A drug repurposing approach to identify therapeutics by screening Medicines for Malaria Ventures exploiting SARS-CoV-2 Main protease

Rashmi Tyagi[1*], Anubrat Paul[1], V. Samuel Raj[1, 3], Krishna Kumar Ojha[2], Manoj Kumar Yadav[1, 3 * #]

[1]Centre for Drug Design Discovery and Development (C4D), SRM University, Delhi-NCR, Sonepat - 131 029, Haryana, India
[2]Department of Bioinformatics, Central University of South Bihar, Gaya-824 236, Bihar, India
[3]Department of Bioinformatics, SRM University, Delhi-NCR, Sonepat - 131 029, Haryana, India

*Shared co-first authorship

[#] **Corresponding Author**

Dr Manoj Kumar Yadav
Department of Bioinformatics,
SRM University, Delhi-NCR, Rajiv Gandhi Education City,
Sonepat - 131 029, Haryana, India
Email: manojiids@gmail.com

Abstract

COVID-19 pandemic makes the human-kind standstill and results in high morbidity and mortality cases worldwide. Still, there are no approved antiviral drugs with proven efficacy nor any therapeutic vaccines to combat the disease as per the current date. In the present study, SARS-CoV-2 main protease (Mpro) has been taken as a potential drug target considering its crucial role in virus propagation. We have used 400 diverse bioactive inhibitors with proven antibacterial and antiviral properties for screening against Mpro target. Our screening result identifies ten compounds with higher binding affinity than N3 (used as a reference compound to validate the experiment). All the compounds possess desire physicochemical properties. Later on, in-depth docking and superimposition of selected complexes confirm that only three compounds (MMV1782211, MMV1782220 and MMV1578574) are actively interacting with the catalytic domain of Mpro.

Furthermore, the selected three molecules complexed with Mpro and N3-Mpro as control are subjected to molecular dynamics simulation study (root means square deviation, root mean square fluctuation, hydrogen bonding, solvent-accessible area and radius of gyration). MMV1782211-Mpro complex shows a strong and stable interaction as compared to others. The MM/PBSA free energy calculation shows the highest binding free energy of –115.8 kJ/mol for MMV1782211 compound also cross-confirms our molecular docking study. Therefore, our *in silico* findings become very interesting towards developing alternative medicine against SARS-CoV-2 Mpro target. So, we can expect prompt actions in this direction to combat the COVID-19.

Keywords: COVID-19, MMV compounds, Virtual Screening, Molecular Docking, ADMET, Molecular Dynamics Simulations

1.   INTRODUCTION

The ongoing outbreak of Coronavirus COVID-19 is responsible for a considerable amount of morbidity, nearly affecting 83 million of total cases and mortality of more than 1.8 million patients worldwide (World Health Organization 2020). The first case reported in December 2019 in the Wuhan City of Hubei Province, China (Chan et al. 2020; Chen et al., 2020). Understanding the severity of the disease, the World Health Organization (WHO) has declared the coronavirus outbreak a Global Public Health Emergency on January 30, 2020. Later on, investigation indicates that the cluster of pneumonia cases is due to severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), and the disease is named as Corona Virus Disease 2019 (COVID-19) as per WHO recommendations. SARS-CoV-2 is an infectious disease whose transmissibility is due to respiratory droplet or aerosols, close contact with an infected person, exposed to coughing, sneezing and likely in oral-faecal (He et al. 2020).

SARS-CoV-2 is a positive-sense single-stranded RNA virus of 26-32 kilobases, and possess 14 open reading frames (ORFs) encoding 27 proteins of variable length. The one-third portion of genome produces an array of structural proteins namely surface glycoprotein (S), a small envelope protein (E), matrix protein (M), and nucleocapsid protein (N); and accessory proteins (3a, 3b, 6, 7a, 7b, 8, 9b, 9c and 10) located in the 3′-terminus of the SARS-CoV-2 genome. The spike protein one. The spike surface glycoprotein is one of the most prominent structural protein that is mainly responsible for the binding of the virus to the host cell. The evolutionary study of SARS-CoV-2 genome sequences reveals its similarity with other severe acute respiratory syndrome coronavirus (SARS-CoV). The study shows that both the viruses share the same cell entry receptor, angiotensin-converting enzyme 2 (ACE2) for entry inside the host cell. Since the spike protein is made early in infection and responsible for the entry of virus inside the host cell, most of the current research is focused on targeting this spike protein either using therapeutics or vaccine treatment.

On the other hand, a two-third portion of the viral genome encodes a replicase polyprotein named polyprotein 1ab (pp1ab) at the 5' end of SARS-CoV-2 genome. The pp1ab contains two overlapping open reading frames (ORFs): ORF1a and ORF1b. Later on, viral proteases cleave these ORFs and result in the generation of 16 non-structural proteins (NSPs). The different NSPs

play a crucial role in establishing the viral replication system and transcription system inside the host cell.

Protease enzyme is essential for viral replication by mediating the maturation of viral replicase complex (Ren et al. 2013). There is still a limited information available to establish these proteases as a key targets for the development of a therapeutics to treat COVID-19. The two proteases namely papain-like protease (PLpro) and main protease (Mpro also called 3Clpro) co-translationally cleaves the two polypeptides (pp1a, and pp1b) into mature non-structural proteins (Jin et al. 2020a; Klemm et al. 2020). The release of functional polypeptides from the polyproteins involve an extensive proteolytic processing, which is mainly accomplished by SARS-CoV-2 Mpro protease. SARS-CoV-2 Mpro processes and digest the polyprotein at 11 conserved sites, starting with the autolytic cleavage of this enzyme itself from pp1a and pp1ab (Hegyi and Ziebuhr 2002; Umesh et al. 2020). Evolutionary study based on SARS-CoV-2 Main protease (Mpro) enzyme reveals their conservation across entire coronaviridae subfamily. The crucial role of Mpro in the viral replication and transcription will establish them as an important factor in the viral life cycle (Jin et al. 2020a). Apart from that the absence of closely related homologues in humans, also establishes them as an attractive target for the design of therapeutics.

Drug repurposing is an interested approach to design and check the efficacy of inhibitors against unknown targets. The authors have applied integrated approach to identify key amino acid residues present at the active site of main protease. In the next step, the authors have chosen to screen library of compounds to test their potential as a potent inhibitor against SARS-CoV-2 main protease.We have utilized 400 diverse compounds with known antibacterial and antiviral activity from Pandemic box for the first time to best of my knowledge. Binding poses of identified compounds are studied using docking experiments. Later on, binding dynamics of identified inhibitors is checked at the active site of the Main protease. The selected compounds are showing promising results against the SARS-CoV-2 Mpro drug target. The inhibition potentials of chloroquine and hydroxychloroquine, anti-malarial drug compounds are already reported to inhibit COVID-19 protease *in vitro*. Due to inherent toxicity and side-effects, however, they are not approved by most of the countries. Therefore, our findings on establishing

MMV compounds as a Mpro inhibitor is significant. Further wet-lab experiments must confirm the efficacy and safety of these compounds.

2.   METHODOLOGY

2.1   Protein structure retrieval:

The three-dimensional crystal structure of the main protease (Mpro) of SARS-CoV-2 co-crystallized with N3 (PDB ID: 6LU7) and the other related information of Mpro's (1WOF and 6YNK) complexed with different inhibitors namely N1 and alpha ketoamide respectively were taken from protein data bank (Protein Data Bank 2019). The ligand-receptor complexes were preprocessed in order to make them suitable for study. The preprocessing steps include removal of solvent molecules and accessory ligands. Later on, different structural features of the Mpro receptor including their superimposition and the involvement of amino acids in defining active site were studied using Chimera, an open-source visualization software (Pettersen et al. 2004).

2.2 Active site prediction

The SARS-CoV-2 Mpro protein interacts with various small molecules at its active-site and performs biological functions. Identification of the main binding site is a crucial step in computer-aided drug design. The possible binding pockets were detected using Computed Atlas of Surface Topography of proteins (CASTp) server (Tian et al. 2018). This is basically a web server that can locate, delineate and measure the geometric and topological properties of a given protein structure.

2.3 Ligand preparation

The pathogen box, available at Medicines for Malaria Venture (MMV) is a collection of over 400 diverse drug compounds which were found to be active against different pathogens. These compounds consist of 201 antibacterials, 153 antivirals and 46 antifungals with a diverse mechanism of action (Rufener et al. 2018). The compounds have not been tested against any target of COVID-19. The 3D structure of MMV's Pathogen Box compounds was constructed taking ACD/ChemSketch public domain software (http://www.acdlabs.com/ resources/ freeware/ chemsketch/) in MOL file. Compounds were edited and optimized using the Avogadro tool (Hanwell et al. 2012). Open Babel, a chemical toolbox, is used to convert ligand 3D structures in suitable file formats required for screening (O'Boyle et al. 2011).

2.4 Virtual screening and ADMET evaluation

Screening is a crucial step in computer-aided drug design to identify hit compounds on the basis of shape, size and ligand-receptor interaction. It records the activity of compounds at the active site of receptor. We have used PyRx virtual screening tool to screen library compounds. PyRx is open source screening software to screen libraries of compounds (Dallakyan and Olson 2015).

The molecular properties and drug-likeness score of MMV compounds were investigated using Molsoft (http://molsoft.com/mprop/P) in order to evaluate their pharmacological and biological properties. Lipinski's rule of five (Ro5) consists of HBA/ HBD value up to 10 and 5, respectively; MW less than 500, LogP value less than 5 and total BBB and drug-likeness score (Chappell and Payne 2020).

2.5 Docking Simulations:

Docking experiments involves different steps of ligand preparation, receptor preparation, grid box creation and docking. AutoDock Vina software was used for simulation purposes (Trott,O., Olson 2019). Ligand preparation step involves the generation of ligand conformers, charges, and identification of a number of rotatable bonds. Mpro receptor protein (PDB ID: 6LU7) is downloaded from RCSB PDB repository (Berman et al. 2000). The crystal structure of COVID-19 main protease in complex with an inhibitor N3, the ligand N3 was removed using Chimera software (Pettersen et al. 2004). Water molecules and heteroatoms were removed from the receptor molecule using ADT (Morris et al 2009). A Grid box is prepared around receptor with dimensions $70 \times 80 \times 70$ xyz points using grid spacing of 1 Å, and grid center is situated at xyz coordinates -27.211, 11.241 and 58.511 respectively. Docking simulation results in receptor-ligand interactions along with binding affinity.

2.6 Molecular dynamics simulation set up

To gain further insight into the protein-ligand interactions, the procured hit compounds from the docking studies were subjected to the molecular dynamics (MD) simulations along with the reference compound N3. The receptor protease (PDB ID:6lu7) was processed with Chimera (Pettersen et al. 2004), and their topologies were generated employing GROMOS53a6 force field implemented in Groningen Machine for Chemical Simulations (GROMACS 5.4.1) software (Pronk et al. 2013). Ligands topologies were generated utilizing the PRODRG server (Lemkul et

al. 2010). The Mpro protein is surrounded by a Dodecahedron box that was created using edit conf module and with 10☐Å distance from the edges. The system was then solvated with simple point charge water model (SPC216) to attain real dynamics followed by neutralizing the system with the counter ions. All the bad contacts were further removed by subjecting the system to pass through steepest descent algorithm at 10,000 steps with an upper limit of the force being lower than 1000 kJ/mol. Following this, the equilibration was conducted by Number of particles, Volume and Temperature (NVT) and Number of particles, Pressure and Temperature (NPT) (Vollmayr-Lee 2020) at 100 ps at 300 k and 100 ps at a pressure of 1 bar maintained by Parrinello-Rahman barostat and allowing the movement of the counterions and the water molecules, constraining the protein backbone. Linear Constraint Solver for Molecular Simulations (LINCS) (Hess et al. 1997) algorithm was used to restrain heavy atom bonds and their respective hydrogen atoms. Particle Mesh Ewald (PME) (Toukmaji et al. 2000) was utilized to compute the long rage electrostatic interaction and a cut-off distance of 12 Å was attributed for Coulombic and van der Waals interactions. MD simulations were performed for 50 ns storing the coordinate data for every 2 fs. We evaluated the corresponding results by employing the XMGRACE (Turner 2005) and Chimera (Pettersen et al. 2004).

3. Results and discussion:

SARS-CoV-2 belongs to a group of viruses that can infect humans and vertebrate animals. COVID-19 infections affect the respiratory, digestive, liver, and central nervous systems of humans and animals. The main protease (Mpro, also called 3CLpro) of SARS-Cov-2 is an attractive drug target due to its dissimilarity with their host, and it also plays an essential role in processing the polyproteins that are translated from the viral RNA. This study focuses on identification and characterization of most probable inhibitors from Pathogen Box available at MMV against a potential therapeutic target, the main protease (3CLpro/Mpro) of SARS-CoV-2 using drug repurposing dimensions.

3.1 Structural Analysis of Mpro drug target

Mpro is the main protease enzyme, crucial for the survival of the virus inside the human cells. SARS-Cov-2 Mpro protein consists of 306 amino acids and shows a 96 % similarity with the

main protease of SARS-CoV (Jin et al. 2020b). The discovery of important amino acid residues of Mpro protease structure in COVID-19 provides an opportunity to identify potential drug candidates for the treatment. The 3D structure of Mpro co-crystallized with Michael acceptor inhibitor, N3 is taken from protein data bank (PDB ID: 6LU7). The structural analysis and visualization show that it exists in a dimer form where each monomer is a protomer and composed of three domains: domain I (8-101 amino acid residues), domain II (102-184) and domain III (201-303). Domain I and II show antiparallel β-barrel structure, and domain III constitutes a globular structure with antiparallel arranged five α-helices. A long loop region usually comprises of 15 amino acid residues connects domain II to domain III.

The other two available 3D structures of Mpro protein (PDB ID-1W0F and 6YNQ) complexed with N1 and α-ketoamide inhibitors are taken from the protein database. We did a structural analysis of available three 3D structures along with their respective ligands using Chimera software. An overlay of the structures of SARS-CoV-2 Mpro-N3 and Mpro-α-ketoamide and SARS-CoV-1 Mpro-N1 show that all the three inhibitors occupy the same binding site of Mpros with nearly similar binding modes as evident from **Figure 1**. The superimposition result shows that the 3D crystal structure of SARS-CoV-2 Mpro (PDB ID-6LU7) is highly similar to that of Mpro complexed with α-ketoamide inhibitor and shows only a root mean square deviation of 0.406 Å, compared to main protease co-crystallized with N1 inhibitors (RMSD:0.671). However, the surface loops and α-helices of domain III are the most variable regions, the perfectly superimposed substrate-binding pockets located in a cleft between domains I and II are still highly conserved among three Mpro structures. It is also evident that N3, N1 and α-ketoamide inhibitors are present in the similar binding mode at the conserved binding pockets. This will pose more confidence that disrupting the Mpro active site by drug repurposed inhibitors may result in a decrease in SARS-CoV-2 activity.
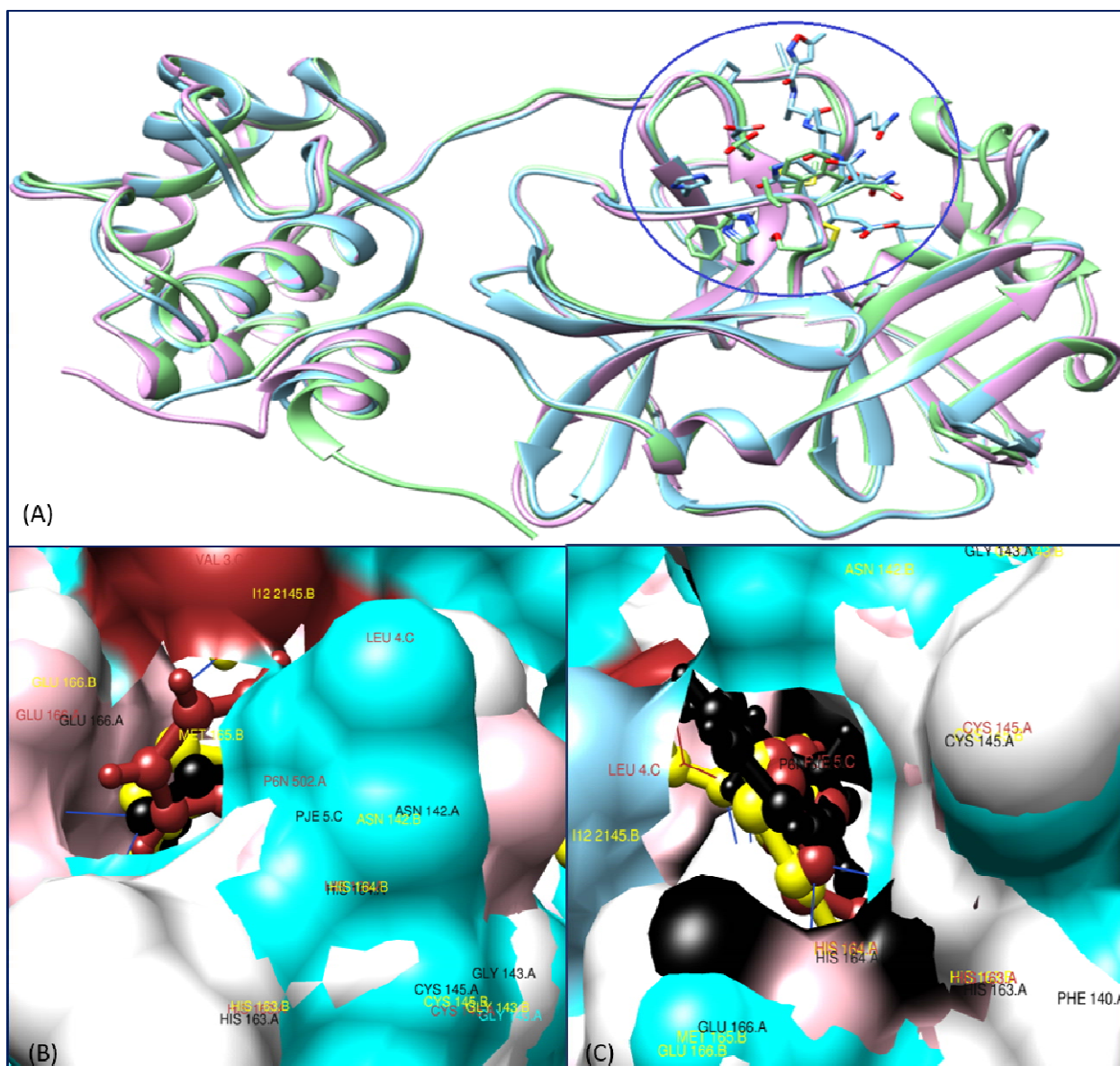
Figure 1: Structural Analysis of SARS-CoV-2 Mpro complexed with inhibitors (A) Superposition of crystal structures of Mpro shown in ribbon form 6LU7 (Pink), 1WOF (Blue), 6YNK (Lime green) co-crystallized with their respective ligands, N3, N1 and alpha ketoamide inhibitor in stick forms with atom level colouring, (B) and (C) Surface view of Mpro active sites

where the active site of 6LU7 (cyan), 1WOF (white) and 6YNK (pink) along with their ligands: N3 (red), N1 (yellow) and alpha ketoamide (black).

3.2 Exploring the potential binding site of Mpro receptor

To bring more robustness in confirming the final, binding site before the screening, we performed a ligand-independent binding site search for Mpro protein using the CASTp server. The output result shows a total of 38 binding pockets and other sub-pockets. The ligand, in general, interacts with the binding pocket having the largest cavity. The solvent-accessible surface area is 224 $Å^2$, with a volume of 180 $Å^3$ for the largest binding pocket (**Figure 2**). The amino acid residues namely THR24, THR25, THR26, LEU27, HIS41, THR45, SER46, MET49, PHE140, LEU141, ASN142, GLY143, SER144, CYS145, HIS163, MET165, GLU166, HIS172 of chain A; and VAL3, LEU4 of chain C are mainly involved in constituting main binding pocket of COVID-19 Mpro receptor. The presence of conserved residues; HIS41 and CYS145 at the catalytic dyad at the substrate-binding site of SARS-CoV-2 Mpro validates our hypothesis of active site prediction.
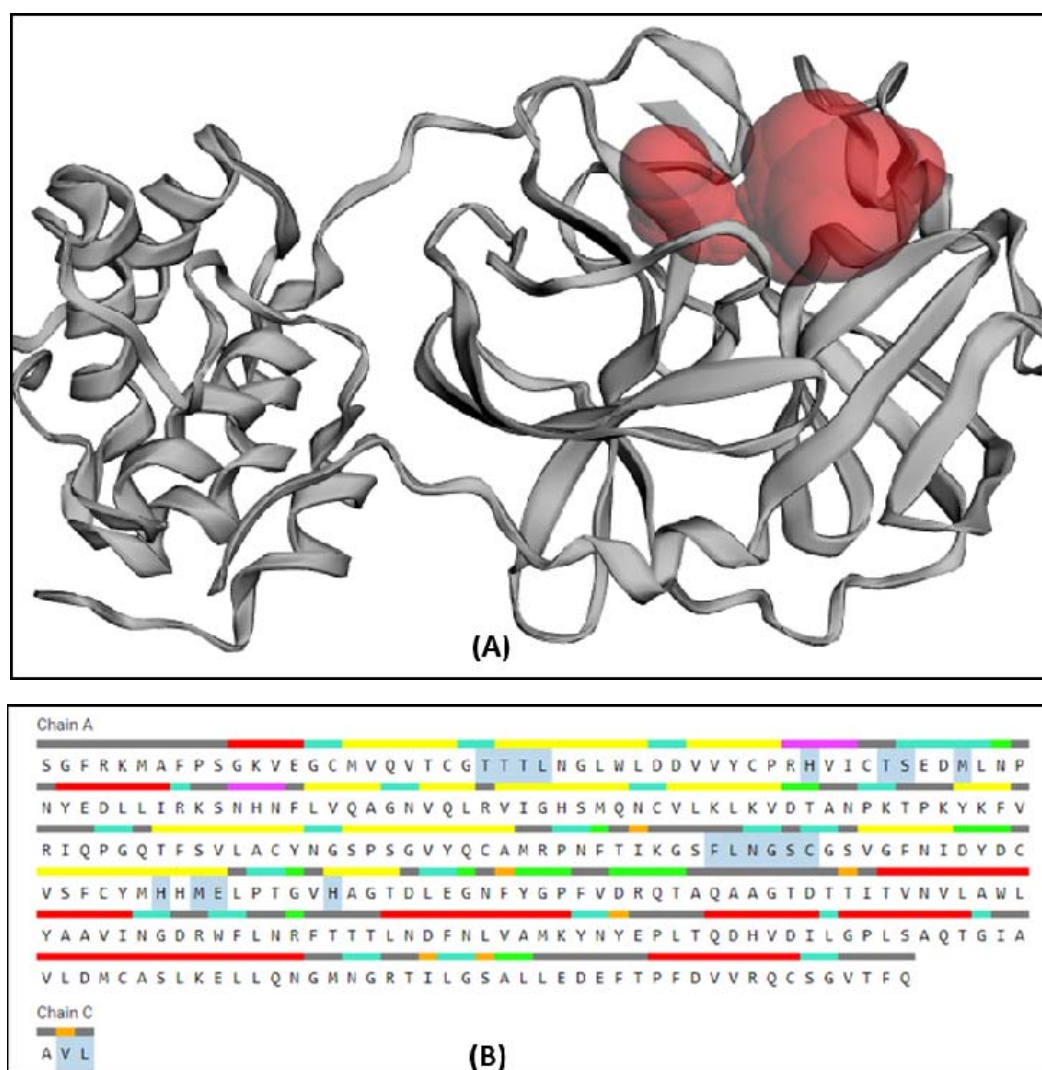
Figure 2: Binding site prediction using the CASTp server (A) Active site pocket of Mpro shown in red spheres on residues (B) secondary structure elements is visible in different colours. Amino acid residues present at the largest binding pocket are highlighted in cyan colour.

3.3 Virtual screening of compounds in Pathogen box

Screening is a computational technique that can identify bioactive hit candidates from a collection of small compound libraries. PyRx is used to identify potential compounds against Mpro enzyme of SARS-CoV-2. We have selected 400 diverse bioactive inhibitors that are previously useful for targeting other diseases from the Pathogen Box available at Medicines for Malaria Venture for screening against SARS-CoV-2 main protease. Jin et al. work shows that small molecule, N3 forms a stable complex with Mpro protein (Jin et al. 2020b). So, we have
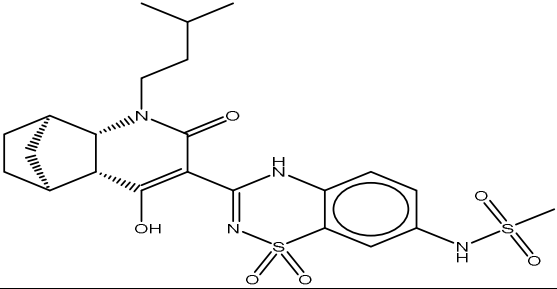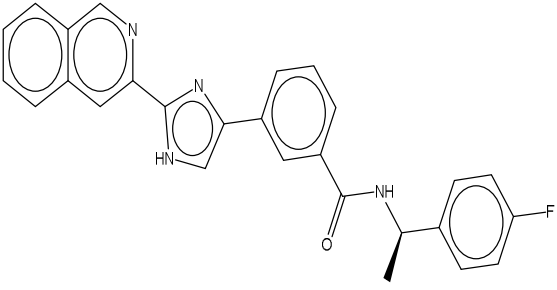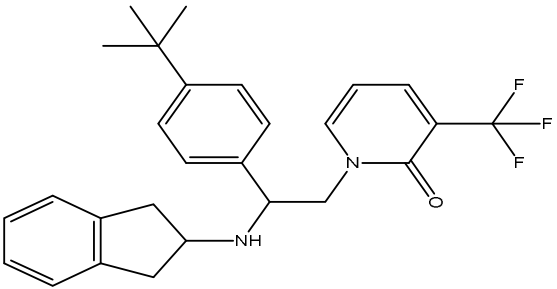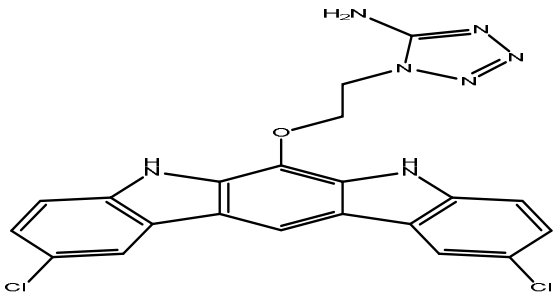
included the N3 inhibitor as a control to screen 400 diverse compounds in order to validate our screening parameters. The presence of N3 inhibitor in the screening result successfully cross validates our hypothesis. Twelve compounds did not show any binding, and remaining other compounds bind at the main cavity with differential binding energies of -4.2 to -9.0 kcal/mol. We have used the binding energy of our control (-6.0 kcal/mol) as a cut off criterion to filter our screening compounds (Table 1). Total 10 compounds having higher binding affinity to our control, are taken further for in-depth study.
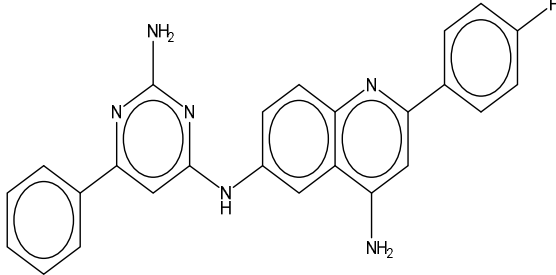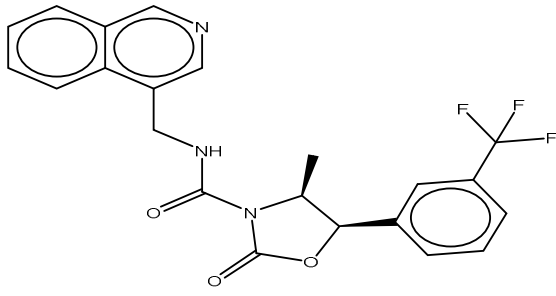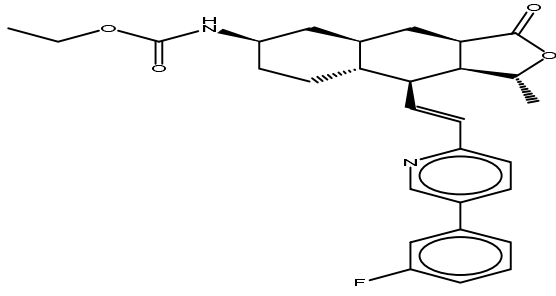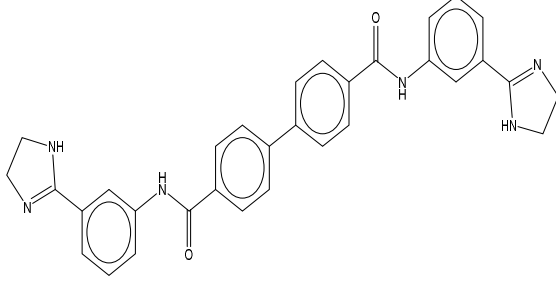
3.4 Study of physicochemical properties of screened compounds:

The interaction of a compound with its physical environment determines its physicochemical properties responsible for the biological activity of compounds. The drug-likeness score is calculated by considering partition coefficient (log P), molecular weight, number of hydrogen donors, number of hydrogen acceptors and number of violations to Lipinski's rule (**Table 1**). The calculated molecular weight, hydrogen bond acceptors, hydrogen bond donors of obtained compounds are 422 to 957, 2 to 7 and 1 to 3 respectively. The molecular polar surface area directly related to the passive transport of drugs through membranes, and their values are lying within the range of 26.92–160.96 $Å^2$. The contribution of each functional group & structural arrangement helps to determine the lipophilic character and is positively associated with the permeability and bioavailability of drugs. The log P values of drugs lie between −0.58 and 6.11. The structural and molecular properties of a particular compound are similar to the known drugs or not, is determined by Drug-likeness score which lies in the range of -0.41 to 2.68. The screened drugs are likely to be orally active as they are in agreement with the Lipinski's rules with fewer violations, and the drug-likeness score. These compounds classify the basic drug criterion and are used in the next step of the drug design process.

Table 1: Calculation of ADMET properties of screened compounds

Compounds	Molecular Properties and Drug-likeness	
	Properties	**Value**
MMV1782220 ($C_{23}H_{30}N_4O_6S_2$) (Antiviral)	Molecular weight	522.16
	Number of HBA	7
	Number of HBD	3

	MolLogP	2.50
	MolPSA	126.41
	Drug-likeness model score	0.75
MMV1782211 (C₂₇H₂₁FN₄O) (Antiviral)	Molecular weight	436.17
	Number of HBA	3
	Number of HBD	2
	MolLogP	5.55
	MolPSA	52.25
	Drug-likeness model score	0.43
MMV1634393 (C₂₇H₂₉F₃N₂O) (Antiviral)	Molecular weight	454.22
	Number of HBA	2
	Number of HBD	1
	MolLogP	6.60
	MolPSA	26.96
	Drug-likeness model score	1.26
MMV1633966 (C₂₁H₁₅Cl₂N₇O) (Antibacterial)	Molecular weight	451.07
	Number of HBA	4
	Number of HBD	4
	MolLogP	4.51
	MolPSA	87.44
	Drug-likeness model score	-0.41
MMV1593541 (C₂₅H₁₉FN₆) (Antibacterial)	Molecular weight	422.17
	Number of HBA	3
	Number of HBD	5

	MolLogP	6.11
	MolPSA	76.64
	Drug-likeness model score	0.14
MMV1593533 (C$_{22}$H$_{18}$F$_3$N$_3$O$_3$) (Antibacterial) 	Molecular weight	429.13
	Number of HBA	4
	Number of HBD	1
	MolLogP	4.68
	MolPSA	55.82
	Drug-likeness model score	0.36
MMV1593515 (C$_{29}$H$_{33}$FN$_2$O$_4$) (Antiviral) 	Molecular weight	492.24
	Number of HBA	5
	Number of HBD	1
	MolLogP	5.85
	MolPSA	63.59
	Drug-likeness model score	0.65
MMV1580853 (C$_{32}$H$_{28}$N$_6$O$_2$) (Antibacterial) 	Molecular weight	528.23
	Number of HBA	4
	Number of HBD	4
	MolLogP	6.12
	MolPSA	90.38
	Drug-likeness model score	0.46
MMV1578574 (C$_{27}$H$_{31}$FN$_4$O$_8$) (Antibacterial)	Molecular weight	558.21
	Number of HBA	10
	Number of HBD	7

	MolLogP	-0.58
	MolPSA	148.71
	Drug-likeness model score	2.68
MMV639951 (C$_{53}$H$_{83}$NO$_{14}$) (Antiviral)	Molecular weight	957.58
	Number of HBA	14
	Number of HBD	3
	MolLogP	5.26
	MolPSA	160.96
	Drug-likeness model score	0.65

3.5 Interaction of screened drugs with Mpro receptor

The selected 10 hits from the screening experiments are further scrutinized by docking simulations using stringent parameters. The receptor comprises of a 3D structure of Mpro enzyme. Autodock Vina software is used for docking purpose. We want to check all the possibilities whether the ligand conformers bind to the main binding site or additional binding site, so we have performed blind docking experiments. The docking result shows a more or less similar trend in binding energy compared to PyRx screening tool (**Table 2**). The docking simulations by Vina sheds more light on binding poses and interactions of ligands at the active site of Mpro receptor in terms of hydrogen bonding and non-bonding interactions.

MMV1782220 ligand docked at the main binding site of Mpro receptor of SARS-CoV-2 with the binding energy of -8.2 kcal/mol. The amino acid residues involved in the hydrophobic interactions are HIS41, THR24, THR25, THR45, GLY143, LEU141, ASN142, GLU166, HIS163, MET165 where THR45 and GLY143 amino acids of receptor show hydrogen bonding with inhibitors. The compound, MMV1782211 interacted with residues THR25, THR26, MET49, ARG188, GLN189, MET165, PRO168 of the receptor with docking affinity -8.9 kcal/mol. In a similar fashion, MMV1634393 (-8.7 kcal/mol) and MMV1633966 (-8.8 kcal/mol) interact with GLN110, VAL202, THR292, PRO293, PHE294, ILE249; and ARG131, LYS137,

ASP197, THR198, THR199, ASN238, LEU286, LEU287, residues respectively, one hydrogen bond generated by MMV1633966 with residue THR199. Another compound, MMV1593541 show interactions with residues namely GLU14, GLN19, GLY70, GLY71, VAL73, GLY120, SER121 with a binding affinity of -8.7 kcal/mol.

The amino acid residues namely LYS137, THR199, TYR237, TYR239, GLU290 are involved in holding MMV1593533 ligand at the binding cavity of Mpro receptor with a binding energy of -8.3 kcal/mol and with one hydrogen bond made by THR199. The interaction of Mpro receptor and MMV1593515 involves TYR237, LYS137, TYR239, ASP289, MET276 with an affinity of -8.2 kcal/mol. MET276 shows hydrogen bonding with MMV1593515.

TYR239, LYS137, LEU287, LEU286, LYS5 interacts with MMV1580853 ligand at the main active site with binding affinity -8.3 kcal/mol. The ligand shows a hydrogen bonding interaction with LEU287.

The docking energy of MMV1578574 with the receptor is -8.4 kcal/mol, and the interacted residues are HIS41, MET49, MET165, HIS164, GLU166, GLY143 along with one hydrogen bonds involving residue HIS164. The close contact amino acid residues namely ARG40, TYR54, GLU55, ASN84, ARG105, ASN153, GLU178, ASN180, ARG188 are showing interactions with MMV639951 compound with -7.9 kcal/mol, and it also forms strong hydrogen bond network with ARG40, TYR54, ARG188.

Table-2: Interaction of selected drug compounds at the active site of Mpro receptor of SARS-CoV-2

S. No.	MMV ID	Binding Affinity	Interacting Residues	No. of Hydrogen Bonds
1	MMV1782220	-8.2	THR45, HIS41, THR25, THR24, GLY143, LEU141, ASN142, GLU166,	THR45, GLY143

			HIS163, MET165	
2	MMV1782211	-8.9	THR25, THR26, MET49, ARG188, GLN189, MET165, PRO168,	0
3	MMV1634393	-8.7	VAL202, GLN110, THR292, PRO293, PHE294, ILE249,	0
4	MMV1633966	-8.8	LEU286, LEU287, THR199, THR198, ASN238, ARG131, ASP197, LYS137	THR199
5	MMV1593541	-8.7	GLU14, SER121, GLY70, GLY71, GLY120, VAL73, GLN19	0
6	MMV1593533	-8.3	TYR237, THR199, TYR239, LYS137, GLU290	THR199
7	MMV1593515	-8.2	TYR237, LYS137, TYR239,ASP289, MET276	MET276
8	MMV1580853	-8.3	TYR239, LYS137, LEU287, LEU286, LYS5	LEU287
9	MMV1578574	-8.4	HIS41, MET49, MET165, HIS164, GLU166, GLY143,	HIS164
10	MMV639951	-7.9	ARG188, ASN153, TYR54,ARG40, GLU55,ASN180, ASN84, ARG105, GLU178	ARG40, TYR54,ARG188
11	N3 compound (control)	-6.0	PHE294, ASN151, SER158, VAL104, GLN110	0

3.6 In-depth binding analysis of docked compounds at the main active site of Mpro receptor protein

The superimposition of all the docked compounds obtained from docking experiment shows that their minimum energy conformers bind at both, the main active site as well as additional binding sites with a differential binding affinity. The in-depth analysis reveals that only three compounds, namely MMV1782220, MMV1782211 and MMV1578574, found to be located at the main active site cavity besides N3 compound as visible from **Figure 3**. Apart from that, other ligands are binding to some other binding cavities, suggesting the presence of additional cavities in the receptor.

The compounds found to be present at the main cavity are further selected for in-depth interaction study. The selected three compounds show a significant binding affinity at the Mpro active site. The network of hydrogen bonding pattern is present between ligands and receptor. The presence of an array of hydrophobic interactions is also responsible for holding these ligands at the main binding pocket of Mpro receptor (**Figure 4**). Hence it is clear from the docking study that the three MMV compounds are present, and showing satisfactory interactions at the main active site of drug target. This experiment needs some additional validation in terms of pose stability and interaction stability. The stability and reliability of complexes are further checked by conducting molecular dynamics simulations.
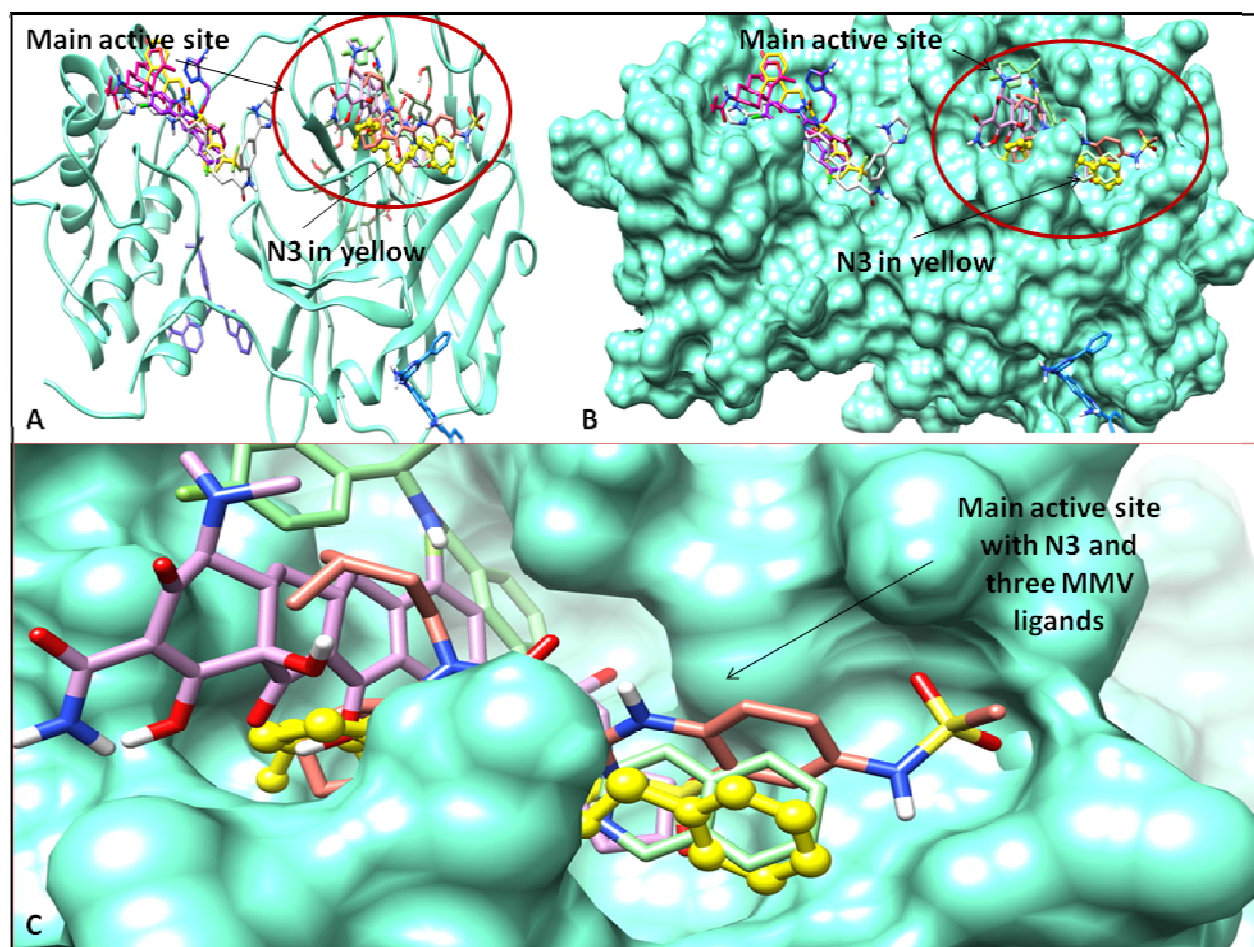
Figure-3: Binding of ligands at the Mpro receptor protein. (A) (A) Secondary structure representation B) Surface view; Only three ligands (MMV1782220 MMV1782211 and MMV1578574) out of 10 binds at the main active site of the receptor (C) enlarged surface view of the main active site showing three ligands at main active site pocket along with N3 (yellow color).
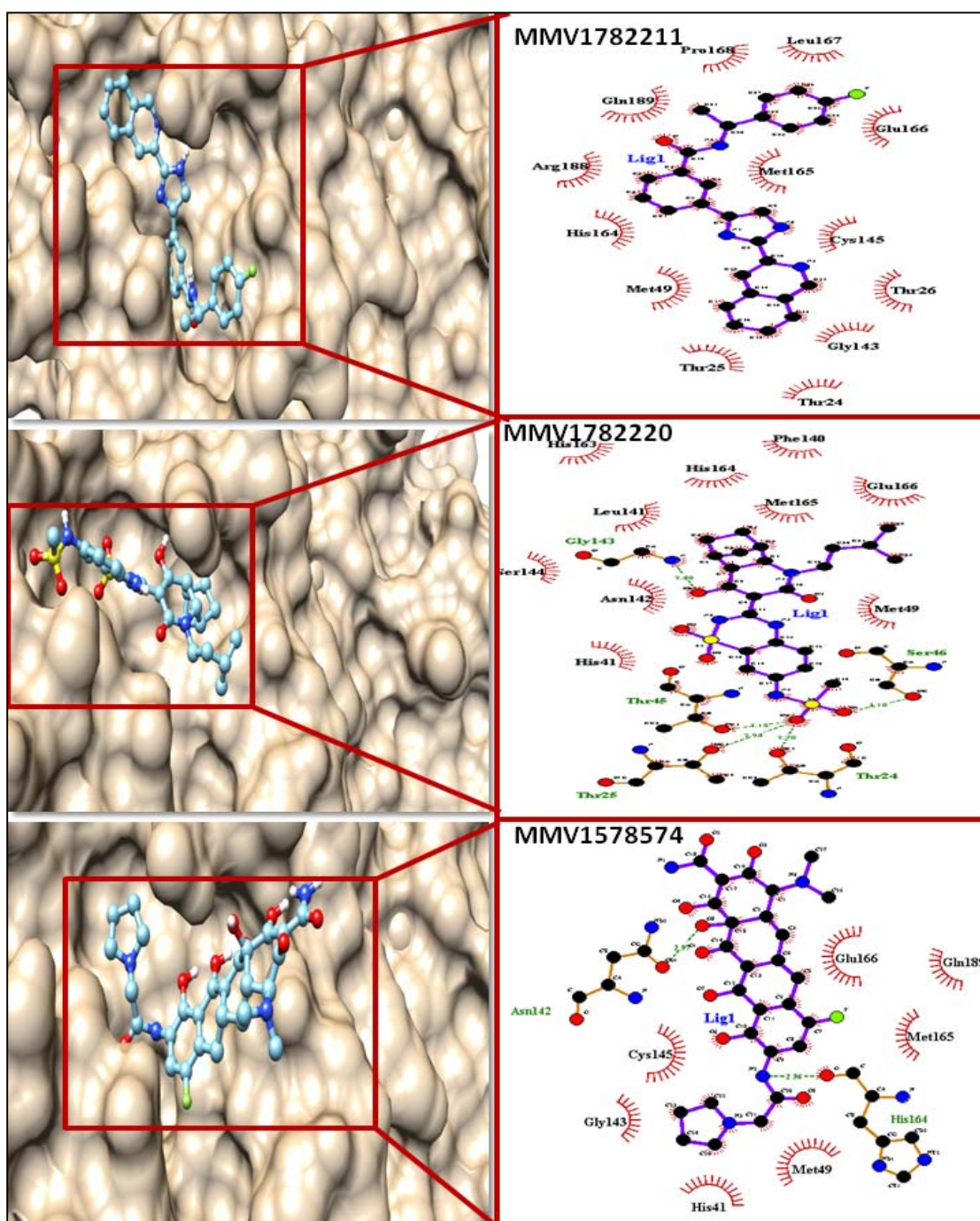
Figure-4: Binding modes of minimum energy conformers after docking experiments of MMV compounds: 3D structure of Mpro protein are shown as molecular surface models in Tan color and ligands are represented as ball and stick models on the left-hand side while ligand-receptor interactions and their close contact residues are visible on the right-hand side pane using LigPlot program where hydrogen bonds are labelled in green colour.

3.7 Molecular dynamics simulation analysis of selected complexes:

The primary objective of performing the MD simulation experiments is to assess the binding stability and dynamics of selected three compounds at the binding pocket of Mpro receptor. In order to provide dynamics information, all the four models, including N3 as a control (Mpro-MMV1782220, Mpro-MMV178221, and Mpro-MMV1578574) are subjected to simulation experiments for 50 ns individually. The output results of MD simulation experiments are examined in terms of their RMSD values, RMSF, hydrogen bond, solvent-accessible area and radius of gyration to assess their individual as well as complex stability. The RMSD profiles indicated that the binding of all selected compounds significantly stabilized the Mpro structure, and their average RMSD varies from 2.2 to 3.7 Å, as shown in **Figure 5(A)**. Our comparative analysis showed that the average RMSD fluctuations of MMV1782211 complex is lowest, and is also showing comparable synchrony with the reference Mpro-N3 complex. The average RMSD fluctuations of this complex are found to be even more stable than the reference one.

The RMSF analysis provides more detailed information about amino acid motions in Mpro receptor upon binding of selected MMV compounds. So, RMSF fluctuations with respect to amino acid residue numbers are plotted from the 50 ns trajectories (**Figure 5(B)**). The convergence of the simulation towards equilibrium can also be inferred from the relaxation of the structure. The RMSF captures, for each atom, the fluctuation about its average position. This gives insight into the flexibility of regions of the protein and corresponds to the crystallographic b-factors (temperature factors). It is also clear from the RMSF profile that the residues in the core region have minimum fluctuations about the average position (0.17 nm) like LEU30, CYS38, ARG40, HIS41, THR45, GLY143, CYS145, VAL 148, MET162, PRO168, ALA206, LEU286. The residues which are in the surface or loop region GLU47, ASP48 MET49, TYR154, ARG222, and ASN277, have high fluctuations. These results have a good coherence with crystallographic data of the protein.
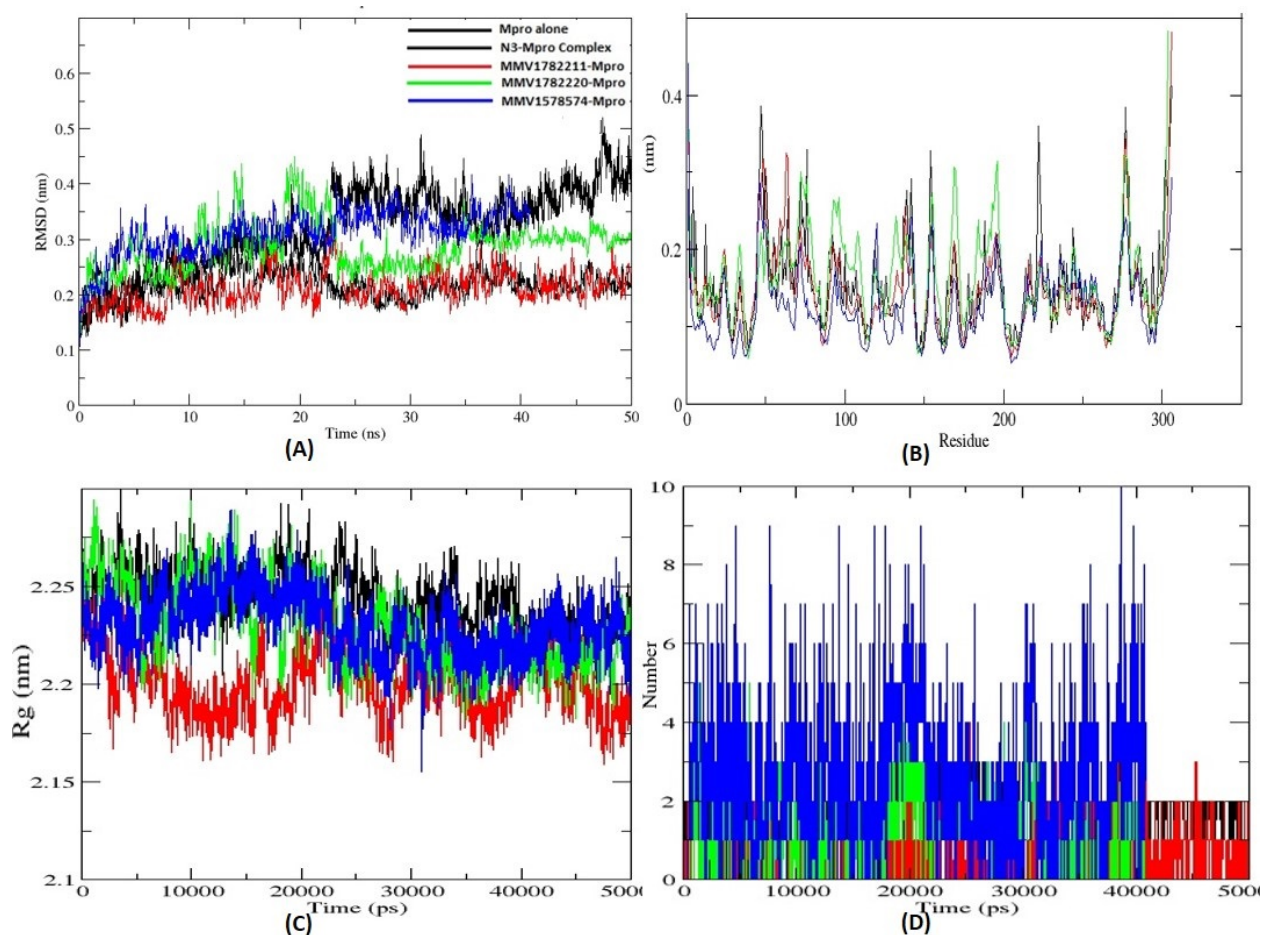
Figure-5: Comparison of molecular dynamics simulation trajectories (A) Root mean square deviation, (B) Root mean square fluctuations, (C) Radius of gyration, and (D) Number of hydrogen bond formation, for Mpro protein docked with the reference ligand N3 (black), MMV178221 (red), MMV1782220 (green), and MMV1578574 (blue) over the 50 ns simulations.

Furthermore, the radius of gyration (R_g) is used to study the overall conformational shape of a protein and its compactness. The radius of gyration showed no abnormal behaviour throughout the simulation (**Figure 5(C)**). The Mpro-MMV1782211 complex shows more compactness in terms of the distribution of mass around the central axis as compared with the reference and other ligands.

The hydrogen bonding is a good measure of the stability of a protein-ligand complex. Intramolecular, as well as intermolecular hydrogen bonds, play a vital role in molecular recognition, stability and overall conformation. The number of hydrogen bonds was analyzed to get insight into the protein-ligand interaction and stability. The number of hydrogen bonds varies from 1-10, with an average of 2.5 during the entire simulation run for ligands and Mpro protein complexes (**Figure-5(D)**). Solvent accessible surface area also varies for different ligand and varies 2-10 nm^2 with an average of 0.6 nm^2 per residue. It should be noted that residues which are in the active site region of the protein-ligand complex have a low solvent accessible surface as compared to others.

3.9 MM-PBSA Free Energy Decomposition

The MM-PBSA is a popular method for predicting free energy of binding due to its good accuracy compared to most other scoring functions of molecular docking methods (Wang et al. 2019). MM-PBSA based binding free energy of all protein-ligands complexes was calculated for the last 50 ns trajectories. The binding free energy was determined using polar and apolar solvation energy. The free binding energy was investigated as electrostatic energy, polar solvation energy, van der Waals energy, SASA energy and average binding energy (**Table 3**). The reference compound N3 showed free binding energy of –115.8 kJ/mol whereas MMV1782211, MMV1782220 and MMV1578574 showed binding energy of –171, –143 and –118 kJ/mol suggesting MMV1782211 have a highest binding affinity for SARS-CoV-2 Mpro protein.

Binding Energy of all ligands with protein during MD simulation showed that the binding energy of MMV1782211 is lowest altogether during the entire simulation run, which is visible in red colour in **Figure 6(B)**.
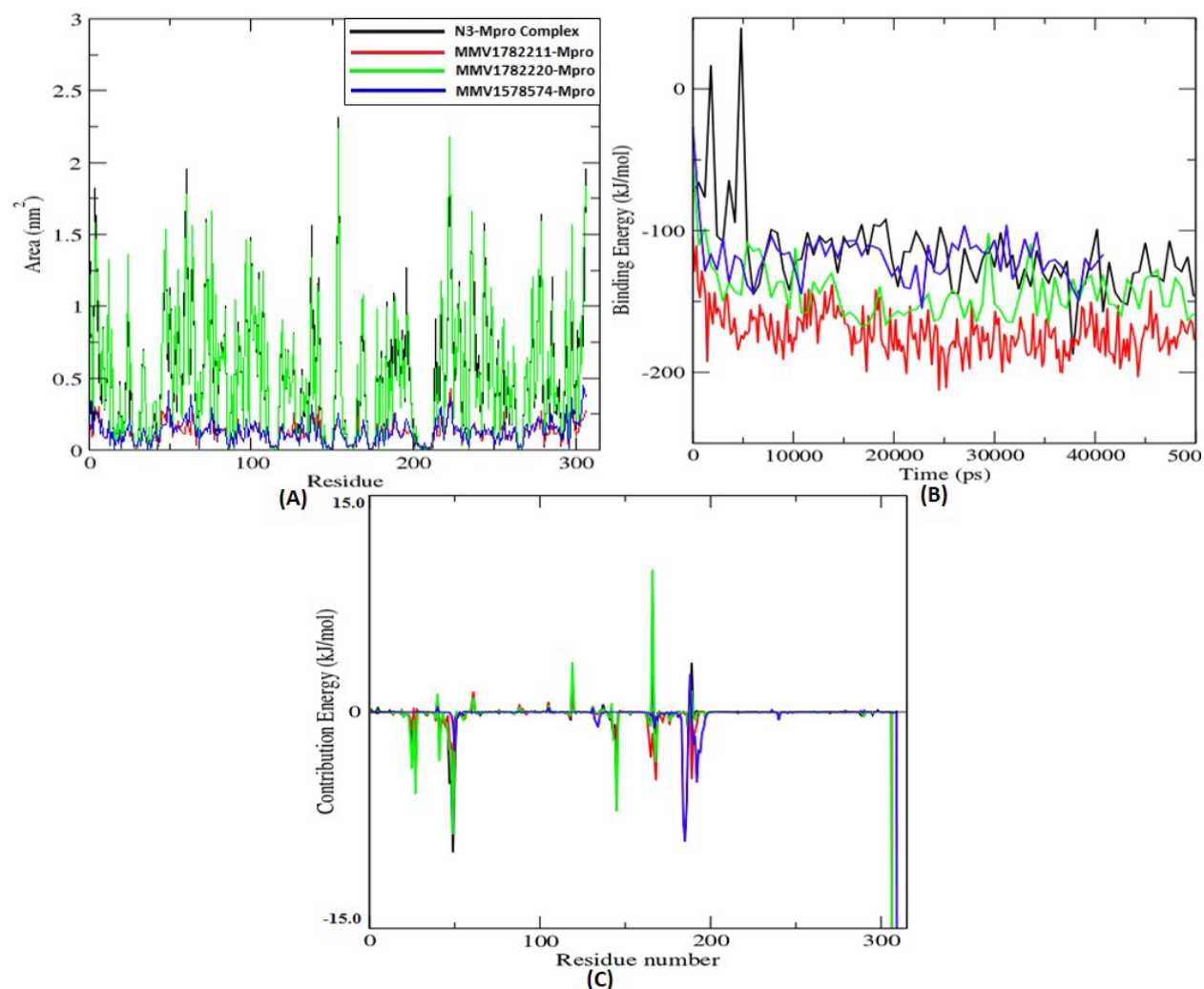
Figure-6: (A) Area per residue over the trajectory (B) Binding free energy components (C) Contribution of residues to the binding energy, of Mpro protein docked with the reference ligand N3, MMV178221, MMV1782220, and MMV1578574.

The energy contribution of individual amino acid residues was also calculated using the MM-PBSA method (**Figure 6(C)**). The residue peaks in negative Y-axis are responsible for the interaction and stability of ligands. The interacting amino acid residues such as THR24-26, HIS41, MET49, PHE140, LEU141, GLY143, SER144, CYS145, HIS163-164, MET165, GLU166, PRO168, HIS172, ASP187, ARG188, GLN189, THR190, ALA191 and GLN192 are found to surround the ligands within a distance of 4 Å.

Table 3 Van der Waals, electrostatic, polar solvation, SASA and binding energy in kJ/mol for each Mpro-drug complex

Compound ID	Van der Waals energy	Electrostatic energy	Polar solvation energy	SASA energy	Binding energy
N3	-181 ± 26.0	-34.4 ± 13.5	118 ± 31.9	-18.4 ± 2.9	-115.8 ± 3.5
MMV1782211	-219 ± 0.88	-22.2 ± 0.4	89.6 ± .74	-19.7 ± .07	-171 ± 0.9
MMV1782220	-195 ± 16.4	-17 ± 8.4	87.9 ± 16.1	-18.5 ± 1.4	-143 ± 19.2
MMV1578574	-158 ± 0.4	-18.8 ± 9.0	72.6 ± 19.0	-14.2 ± 2.0	-118 ± 17.3

4.   Conclusion

COVID-19 has emerged as a pandemic and responsible for enormous mortality and morbidity in the human population worldwide. However, no approved therapeutic drugs currently exist to treat the disease, the prophylactic vaccines are the only available options, but its efficacy and safety are still concerns. We aim to combat the COVID-19 crisis by utilizing the potential of SARS-CoV-2 main protease (Mpro) as a drug target. The superimposition of 3D structures of Mpro provides information on key amino acids involved in the interaction at the main binding pocket. We have screened 10 out of 400 diverse, drug-like molecules of Pathogen box on the basis of binding affinity compared to N3, which is used as a control to cross validates screening protocol. The physicochemical properties of these compounds lie in the permissible range. Later on, these compounds are subject to docking experiments using stringent parameters where their binding affinity lies in the range of -7.9 to -8.9 kcal/mol. The structural and superimposition analysis of selected complexes reveal that a total of only 3 compounds, namely, MMV1782211, MMV1782220, and MMV1578574, were found to interact at the main active site pocket of

Mpro. The selected complexes were further evaluated in terms of intermolecular interactions, complex stability and binding affinity against SARS-CoV-2 Mpro-N3 inhibitor as a reference complex using molecular dynamics simulation and MM/PBSA study. The molecular docking, simulation and MM/PBSA study shows that MMV1782211-Mpro complex is the most stable configuration with high free binding energy compared to reference compound N3. This compound may have the ability to block the expression of Mpro protein, results in disruption of their replication mechanism. Hence, this compound can be further evaluated as a SARS-CoV-2 Mpro inhibitor using *in vitro* and *in vivo* model studies.

Conflict of Interest

All the authors declared no conflict of interests.

References:

Berman HM, Westbrook J, Feng Z, et al (2000) The protein data bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235

Chan JFW, Kok KH, Zhu Z, et al (2020a) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. https://doi.org/10.1080/22221751.2020.1719902

Chan JFW, Yuan S, Kok KH, et al (2020b) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. https://doi.org/10.1016/S0140-6736(20)30154-9

Chappell M, Payne S (2020) Pharmacokinetics. In: Biosystems and Biorobotics

Chen N, Zhou M, Dong X, et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. https://doi.org/10.1016/S0140-6736(20)30211-7

Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol. https://doi.org/10.1007/978-1-4939-2269-7_19

Hanwell MD, Curtis DE, Lonie DC, et al (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4:17. https://doi.org/10.1186/1758-2946-4-17

He X, Lau EHY, Wu P, et al (2020) Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 26:672–675. https://doi.org/10.1038/s41591-020-0869-5

Hegyi A, Ziebuhr J (2002) Conservation of substrate specificities among coronavirus main proteases. J Gen Virol 83:595–599. https://doi.org/10.1099/0022-1317-83-3-595

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A Linear Constraint Solver for molecular simulations. J Comput Chem. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Huang C, Wang Y, Li X, et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. https://doi.org/10.1016/S0140-6736(20)30183-5

Jin Z, Du X, Xu Y, et al (2020a) Structure of M pro from COVID-19 virus and discovery of its inhibitors. Nature. https://doi.org/10.1101/2020.02.26.964882

Jin Z, Du X, Xu Y, et al (2020b) Structure of Mpro from COVID-19 virus and discovery of its inhibitors. Nature. https://doi.org/10.1038/s41586-020-2223-y

Klemm T, Ebert G, Calleja DJ, et al (2020) Mechanism and inhibition of the papain□like protease, PLpro, of SARS□CoV□2. EMBO J 39:1–17. https://doi.org/10.15252/embj.2020106275

Lemkul JA, Allen WJ, Bevan DR (2010) Practical considerations for building GROMOS-compatible small-molecule topologies. J Chem Inf Model. https://doi.org/10.1021/ci100335w

O'Boyle NM, Banck M, James CA, et al (2011) Open Babel: An Open chemical toolbox. J Cheminform. https://doi.org/10.1186/1758-2946-3-33

Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem. https://doi.org/10.1002/jcc.20084

Pronk S, Páll S, Schulz R, et al (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. https://doi.org/10.1093/bioinformatics/btt055

Protein Data Bank (2019) RCSB PDB: Homepage. Rcsb Pdb

Ren Z, Yan L, Zhang N, et al (2013) The newly emerged SARS-Like coronavirus HCoV-EMC also has an "Achilles" heel": Current effective inhibitor targeting a 3C-like protease." Protein Cell. https://doi.org/10.1007/s13238-013-2841-3

Rufener R, Dick L, D'Ascoli L, et al (2018) Repurposing of an old drug: In vitro and in vivo efficacies of buparvaquone against Echinococcus multilocularis. Int J Parasitol Drugs Drug

Resist. https://doi.org/10.1016/j.ijpddr.2018.10.011

Tian W, Chen C, Lei X, et al (2018) CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Res. https://doi.org/10.1093/nar/gky473

Toukmaji A, Sagui C, Board J, Darden T (2000) Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. J Chem Phys. https://doi.org/10.1063/1.1324708

Trott,O., Olson AJ (2019) Autodock vina: improving the speed and accuracy of docking. J Comput Chem. https://doi.org/10.1002/jcc.21334.AutoDock

Turner P (2005) XMGRACE, Version 5.1. 19. Cent Coast Land-Margin Res Oregon Grad Inst Sci Technol Beavert. https://doi.org/10.1163/_q3_SIM_00374

Tyagi R, Srivastava M, Jain P, et al (2020) Development of potential proteasome inhibitors against Mycobacterium tuberculosis. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2020.1835722

Umesh U, Kundu D, Selvaraj C, et al (2020) Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target. J Biomol Struct Dyn 0:1–7. https://doi.org/10.1080/07391102.2020.1763202

Vollmayr-Lee K (2020) Introduction to molecular dynamics simulations. Am J Phys. https://doi.org/10.1119/10.0000654

Wang E, Sun H, Wang J, et al (2019) End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem. Rev.

World Health Organization (2020) WHO Coronavirus Disease (COVID-19) Dashboard. In: World Heal. Organ.