HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

الجمعية السعودية لعلوم الحياة
SAUDI BIOLOGICAL SOCIETY

Original article

Finding potential inhibitors for Main protease (Mpro) of SARS-CoV-2 through virtual screening and MD simulation studies

N. Anis Ahamed *, Ibrahim A. Arif

Department of Botany and Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Keywords:
SARS-CoV-2
Main protease (Mpro)
Covid-19
Virtual screening
MD simulation

ABSTRACT

SARS-CoV-2 is a highly hazardous species that can infect people with Covid-19 disease, dramatically increasing mortality rates worldwide. Plenty of researches have been done to find drugs or inhibitors, with this study aiming to identify an inhibitor within the ChEMBL database using computational approaches. From the ChEMBL library, 19,43,048 compounds which are known type of small compounds and proteins were downloaded and docked with the Main protease (M^{pro}). After performing compound screening using Lipinski's rule, Qikprop analysis following with virtual Screening, Induced Fit Docking (IFD) and MM-GBSA analysis with the Glide and Prime modules of Schrödinger, the best complex was subjected to MD simulation with Desmond. According to the docking results, small protein 2,371,668 and compound 1,090,395 were docked with Main protease with −12.6, −12.0 kcal/mol dock score and interacted with the functional site residues His 41 and Cys 145, as well as the binding site residues Thr 26, Phe 140, Asn 142, Gly 143, Glu 166, and Gln 189. Complex structures were shown to be steadier by the MD simulation study because both the ligands heavy atoms and the protein Cα atoms' RMSD values fell within acceptable ranges. As a result, this research suggests that the molecule CHEMBL2371668 and the compound CHEMBL1090395 may inhibit the activity of Main protease, and the usefulness of these molecules will be examined further through experimental research.

1. Introduction

The Coronaviridae family includes the coronaviruses, Severe Acute Respiratory Syndrome 2 (SARS-CoV-2) that causes COVID-19. It encompasses the big positive-sense and single stranded RNA, it has a genome made up of about 30,000 nucleotides (Nga et al., 2011). Most commonly, it causes cough, fever, severe body pain, loss of taste or smell, tiredness, difficulty breathing, and ADRS that may lead to death (Jin et al., 2020a, 2020b).

The main pandemic occurrence of the 21st century because of SARS-CoV-2 has become a worldwide hazard to people healthiness by way of its extreme rate of infection resulting in mortality (Mengist et al., 2021). As of December 2020, there were 64,326,880 reported cases and 1,488,992 deaths worldwide (Mody et al., 2021). According to the World Health Organisation, 6,953,743 people died and 768,983,095 people were afflicted globally as of July 2023. The severity of the sickness, the viral infection and transmission, and the emergence of immunity against the virus, including response to vaccination, may all be impacted by SARS-CoV-2-specific pathogenic characteristics (Kumar et al., 2020).

A specific form of cysteine protease known as Main Protease (Mpro) is otherwise known as 3-Chymotrypsin-like protease. (3CLpro), which generates functional proteins through hydrolyses the viral polyproteins (Jo et al., 2020). Mpro is a conserved enzyme in the Coronaviridae family with His41 and Cys145 forming its catalytic dyad (Hu et al., 2022). In coronaviruses, the Mpro is one of the well-characterized healing target (Anand et al., 2003) and it plays major role for coronavirus replication and it's thought of a crucial target for disorders brought on by coronaviruses, such as COVID-19 (Needle et al., 2015). Currently there are no known human host-cell proteases with equivalent selectivity suggests that choosing Mpro as an outstanding target for drugs (Ullrich and Nitsche, 2020).

Tremendous varieties of revisions are current to work out the therapeutic use of antivirals and immune modulators to resolve Covid-19. The effective treatment for COVID-19 and coming up with prospective medication that might stop infection and disease progression is important (Connelly, 2020; Chaudhary et al., 2021). This study focused on the natural and synthetic compounds of the ChEMBL database to find the

Peer review under responsibility of King Saud University.
* Corresponding author.
E-mail addresses: anazeer@ksu.edu.sa (N.A. Ahamed), iaarif@ksu.edu.sa (I.A. Arif).

https://doi.org/10.1016/j.sjbs.2023.103845
Received 23 August 2023; Received in revised form 9 October 2023; Accepted 15 October 2023
Available online 20 October 2023

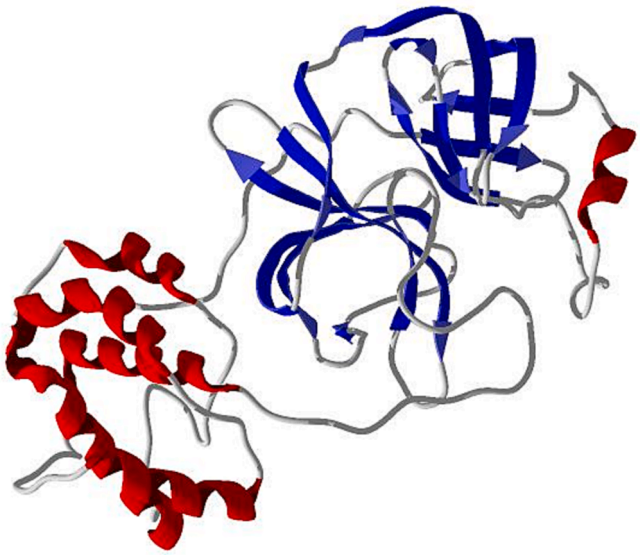


Fig. 1. Crystal structure of Main Protease of SARS-CoV-2 (6LU7).

inhibitor against the Main protease of SARS-CoV-2 through *in silico* approaches virtual screening, binding free energy calculation and molecular dynamics simulation studies.

2. Materials and methods

2.1. Protein structure retrieval

Crystal structure of Main protease (Mpro) enzyme of SARS-CoV-2 was retrieved from the PDB database (www.rcsb.org), which comprises evidence about experimentally solved structures, especially protein macromolecules elucidated by the various techniques (Berman et al., 2000).

2.2. Protein domain and active site analysis

Protein structural and functional information was studied in detail using Uniprot (www.uniprot.org) database (Apweiler et al., 2004). Protein domain analysis was carried out using Pfam and Interpro (https://www.ebi.ac.uk/interpro) databases. Commonly, proteins consist of one or more functional sections known as domains. (Finn et al., 2016). Protein active site was found using Uniprot database and inhibitor site was predicted using Protein Data Bank. Consurf (http://consurf.tau.ac.il) is another important server to predict the residues that are functionally and structurally significant in a protein (Celniker et al., 2013).

2.3. Preparation of compound structures

The compounds retrieved from the ChEMBL (www.ebi.ac.uk/chembl/) database for this study (Gaulton et al., 2017). Further these compounds were analysed for finding their physico-chemical

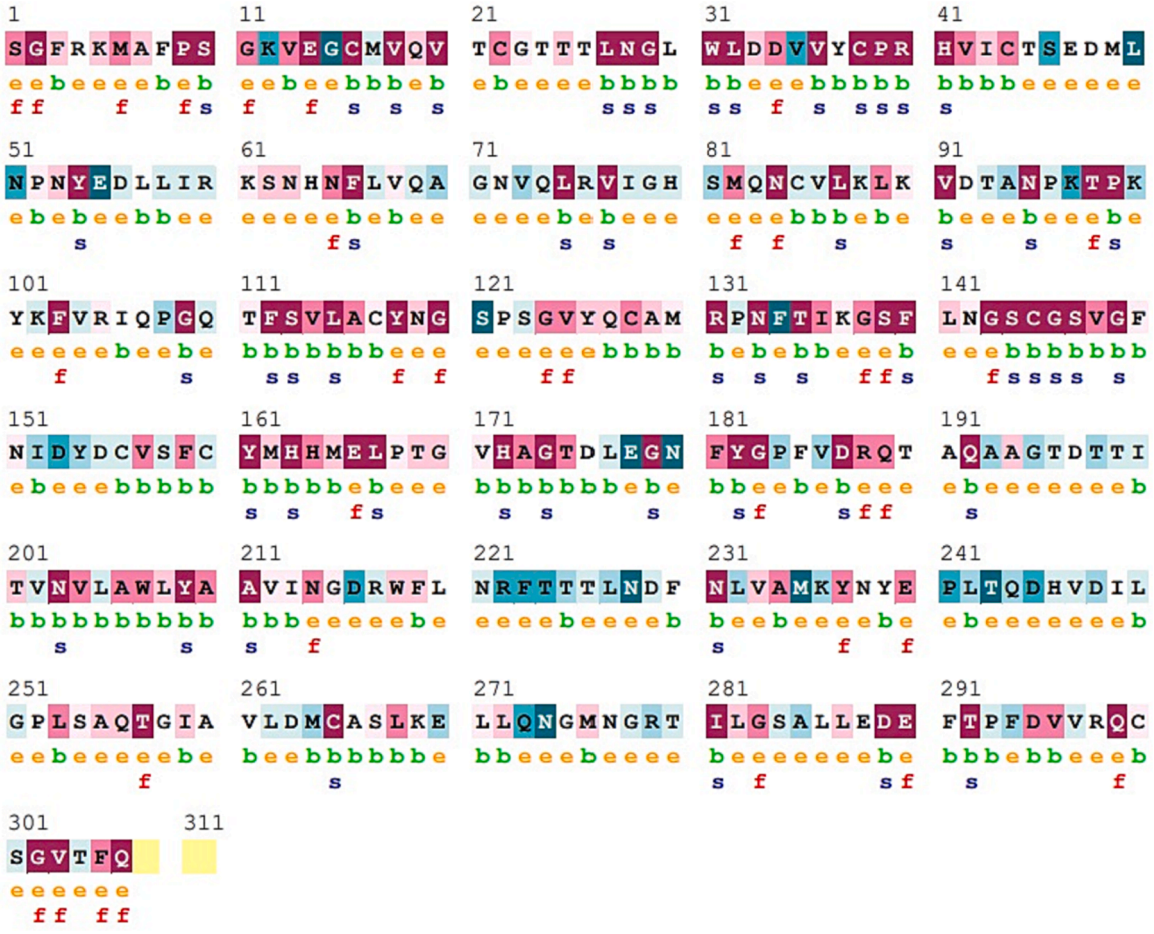


Fig. 2. Functionally and structurally important residues of Main protease (Mpro). Highlighted in dark red is conserved residues, white is an average and blue is variable. The letter denoted 'b' is buried residues and "e' is exposed residues. Then 's' is structurally conserved and 'f' is functionally conserved residues.

Table 1
Selected top 10 compounds from virtual Screening process with Mpro.

S.No	Compound Id	Dock Score (kcal/mol)	Interacting Residues	Bond Length (Å)
1.	2,371,668	−9.6	Thr26, Phe 140, Gly 143, Glu 166(3), Gln 189	1.73, 1.81, 1.84, 3.47, 1.87, 1.97, 1.77
2.	1,090,395	−9.4	Thr 26(2), Leu 141, Gln 189	1.75, 1.98, 2.19,0.2.12
3.	4,127,290	−9.3	Thr 26, Leu 141, Asn 142, Gly 143	1.90, 1.95, 1.86, 1.90
4.	3,941,250	−9.1	Thr 26, Gly 143, Glu 166(2), Gln 189	1.91, 1.80, 2.04, 2.13, 2.73
5.	4,285,498	−9.0	His 164, Glu 166(2), Thr 190	2.04, 2.11, 1.89, 1.77
6.	3,634,575	−8.9	Thr 26, His 41, Gly 143, Gln 189	1.84, 4.14, 1.67, 2.13
7.	3,408,435	−8.9	Leu 141, Gly 143, His 164, Gln 189	2.09, 1.84, 2.14, 1.97
8.	592,991	−8.8	Phe 140, Gly 143, Glu 166	1.79, 1.83, 1.61
9.	3,408,436	−8.7	Leu 141, Gly 143, His 164, Gln 189	2.11, 1.85, 2.12, 1.97
10.	1,872,577	−8.7	His 41, Leu 141, Gly 143, His 164, Gln 189	4.89, 1.82, 1.98, 1.80, 1.96

properties using Lipinski's rule (Petit et al., 2012) and Qikprop analysis (Ricci-López et al., 2019).

2.4. Virtual Screening

Virtual Screening process was executed by Glide module of Schrödinger (Fu et al., 2020). Primarily, the protein molecule was prepared by performing the following procedures with the protein preparation wizard 1. Preprocess the protein molecule, 2. Protein optimization and 3. Protein minimization. In preprocessing, importantly protein molecule is adding with the hydrogens and removed water molecules from it. For minimizing the protein molecule, OPLS3e force field was used (Velusamy et al., 2023). Then grid was prepared using Receptor Grid Generation option by adding the highly important residues of the protein. Before starting the preparation of compounds, the process was started with the screening of compounds with the analysis of Qikprop, Lipinski's rule and reactive functional groups. After these processes, further the selected compounds were taken for ligand preparation. Further these compounds were processed for molecular docking using High Throughput Virtual Screening (HTVS). Then, top 10 percent of all stated compounds were further processed for molecular docking using Standard Precision (SP) method with flexible approach. Again, top 10 percent of the good scored compounds from SP method were further processing with Extra Precision (XP) docking method (Yadav et al., 2022). Finally, top 10 percent of best scored compounds were identified form the 19,43,048 compounds.

2.5. Induced Fit docking (IFD)

This method was specially executed because it makes the protein molecule with more flexibility especially the binding site was reformed based on the type of ligand binding (Miller et al., 2021). The same processes, protein preparation, Receptor grid generation, ligand preparation were done. The selected best 10 compounds from XP method was used for this process. IFD produced the various group of ligand poses, and the optimal one was selected according to the dock score, h-bond interactions with the highly important residues.

2.6. Binding free energy analysis

Using the Prime module, Schrödinger (Abuthakir et al., 2021), the binding free energy of the selected compound with Mpro was computed. The progression of the MM-GBSA study was done using the OPLS3e force field and the VSGB solvation model. The following formula was used to compute binding free energy,

$$\Delta G\,\text{binding} = \Delta G\,(\text{complex}) - \Delta G\,(\text{Protein}) - \Delta G\,(\text{Ligand})$$

2.7. MD simulation studies

In order to confirm that the complex was stable, the best-docked complex structure was subsequently executed for molecular dynamics simulation studies utilising the Desmond module, Schrödinger

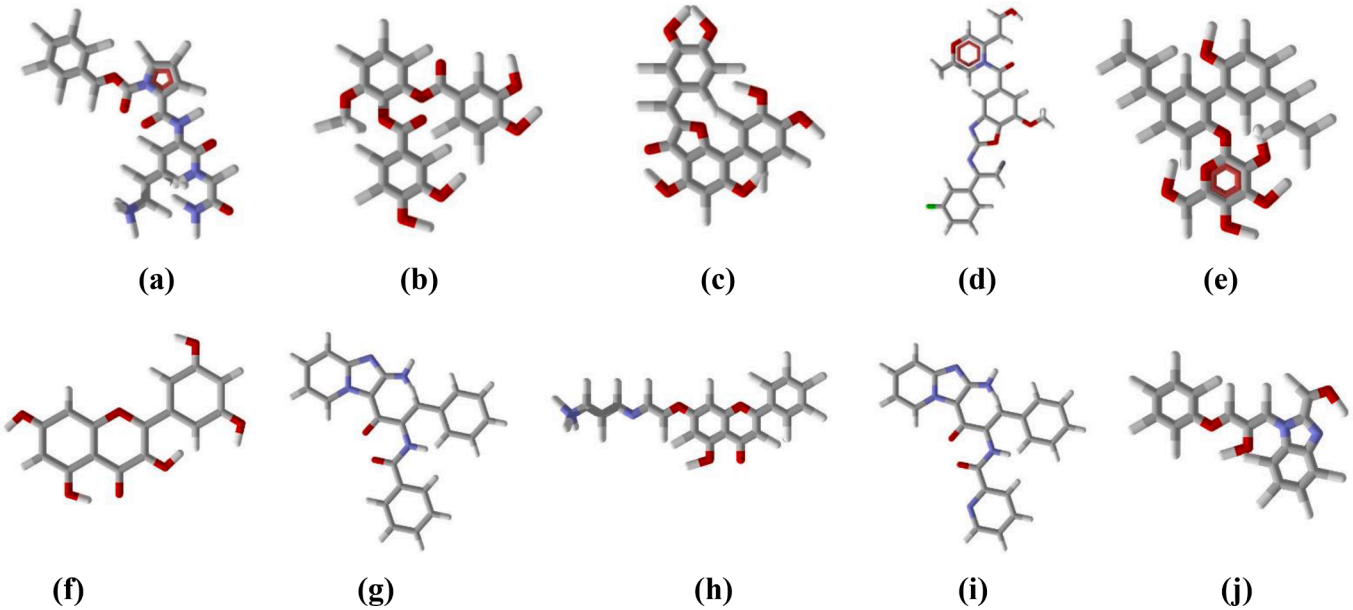


Fig. 3. 3D structures of top 10 compounds selected from the virtual Screening process 2371668, b) 1090395, c) 4127290, d) 3941250, e) 4285498, f) 3634575, g) 3408435, h) 592991, i) 3408436, j) 1,872,577.

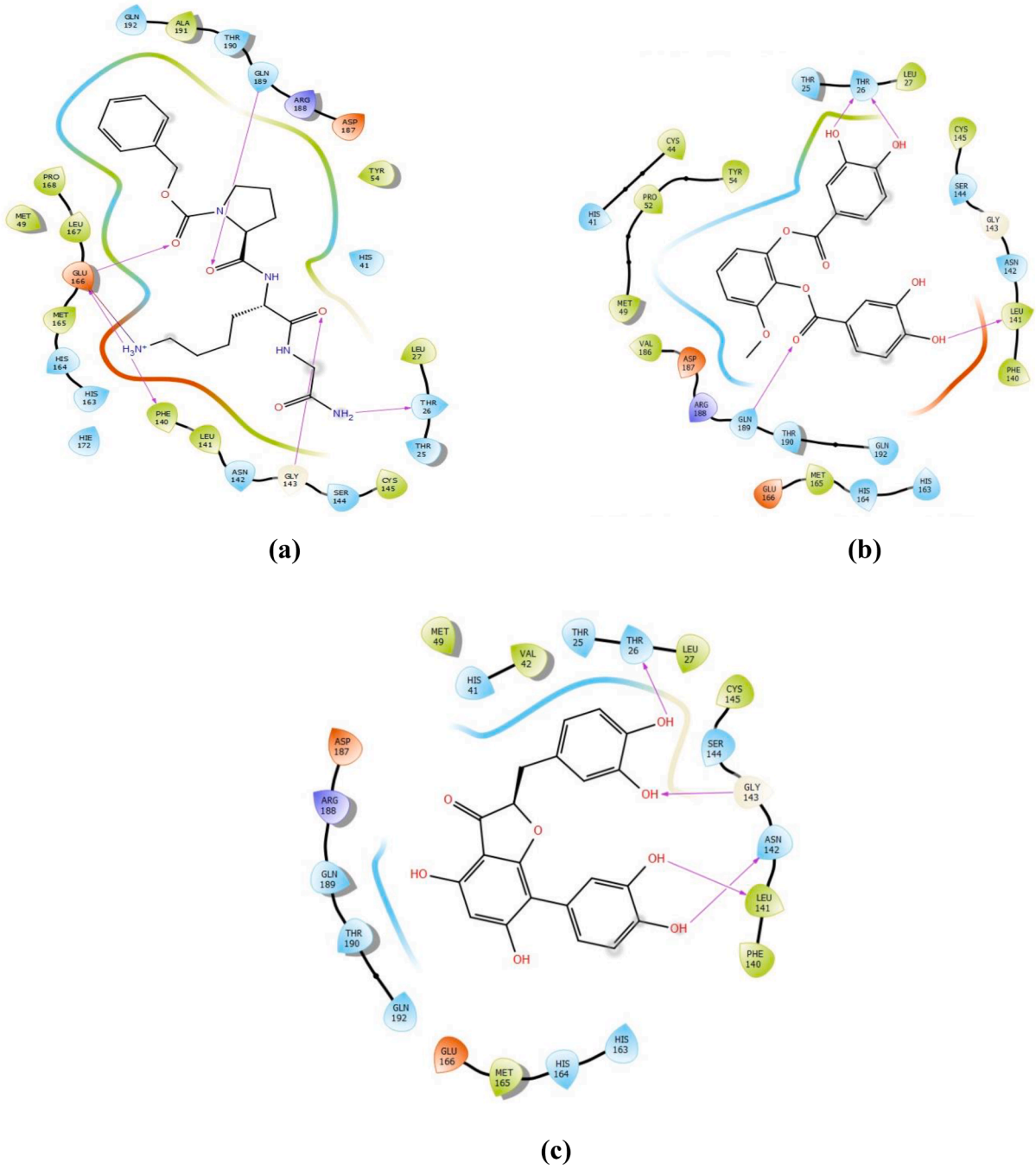


Fig. 4. Docked complex structures of Main protease (MPro) with a) 2371668, b) 1,090,395 and c) 4,127,290.

(Gopinath and Kathiravan, 2021). For MD simulation with a time of 100 ns, the docked complex of Mpro with 2,371,668 and Mpro with 1,090,395 were employed. The solvent model, boundary conditions and force field were set up in the system builder before to starting the MD simulation.

The system filled with TIP3P water model, OPLS 2005 force field was selected, and the complex was centred with boundary conditions of 10 Å distance in a cubic box and ions were added (Balakrishnan et al., 2022). The complex structure's energy was reduced; the NPT ensemble class was employed to simulate complicated structures. The temperature at 300 K, and used to fix the 1 bar pressure coupling (Azam et al., 2020).

The RMSD, RMSF, and protein–ligand bindings were calculated using simulation interaction analysis in order to study the integrity of the ligand inside the protein.

3. Results

3.1. Protein Crystal structure

Crystal structure of Main Protease (Mpro) enzyme (Fig. 1) was retrieved from the PDB database, protein id is 6LU7. The crystal structure was experimentally produced using X-ray diffraction method with 2.16 Å resolution and this protein found in cytoplasm and golgi apparatus region.

3.2. Protein domain and active/inhibitor site analysis

The domain Peptidase_C30, which spans the range of 3264–3569, is present in the main protease enzyme. His 41, Cys 145 were active site

Table 2
IFD result of top 10 compounds docked with the Mpro.

S. No	Compound Id	Dock Score (kcal/mol)	Interacting Residues	Bond Length (Å)
1.	2,371,668	−12.6	Thr 26, His 41, Phe 140, Asn 142, Gly 143, Cys 145, Glu 166, Gln 189	1.93, 2.69, 1.90, 1.96, 1.94, 2.09, 2.22, 1.72
2.	1,090,395	−12.0	Thr 26, His 41 (2), Asn 142, Glu 166(2), Gln 189	1.91, 2.00, 2.24, 2.09, 1.82, 1.94, 1.96
3.	4,127,290	−11.8	Thr 26, His 41, Phe 140, Gly 143, His 164, Glu 166, Asp 187	1.81, 3.85, 1.81, 1.84, 2.30, 2.00, 1.89
4.	3,941,250	−11.4	His 41(2), Gly 143, Cys 145, Glu 166(2)	1.90, 2.01, 1.77, 2.64, 2.10, 2.37
5.	4,285,498	−10.9	Phe 140, His 163, Glu 166 (2), Gln 189	1.81, 2.04, 1.69, 2.15, 2.06
6.	1,872,577	−9.6	His 41, Asn 142, Gly 143, Gln 189	4.88, 1.66, 2.05, 1.84
7.	3,408,435	−9.3	His 41(2), His 164, Glu 166, Gln 189	3.95, 4.75, 2.27, 2.61, 1.81
8.	3,634,575	−9.1	His 41, Leu 141, Gly 143, Glu 166, Asp 187	5.01, 1.78, 2.03, 2.22, 1.90
9.	592,991	−9.1	Phe 140, Gly 143, His 163, Glu 166(3)	2.08, 1.91, 5.38, 1.57, 2.59, 2.78
10.	3,408,436	−9.0	His 41, Leu 141, His 163, His 164, Glu 166, Gln 189	3.97, 2.13, 2.00, 2.01, 2.02, 1.87

residues a nd the inhibitor N3 interacted with the residues Phe 140, Gly 143, Cys 145, His 164, Glu 166, Gln 189, and Thr 190 in the crystal structure deposited in the database (Jin et al., 2020a, 2020b). In addition to that, Consurf server predicted the functionally and structurally important residues by comparing more than 150 similar sequences (Fig. 2).

3.3. Compounds preparation

The compounds were downloaded from the ChEMBL library, totally 19,43,048 compounds were downloaded. Among these, some of these compounds have known clear activity and some are not known. These compounds were further evaluated for its physico-chemical properties.

3.4. Virtual Screening

The prepared protein was docked with 19,43,048 compounds using virtual Screening workflow by Glide module of Schrödinger. By the analysis of compounds properties using Lipinski's rule, Qikprop properties such as CNS, metabolism, Human oral absorption, BBB, Caco permeability, water solubility, human serum albumin binding, skin absorbency and reactive functional groups. From this analysis, 7,59,804 compounds were taken for further analysis because these compounds only satisfied the various properties of Lipniski's rule and Qikprop, rest of the compounds were eliminated. There are three types of molecular docking methods used in the virtual Screening workflow, 1. HTVS, 2. SP docking and 3. XP docking. At the end of HTVS process 75, 980 compounds (Supp.Table 1) were selected as the top 10 percent based on dock score from the 7,59,804 compounds. Further, SP and XP methods were processed, through these 7,598 (Supp.Table 2) and 759 (Supp.Table 3) compounds respectively, were selected as top 10 percent from each process.

The overall result of virtual Screening process was shown in Fig. 6. From the virtual Screening analysis, 592991, 1090395, 1872577, 2371668, 3408435, 3408436, 3634574, 3634575, 3941250, 4127290, 4,285,498 were chosen as top ranked compounds based on dock score and interactions between ligand with highly important residues of the protein (Table 1). The compounds 3d structures that were selected as top 10 by virtual Screening process as shown in Fig. 3 and top 3 docked complex structures were shown in Fig. 4.

3.5. Induced fit docking

From the result of IFD analysis (Table 2), the compound 2,371,668 produced better result compare with other compounds, it had −12.6 kcal/mol dock score and bonding with the active site residues His 41, Cys 145, also it is interacted with Thr 26, Phe 140, Asn 142, Gly 143, Glu 166 and Gln 189 (Fig. 5a). The compound 1,090,395 had interacted with the residues Thr 26, His 41, Asn 142, Glu 166, Gln 189 with dock score −12.0 Kcal/mol (Fig. 5b). The compound 4,127,290 had dock score −11.8 kcal/mol and bonding with the residues Thr 26, His 41, Phe 140, Gly 143, His 164, Glu 166 and Asp 187 (Fig. 5c). From the overall docking analysis, the small molecule 2,371,668 and compound 1,090,395 had better result than other compounds, hence this docked complexes further evaluate for computational molecular dynamics simulation analysis.

3.6. Binding energy analysis

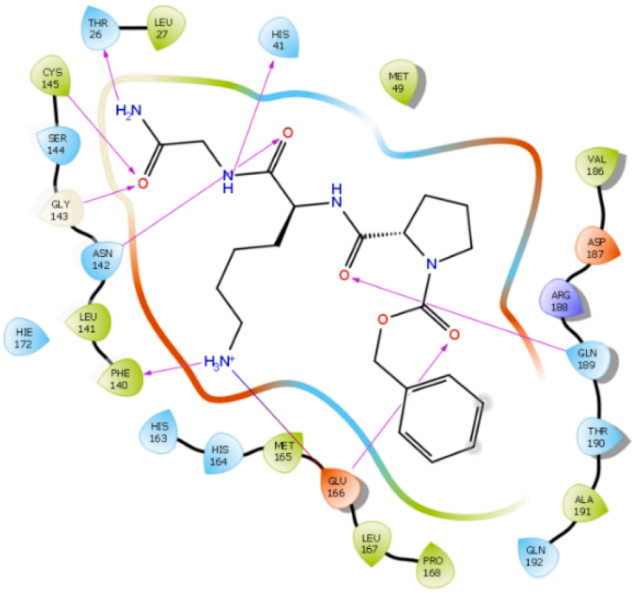
The binding energy of each compound with the receptor was calculated through MM-GBSA analysis. From the result, the compound 2,371,668 had very efficient binding affinity with the protein Main Protease (Mpro), it showed the binding free energy value as −96.37 Kcal/mol. Also, the compounds 1090395, 4127290, 3941250, 4,285,498 had showed binding free energy value as −80.83, −81.98, −84.23, −84.46 kcal/mol, respectively (Table 3).

3.7. MD simulation studies

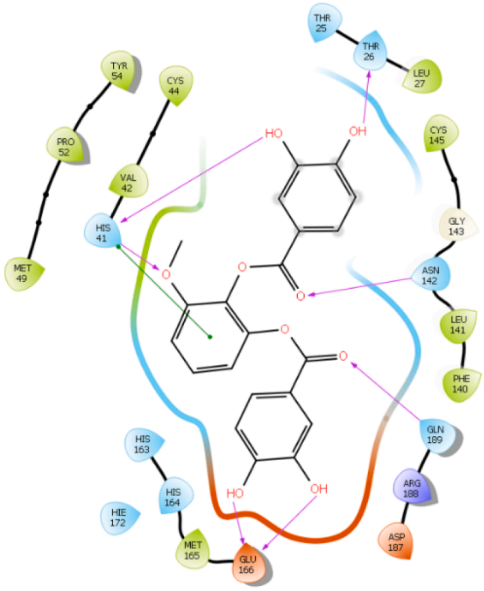
Molecular dynamics simulation studies for complex structure of Main protease (Mpro) with 2,371,668 was executed for examining steadiness of the complex up to 100 ns. The result was analysed using the various plot results like RMSD, RMSF, protein–ligand contacts, 2D diagram of interactions between Main protease enzyme and compound 2371668. RMSD plot directed that the docked complex structure was stable, because the protein and ligand fluctuations were within the acceptable range (1–3 Å), especially from the starting to around 65th ns, the fluctuation range of both protein C alpha atoms and ligand heavy atoms were within 3 Å and both the atoms were arranged very closer (Fig. 7a). The values of ligand heavy atoms were not larger than the protein, hence RMSD plot evidenced that the complex was stable.

RMSF plot exposed that the residues other than N and C terminal were not fluctuated more over the period of simulation. The residues around the 50th position were slightly fluctuated among all the residues, even though they were in acceptable range. The protein structure majorly contains beta strands and alpha helices from 0 to 170 and 200 to 270 respectively, also the loops region present with in these regions. Usually these secondary structure elements were more rigid and RMSF plot observes the fluctuation of these regions were lesser than the other regions. Also, the active site and binding site residues His 41, Phe 140, Gly 143, Cys 145, His 164, Glu 166, Gln 189, and Thr 190 have lesser fluctuation over the simulation period (Fig. 7b). The atoms of 2,371,668 were with within the acceptable range except few atoms (Fig. 7c).

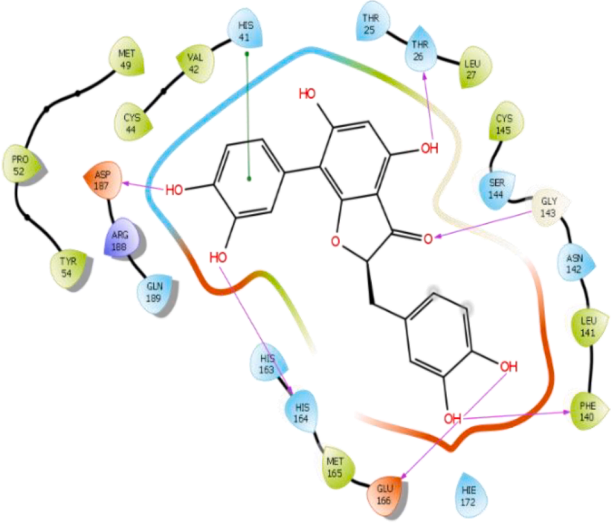
The plot of protein–ligand contacts displayed that the several interactions like hydrogen, hydrophobic, ionic, water bridges between the protein Mpro and ligand 2371668. The residues Thr 26, His 41, Cys 44, Ser 46, Glu 47, Asn 142, Gly 143, Ser 144, Cys 145, His 164, Glu 166, Gln 189 have hydrogen bond and water bridges with the protein Mpro (Fig. 8a). Met 49, Met 165, Leu 167, Pro 168 have hydrophobic interactions with the protein Mpro. Especially, the functionally important residues Gly 143, Glu 166 and Gln 189 have hydrogen bond interactions with the protein and these interactions maintained 34 %, 93 % and 89 %



(a)



(b)



(c)

Fig. 5. Docked complex structure of Main protease (Mpro) with small molecule (a) 2,371,668 and compounds (b) 1,090,395 (c) 4,127,290.

respectively, throughout the simulation time. Ser 46 and structurally conserved residue His 41 have hydrogen bond interaction and it is maintained more than 50 % time of total simulation period. In addition, Thr 26 that is structurally conserved residue have hydrogen bond interaction and it is maintained 30 % of simulation time (Fig. 8b, c).

3.8. MD simulation studies of compound 1,090,395 with Mpro

The result of MD simulation reveals that the complex structure of compound 1,090,395 with Main protease was stable. The RMSD plot expressed that the equilibrium of C alpha atoms of Main protease and heavy atoms of compound 1090395, because the fluctuations of both were within the 3 Å throughout the simulation period of 100 ns (Fig. 9a). RMSF plot showed the equilibrium of protein, from the analysis the residues of Main protease were not fluctuated more except the residues of N and C terminal (Fig. 9b). Especially fluctuations of the residues His 41, Phe 140, Gly 143, Cys 145, His 164, Glu 166, Gln 189, and Thr 190 were within 2 Å up to the completion of the simulation, hence the

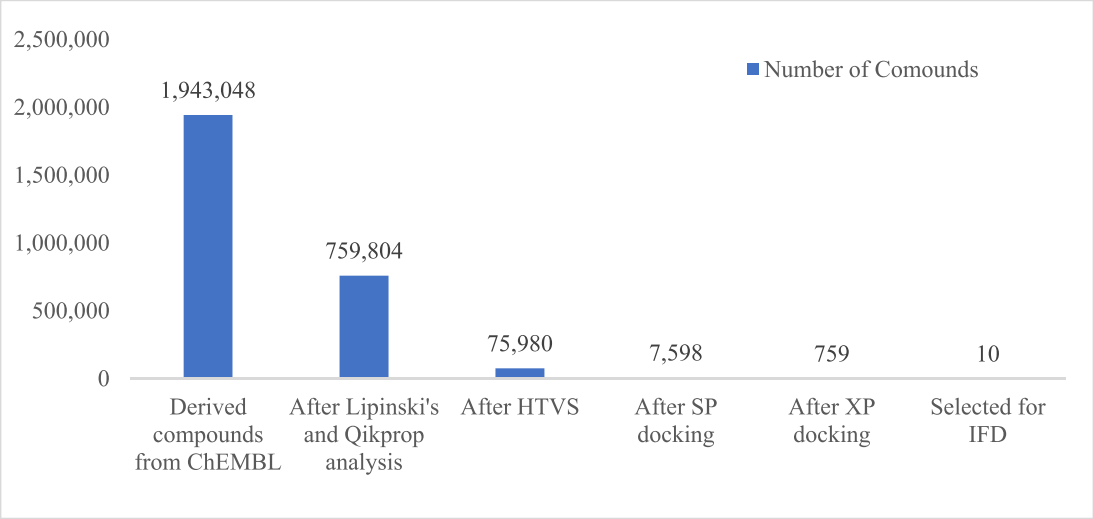


Fig. 6. Output of systematic process of Virtual Screening and IFD.

Table 3
MM-GBSA Binding free energy calculation of complex structures.

S. No	Compound Id	MM-GBSA dG Bind (kcal/mol)
1.	Benzyl (2*S*)-2-[[(2*S*)-6-amino-1-[(2-amino-2-oxoethyl) amino]-1-oxohexan-2-yl]carbamoyl]pyrrolidine-1-carboxylate (2371668)	−96.37
2.	[2-(3,4-Dihydroxybenzoyl)oxy-3-methoxyphenyl] 3,4-dihydroxybenzoate (1090395)	−82.83
3.	7-(3,4-dihydroxyphenyl)-2-[(3,4-dihydroxyphenyl) methyl]-4,6-dihydroxy-1-benzofuran-3-one (4127290)	−81.98
4.	(2-{[1-(3-Chlorophenyl)-2-fluoroethyl]amino}-7-methoxy-1,3-benzoxazol-5-yl)[5-(2-hydroxyethyl)-2-methylmorpholin-4-yl]methanone (3941250)	−84.23
5.	2-(beta-D-Glucopyranosyloxy)-5,5′-diallylbiphenyl-2′-ol (4285498)	−84.46
6.	2-methylsulfanylcyclohexa-2,5-diene-1,4-dione (592991)	−66.41
7.	4-Bromo-2-[[2-chloro-5-(trifluoromethyl)phenyl] iminomethyl]-6-nitrophenolate (3408436)	−76.28
8.	(2,2-Dichloro-1-{[(4-methylphenyl)carbonyl]amino} ethenyl) (triphenyl)phosphonium (3634575)	−77.84
9.	N-[(2*S*)-1-(2-aminoimidazo[1,2-a]pyridin-3-yl)-1-oxo-3-phenylpropan-2-yl]benzamide (3408435)	−77.25
10.	1-[2-(hydroxymethyl)-1H-1,3-benzodiazol-1-yl]-3-phenoxypropan-2-ol (1872577)	−65.98

protein was steady. The heavy atoms of ligand 1,090,395 were within the 1.5 Å, it shows the ligand was more stable (Fig. 9c).

3.9. Protein-Ligand interactions

The bar diagram of protein–ligand contacts showed that, the compound 1,090,395 had H-bond interactions with Gly 143, Cys 145, His 164, Glu 166 and Gln 189 which are interacted residues of N3 inhibitor. In addition to that, the compound had H-bond interactions with Thr 26 and structurally important residue Ser 144. Further, the compound had hydrophobic interactions with Met 49, Met 165 and active site residue His 41. Then, water bridges with Thr 24, Thr 25, Leu 141, Glu 166 and Gln 189 (Fig. 10a). Timeline representation image shows that the interaction of compound 1,090,395 with various residues of Main protease throughout the simulation period. The compound 1,090,395 had strong interactions with Thr 26, Gly 143, Ser 144, Glu 166 and Gln 189, in addition to that compound has interactions with Met 49, Met 165 and active site residue His 41, Cys 145 (Fig. 10b). 2D image of interactions diagram represents that at the 100th ns, the compound has good interactions with the N3 inhibitor site residues Gly 143, Glu 166, Gln 189 and structurally conserved residue Gly 144. The compound has hydrogen bond interactions with residue Gly 143 with 85 % of the simulation period, 58 % with Glu 166, multiple interaction with Ser 144, 95 % with Gln 189 and 59 % with Thr 26 (Fig. 10c).

From the overall result of simulation studies, the docked structure of small protein molecule 2,371,668 with Mpro and compound 1,090,395 with Mpro maintained theirs stability along with the simulation period and retained the interactions of active site residue His 41, functionally conserved residues Gly 143, Glu 166, Gln 189, structurally important residue Thr 26. Hence, this simulation analysis recommends that the small protein molecule 2,371,668 and compound 1,090,395 have better action against SARS Corona Virus-2.

4. Discussion

Covid-19 disease produced by SARS-CoV-2 (Severe Acute Respiratory Syndrome–Corona Virus-2) and it is very threatening for humanity due to its mortality rate (Gupta, 2020). Continuous research work going on Covid-19 disease at global level for finding drugs and vaccines (Bahrami et al., 2022).

The protein Main protease (Mpro) found in the cytoplasm and golgi apparatus region. Hence, this protein is very good drug target because drug molecules aim to bind with the proteins that are found in the cytoplasm (Abuthakir et al., 2020). Potential approaches to current treatments for the Covid-19 include medicines that directly bind to and block Main protease. The Mpro is crucial for the virus replication (Kneller et al., 2022) and it is in charge of maturing the other protease of coronaviruses and additional significant polyproteins (Ziebuhr et al., 2000). According to Stoermer, (2020), the Mpro's sequence of SARS-CoV-2 is closely similar to Mpro of other species; especially it is nearly 96 % similar with Mpro of SARS-CoV and higher than 50 % similar to Mpro of MERS-CoV. For these reasons, the Mpro considered as a potential drug target for new drug discovery.

Among top 10 compounds, 9 compounds were mainly interacted with the active site residue His 41 and structurally conserved residue Glu 166. The residue Glu 166 is regarded as being crucial because it is necessary for the homodimerization of 3CLpro in SARS-CoV-2. Cheng et al, 2010 study revealed that the formation of dimerization of SARS Coronavirus main protease blocked when the residue Glu166 was mutated. The dimer formation is a significant for the enzymatic activity of protease enzyme hence any H-bond interaction to the Glu 166 can result in the deformation of dimer and it is suffer its enzymatic reactions (Zhang et al., 2020).

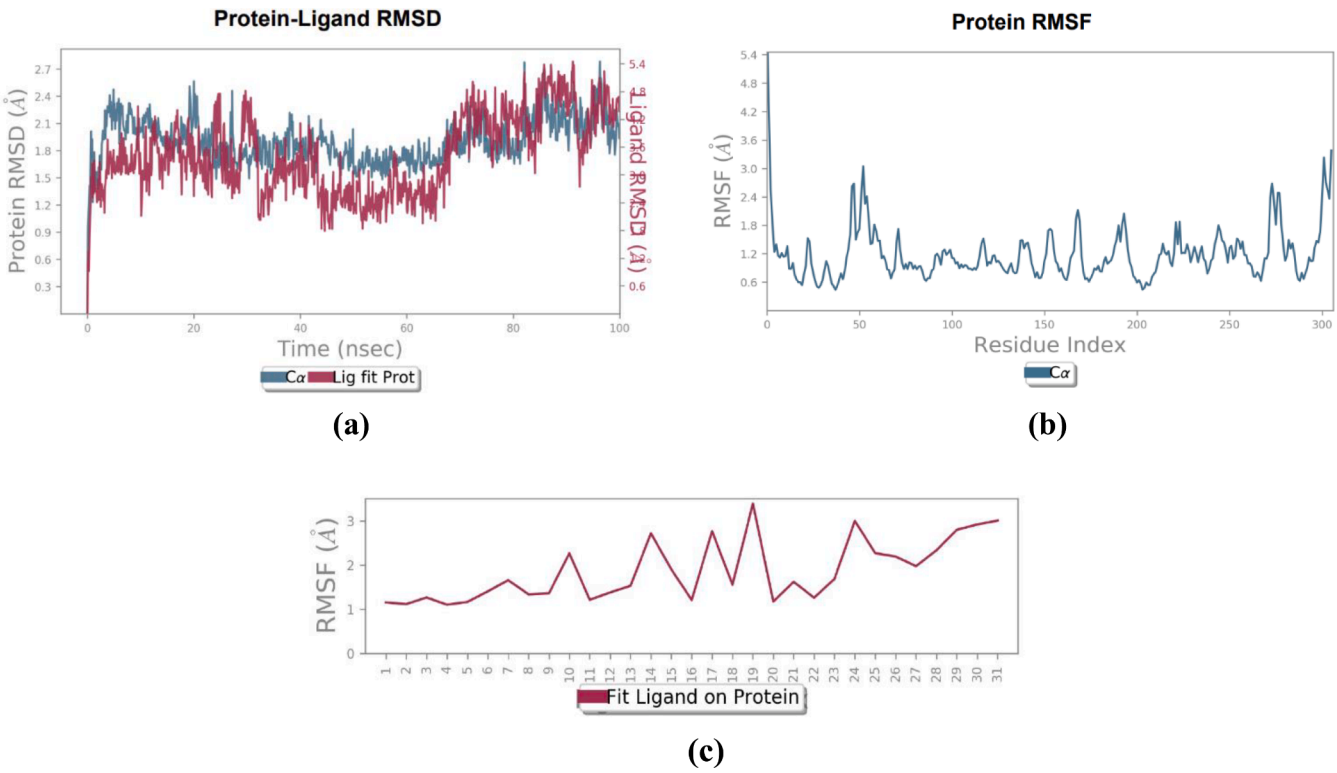


Fig. 7. **(a)** RMSD plot of Docked complex structure of Main protease (Mpro) with 2,371,668 and RMSF plot of **(b)** protein Main Protease, **(c)** 2,371,668.

The smaller protein molecule 2,371,668 had interaction with the dynamic catalytic site residue His 41 and Cys 145. MPro contains the His 41-Cys 145 catalytic dyad and this dyad is very vital for catalyses the cleavage of the protein substrate (Zhang et al., 2010). According to Ferreira et al (2020), the residues of catalytic dyad His 41-Cys 145 had lost their catalytic activity of Mpro due to alterations with Alanine residue in the dyad. In addition to that, Jin et al (2020b) showed that the anti-cancer medication Carmofur binds to the Mpro's Cys145 catalytic dyad with a potential experimental suppression of virus copying. According to Nguyen et al (2020), the ligands make hydrogen bonds with the most desirable residue Gly 143 of Main protease preceded by Cys145, His163, and Glu166. Among 10 compounds, 2371668, 4127290, 3941250, 592991, 3634575, 1,872,577 were made hydrogen bond interactions with the residue Gly 143.

The smaller protein 2,371,668 and compounds 1090395, 4,127,290 interacted with the residue Thr 26 which is a weighty amino acid in Mpro's binding pocket. According to the Mengist et al (2021), the drug molecules Lopinavir and ritonavir were interacted to the residues Thr 24, Thr 26, Asn 119 that were the significant amino acids in the active pocket. Various natural compounds interact with the binding site residues Thr 26, His 41, Cys 145, Glu 166, Gln 189 (Antonopoulou et al., 2022).

The smaller protein 2,371,668 had showed the interactions with the residues Thr 26, His 41, Gly 143, Glu 166 and Gln 189 in XP, IFD docking studies were again retrained in the MD simulation studies. The binding free energy of 2,371,668 is −96.37 kcal/mol and 1,090,395 has −82.83 kcal/mol. Previous research demonstrated that rescoring docked complexes based on MM-GBSA produce positive outcomes with experimental binding affinities (Zhang et al., 2015). MD simulation studies showed the steadiness of the docked complex structure of 2,371,668 with main protease, RMSD plot proved that the complex structure had strong stability because of lesser deviation between protein and ligand. Smaller deviation between protein and ligand indicates that the structure is more stable (Aier et al., 2016).

The compound 3,941,250 only already reported compound among top 10 compounds. It is reported that it has the activity of inhibit the function of protein thrombin which is otherwise known as coagulation factor II which contributes an significant part in thrombosis and haemostasis by transforming fibrinogen into fibrin during the development of blood clots, by encouraging platelet aggregation, and by triggering additional coagulation factors (Moser and Patterson, 2003). In addition to that maintaining arterial integrity during development and after birth, thrombin also helps in cell growth, tissue repair, and angiogenesis. Biological activities of other compounds are still not known because there is no experimental proof about active against the any protein molecule.

5. Conclusion

SARS-Cov-2 is a dangerous virus and Main protease (Mpro) this virus is much conserved. The developing inhibitors against this virus is an important need for human being for future. This study finds the potential inhibitor from the database of ChEMBL that contains more than 2 million of compounds. The smaller protein molecule 2,371,668 and compounds 1090395, 4127290, 3941250, 4,285,498 have potent activity by scored better dock score and interacted with the functionally, structurally important residues of Mpro. Further, the experimental study will carry out for checking the efficiency of these compounds especially the smaller molecule 2,371,668 and compound 1,090,395 against the Main protease (Mpro).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

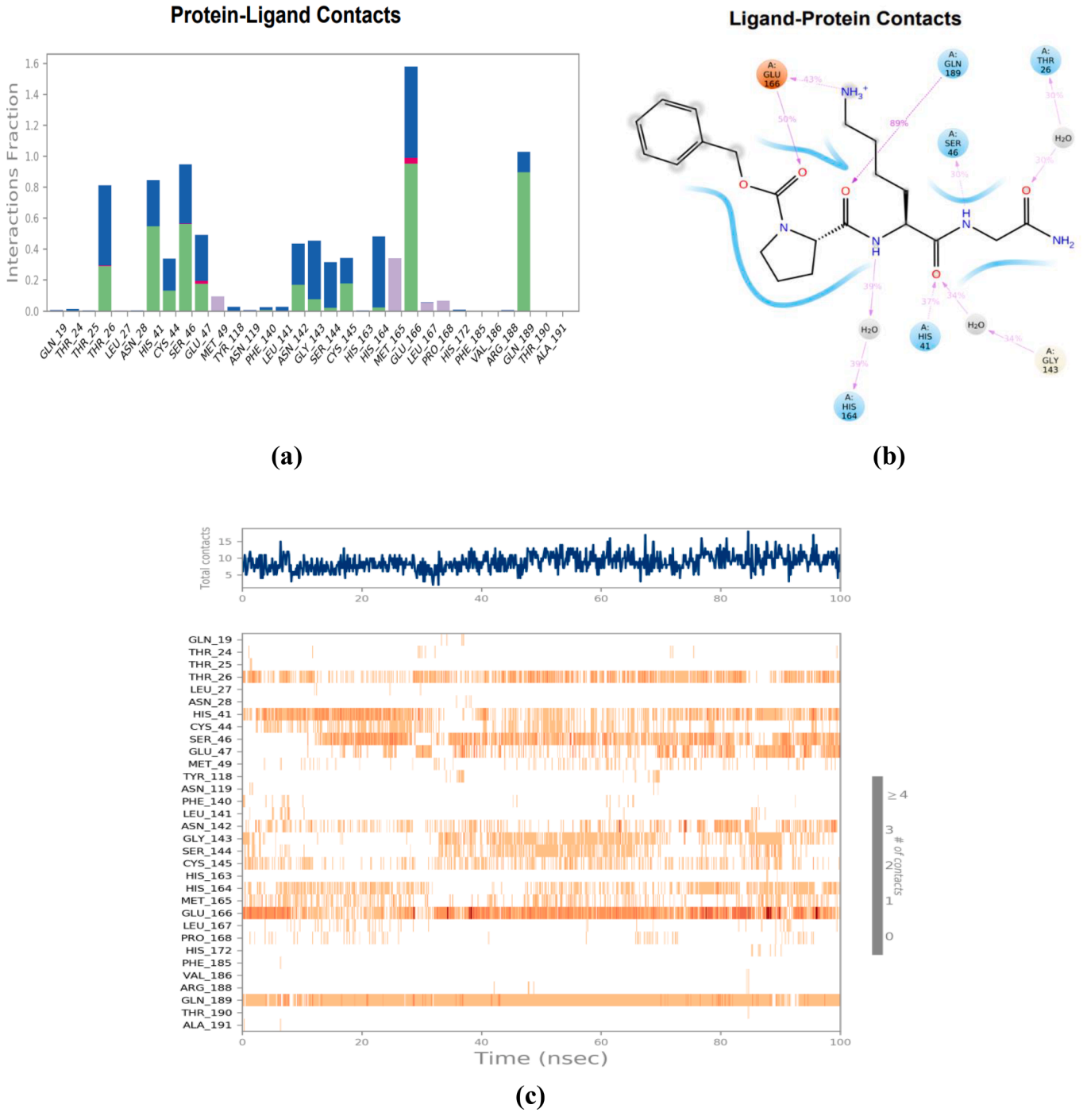


Fig. 8. **(a)** Various interaction types, green color - Hydrogen bonds, gray – hydrophobic, red – ionic interactions and blue – water bridges **(b)** 2D diagram of protein–ligand interactions and **(c)** Protein-Ligand contacts of compound 2,371,668 with the residues of Main protease (Mpro) up to 100 ns (darken lines indicates more than one interactions with those residues).

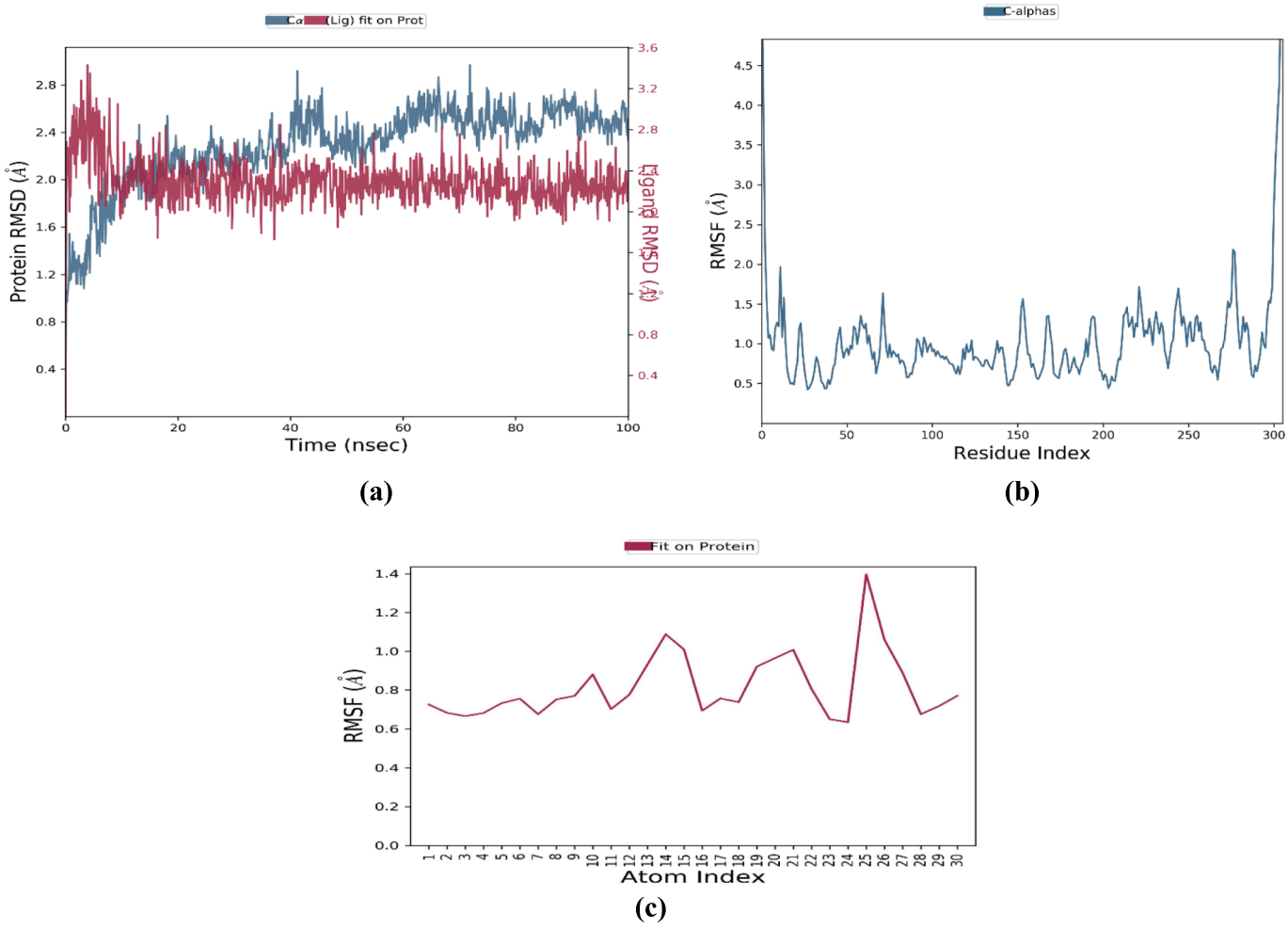
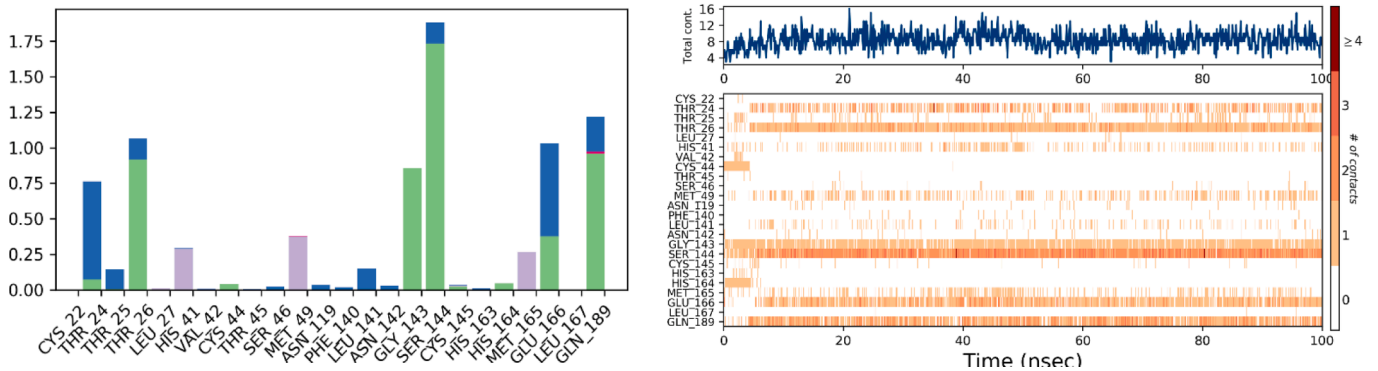
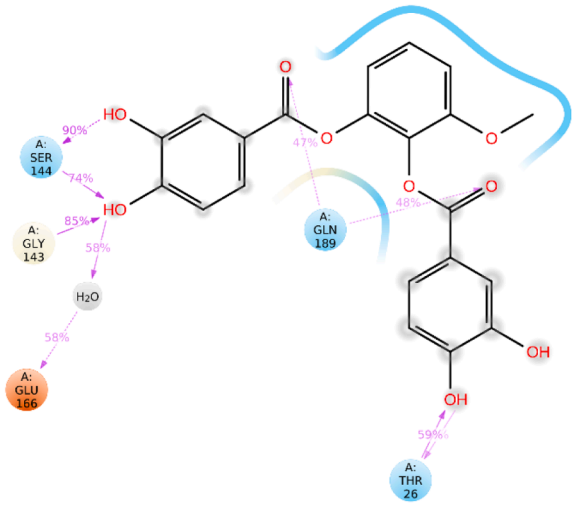


Fig. 9. **(a)** RMSD plot of docked structure of Main protease (Mpro) with compound 1090395, RMSF plot of **(b)** Main Protease protein and **(c)** compound 1,090,395.



(a) (b)



(c)

Fig. 10. **(a)** Various interaction types, green color - Hydrogen bonds, gray – hydrophobic, red – ionic interactions and blue – water bridges **(b)** Protein-Ligand contacts of compound 1,090,395 with the residues of Main protease (Mpro) and **(c)** 2D image of protein–ligand interactions at the 100th ns (darken lines indicates more than one interactions with those residues).

Acknowledgement

The authors extend their appreciation to the Deputyship for Research and Innovation "Ministry of Education" In Saudi Arabia for funding this research (IFKSUOR3-534).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2023.103845.

References

Abuthakir, M.H.S., Jebastin, T., Sharmila, V., Jeyam, M., 2020. Putative drug target identification in tinea causing pathogen Trichophyton rubrum using subtractive proteomimcs approach. Curr. Microbiol. 77 (10), 2953–2962.

Abuthakir, M.H.S., Sharmila, V., Jeyam, M., 2021. Screening Balanites aegyptiaca for inhibitors against putative drug targets in Microsporum gypseum – Subtractive proteome, docking and simulation approach. Infection Genetics Evolut. 90, 104755.

Aier, I., Varadwaj, P.K., Raj, U., 2016. Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. Sci. Rep. 6, 34984.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., Hilgenfeld, R., 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 300, 1763–1767.

Antonopoulou, I., Sapountzaki, E., Rova, U., Christakopoulos, P., 2022. Inhibition of the main protease of SARS-CoV-2 (Mpro) by repurposing/designing drug-like substances and utilizing nature's toolbox of bioactive compounds. Comput. Struct. Biotechnol. J. 20, 1306–1344.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., et al., 2004. Uniprot: the universal protein knowledgebase. Nucleic Acids Res. 1 (32), D115–D119.

Azam, F., Taban, I.M., Eid, E.E.M., Iqbal, M., Alam, O., Khan, S., Mahmood, D., Anwar, M.J., Khalilullah, H., Khan, M.U., 2020. An *in-silico* analysis of ivermectin interaction with potential SARS-CoV-2 targets and host nuclear importin α. J. Biomol. Struct. Dyn. 40 (6), 2851–2864.

Bahrami, A.A., Azargoonjahromi, A., Sadraei, S., Aarabi, A., Payandesh, Z., Rajabibaz, M., 2022. An overview of current drugs and prophylactic vaccines for coronavirus disease 2019 (COVID-19). Cell. Mol. Biol. Lett. 27 (38), 1–256.

Balakrishnan, V., Ganapathy, S., Veerasamy, V., Subramaniyan, S., Hussain, S.A.M., Duraisamy, R., 2022. Modifying effects of nerolidol on cell surface glycoconjugates and suppressed inflammation during DMBA-induced oral carcinogenesis: an in vivo and in silico. Biologia 78, 529–541.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bpurne, P.E., et al., 2000. The protein data bank. Acta Crystallogr. D Biol. Crystallogr. 58, 899–907.

Celniker, G., et al., 2013. ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr. J. Chem. 53, 199–206.

Chaudhary, J.K., Yadav, R., Chaudhary, P.K., Maurya, A., Roshan, R., Azam, F., Metha, J., et al., 2021. Host Cell and SARS-CoV-2-associated molecular structures and factors as potential therapeutic targets. Cells 10 (9), 2427.
Cheng, S., Chang, G., Chou, C., 2010. Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. Biophys. J. 98 (7), 1327–1336.
Connelly, D., 2020. Targeting COVID-19: the drugs being fast-tracked through clinical trials and how they work. Pharm. J. 304 (7937), 312–313.
Ferreira, R.J., Buttell, F., Cannon, C., 2020. COVID-19: immediate predictors of individual resilience. Sustainability 12 (6), 6495.
Finn, R.D., Coggil, P., Eberhardt, R.Y., et al., 2016. The Pfam protein families' database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285.
Fu, Y., Ye, T., Liu, Y., Wang, J., Ye, F., 2020. Based on the virtual screening of multiple pharmacophores, docking and molecular dynamics simulation approaches toward the discovery of novel HPPD inhibitors. Int. J. Mol. Sci. 21 (15), 5546.
Gaulton, A., Hersey, A., Nowotka, M., Bento, A.P., Chambers, J., Mendez, D., Mutowo, P., et al., 2017. The ChEMBL database in 2017. Nucleic Acids Res. 45, D945–D954.
Gopinath, P., Kathiravan, M.K., 2021. Docking studies and molecular dynamics simulation of triazole benzene sulfonamide derivatives with human carbonic anhydrase IX inhibition activity. RSC Adv. 11, 38709–138093.
Gupta, S.D., 2020. Coronavirus pandemic: a serious threat to Humanity. J. Health Manag. 22 (1), 1–2.
Hu, Q., Xiong, Y., Zhu, G.H., Zhang, Y.N., Zhang, Y.W., Huang, P., et al., 2022. The Sars-Cov-2 main protease (M(pro)): structure, function, and emerging therapies for Covid-19. Medcomm 3, E151.
Jin, Y.H., Cai, L., Cheng, Z.S., Cheng, H., Deng, T., Fan, Y.P., 2020a. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia. Mil. Med. Res. 7, 4.
Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., et al., 2020b. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. Nat. Struct. Mol. Biol. 27 (6), 529–532.
Jo, S., Kim, S., Shin, D.H., Kim, M.S., 2020. Inhibition of SARS-CoV 3CL protease by flavonoids. J. Enzyme Inhibition Medicinal Chem. 3, 145–151.
Kneller, D.W., Li, H., Phillips, G., Weiss, K.L., Zhang, Q., Arnould, M.A., Jonsson, C.B., Surendranathan, S., et al., 2022. Covalent narlaprevir- and boceprevir-derived hybrid inhibitors of SARS-CoV-2 main protease. Nat. Commun. 2268.
Kumar, A., Prasoon, P., Kumari, C., Pareek, V., Faiq, M.A., Narayan, A.K., Klandhasamy, M., Kant, K., 2020. SARS-CoV-2 specific virulence factors in COVID-19. J. Medical Virol. 93 (3), 1343–1350.
Mengist, H.M., Dinnessa, T., Jin, T., 2021. Structural basis of potential inhibitors targeting SARS-CoV-2 Main protease. Front. Chem. 9, 1–19.
Miller, E.B., Murphy, R.B., Sindhikara, D., Borrelli, K.W., Grisewood, M.J., Ranalli, F., Dixon, S.L., Jerome, S., Boyles, N.A., Day, T., Ghanakota, P., et al., 2021. Reliable and accurate solution to the Induced Fit Docking problem for protein-ligand binding. J. Chem. Theory Comput. 17 (4), 2630–2639.
Mody, V., Ho, J., Wills, S., Mawri, A., Lawson, L., Ebert, M., Fortin, G.M., Rayalam, S., Taval, S., 2021. Identification of 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents. Nat. Commun. Biol. 4, 93.
Moser, M., Patterson, C., 2003. Thrombin and vascular development. Arterioscler. Thromb. Vasc. Biol. 23, 922–930.
Needle, D., Lountos, G.T., Waugh, D.S., 2015. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. Acta Crystallogr. D Biol. Crystallogr. 71, 1102–1111.
Nga, P.T., Parquet, M.C., Lauber, C., Parida, M., Nabeshima, T., Yu, F., 2011. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. PLoS Pathog. 7.
Nguyen, D.D., Gao, K., Chen, J., Wang, R., Wei, G.-W., 2020. Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning. Chem. Sci. 11 (44), 12036–12046.
Petit, J., Meurice, N., Kaiser, C., Maggiora, G., 2012. Softening the Rule of Five where to draw the line? Bioorg. Med. Chem. 20, 5343–5351.
Ricci-López, J., Vidal-Limon, A., Zuñiga, M., Jimènez, V.A., Alderete, J.B., Brizuela, C. A., et al., 2019. Molecular modeling simulation studies reveal new potential inhibitors against HPV E6 protein. PLoS One 14 (3), e0213028.
Stoermer, M., 2020. Homology models of coronavirus 2019-nCoV $3CL^{pro}$ protease. ChemRxiv.
Ullrich, S., Nitsche, C., 2020. The SARS-CoV-2 main protease as drug target. Bioorg. Med. Chem. Lett. 30 (17), 127377.
Velusamy, S., Husain, S.A.M., Alarifi, Saud, Arokia, V.A.M., Muthusamy, J., 2023. Screening potential inhibitor from actinomycetes for Dihydrofolate reductase of Staphylococcus aureus – In vitro and in silico studies. J. King Saud Univ. Sci. 35 (6), 102762.
Yadav, R., Dhiman, U., Parihar, R.D., Upadhyay, S.K., 2022. Virtual screening of potential drug molecules against Covid-19 Targets: a drug repurposing approach. Lett. Appl. Nano Biosci. 11 (1), 2965–2980.
Zhang, L., et al., 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science 368, 409–412.
Zhang, X., Perez-Sanchez, H., Lightstone, F.C.A., 2015. Comprehensive docking and MM/GBSA rescoring study of ligand recognition upon binding antithrombin. Curr. Top. Med. Chem. 17 (14), 1631–1639.
Zhang, S., Zhong, N., Xue, F., Kang, X., Ren, X., Chen, J., 2010. Three-dimensional domain swapping as a mechanism to lock the active conformation in a super-active octamer of SARS-CoV main protease. Protein Cell 1, 371–383.
Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. 81 (4), 853–879.

Further Reading

Marpaung, D.N., Siregar, L.F., Pongkendek, J.J., 2020. Effect of using Chemsketch on teaching molecular shape of of hydrocarbon to increase student's achievement. J. Phys. Conf. Ser. 1569 (4), 042101.
Priya, K., Manandhar, S., Snakhe, R., Setty, M.M., Babu, U.V., Pai, S.R., 2022. Structure based virtual docking and molecular dynamics guided identification of potential phytoconstituents from traditionally used female antifertility plant. J. Pharm. Sci. 28 (2), 285–294.