and prevent the spread of its infection [7]. Evolutionarily, features of the enzymes are preserved [8-10], and a comprehensive analysis of such features can help gain a deeper understanding of the protein functions and its roles. The analysis of unknown atomic resolution model of SARS-CoV-2 viral protein (target protein) from its amino acid sequence and 3D structure of a related homologous protein (template protein) are presented. A comparative model is highly dependent on identifying at least one known protein structure that may resemble the structure of a query sequence (target protein), which will then produce an alignment that maps the residues in the query sequence to residues in the template sequence. Research has shown that three dimensional (3D) protein structure is more conserved than protein sequences by evolution among homologues. However, sequences below the range of 20% sequence identity usually possess structures that are highly different [11]. Homologous proteins that occur naturally contain similar structures just as related proteins contain similar sequences [8]. Since 3D structures of proteins in a family are jealously preserved than their sequences [9], protein structures are more conserved than DNA sequences. Detectable levels of sequence similarity usually imply significant structural similarity [10]. Sequence alignment and template structure are then deployed to produce a structural model of the target.

The first important contribution of this study to the existing literature is the novel identification of the binding site of COVID-19 main protease and surface cavity of binding pocket of the virus using various simulation and analysis methods in the field of bioinformatics for possible target discovery. The other essential contribution is the presentation of 3D structure of COVID-19 main protease. Since the influx of the virus, many studies have been carryout in different domains, but a lot is not known about its protease yet, which presents an apparent chasm in the literature that this study is making a distinctive contribution.

2. MATERIALS AND METHODS

Materials for this study include Uniprot database, Swiss-Model template library, basic local alignment search tool (BLAST), HHBlits and protein database (PDB), which were used to build Mpro tertiary structure. ProMod3 3.0.0 program on Swiss-Model workspace, MolProbity program, PROCHECK tool, Ramachandran plot, Verify3D, and QMEAN were exploited for validating the reliability of the COVID-19 structural model.

A self-optimized prediction method with alignment (SOPMA) was applied to compute the features of the secondary structure while DeepSite algorithm was used to elucidate the binding site area. In-silico procedure for 3D-model generation began with downloading Mpro (target) sequence from Uniprot database with ID P15423. This was followed by selection of template, extraction of template structure on PDB database and entry ID 6u7h.1.A was completed in a template alignment with target sequence. Model building involving a tertiary structure, optimization and validation of the generated structure was implemented in the homology modeling procedure. The PDB file of target protein COVID-19 Mpro was downloaded [5]. The visual molecular dynamics (VDM) visualization program was used to view the secondary features of the

protein [12]. FASTA format of the Mpro sequence with accession ID P15423 was exhumed from the famous UNiVersal PROTein (UniProt) database [13, 14]. The Swiss-Model template library (SMTL) version 2020-03-25 with PDB release 2020-03-20 [15] was extensively screened using the BLAST server for related structures that matched the target sequence. The ProMod3 3.0.0 program [16] was used to model the COVID-19 Mpro as a PDB file described in homology modeling methods. The Swiss-Model workspace [17-20] produced the 3D-structure of the target protein.

The MolProbity program [21] was used to evaluate the reliability of 3D atomic models of COVID-19 Mpro. The program checks for validity of all experimental measurements of the molecular atoms, consistency of the atomic model when compared to the experimental data and consistency of the model with known physical and chemical properties. PROCHECK [22] was deployed to verify the stereochemical quality of the models and selected the best model. The verify3D structure evaluation server [23, 24] and QMEAN [25, 26] were employed for 3D profiling of the residue. SOPMA [27, 28] was used to analyze the secondary structure evolution. The identification of active site was accomplished using the DeepSite, which is a state-of-the-art machine learning algorithm based on deep convolution neural network method that uses a volumetric overlap and a distance approach for determining the ligand binding site in a protein [29].

3. RESULTS AND DISCUSSION

The structure of COVID-19 Mpro was constructed with the template sequence of chain A as shown in Figure 1. The selected template is an experimentally generated electron microscopy structure of spike glycoprotein HCoV-229E (PDB ID: 6u7h.1.A) [30-32].

```

MFVLLVAYALLHIAGCQTTLNGLNTSY SVCNGCVGYSEN
VFAVESGGYIPSDFAFNNWELLTNTSSVVDGVVR SFQPL
LLNCLWSVSGLRFTTGFVYFNGTGRGDC KGFSSDVLSD
VIRYNLNFEENLRRGTILFKTSYGVVVFC TNNTLVSG
DAHIPFGTVLGNFYCFVNTTIGNETTSAFVGALPKTVR
EFVISR TGHFYINGYRYFTLGNVEAVNFNV TTAETDF
CTVALASYADVLNVVQSQTIANIYCNSVINRLRCDQLSF
DVPDGFYSTSPIQSVELPVSIVSLPVYHKHTFIVL YVDFK
PQSGGGKCFENCY PAGVNIILANFNETKGPLCVDTSHTF
TKYVAVYANVGRWSASINTGNCPFSFGKVN NFVKFGS
VCFSLKDIPGGCAMPIVANWAYS KYYTIGSLYVSWSDG
DGITGVQPVEGVSSFMNVTLDKCTKYNIYDVS GVGVI
RVSNDITLNGITYTSTSGNLLGFKDVT KGTIYSITPCNPP
DQLVVYQQA VVGAMLSENF TSYGFSNVVELPKFFYASN
GTYNCTDAVLTYSFSGVCADGSHIIVQPRNVSY DSVSAI
VTANLSIPSNWTTSVQVEYLQITSTPIV VDCSTYVCNGN
VRCVELLKQYTSACKTIEDALRNSARLESADVSEMLTFD
KKAFTLANVSSFGDYNLSSVIPSLPTS GSRVAGRSIAIEDIL
FSKLVTSGLGTVDADYKCKTKGLSIADLAC AQYYNGIM
VLPGVADAERMAMYTGSLIGGIALGGLTSAVSIPFLAI
QARLNYVALQTDVLQENQKILAASFNKAMTNIVDAFT
GVNDAITQTSQALQTVA TALNKIQD VVNQQGNSLNHL
TSQLRQNFQAISSSIQAIYDRLDTIQADQQVDR LITGRLA
ALNVFVSHTLTKYTEVRASRLAQQKVN ECVKSQSKRY
GFCNGTHIFSIVNAAPEGLVFLHTVLLPTQYKDVEAW
SGLCVDGTNGYVLRQPNLALYKEGNYRITSRIMFEPRI
PTMADFVQIENCNVTFVNISRSELQTI VPEYIDV NKTLLQ
ELSYKLPNYTVPDLVVEQYNQTI LNLTS EISTLENKSAE
LNYTVQKQLTIDNINSTLVDLKWLN RVETIYKWPWW
VWLCISVVLIFVVSMLLLCCSTGCCGFFSCF ASSIRGCC E STKLPYYDVEKIHQ

```

Figure 1. COVID-19 Mpro amino acid sequence (target sequence)

The homology modeling of 3D structure of COVID-19 Mpro was completed as shown in Figure 2 with PDB template and Uniprot target sequence. The structure is based on linear combination of the target template alignment coverage of 99.10% sequence identity, 0.95 coverage with homotrimer oligo-state 3.1A resolution and sequence similarity of 0.60.

The position specific iterated (psi) against pattern heat initiated (phi) distribution values of non-glycine and non-proline residues on Ramachandran plot [33] are shown in Figure 3 and Table 1. During the BLAST search, one hundred and seventeen (117) template matches were established. The quality feature of a template was predicted for the templates identified by BLAST as obtained from the target template alignment. The template with the highest quality feature of HCoV-229E belonging to the human coronavirus S-protein family was selected for model building. This family is known to mediate receptor binding, fusion of viral and host cell membrane, but maintaining the functions of the S-protein is yet undiscovered [30]. The availability of 229E S-protein electron cryomicroscopy structure of in-silico 3D structure prediction may

provide a broader intuition into immune evasion, cross-species and host-to-host transmission of COVID-19. The Ramachandran plot of the model depicts 90.3% amino acid deposit in the favorable region and 0.1% amino acid deposit in the disallowed region with a total residue deposit of 99.9% in favored and allowed regions. However, classification of residues is based on quadrangular regions in the Ramachandran plot analysis. The red area in the graph indicates the most allowed regions while yellow region shows the allowed regions. The triangular signs represent glycine while squares represent other residues as shown in Figure 3 and Table 1. PROCHECK analysis provides an evidence that all residues fell within the expected limits of the Ramachandran plot and is appraised a good model.



Figure 2. The 3D model of COVID-19 Mpro by comparative modeling

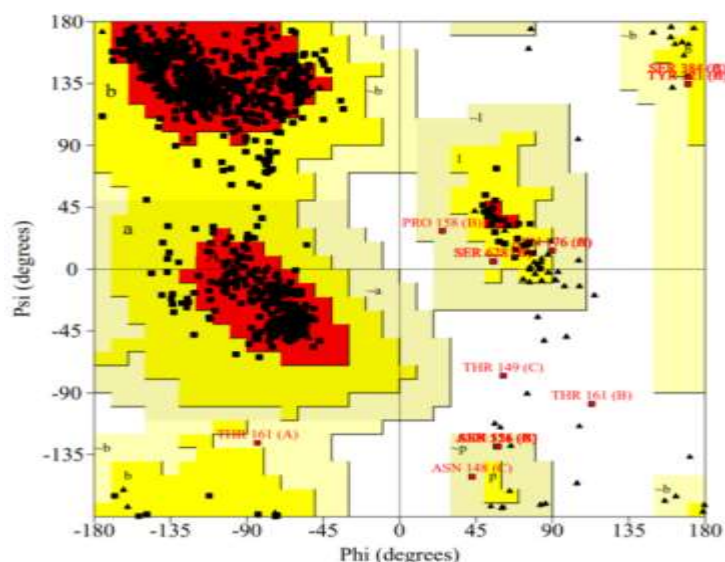


Figure 3. Ramachandran plot of COVID-19 Mpro

The Verify3D ascertains whether the 3D atomic model is compatible with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar) and comparing the results to good structures. Figure 4 verifies the compatibility level of the COVID-19 model [26, 34]. The Verify3D graph shown in Figure 4 rated the model a PASS grade corresponding to the acceptable environment of the model. Moreover, a rating of the model with 90.87% of a residue above the required average was obtained, meaning that at least 80% of the amino acids have scored at least 0.2 in the 3D/1D profile [35].

Table 1. Ramachandra plot statistics of COVID-19 Mpro

Statistics	Residue Number	Percentage (%)
Residues in the most favored regions [A, B, L]	2411	90.3
Residues in the additional allowed regions [a, b, l, p]	237	8.9
Residues in the generously allowed regions [-a, -b, -l, -p]	20	0.7
Residues in the disallowed regions	2	0.1
Number of non-glycine and non-proline residues	2670	100.0
Number of end-residues (excl. Gly and Pro)	6	-
Number of glycine residues (shown in triangles)	216	-
Number of proline residues	99	-
Total number of residues	2991	-

The global model quality estimate (GMQE) of the target sequence model was evaluated using the QMEAN4 based on four structural descriptors and results are presented in Figure 5 and Table 2. The GMQE metric considers reliability scores ranging from 0 to 1 for the predicted model [25] with COVID-19 Mpro model recording a raw score of 0.80. Good models are usually situated within the dark zone. The QMEAN scoring function estimates the GMQE of the model on the basis of a linear combination of four structural

descriptors, which are the statistical potentials of the mean force. The model is affirmed to be of a higher quality because the Z-scores of C β interaction energy, all atom pairwise energy, solvation and torsion are -0.65 -0.84, -0.90 and -0.57 respectively as shown in Figure 5 and Table 2. The predicted global model reliability is ranging from 0 to 1 [34]. The COVID-19 Mpro has a possibility of being a better model with the global GMQE score of 0.80.

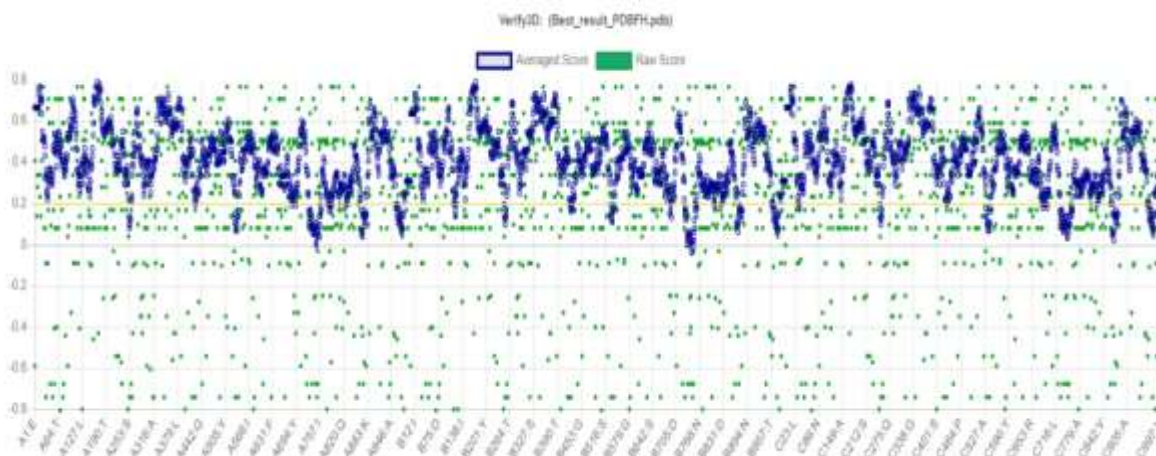


Figure 4. Verify3D graph of COVID-19 Mpro model

The local geometry is analyzed by a torsion angle potential over three consecutive amino acids. Two distance-dependent interaction potentials based on C β atoms and all atoms are used respectively to assess long-range interactions [34]. A solvation potential describes the burial status of the residues. Cheng et al. [36] reported two terms reflecting the agreement between the predicted and calculated secondary structure and solvent accessibility [37]. The QMEAN Z-score measures the features of the degree of nativeness observed in a structural model and shows the model is comparable to existing experimental structures [25]. The average Z-score for COVID-19 Mpro 3D-model is -1.01 according to the analysis as shown in Table 2 and Figure 6.



Figure 5. Global model quality estimate graph (with GMQE=0.80)

Table 2. Global quality estimate (Z-score) of COVID-19 Mpro

Scoring function terms	Z-score
QMEAN score	-1.01
C β interaction energy	-0.65
All-atom pairwise energy	-0.84
Solvation energy	-0.90
Torsion angle energy	-0.57

Figure 6 presents the result of the average Z-score and the position of our generated target model is indicated by the red star, which is considered a good model because that position falls within the dark zone region with Z-score less than 1. Consensus prediction from analysis of multiple alignments of secondary structures is needed to determine if such amino acid lies within a helix strand or coil. Structural features predicted by improvement in protein secondary structural based on SOPMA [28, 38-40] are represented in Table 3. The enzyme was modeled with the standard conventions of the tool with an evidence that α -helix, β -sheets and loops are components of enzyme. The homology modeling algorithm has proved to be a vigorous bioinformatics modeling tool. The α -helix dominated 32.05% amongst the secondary structure elements quantified, followed random coils of 37.17% while β -sheets (turns) dominated 4.86% of the secondary element quantification as represented on Table 3.

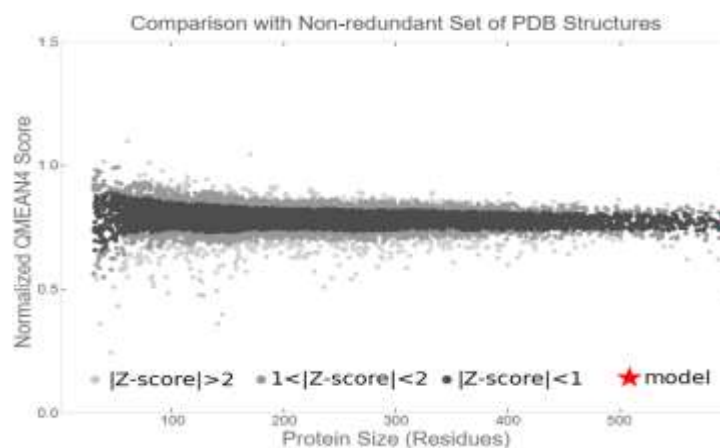


Figure 6. Absolute quality estimation graph for COVID-19 Mpro model

Table 3. Computed secondary structure elements of COVID-19 Mpro

Secondary structure	Element quantity	Percentage (%)
Alpha helix	376	32.05
3_{10} helix	0	0.00
P_i helix	0	0.00
Beta helix	0	0.00
Extended strand	304	25.92
Beta turn	57	4.86
Bend region	0	0.00
Random coil	436	37.17
Ambiguous state	0	0.00
Other states	0	0.00

ProMod-3 3.0.0 server models the coordinates of enzyme. These are conserved areas between the target and the template, which are then copied from the template to the model. Fragment library was used in modeling insertion and deletion operations. Furthermore, side chains are rebuilt while force field is used in regularizing the geometry of the built model [16]. Default parameters used in this prediction are window width of 17, similarity threshold of 8 and number of state is 4 (Figure 7).

This study reports a suitable binding site of the experimental protein along with the centroid and predicts the binding pocket involved in the active site formation as shown in Figure 8 [29]. The highest percentage commanded by the coiled regions is interpreted to be the high level of conservation and stability of the modeled structure [41]. The 3D image is distinguished by its grid of $1 \times 1 \times 1 \text{ \AA}^3$ buffer to account for pockets located to the protein edges sized voxels. This is a collection of atomic-based pharmacophoric effects defined for each voxel and a voxel tenancy is dependent on the atoms in the protein, their volume and seven other atom properties known as channels. The DeepSite server calculated the occupancy for each property of each voxel of the 3D image of COVID-19 Mpro submitted as the maximum contribution of all atoms belonging to that channel at its center. The active site position of the building protein lies within the following scores and centroids: 0,0.9994552731513977 [171.4 - 143.6 - 133.6], 1,0.9972174167633057, [197.4 - 163.6 - 171.6] and 2,0.9986205101013184, [147.4 - 181.6 - 179.6]. This location is the gold colored area depicted in Figure 8.

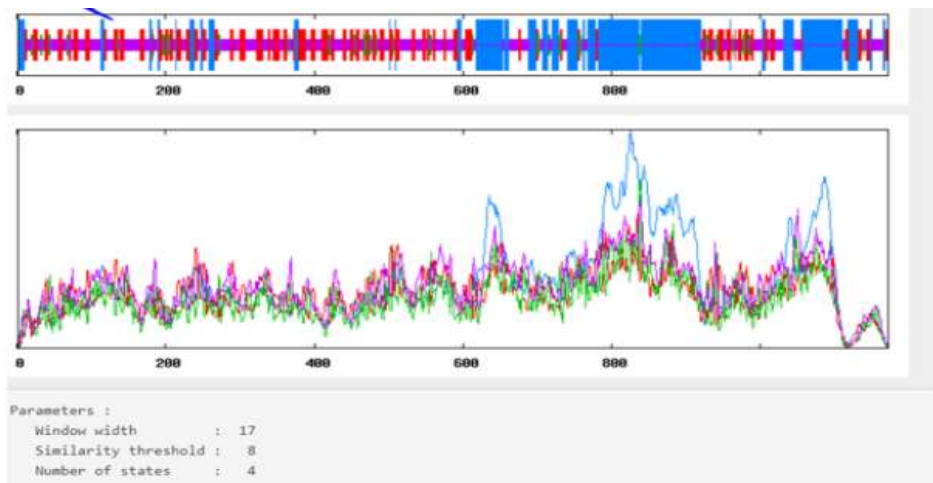


Figure 7. Secondary structure element distribution of COVID-19 Mpro

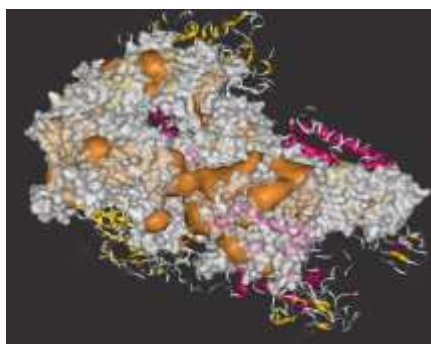


Figure 8. Cavity predictions by DeepSite represented by the gold-colored volume

The DeepSite takes as input the atomic properties and features (channels) of the sub-grids and generates binding site that captures local patterns in the structure as output of the grid, which makes the binding pocket prediction possible [29]. Consequently, this work gives the major coverage area and volume of amino acid involved in predicting the binding site and locations for local pattern in the structure of the grid for the characterization of binding pockets. The aromatic surface cavity of the binding pocket for the predicted binding site of COVID-19 Mpro is represented by the green colored area in Figure 9. Further experimental analysis of the protein will be necessary to get the drug design and discovery phase of this study ascertained.

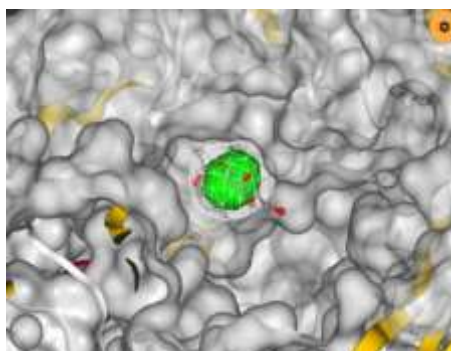


Figure 9. Aromatic surface of the binding pocket of COVID-19 Mpro

4. CONCLUSION

The purpose of this work was to identify the binding site in-silico and present a standard 3D structure of COVID-19 main protease (Mpro) by homology modeling. A targeted effort to support global response to combatting the SARS-2 COVID-19 pandemic has motivated the modeling of the main protease of the disease as reported in this paper. The structural analysis of COVID-19 Mpro was completed with application of diverse tools and software systems. The study findings have revealed that further characterization of COVID-19 Mpro will be most useful for an expedient drug design based intervention, regulation and control of the disease. Hence, our thought is that Mpro of COVID-19 may be considered a potential drug target. Nevertheless, more experimental analyses, verification and validation experiments will be required as a targeted drug or vaccine design against COVID-19 virus.

ACKNOWLEDGEMENTS

The first author would like to appreciate the postgraduate research directorate of Durban University of Technology in South Africa for the postdoctoral fellowship award to conduct this study.

DATA AVAILABILITY

For this study, COVID-19 Mpro (target) sequence was obtained from the universal protein knowledgebase (UniProtKB), while the X-ray crystallography based experimental structure and sequence of the template protein used in this research were extracted from the Protein database (PDB), a database of experimentally acquired structures and sequences.

CONSENT

All participants provided written informed consent to publish this study.

CONFLICTS OF INTEREST

Authors declare that they have no conflicts of interest.

REFERENCES

- [1] E. Dong, H. Du, L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time", *The Lancet Infectious Diseases*, vol 20, 2020. doi.org/10.1016/S1473-3099(20)30120-1.
- [2] L. Bender, "Key Messages and Actions for COVID-19 Prevention and Control in Schools", *WHO documents / Education UNICEF NYHQ*.
- [3] P. Kumar, Y. Wang, Z. Zhang, Z. Zhao, G. D. Cymes, E. Tajkhorshid, and C. Grosman, "Cryo-EM structures of a lipid-sensitive pentameric ligand-gated ion channel embedded in a phosphatidylcholine-only bilayers", *Proceedings of the National Academy of Science USA*, vol 117, pp. 1788-1798, 2020.
- [4] M. Ullah et al., "A bioinformatics approach for homology modeling and binding site identification of triosephosphate isomerase from *Plasmodium falciparum* 3D7", *J Young Pharmacists*, vol. 4, pp. 261-266, 2012.
- [5] Z Jin et al., "Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors", *10.2210/pdb6lu7/pdb*.
- [6] M. Adebisi et al., "Computational prediction of corynebacterium matruchotii Protein's 3D structure reveals its capacity to bind to DNA domain site in the Malaria Vector, Anopheles", *International Journal of Engineering Research and Technology*, vol. 12, pp. 1935-940, 2019.
- [7] Z Jin, et al., "Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors", *Nature*, vol. 582, pp. 289-293, 2020. doi.org/10.1038/s41586-020-2223-y.
- [8] S. Kaczanowski & P. Zielenkiewicz, "Why similar protein sequences encode similar three-dimensional structures?", *Theoretical Chemistry Accounts*, vol. 125, pp. 543-550, 2010.
- [9] C. Lee, "Testing homology modeling on mutant proteins: predicting structural and thermodynamic effects in the Ala98 → Val mutants of T4 lysozyme", *Folding Design*, vol. 1, pp. 1-12, 1995.
- [10] M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, "Comparative protein structure modeling of genes and genomes", *Annu Rev Biophys Biomol Struct*, vol. 29, pp. 291-325, 2000.
- [11] C. Chothia & A. Lesk, "The relation between the divergence of sequence and structure in proteins", *The EMBO journal*, vol. 5, pp. 823-826, 1986.
- [12] S. Ciudad et al., "Aβ(1-42) tetramer and octamer structures reveal edge conductivity pores as a mechanism for membrane damage", *Nature Communications*, vol. 11, 2020.
- [13] T. Raabe, B. Schelle-Prinz, S. Siddell, "Nucleotide sequence of the gene encoding the spike glycoprotein of human coronavirus HCV 229E", *Journal of General Virology*, vol 71, pp. 1065-1073, 1990.
- [14] The UniProt Consortium, "UniProt: the universal protein knowledgebase", *Nucleic Acids Research*, vol. 45, pp. D158-D169, 2017. https://doi.org/10.1093/nar/gkw1099.

- [15] C. Camacho et al., "BLAST+: architecture and applications", *BMC Bioinformatics*, vol. 10, pp. 421-430, 2009.
- [16] N. Guex et al., "Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective", *Electrophoresis*, vol. 30, pp. S162-S173, 2009.
- [17] A. Waterhouse et al., "SWISS-MODEL: homology modelling of protein structures and complexes", *Nucleic Acids Res*, vol. 46, pp. W296-W303, 2018.
- [18] S. Bienert et al., "The SWISS-MODEL Repository - new features and functionality", *Nucleic Acids Res*, vol. 45, pp. D313-D319, 2017.
- [19] L. Bordoli et al., "Protein structure homology modeling using SWISS-MODEL workspace", *Nat Protoc*, vol. 4, pp. 1-13, 2009.
- [20] K. Arnold et al., "The SWISS-MODEL workspace: web-based environment for protein structure homology modelling", *Bioinformatics*, vol. 22, pp. 195-201, 2006.
- [21] Williams et al., "MolProbity: More and better reference data for improved all-atom structure validation", *Protein Science*, vol. 27, pp. 293-315, 2018.
- [22] R. Laskowski et al., "PROCHECK - a program to check the stereochemical quality of protein structures", *J. App. Cryst.*, vol. 26, pp. 283-291, 1993.
- [23] J. U. Bowie, R. Luthy, D. Eisenberg, "Method to identify protein sequences that fold into a known three-dimensional structure", *Science*, vol. 253, pp. 164-170, 1991.
- [24] R. Luthy, J. Bowie, D. Eisenberg, "Assessment of protein models with three-dimensional profiles", *Nature*, vol. 356, pp. 83-85, 1992.
- [25] P. Benkert et al., "Toward the estimation of the absolute quality of individual protein structure models", *Bioinformatics*, vol. 27, pp. 343-350, 2011.
- [26] M. Bertoni et al., "Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology", *Scientific Reports*, vol 7, 2017.
- [27] C. Colovos, & T. Yeates, "Verification of protein structures: patterns of nonbonded atomic interactions, *C++ software*. 1993.
- [28] C. Geourjon, & G. Deléage, "SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments", *Cabios*, vol. 11, pp. 681-684, 1995.
- [29] J. Jiménez, S. Doerr, G. Martínez-Rosell, A. S. Rose, G. De Fabritiis, "DeepSite: protein-binding site predictor using 3D-convolutional neural networks", *Bioinformatics*, vol. 33, pp. 3036-3042, 2017. doi.org/10.1093/bioinformatics/btx350
- [30] Z. Li et al., "The human coronavirus HCoV-229E S-protein structure and receptor binding", *eLife*, vol. 8, 2019.
- [31] J. Lan, et al., "Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor", *Nature*, pp. 1-20, 2020. doi:10.1101/2020.02.19.956235.
- [32] Z. Li et al., "The human coronavirus HCoV-229E S-protein structure and receptor binding", 2019.
- [33] G.N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, "Stereochemistry of Polypeptide Chain Conformations", *J. Mol. Biol.*, vol. 7, pp. 95-99, 1968.
- [34] P. Benkert, M. Kunzli, T. Schwede, "QMEAN server for protein model quality estimation", *Nucleic Acids Res*, vol. 37, pp. 10-14, 2009.
- [35] R.W.W. Hoofst et al., "Errors in protein structures", *Nature*, vol. 381, 1996.
- [36] J. Cheng et al., "SCRATCH: A protein structure and structural feature prediction server", *Nucleic Acids Res*, vol. 33, pp. 72-76, 2005.
- [37] D. T. Jones, "Protein secondary structure prediction based on position-specific scoring matrices", *J Mol Biol.*, vol. 292, 195-202, 1999.
- [38] G. Deléage, "An interactive 3D viewer of molecules compatible with the suite of ANTHEPROT programs", *Journal of Biophysical Chemistry*, vol. 3, pp. 35-38, 2012.
- [39] G. Deléage, C. Combet, C. Blanchet, C. Geourjon, "ANTHEPROT: An integrated protein sequence analysis software with client/server capabilities", *Computers in Biology and Medecine*, vol. 31, pp. 259-267, 2001.
- [40] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, "NPS@: Network Protein Sequence Analysis", *TIBS*, March vol. 25, pp. 47-150, 2000.
- [41] E. Neelamathi, E. Vasumathi, S. Bagyalakshmi, R. Kannan, "Insilico prediction of structure and functional aspects of a hypothetical protein of *Neurospora crassa*", *J of Cell Tissue Re*, vol. 9, pp. 1889-1894, 2009.