Article

Not peer-reviewed version

Structure Based Identification of Some Potential Pan-Coronavirus Main Protease Inhibitors via Pharmacophore Modeling and Molecular Dynamics Simulation Within a One Health Framework

Mohd Yasir Khan , Farah Maarfi , Abid Ullah Shah , Nithyadevi Duraisamy , Mohammed Cherkaoui ,
Maged Gomaa Hemida *

Posted Date: 20 May 2026

doi: 10.20944/preprints202605.1352.v1

Keywords: coronaviruses; main protease (Mpro); pharmacophore modeling; virtual screening; molecular docking; molecular dynamics simulation; catalytic dyad

Article

Structure Based Identification of Some Potential Pan-Coronavirus Main Protease Inhibitors via Pharmacophore Modeling and Molecular Dynamics Simulation Within a One Health Framework

Mohd Yasir Khan [1], Farah Maarfi [1], Abid Ullah Shah [2], Nithyadevi Duraisamy [1], Mohammed Cherkaoui [1] and Maged Gomaa Hemida [2,*]

[1] Department of Digital Engineering and Artificial Intelligence, College of Science, Long Island University, Brooklyn, NY 11201, USA
[2] Department of Veterinary Biomedical Sciences, Lewyt College of Veterinary Medicine, Long Island University, 720 Northern Boulevard, Brookville, NY 11548, USA
* Correspondence: maged.hemida@liu.edu; Tel.: +1-516-299-3650

Abstract

Background. The main protease (Mpro/3CLpro) of coronaviruses (CoVs) is an essential enzyme involved in viral replication and represents an attractive target for antiviral drug discovery. Based on the similar binding pocket residues within the Mpro of different CoVs, the study aimed to identify potential inhibitors of Mpro from PDB ID 6M2N, using integrated computational approaches. **Methods.** Structure-based pharmacophore modeling, virtual screening, molecular docking, MM-GBSA binding energy calculation, and molecular dynamics (MD) simulation were performed using BIOVIA Discovery Studio. The validated pharmacophore model was utilized to screen the ZINC database, followed by docking and 100 ns MD simulation analyses of the top-ranked compounds. **Results.** The pharmacophore model 01 demonstrated favourable predictive performance (AUC = 0.781). Virtual screening identified 483 compounds, from which 21 compounds were selected for docking studies. Among them, ZINC95473654 (Lig-1), ZINC95473725 (Lig-2), and ZINC08792368 (Lig-3) exhibited strong binding affinity toward Mpro. Lig-1 demonstrated the best docking score and binding free energy along with stable interactions with key catalytic residues HIS41, CYS145, and GLU166. MD simulation analyses further confirmed that Lig-1 and Lig-2 maintained stable conformations and persistent intermolecular interactions throughout the 100 ns simulation period. **Conclusion.** The findings suggest that Lig-1, followed by Lig-2, may serve as promising CoVs Mpro inhibitors and warrant further experimental validation. Further experimental validation are required to consolidate the identified compounds as universal inhibitors of the CoVs-Mpro enzyme.

Keywords: coronaviruses; main protease (Mpro); pharmacophore modeling; virtual screening; molecular docking; molecular dynamics simulation; catalytic dyad

1. Introduction

Coronaviruses (CoVs) are a diverse group of enveloped, (+) sense RNA viruses that infect a wide range of mammalian and avian hosts, causing respiratory, enteric, hepatic, and neurological diseases [1–5].

They are taxonomically classified into four genera (Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus) based on the phylogenetic and genomic characteristics [1–5]. The emergence of highly pathogenic human CoVs, including SARS-CoV in 2002, MERS-CoV in 2012, and SARS-CoV-2 in 2019, has underscored the global health threat posed by zoonotic spillover events and highlighted the urgent need for broad-spectrum antiviral therapeutics [2,6–9]. One of the novel approaches in the vaccine and drug designs is to monitor the genomic makeup of the pathogens and to ensure the designed vaccines/drugs are targeting the currently circulating strains of viruses infecting humans, various species of animals and birds [6,8,10–20]. The CoVs gnome structure and organization is highly conserved among various groups including the 5′ two thirds encompassing the nonstructural proteins (NSPs). These NSPs are encoded as large polyprotein which consists of two overlapping open reading frames (ORFs) with a ribosomal frameshifting in between. These ORFs are cleaved into 16 non-structural proteins (NSPs). There are two viral encoded proteases which cleave the long polypeptide into 16 NSPs. The NSp3 encodes the papain like protease (PLp) while the main viral protease is encoded by the NSP5 called 3C Like protease or the main protease (Mpro). The 3′ one third of the genome encodes eh major structural proteins. The full-length genomes of CoVs are flanked by two untranslated regions (UTRs) at both ends. A central feature shared across all known coronaviruses is the presence of a highly conserved main protease (Mpro, also known as 3CLpro), a cysteine protease indispensable for viral replication. MPro is responsible for the proteolytic cleavage of the large viral polyproteins pp1a and pp1ab into several functional non-structural proteins that are essential for viral replication, transcription, and assembly [21]. Because MPro is somewhat conserved among human and animal coronaviruses and lacks closely related homologs in mammalian host cells, it has emerged as one of the most attractive antiviral targets for the development of broad-spectrum and pan-coronavirus therapeutics within a One Health framework. The conservation of the catalytic dyad composed of His41 and Cys145 across diverse coronaviruses further supports the feasibility of designing universal inhibitors effective against multiple zoonotic and animal-origin coronaviruses [22].

Computational drug discovery has emerged as a powerful strategy for accelerating antiviral development. Structure-based pharmacophore modeling, virtual screening, molecular docking, and molecular dynamics (MD) simulations enable rapid identification of compounds with favorable binding characteristics prior to experimental validation. Based on the similar binding pocket residues within the Mpro of different CoVs, the study aimed to identify some potential inhibitors using integrated computational approaches. These in-silico methods are particularly valuable for conserved viral targets like Mpro, where subtle differences in binding pocket architecture can be exploited to design inhibitors with broad antiviral potential [23–25] .

Among the important amino acids of Mpro binding pocket, His41 and Cys145 belong to the catalytic dyad of Mpro, which has been reported to play an important role in inhibiting SARS-CoV-2 replication. This catalytic dyad (HIS41–CYS145) is responsible for cleaving the viral polyproteins pp1a and pp1ab at more than 11 conserved sites to release non-structural proteins (NSPs) essential for assembly of the replication transcription. Because humans lack closely related homologs of this protease, Mpro is considered one of the most attractive antiviral drug targets for CoVs [26,27].

Structural studies have demonstrated that Mpro is a homodimeric enzyme whose activity depends on proper dimerization and the integrity of key residues such as HIS41, CYS145, and GLU166, the latter stabilizes the S1 substrate binding pocket and supports dimer formation. Consistent with this, our study shows that the amino acid (Glu166) plays an important role in the dimerization of the Mpro of CoVs. The multiple sequence alignment of the Mpro from representative α- and β-coronaviruses in our study revealed that HIS41, CYS145, and GLU166 are fully conserved, reinforcing the feasibility of developing pan-coronavirus inhibitors that target this protease [28].

The therapeutic relevance of Mpro is further supported by the success of protease inhibitors in other RNA viruses, including HIV and HCV. For SARS-CoV-2 specifically, several small-molecule inhibitors such as nirmatrelvir, the active component of Paxlovid have demonstrated clinical efficacy by directly targeting Mpro. However, the rapid evolution of CoVs in both humans and animals, combined with the ongoing risk of zoonotic transmission, necessitates the discovery of broad spectrum, cross-species Mpro inhibitors that align with a One Health approach [29].

Among the most important veterinary coronavirus protease inhibitors is GC376, which was initially developed against feline infectious peritonitis virus (FIPV), a highly fatal feline coronavirus affecting domestic and wild cats. GC376 demonstrated potent inhibition of FIPV replication both in vitro and in naturally infected cats through direct targeting of the viral MPro enzyme. Treatment with

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2026 doi:10.20944/preprints202605.1352.v1

GC376 resulted in significant clinical improvement and prolonged survival in cats with naturally occurring feline infectious peritonitis (FIP). Importantly, subsequent structural and biochemical studies revealed that GC376 also effectively inhibits the MPro enzymes of SARS-CoV, MERS-CoV, and SARS-CoV-2, highlighting its broad-spectrum pan-coronavirus potential. Because feline coronaviruses share substantial structural similarities with human coronaviruses, GC376 became one of the earliest examples illustrating how antiviral discoveries in veterinary medicine can directly contribute to human pandemic preparedness [30,31]. In this study, we employed a comprehensive computational pipeline including pharmacophore modeling, ZINC database virtual screening, molecular docking, MM-GBSA binding energy estimation, and 100-ns MD simulations to identify novel small molecule inhibitors capable of targeting conserved catalytic residues of coronavirus Mpro. Our findings highlight several promising compounds, particularly ZINC95473654 (Lig-1) and ZINC95473725 (Lig-2), which demonstrated strong binding affinity, stable interactions with HIS41, CYS145, and GLU166, and favorable dynamic stability throughout simulation. These results support their potential as pan-coronavirus Mpro inhibitors, warranting further biochemical and antiviral evaluation. The experimental studies of our findings are highly encouraged to set up the pipeline for validation of these compounds as strong anti-CoV-Mpro inhibitors and to move forward toward the clinical trials of these compounds as potential drugs for most CoVs from a One health perspective [31].

2. Results

2.1. Multiple Sequence Alignment Using Mpro from Repreentative Coronaviruses

Among the important amino acids of Mpro binding pocket, His41 and Cys145 belong to the catalytic dyad of Mpro, which has been reported to play an important role in inhibiting SARS-CoV-2 replication [29,32,33]. The Glu166 plays an important role in the dimerization of Mpro [34,35].

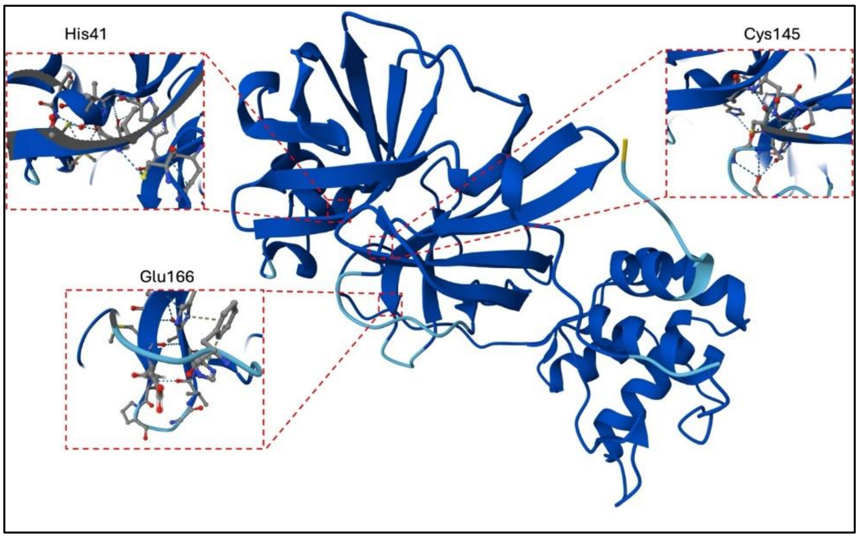


Figure 1. 3D structure of SARS-CoV-2 Mpro (PDB ID:6M2N). Three amino acids, His41, Cys145, and Glu166 are highlighted in red dotted boxes in the 3D protein structure of 3CL-Pro (Mpro).

To highlight these residues of binding pocket of Mpro, we performed multiple sequence alignment using mpro from 10 different coronaviruses. Among these 10 coronaviruses, four coronaviruses including HCoV-NL63, HCoV-229E, FCoV, and PEDV belongs to alphacoronvirus group. While six coronaviruses including SARS-CoV-2, SARS-CoV, HCoV-HKU1, HCoV-OC43, BCoV, and MERS belongs to betacoronaviruses. Results showed that His41, Cys145, and Glu166 amino acids are conserved among all ten coronaviruses. However, Met49 was only conserved among the coronaviruses of betacoronvirus family except MERS (Figure 2).

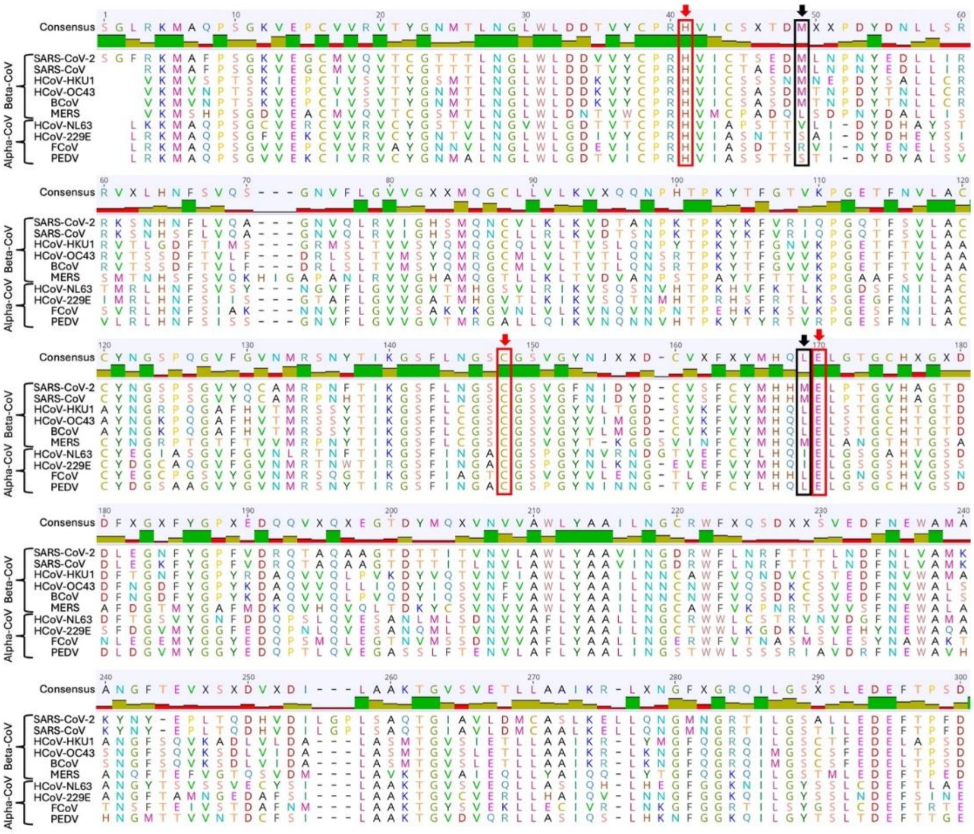


Figure 2. Multiple sequence alignment of the MPro protein of different alphacoronaviruses and betacoronaviruses. Mpro of six betacoronaviruses and four alphacoronaviruses were used to perform multiple sequence alignment. Among the five important amino acids interacting with compounds, three amino acids, which are conserved among all ten coronaviruses, are indicated with a red box and red arrowhead. Two amino acids that are not conserved are shown with a black box and black arrowhead. The multiple sequence alignment was performed using Geneious Prime V.9.0.

2.2. Validation of the Generated Pharmacophore Model

The fundamental principles behind pharmacophore features in molecular three-dimensional visualization elucidate the methodology for evaluating the capacity to bind to a designated target protein. A five-point pharmacophore hypothesis AAHaromHarmR with A as acceptor, H as hydrophobic aromatic, and R as aromatic rings, generated based on co-crystal ligands present in the protein binding pocket. Total ten pharmacophore models were constructed, comprising combinations of hydrogen bond acceptors, hydrophobic aromatic regions, and aromatic interaction features. The generated models were evaluated using a validation dataset consisting of known active and inactive compounds (Figure S1). Among the models, Pharmacophore model_01 with number of features five (AAHaromHaromR) demonstrated the best predictive performance with an AUC value of 0.781, indicating good discrimination capability (Table 1).

Table 1. Results of the Receiver operating characteristic (ROC) curves for selected pharmacophore models demonstrating their ability to discriminate active and inactive compounds.

Pharmacophore model Validation with Known Actives/Inactive									
Pharmacophore model (PPM)	Total Actives	Total Inactives	True Positives	True Negatives	False Positives	False Negatives	Sensitivity	Specificity	AUC ROC

PPM-1	38	39	31	24	15	7	0.81579	0.61538	0.786
PPM-2	38	39	30	21	18	8	0.78947	0.53846	0.738
PPM-3	38	39	37	5	34	1	0.97368	0.12821	0.513
PPM-4	38	39	37	6	33	1	0.97368	0.15385	0.562
PPM-5	38	39	37	7	32	1	0.97368	0.17949	0.521
PPM-6	38	39	36	8	31	2	0.94737	0.20513	0.678
PPM-7	38	39	37	7	32	1	0.97368	0.17949	0.529
PPM-8	38	39	32	19	20	6	0.84211	0.48718	0.686
PPM-9	38	39	34	9	30	4	0.89474	0.23077	0.508
PPM-10	38	39	30	19	20	8	0.78947	0.48718	0.700

This model Pharmacophore_01 exhibited a balanced sensitivity (0.815) and specificity (0.615), suggesting its robustness in identifying active compounds while minimizing false positives. The dominant features observed across the top-performing models included hydrogen bond acceptors and hydrophobic interactions, along with aromatic ring contributions, highlighting their critical role in stabilizing ligand binding within the active site (Figure 3). These findings suggest that both polar interactions and hydrophobic contacts are essential determinants of ligand recognition and binding affinity.

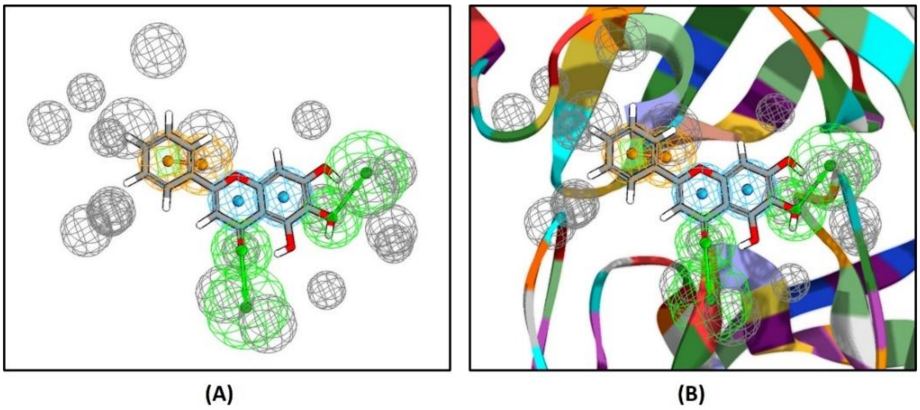


Figure 3. Structure-based pharmacophore modeling and validation. (A) 3D representation of the optimized pharmacophore model generated from ligand–receptor interactions, highlighting key features including hydrogen bond acceptors (blue), hydrophobic (green), aromatic regions (yellow), and exclusion volumes (grey). (B) Pharmacophore model_01 in the interaction site of ligand binding cavity from protein Mpro.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2026 doi:10.20944/preprints202605.1352.v1

Overall, the selected pharmacophore model 01 provides a reliable structural framework for virtual screening and rational drug design, facilitating the identification of novel compounds with improved binding characteristics.

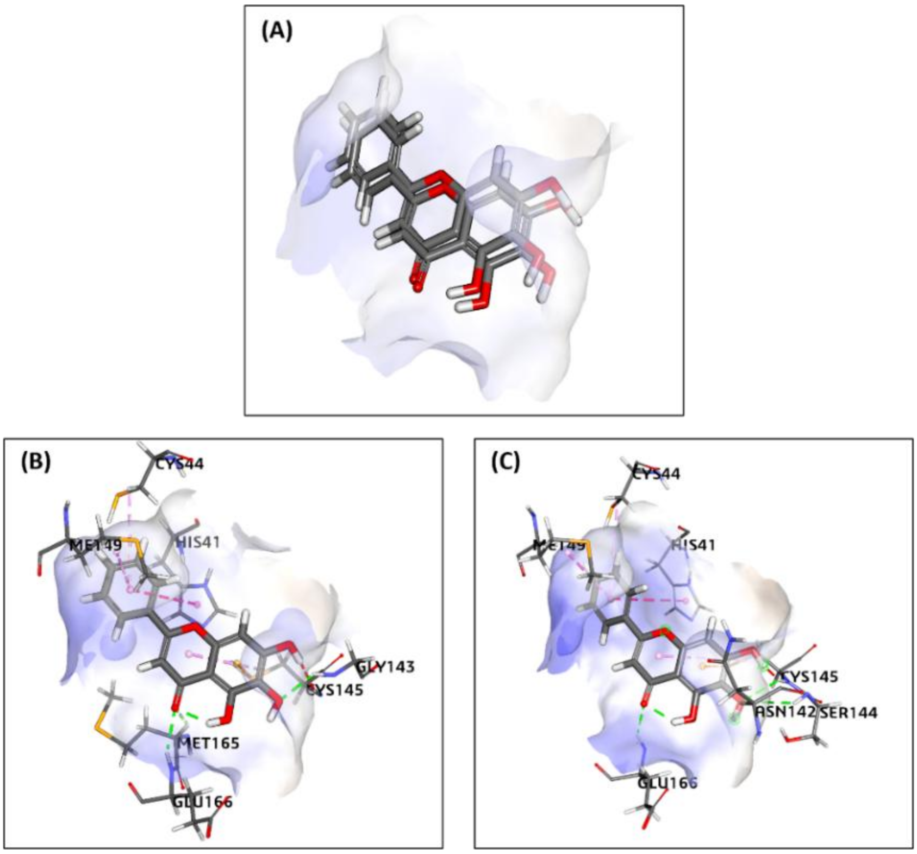
2.3. Virtual Screening Based on Pharmacophore Model

ZINCPharmer used best pharmacophore model (Pharmacophore 01) to filter out only compounds with a molecular weight between 350 and 500 Da. This search yielded 483 hit compounds, which were subsequently prepared, duplicates removed, and energy minimized to assess their potential as lead candidates for Mpro protein inhibition, left with 49 compounds [36].

Based on the pharmacophore model mapped ligand fit value, compounds with fit value between 3.5 (maximum) and 1 (minimum) were selected. Out of 49, only 21 curated ligands based on fit value, were further utilized for comprehensive docking with Mpro protein (Table S1).

2.4. Redocking Validation for the Molecular Docking Process

To validate further docking results of selected ligands from virtual screening method, the Mpro receptor protein was re-docked with native Baicalein (standard ligand). The redocking's binding site area was x: - -33.218228, y: -65.348922, z: 41.208347, with 6.85 sphere radius. The parameter of the validation method is RMSD. The RMSD analysis showed the degree of deviation from experimental ligand docking results to the native ligand (standard) attached with Mpro protein at the same binding site. The higher the RMSD value, the greater the deviation, which indicates the higher prediction error of ligand-protein interactions. Conversely, the low RMSD value is attributed to better conformation because the redocking ligand position is closer to the ligand position resulting from crystallography. The redocking results indicated a 0.683 Å RMSD value from the native ligand with the Mpro receptor (Figure 4). Therefore, based on the low RMSD value between crystallized native and re-docked ligand, the method used for redocking in this study is valid and can be used against tested ligands with the same binding site area.



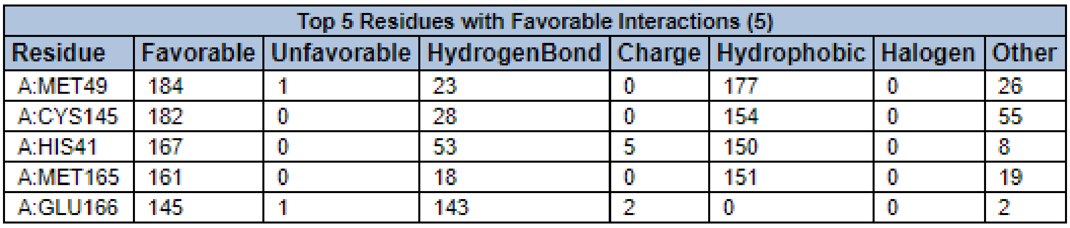
doi:10.20944/preprints202605.1352.v1

Figure 4. Redocking validation of the co-crystallized Baicalein inhibitor with Mpro protein (PDB ID: 6M2N). (A) Superimposition of the redocked and native co-crystallized Baicalein within the active-site pocket. (B) Interaction profile of the native Baicalein–Mpro complex showing interactions with HIS41, MET49, GLY143, CYS145, MET165, and GLU166 residues. (C) Interaction profile of the redocked Baicalein–Mpro complex demonstrating conserved interactions With HIS41, MET49, SER144, CYS145, ASN142, and GLU166 active-site residues.

2.5. Mpro-Ligands Interaction and Binding Energy (MM-GBSA)

Mpro in complex with novel inhibitor **Baicalein** compound in X-ray crystal structure from RCSB PDB database (PDB ID: 6M2N; 2.20 Å resolution) was chosen as the receptor for different ligand docking. The Zinc ID for all ligands is given in Table 1. The residues included HIS41 and CYS145, forms essential catalytic dyad in the structure of Mpro. The interaction pattern of all selected ligands from docking results clearly exhibited that most of the ligands interacted with catalytic dyad residues HIS41 and CYS145 and other important residues of binding pocket of Mpro (Table 2 and Figure 5).

Table 2. Top residues of the binding pocket include catalytic dyad residues CYS145 and HIS41 exhibiting favorable interactions with the 21 selected ligands.



Top 5 Residues with Favorable Interactions (5)

Residue	Favorable	Unfavorable	HydrogenBond	Charge	Hydrophobic	Halogen	Other
A:MET49	184	1	23	0	177	0	26
A:CYS145	182	0	28	0	154	0	55
A:HIS41	167	0	53	5	150	0	8
A:MET165	161	0	18	0	151	0	19
A:GLU166	145	1	143	2	0	0	2

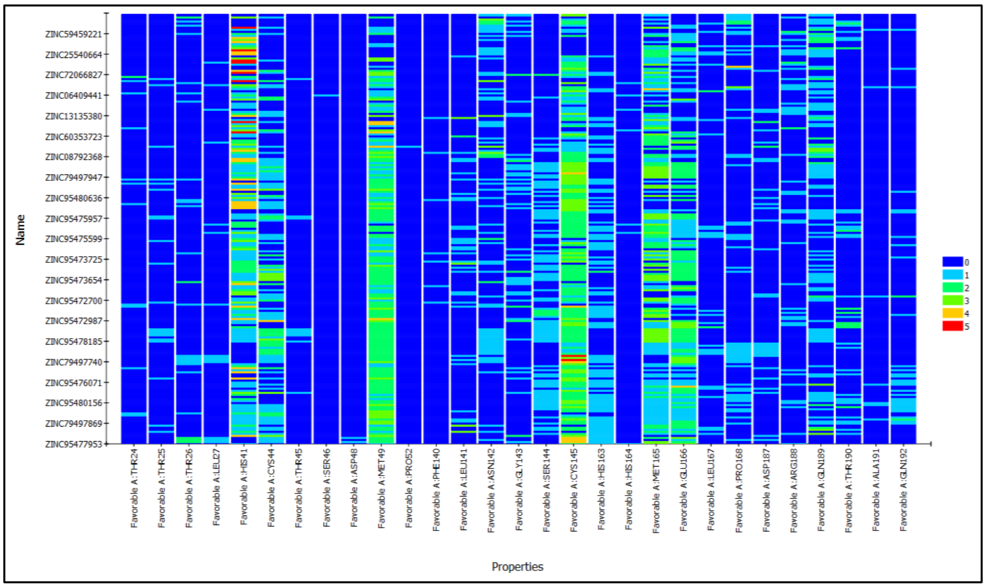
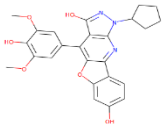
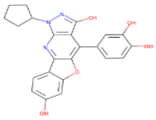
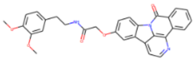
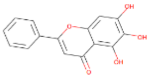


Figure 5. Heatmap exhibiting interaction patterns of selected 21 ZINC compounds with key binding-site residues. Rows represent ligands and columns represent residue-specific interactions. The color scale (blue to red) indicates interaction intensity, from low/no interaction (blue) to strong interaction (red). Key residues such as HIS41, CYS145, MET49, MET165, and GLU166 exhibit prominent interaction hotspots, highlighting their importance in ligand binding.

In the attempt to filter suitable compounds with the best binding to protein Mpro, our molecular docking studies revealed a strong interaction of nearly all the 21 selected compounds with -CDOCKER interaction energy score (-CDOCK Score) as binding affinity score from 53 to 28 kcal/mol and binding energy (MM-GBSA) from -102 to -23 kcal/mol against the binding pocket residues of protein Mpro (Table S1).

Among all compounds, **ZINC95473654, ZINC95473725** and **ZINC08792368** topped the affinity score (-CDOCK Score: 54.60, 50.83 and 52.73, respectively) and binding energy score (MM-GBSA: -121.86, -102.47 and -97.73 kcal/mol), shared high chemical structural similarity exhibiting a comparable strong pattern of interaction with main protease, which indicates a valid docking approach (Table 3). In contrast the, the best docking pose of reference ligand uses as standard (**Baicalein**) exhibited a less binding affinity (CDOCK Score: 34.23) to Mpro with low binding energy (MM-GBSA: -90.11 Kcal/mol) (Figure 6 and Table 3).

Table 3. Comparative docking scores, pharmacophore fit values, RMSD, and MM-GBSA binding free energy of top three ZINC compounds out of 21 selected compounds.

S. No.	**Compound ID**	**Structure**	**Fit Value**	**RMSD (Å)**	**-CDOCK Score**	**MM-GBSA (Kcal/mol)**
1	ZINC95473654		2.69159	1.25	54.60	-121.86
1	ZINC95473725		2.63669	1.52	50.83	-102.47
2	ZINC08792368		1.96643	1.43	52.73	-97.73
4	**Standard** (Baicalein**)**		-	0.68	34.23	-90.02

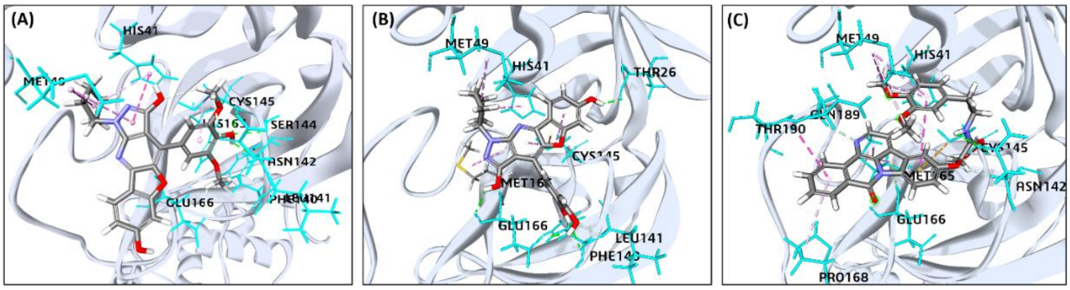


Figure 6. Protein–ligand interaction profiles of (A) Lig-1, (B) Lig-2, and (C) Lig-3 complexes within the SARS-CoV-2 Mpro active-site cavity showing interactions with key catalytic and binding pocket residues including HIS41, CYS145, GLU166, MET49, MET165, ASN142, SER144, THR190, and PRO168.

Further, the stability of the **ZINC95473654–Mpro** complex is primarily governed by interactions with key active-site residues of Mpro binding pocket with catalytic dyad residues, including HIS41, MET49, PHE140, LEU141, CYS145, SER144, ASN142, HIS163, and GLU166. Importantly, SER-144 and CYS-145 were found to form strong conventional hydrogen bond interactions, while LEU-141 contributed to hydrophobic interactions and HIS-41 showed π–π stacking interactions.

The stability of the ZINC95473654–Mpro complex is mediated by interactions with key active-site residues, including HIS41, MET49, PHE140, LEU141, CYS145, SER144, ASN142, HIS163, and GLU166. Conventional hydrogen bonds formed with SER144 and CYS145 play a critical role in anchoring the ligand within the binding pocket. In addition, π–π stacking interactions involving HIS41 and hydrophobic contacts with MET49 and LEU141 further contribute to ligand stabilization. Surrounding residues such as PHE140, ASN142, HIS163, and GLU166 provide additional support through van der Waals interactions, collectively ensuring a stable ligand–protein complex.

However, in case of **ZINC95473725-Mpro,** the amino acid residues involved in the complex's stability were THR26, HIS41, MET49, PHE140, LEU141, CYS145, MET165, and GLU166. In this complex**,** THR26, PHE140, GLU166 and CYS145 were found to form strong conventional hydrogen bond interactions, while HIS-41, MET-49 and MET-165 contributed to hydrophobic and π–alkyl interactions, with LEU-141 participating in additional van der Waals contacts.

The stability of the ligand–Mpro complex is governed by key interactions with active-site residues including THR26, PHE140, GLU166, CYS145, HIS41, MET49, MET165, and LEU141. Strong conventional hydrogen bonds formed with THR26, PHE140, GLU166, and CYS145 play a crucial role in anchoring the ligand within the binding pocket. Additionally, hydrophobic and π–alkyl interactions involving HIS41, MET49, and MET165 contribute to ligand stabilization and proper orientation within the active site. Residues such as LEU141 further support the complex through van der Waals interactions. The presence of multiple hydrogen bonds along with complementary hydrophobic contacts suggests a stable and well-accommodated ligand within the Mpro binding cavity.

The residues involved in the **ZINC08792368-Mpro** complex's stability were HIS41, MET49, ASN142, CYS145, MET165, GLU166 and PRO168. The residue CYS145 and ASN142 were found to form strong hydrogen bond interactions, while PRO-168 and MET-165 contributed to hydrophobic and π–π stacking interactions.

The interaction analysis revealed that the ligand forms conventional hydrogen bonds with residues CYS145 and ASN142, contributing to binding stability. Additional stabilization is provided by amide–π stacking interactions with PRO168 and hydrophobic contacts involving MET49, MET165, and HIS41. However, the absence of strong hydrogen bonding interactions with key catalytic residues such as HIS41 and GLU166 suggests comparatively weaker binding affinity relative to more optimal ligands. Weak interactions, including carbon hydrogen bonds and van der Waals contacts, further support ligand positioning within the binding pocket. The interaction pattern of the top three ligands **ZINC95473654**, **ZINC95473725**, **ZINC08792368,** and standard **Baicalein** is also given in 2D (Figure S2).

2.6. Stability of the Protein Mpro–Ligand Complexes during Molecular Dynamics Simulation

RMSD analysis was carried out over a 100 ns molecular dynamics simulation to evaluate the structural stability of the Mpro protein in complex with compound **ZINC95473654** as Ligand-1 (**Lig-1**), **ZINC95473725** as Ligand-2 (**Lig-2**), and **ZINC08792368** as Ligand-3 **(Lig-3**), in comparison with the standard (**Baicalein-Mpro)** complex. As shown in Figure 5A, all Ligands-Mpro complex systems exhibited an initial equilibration phase during the early stage of the simulation, followed by relatively stable confirmations throughout the remaining simulation period. The Lig-1 complex demonstrated the highest structural stability, maintaining an average RMSD between approximately 1.2 and 1.6 Å with only minor fluctuations across the simulation trajectory. Similarly, the Lig-2 complex exhibited a stable RMSD profile ranging from nearly 1.2 to 1.7 Å, indicating stable binding and minimal conformational deviation of the protein–ligand complex. The standard complex also maintained a comparatively stable trajectory with RMSD values fluctuating around 1.3–1.9 Å during the 100 ns simulation. Among all three selected ligands, the Lig-3 complex displayed relatively higher fluctuations, particularly after 30 ns, where RMSD values increased to approximately 2.0–2.5 Å with occasional transient peaks. Despite these fluctuations, the complex remained within an acceptable RMSD range, suggesting conformational adaptability while retaining overall structural integrity.

Overall, the RMSD profiles indicate that the Lig-1 and Lig-2 complexes exhibited greater structural stability compared to Lig-3 and showed stability comparable to or better than the standard complex throughout the 100 ns simulation period.

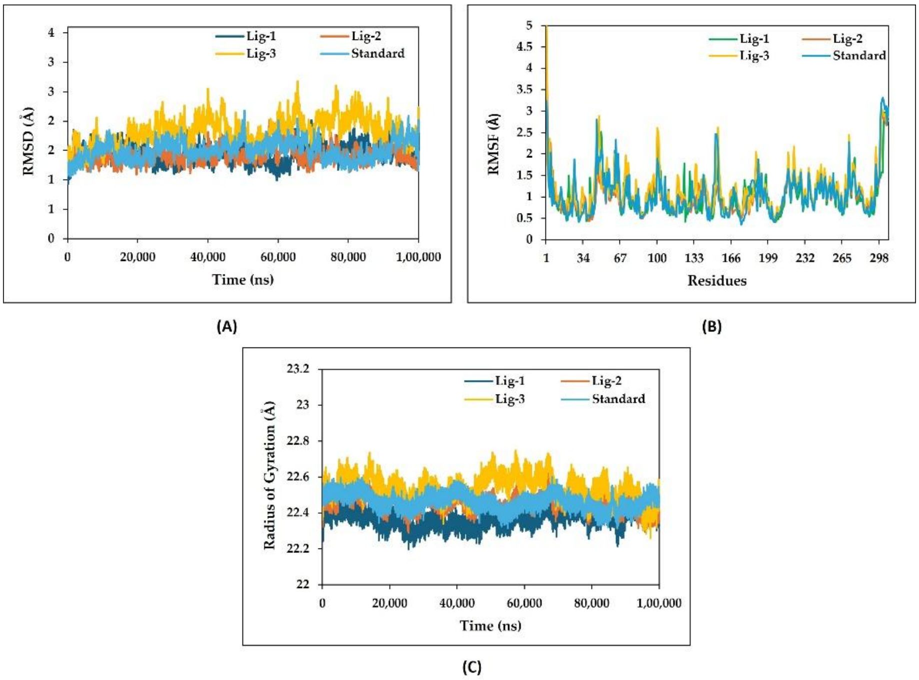


Figure 5. Molecular dynamics simulation analysis of the Standard, Lig-1, Lig-2, and Lig-3 complexes over 100 ns. **(A)** RMSD analysis showing structural stability of the complexes during the simulation period. **(B)** RMSF analysis representing residue-wise flexibility of the protein complexes. **(C)** Radius of gyration (RG) analysis indicating the compactness and conformational stability of the complexes throughout the simulation trajectory.

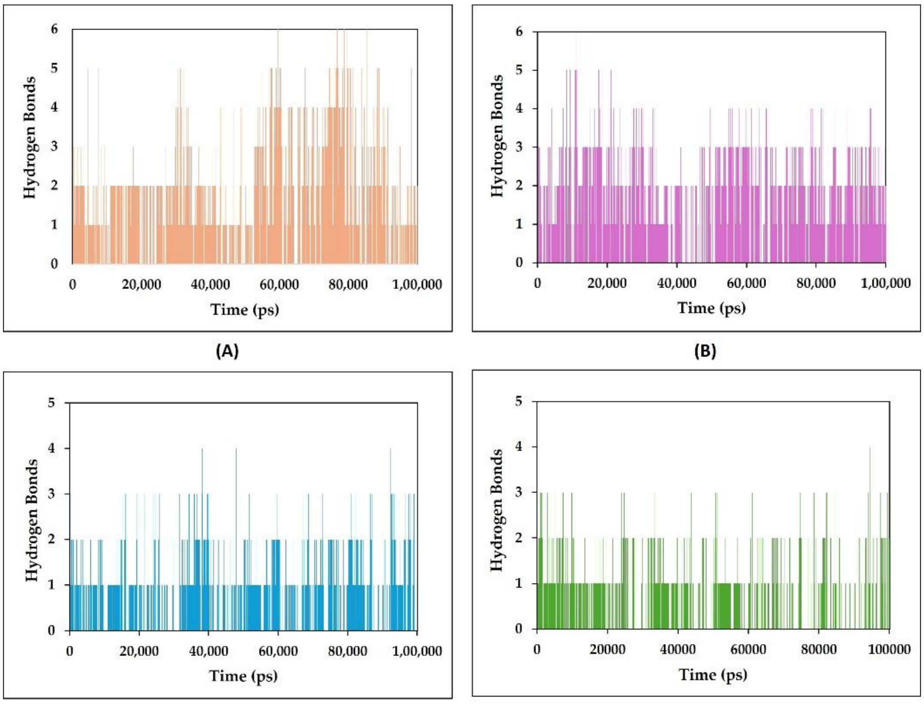
The RMSF analysis was performed by molecular dynamics simulation to verify the residue-wise flexibility of the protein Mpro in complex with Lig-1, Lig-2, Lig-3, and the Standard compound (Figure 5B). Most residues in all complexes fluctuated within approximately 0.5–2.5 Å, indicating overall structural stability throughout the simulation period. Among all systems, the Lig-1-Mpro exhibited lower residue fluctuations, suggesting enhanced conformational stability and stable ligand accommodation within the active-site cavity. Lig-2 also showed relatively stable fluctuation patterns comparable to the Standard-Mpro complex, although slightly higher fluctuations were observed compared to Lig-1. Importantly, the key active-site residues CYS145 and GLU166 exhibited lower fluctuations in both Lig-1 and Lig-2 complexes, indicating stable intermolecular interactions during the simulation. In contrast, the Lig-3 complex displayed comparatively higher fluctuations, particularly near terminal and loop regions, reflecting greater conformational flexibility. A prominent fluctuation peak reaching approximately 4–5 Å was observed near the terminal residue regions in all systems, corresponding to the naturally flexible behavior of these regions. Further, the RMSF analysis suggested that Lig-1 maintained the most stable residue fluctuation profile, while Lig-2 also demonstrated prominent structural stability throughout the 100 ns simulation period.

Radius of gyration (RG) analysis was performed over a 100 ns molecular dynamics simulation to evaluate the overall compactness and conformational stability of the protein in complex with Lig-1, Lig-2, Lig-3, and the standard compound. All the complexes maintained relatively stable RG trajectories throughout the simulation period, indicating preservation of the protein's folded architecture under solvated conditions (Figure 5C). The Lig-1 complex demonstrated a highly stable compactness pattern, maintaining an average RG around value 22.2–22.4 Å with minor fluctuations throughout the simulation. Similarly, the Lig-2 complex exhibited stable conformational stability with RG values fluctuating approximately between 22.3 and 22.5 Å, suggesting maintenance of structural compactness and stable ligand accommodation within the binding pocket. The standard complex also demonstrated consistent compactness throughout the simulation, with RG values remaining nearly between 22.4 and 22.6 Å, indicating structural stability comparable to the ligand-bound systems. In contrast, the Lig-3 complex displayed relatively higher Rg fluctuations, particularly between 40 and 80 ns, where values transiently increased up to approximately 22.7 Å. These fluctuations suggest moderate conformational flexibility and slight expansion of the protein structure during simulation; however, the system remained within an acceptable and stable compactness range.

Collectively, the RG profiles indicated that the Lig-1 and Lig-2 complexes preserved a compact and conformationally stable architecture throughout the 100 ns simulation, showing behavior comparable to the standard complex. Notably, the Lig-1 complex exhibited the least fluctuation in RG values among all systems, reflecting enhanced structural rigidity and a well-maintained protein–ligand conformation under dynamic simulation conditions.

2.7. Stability of Complexes Assessed with Hydrogen Bond Monitoring

The evaluation of hydrogen bond monitor suggests bonding between residues of protein and selected ligands. Hydrogen bond monitoring analysis was carried out during the 100 ns molecular dynamics simulation to evaluate the persistence and stability of intermolecular interactions formed between the protein and the Standard, Lig-1, Lig-2, and Lig-3 complexes (Figure 6A–D). The number of hydrogen bonds maintained during the simulation is an important indicator of binding stability, where a higher and more persistent hydrogen bond count generally corresponds to stronger and more stable protein–ligand interactions.



(A) (B)

(C) (D)

Figure 6. Hydrogen bond monitoring analysis of the Standard **(A),** Lig-1 **(B),** Lig-2 **(C)**, and Lig-3 **(D)** complexes with Mpro protein showing the number of intermolecular hydrogen bonds maintained during the 100 ns molecular dynamics simulation.

The standard complex (Figure 6A) maintained approximately two to four hydrogen bonds throughout most of the simulation period, with several intervals showing transient increases up to six hydrogen bonds. The continuous presence of hydrogen bonds with fewer dissociation events indicated stable intermolecular interactions within the binding pocket. Similarly, the Lig-1 complex (Figure 6B) demonstrated comparatively stable hydrogen bond occupancy during the entire

doi:10.20944/preprints202605.1352.v1

simulation period. The complex predominantly maintained three to five hydrogen bonds, while transient increases reaching six to seven hydrogen bonds were also observed during several intervals of the trajectory. The sustained presence of multiple hydrogen bonds suggested strong and persistent intermolecular interactions, thereby contributing to enhanced stability of the Lig-1 complex. However, the Lig-2 complex (Figure 6C) showed lower hydrogen bond occupancy as compared to Standard, generally maintaining one to three hydrogen bonds throughout the simulation. Although occasional increases in hydrogen bond number were observed, the complex showed more frequent dissociation events and reduced interaction persistence relative to the Standard and Lig-1 complexes, indicating comparatively weaker binding stability. Likewise, the Lig-3 complex (Figure 6D) maintained predominantly one to two hydrogen bonds during most of the simulation trajectory, with only occasional transient increases. Frequent fluctuations and intermittent reductions in hydrogen bond count suggested comparatively less stable intermolecular interactions and weaker binding persistence within the active-site region.

Overall, the hydrogen bond monitoring analysis revealed that Lig-1 and Lig-2 maintained comparatively higher and more persistent hydrogen bond occupancy throughout the 100 ns simulation period, suggesting stable intermolecular interactions and improved protein–ligand complex stability under dynamic simulated conditions.

2.7. Hydrogen Bonds Distance Variation analysis for complex stability

Variation in hydrogen bond (HB) distance analysis for amino acid residues contributing to substantial hydrogen bonds, was performed throughout the 100 ns molecular dynamics simulation to investigate the strength and stability of intermolecular interactions formed between the protein with standard and other ligand-bound complexes. The hydrogen bond distance variation for the native standard (Baicalein), Lig-1, Lig-2, and Lig-3-Mpro complex, respectively (Figure 7A–D).

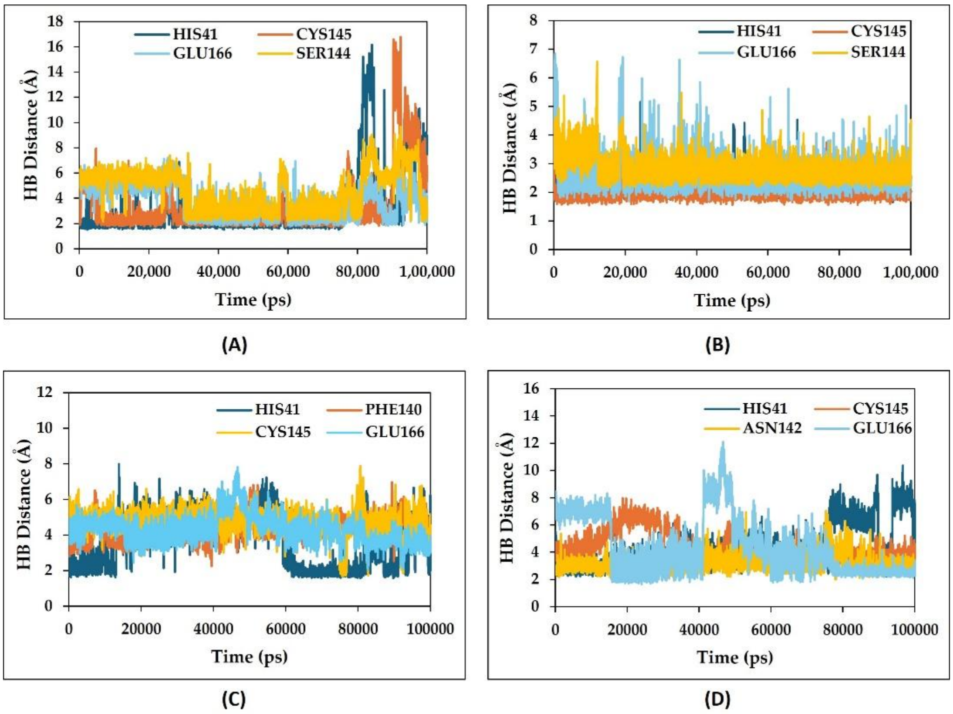


Figure 7. Hydrogen bond (HB) distance analysis of the Standard **(A),** Lig-1 **(B),** Lig-2 **(C),** and Lig-3 **(D)** complexes with Mpro protein showing interactions with key active-site residues during the 100 ns molecular dynamics simulation.

The standard compound formed hydrogen bond interactions mainly with HIS41, GLU166, CYS145, and SER144 residues (Figure 7A). The HIS41 and GLU166 residue average HB distances were around ~1.8–2.5 Å during most of the simulation, indicating relatively strong and stable

doi:10.20944/preprints202605.1352.v1

hydrogen bonding interactions. In contrast, SER144 and CYS145 showed comparatively higher bond distances ranging approximately from 3–6 Å, with transient fluctuations in bond distance, which reached up to ~15–16 Å in the later stages of the simulation, suggesting comparatively weaker and unstable intermolecular interactions. However, the Lig-1-Mpro complex demonstrated comparatively stable hydrogen bonding interactions throughout the simulation time (Figure 7B). To Lig-1, the Mpro interacting residues HIS41, GLU166, CYS145, and SER144 maintained average HB distances of approximately 2.5–3.5 Å. Among these, CYS145 showed the shortest and most consistent bond distance around ~1.7–2.0 Å, indicating a strong and persistent interaction with Lig-1. Although GLU166 and SER144 exhibited occasional fluctuations reaching ~4–6 Å, but most of the interactions remained within favorable hydrogen bonding distance ranges (~1.7–3 Å), supporting stable ligand binding. Similarly, the Lig-2-Mpro complex (Figure C) demonstrated moderate and relatively stable hydrogen bond interactions with key active-site residues HIS41, PHE140, CYS145, and GLU166 throughout the 100 ns simulation period. The average hydrogen bond distances of HIS41 and GLU166 remained approximately within the range of 2–5 Å, indicating stable intermolecular interactions during most of the trajectory. Similarly, CYS145 and PHE140 exhibited bond distances predominantly around 4–6 Å with moderate fluctuations and occasional transient increases. Compared to Lig-1, the Lig-2 complex displayed slightly higher variations in hydrogen bond distances, suggesting comparatively reduced interaction stability within the active-site cavity. However, the majority of interactions remained within an acceptable hydrogen bonding range, supporting stable ligand accommodation throughout the simulation period. In contrast, the Lig-3 complex (Figure 7D) exhibited comparatively less stable hydrogen bonding interactions. The interacting residues HIS41, CYS145, ASN142, and GLU166 showed substantial fluctuations in HB distances throughout the simulation. Average bond distances varied approximately between 3 and 8 Å, while certain interactions transiently increased up to ~10–12 Å, particularly for GLU166 and HIS41 during the later stages of the total simulation time. These pronounced fluctuations suggest weaker interaction persistence and comparatively reduced binding stability of the Lig-3 complex under dynamic simulation conditions. Overall, the HB distance analysis indicated that Lig-1 maintained the most stable and consistent hydrogen bonding interactions among all investigated complexes, while Lig-2 also demonstrated prominent and relatively stable intermolecular interactions, although with slightly higher fluctuations compared to Lig-1, throughout the 100 ns simulation period.

3. Discussion

Coronaviruses are broadly classified into four genera, namely α-, β-, γ-, and δ-coronaviruses, based on genomic organization and serological characteristics [37].

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a member of the family *Coronaviridae* and subfamily *Orthocoronavirinae*, belongs to the β-coronavirus genus. The Coronaviridae family, which also includes viruses like SARS-CoV, MERS-CoV, and SARS-CoV-2. Among humans and animals, coronaviruses exhibit the presence of similar structurally related functional proteins, such as the main protease (Mpro) (Figure 2), also known as 3CL protease (3CLpro), which plays a crucial role in viral replication [38].

The viral proteins encoded by SARS-CoV-2, the Mpro has emerged as one of the most promising therapeutic targets due to its important role in viral replication and pathogenesis. During viral infection, SARS-CoV-2 synthesizes two large polyproteins, pp1a and pp1ab, which undergo proteolytic cleavage to release 16 non-structural proteins (nsps) essential for viral replication and transcription. Mpro performs cleavage at 11 conserved sites within these polyproteins, thereby playing a central role in the maturation of the viral replication machinery [39].

Structurally, Mpro is a conserved dimeric cysteine protease characterized by a catalytic dyad formed by HIS41 and CYS145, residues that are critically involved in substrate recognition and catalytic activity [40,41].

Due to the absence of closely related human homologs and its indispensable role in the viral life cycle, Mpro represents an attractive target for antiviral drug development. The therapeutic success of protease inhibitors against other RNA viruses, including HIV and HCV, further supports the potential of targeting Mpro for SARS-CoV-2 treatment [42].

Beyond viral polyprotein processing, Mpro also promotes viral survival by suppressing host antiviral immune signaling and altering host metabolic pathways, including lipid metabolism, to facilitate efficient SARS-CoV-2 replication [39,43].

Considering the multifunctional role of Mpro in both viral replication and immune evasion, inhibition of this protease could simultaneously suppress viral maturation and interfere with mechanisms that support viral persistence within the host. In the present study, molecular docking and molecular dynamics simulations were employed to evaluate the stability and inhibitory potential of the selected ligands against the Mpro active site. The investigated complexes demonstrated stable binding interactions with key catalytic residues, particularly HIS41, CYS145, and GLU166, which are known to play essential roles in ligand recognition and enzymatic activity.

The generated pharmacophore models highlighted the importance of hydrogen bond acceptors, hydrophobic aromatic regions, and aromatic ring interactions for stable ligand binding within the SARS-CoV-2 Mpro active-site cavity. Among the generated hypotheses, the five-feature pharmacophore model AAHaromHaromR demonstrated the best predictive performance with favorable AUC, sensitivity, and specificity values, indicating reliable discrimination between active and inactive compounds. The validated model was subsequently utilized for virtual screening using the ZINCPharmer database, resulting in 483 initial hit compounds [44].

Following duplicate removal, energy minimization, and pharmacophore fit-value filtering, 21 compounds exhibiting favorable alignment with the essential pharmacophoric features were selected for further docking analysis. These findings suggest that the selected pharmacophore model provides a reliable framework for identifying novel Mpro inhibitors with promising binding characteristics [45,46].

The molecular docking of selected ligands and Mpro, revealed that Lig-1 and Lig-2 exhibited stronger binding affinity and more favorable interaction profiles within the SARS-CoV-2 Mpro active-site cavity compared to the other investigated compounds. Both ligands established stable interactions with crucial catalytic residues including HIS41, CYS145, GLU166, MET49, and MET165, which are essential for the proteolytic activity of Mpro. Among the investigated compounds, Lig-1 demonstrated the most stable binding orientation and interaction persistence, suggesting enhanced inhibitory potential against Mpro activity. The stable interactions of Lig-1 and Lig-2 with catalytic pocket residues were further supported by molecular dynamics simulation analyses, indicating stable complex formation throughout the 100 ns simulation period. The interaction of ligands with binding pocket residues including catalytic dyad residues would impact cleavage of viral polyprotein. Owing to the determining role in the proteolysis of viral polyproteins, the Mpro of the FIPV has been established as the preferred target combating its virulence [38].

The Mpro catalytic dyad is pivotal for the enzymatic activity of Mpro, facilitating the cleavage of viral polyproteins into functional proteins essential for viral replication and assembly. Upon host's transcriptional machinery takeover, host cell synthesize two overlapping polyproteins, pp1a and pp1ab, which cleaves through coronavirus-encoded proteases—papain-like protease (PLpro) and Mpro. The cleavage of pp1a and pp1ab forms 16 nonstructural proteins play roles in viral replication [47–49].

During molecular dynamic simulation, the RMSD, RMSF, and RG analyses collectively demonstrated that Lig-1 and Lig-2 maintained stable conformations and compactness throughout the 100 ns simulation period. Notably, Lig-1 exhibited lower structural fluctuations, reduced residue flexibility, and more persistent hydrogen bonding interactions compared to the other investigated complexes, indicating enhanced conformational stability within the active-site cavity.

Ideally, RMSD values would be zero; however, due to statistical uncertainties, it is not possible for a protein to have an RMSD of zero [50].

The RMSF quantifies the overall structural deviations over time and measures the root mean square fluctuations of individual amino acid residues, providing insight into the dynamic behavior and flexibility of these residues during simulations [51].

The magnitude of RG inversely correlates with protein stability, where a larger RG indicates a less stable, more expanded structure. The RG value GS-441524-Mpro complex was 0.2–0.4 Å low comparable to that of the N3-Mpro complex.

Therefore, the RMSF, RMSD and RG profiles of all investigated complexes were comparable to those of the reference Lig-1–Mpro and Lig-2-Mpro complex, exhibiting relatively lower residue fluctuations and indicating favorable structural stability throughout the simulation period [52].

Furthermore, hydrogen bond monitoring and hydrogen bond distance analyses revealed that Lig-1 consistently maintained stable intermolecular interactions with catalytic pocket residues, suggesting stronger binding persistence under dynamic simulated conditions. However, the Lig-2 also exhibited considerable interaction stability; however, slightly higher fluctuations and lower hydrogen bond occupancy were observed compared to Lig-1. Conversely, Lig-3 showed comparatively weaker interaction retention and increased conformational flexibility during the simulation period. Greater variation in hydrogen bond distances reflects reduced stability of ligand interactions within the active-site pocket, whereas hydrogen bond distances close to or below 2.5 Å are typically associated with stronger and more stable protein–ligand binding interactions [53,54].

In conclusion, the present computational investigation identified Lig-1, followed by Lig-2, as promising potential inhibitors of CoVs Mpro based on their favorable binding affinity, stable interactions with key catalytic residues, and sustained structural stability throughout the 100 ns molecular dynamics simulation. The combined pharmacophore modeling, molecular docking, and molecular dynamics analyses demonstrated that these compounds effectively maintained stable occupancy within the active-site cavity of Mpro, suggesting their potential to interfere with viral proteolytic activity and replication. Although these computational findings provide valuable insights for hypothesis generation and rational antiviral drug discovery, further in vitro and in vivo experimental studies are required to validate their biological activity, therapeutic efficacy, and translational significance as potential anti-CoVs agents.

4. Materials and Methods

BIOVIA Discovery Studio Client (v24.1.0.321712, Dassault Systèmes) was used for all computational investigations. Molecular dynamics simulations were performed using the simulation protocols available within the BIOVIA Discovery Studio platform.

4.1. Multiple Sequence Alignment

The protein sequence of Mpro from ten coronaviruses were retrieved from NCBI with the accession numbers (HCoV-NL63 Accession No. XPR31551, HCoV-229E Accession No. WDE18043, FCoV Accession No. QSL97047, PEDV Accession No. WJJ67095, SARS-CoV-2 Accession No. 5R7Z_A, SARS-CoV Accession No. AFR58685, HCoV-HKU1 Accession No. WDE18948, HCoV-OC43 Accession No. YP_009555238, BCoV Accession No. UZT75375, and MERS Accession No. AGV08377). Multiple sequence alignment was performed using Geneious Prime V.9.0. The 3D protein structure was designed using Alphafold3 online server [55].

4.2. Preparation of Protein, Energy Optimization and Sphere Generation

X-Ray crystallographic structure of SARS-CoV-2 3CL protease (PDB ID: 6M2N) has been obtained from the RCSB Protein Data Bank (RCSB PDB accessed on 10. 10. 2025, https://www.rcsb.org/structure/6M2N). The Mpro protease (3CL protease) protein is a highly conserved protease which is indispensable for coronaviruses (CoVs) replication and is a promising target for development of broad-spectrum antiviral drugs. The 6M2N 3CL protein is co-crystal complex with novel inhibitor ligand Baicalein (5,6,7-trihydroxy-2-phenyl-4H-chromen-4-one) in the binding site for inhibition property. The Mpro of severe acute respiratory syndrome coronavirus 2, with sequence length of 306, devoid of mutations, with a resolution of 2.20 Å. Following the selection of Mpro, the structure of protein was prepared by applying the CHARMm force field, using BIOVIA Discovery Studio (BIOVIA Discovery Studio Client v24.1.0.321712). During protein preparation, unnecessary chains, heteroatoms, ligands not relevant to the study, and water molecules water molecules beyond 5 Å, were removed. Hydrogen atoms were added, and missing atoms or residues were filled and fixing the unresolved residues/side chains, using the prepared protein method for the minimization process to be executed [56].

After the minimization of the pharmacological target (PDB: 6M2N), the *Define and edit binding site* method was employed for the generation of a receptor binding site sphere that was used for developing a binding sphere in the catalytic binding site [38].

The binding site spheres was the centroid of the bound co-crystallized ligand was considered for binding site sphere generation with the following coordinates: Mpro—XYZ: (-33.2182, -65.3489, 41.2083), radius: 6.854 A. The binding sphere was carried out by excluding co-crystal ligands so that the other ligands could bind in the selected grid during the receptor ligand docking process. Moreover, the docking protocol validation (Re-Docking of Co-Crystallized Ligand) was conducted through the ligand receptor docking of the co-ligand and the protein. To validate the docking protocol, it is necessary for Root Mean Square deviation (RMSD) between the co-crystal ligand and the Re-docked co-crystal ligand to be obtained ≤2 Å.

4.3. Pharmacophore Model Development and Validation

In this study, a structure-based pharmacophore (SBP) modeling approach of Biovia DS software (BIOVIA Discovery Studio v24.1.0.321712) was employed, which relies on information derived from the ligand binding to the receptor's active site to generate pharmacophore features. Specifically, the known 3CL Pro–co crystalized ligand interaction interface was utilized, centered at the coordinates (-33.2182, -65.3489, 41.2083) with a sphere radius of 6.854 Å —corresponding to the ligand binding region, as described previously [57].

Validation of the generated pharmacophore model is essential to ensure its reliability in distinguishing active-like ligands from putative inactive controls and in capturing meaningful geometric and chemical features relevant to true protein–ligand interactions. This step also confirms the quality of the pharmacophore features and their suitability for downstream virtual screening. The general validation strategy was adapted from established pharmacophore-based screening workflows which emphasize the use of ligand sets to assess model discrimination [57,58].

To evaluate model specificity, a set of 39 structurally diverse protease inhibitors with no known or predicted interaction with ASIP was used as putative inactive controls. Together, these 67 ligands (38 putative active compounds + 39 putative inactive compounds) formed the validation dataset used to assess pharmacophore performance. Pharmacophore validation was conducted using the Ligand Pharmacophore Mapping protocol in BIOVIA Discovery Studio. Screening parameters were set as follows: maximum features = 8, minimum features = 4, and maximum pharmacophores = 10.

According to Discovery Studio's mapping criteria, a ligand was required to match at least four pharmacophore features to be considered a valid hit. This internally generated, docking-driven validation framework enabled assessment of the pharmacophore model's ability to preferentially recognize predicted ligand binders while excluding structurally unrelated non-binders, thereby establishing its suitability for subsequent large-scale virtual screening.

4.4. Pharmacophore Model Based Virtual Screening of Compounds

To perform virtual screening, the ZINCPharmer web tool (http://zincpharmer.csb.pitt.edu/pharmer.html, accessed on 13 December 2025) was utilized to screen the compounds from the ZINC purchasable database. ZINCPharmer is an online database that can search approximately 230 million conformations against the validated pharmacophore model in just a few minutes by using the Pharmer pharmacophore search technology [59].

doi:10.20944/preprints202605.1352.v1

4.5. Molecular Docking and Binding Energy Calculation (MM-GBSA) of selected ligands

To evaluate the binding affinities and interaction profiles of 21 curated drug-like small molecules that passed prior fit value filters with Mpro protein, molecular docking was performed using the −CDOCKER module (-CDOCK) of BIOVIA Discovery Studio v24.1.0.321712. −CDOCK is a CHARMm-based molecular docking tool that enables full ligand flexibility, accounting for variations in torsion angles, bond angles, and bond lengths, thereby generating highly accurate docked conformations. Each ligand was allowed to generate up to ten conformations (poses), and the −CDOCKER interaction energy score (−CDOCK Score) binding was computed for each pose. The pose exhibiting the most positive –CDOCK score was considered the most favorable and stable binding conformation. Docked complexes were visualized in Discovery Studio, and interaction maps were generated to examine key interactions such as hydrogen bonds, hydrophobic contacts, π–π stacking and van der Waals forces. To describe the binding site in protein, the interaction binding affinity score (−CDOCK Score) and calculation of pre molecular dynamic (MD) MM-GBSA (ΔG) are likely internal steps within BIOVIA Discovery Studio v24.1.0.321712 [24,25,38,57].

4.6. Molecular Dynamic Simulation (MDS) and Binding Energy Calculation (MM-GBSA)

Molecular dynamics (MD) simulations were carried out using BIOVIA Discovery Studio v24.1.0.321712. Based on docking score, binding energy, and conformational pose analysis, the best complexes (MPro–ligand) were selected for molecular dynamics (MD) simulation. The best predicted top hits of ligands with respect to Mpro were selected to perform 100 nanoseconds (ns) through the standard dynamics cascade method. To generate the molecular topology files for the Mpro and Mpro–ligand complex and to create the topology of ligands, the CHARMm force field was used. The solvation system consists of an explicit boundary TIP 3-point water solvent model, an orthorhombic box with a minimal distance of 7 nm between the protein surface and the edge of the box neutralized with the inclusion of cation-type sodium (Na) and anion-type chloride (Cl) counter ions [60].

After solvation of the protein and complex, the standard dynamics cascade method was employed. For energy minimization, the steepest descent (minimization 1) for 1000 steps with RMS gradient 1 and conjugate gradient (minimization 2) for 4000 steps with RMS gradient 0.1. Both minimization algorithms were used for a total of 5000 steps. The heating phase was performed using a simulation time of 50 picoseconds (ps) with a time step of 2 fs; for immersion, the initial temperature is 50, and the target temperature is 300 K with a save results interval of 10 ps. The reference temperature was 300 K, and the reference pressure was 1.0 bar for the NPT (isothermal–isobaric) ensemble. The equilibration phase was carried out for a 200 ps simulation run with a 2 fs (femtosecond) time step, and the save result interval is 10 ps to generate the restart file for further equilibration of the complex. The Particle_Mesh_Ewald (PME) algorithm was used for long-range electrostatic interactions with fourth-order cubic interpolation and a kappa 0.34 A grid spacing. The advanced dynamic integrator used the Leapfrog Verlet algorithm with the applied shake constraint. The explicit solvent model was used with a dielectric constant of 1, a nonbond list radius cutoff of 14 A, in which the nonbond higher cutoff distance is 12 A and the nonbond lower cutoff distance is 10 A. The production step of a standard dynamic cascade of MD simulation was carried out for 100 ps with a save result interval step of 10 ps, as described previously.

Several key parameters were analyzed to understand the dynamics of the system. The root mean square deviation (RMSD), root means square fluctuation (RMSF), and radius of gyration (RG) were calculated to measure the overall stability of the protein in the absence and presence of ligand by tracking how much the atoms moved relative to a reference structure [61,62].

The stability of the complex in each frame was further confirmed by monitoring established hydrogen bonds between protein and ligand interacting atoms. The stability of the complex is indicated by the highest potential inhibitor from the stable complex protein–ligand through BIOVIA Discovery Studio v24.1.0.321712 [38].

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org, **Figure S1.** Validation of generated pharmacophore hypotheses showing receiver operating characteristic (ROC) curves and predictive performance metrics including sensitivity, specificity, and area under the curve (AUC). Among the generated models, Pharmacophore_01 (AAHaromHaromR) demonstrated the best predictive performance with an AUC value of 0.781 and was selected for further virtual screening studies. **Table S1.** Pharmacophore-based selection of 21 ZINC compounds from 49 screened compounds based on fit values ranging from 3.5 to 1.0 for further molecular docking analysis against SARS-CoV-2 Mpro. The table includes compound IDs, chemical structures, fit values, and RMSD (Å). **Figure S2.** Two-dimensional (2D) protein–ligand interaction diagrams of the standard and selected ZINC compounds docked within the active-site cavity of SARS-CoV-2 Mpro. (A) Standard reference ligand Baicalein, (B) ZINC95473654, (C) ZINC95473725, and (D) ZINC08792368 showing hydrogen bonding, π–π stacking, π–alkyl, π–sulfur, carbon hydrogen bonding, and van der Waals interactions with key active-site residues.

Author Contributions: Conceptualization, F.M. and M.Y.K.; methodology, A.U.S.; software, M.C.; validation, M.Y.K., F.M. and A.U.S.; formal analysis, M.Y.K. and A.U.S. investigation, M.Y.K.; Resources. N.D.; data curation, M.Y.K writing—original draft preparation, M.Y.K. and A.U.S. writing—review and editing, M.Y.K.; visualization, F.M.; supervision, M.Y.K.; project administration, M.C. and M.G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by a seed grant (PI: MGH) from Long Island University (Grant no: 40262) and funds from the USDA-NIFA Animal Health and Disease Research grant (NI26AHDRXXXXG063-0001).

Informed Consent Statement: Not applicable

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: The authors would like to thank the Long Island University-Dassault Systèmes Center of Excellence.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ba Abduallah, M.M. and M.G. Hemida, *Comparative analysis of the genome structure and organization of the Middle East respiratory syndrome coronavirus (MERS-CoV) 2012 to 2019 revealing evidence for virus strain barcoding, zoonotic transmission, and selection pressure.* Rev Med Virol, 2021. **31**(1): p. 1-12.
2. Hemida, M.G., D. Egamberdieva, and Y.S. Malik, *Editorial: Coronaviruses from the One Health perspective.* Front Microbiol, 2023. **14**: p. 1338529.
3. Shah, A.U., et al., *Comparative Genome Sequencing Analysis of Some Novel Feline Infectious Peritonitis Viruses Isolated from Some Feral Cats in Long Island.* Viruses, 2025. **17**(2).
4. Shah, A.U., P. Gauger, and M.G. Hemida, *Isolation and molecular characterization of an enteric isolate of the genotype-Ia bovine coronavirus with notable mutations in the receptor binding domain of the spike glycoprotein.* Virology, 2025. **603**: p. 110313.
5. Shah, A.U., et al., *Some novel field isolates belonging to lineage-1 of the genotype GI-avian infectious bronchitis virus (AIBV) show strong evidence of recombination with field/vaccinal strains.* Infect Genet Evol, 2025. **129**: p. 105723.
6. Hemida, M.G. and H.M. Al-Ghadeer, *Evidence of Peste des petits Ruminants' Virus in Dromedary Camels in the Kingdom of Saudi Arabia between 2014 and 2016.* Vet Med Int, 2019. **2019**: p. 4756404.
7. Hemida, M.G. and A. Alnaeem, *Some One Health based control strategies for the Middle East respiratory syndrome coronavirus.* One Health, 2019. **8**: p. 100102.
8. Hemida, M.G. and M.M. Ba Abduallah, *The SARS-CoV-2 outbreak from a one health perspective.* One Health, 2020. **10**: p. 100127.
9. Mubarak, A., W. Alturaiki, and M.G. Hemida, *Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Infection, Immunological Response, and Vaccine Development.* J Immunol Res, 2019. **2019**: p. 6491738.

10. Al-Mubarak, A.I.A., et al., *Detection of Avian Orthoavulavirus-1 genotypes VI.2.1 and VII.1.1 with neuro-viscerotropic tropism in some backyard pigeons (Columbidae) in Eastern Saudi Arabia.* Front Vet Sci, 2024. **11**: p. 1352636.
11. Al-Mubarak, A.I.A., et al., *Evidence of the circulation of avian metapneumovirus in domestic backyard chickens in Eastern Saudi Arabia in 2019.* Vet World, 2023. **16**(6): p. 1246-1251.
12. Alsultan, M.A., M.A. Alhammadi, and M.G. Hemida, *Infectious bronchitis virus from chickens in Al-Hasa, Saudi Arabia 2015-2016.* Vet World, 2019. **12**(3): p. 424-433.
13. Chu, D.K.W., et al., *MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity.* Proc Natl Acad Sci U S A, 2018. **115**(12): p. 3144-3149.
14. Hemida, M.G., *The next-generation coronavirus diagnostic techniques with particular emphasis on the SARS-CoV-2.* J Med Virol, 2021. **93**(7): p. 4219-4241.
15. Hemida, M.G., et al., *Molecular characterization and phylogenetic analyses of virulent infectious bronchitis viruses isolated from chickens in Eastern Saudi Arabia.* Virusdisease, 2017. **28**(2): p. 189-199.
16. Hemida, M.G., et al., *Prevalence and molecular characterization of some circulating strains of the peste-des-petits-ruminants virus in Saudi Arabia between 2014-2016.* PeerJ, 2020. **8**: p. e9035.
17. Hemida, M.G. and A.A. Alnaeem, *Betaretrovirus infections in dromedary camels in Saudi Arabia.* Vet Med Sci, 2022. **8**(3): p. 1330-1336.
18. Hemida, M.G., et al., *Coronavirus infections in horses in Saudi Arabia and Oman.* Transbound Emerg Dis, 2017. **64**(6): p. 2093-2103.
19. Hemida, M.G., et al., *Evidence of equine influenza A (H3N8) activity in horses from Eastern and Central Saudi Arabia: 2013-2015.* Equine Vet J, 2019. **51**(2): p. 218-221.
20. Hemida, M.G., et al., *Foot-and-mouth disease virus O/ME-SA/Ind 2001 lineage outbreak in vaccinated Holstein Friesian cattle in Saudi Arabia in 2016.* Vet Q, 2018. **38**(1): p. 88-98.
21. Yang, K.S., et al., *A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors*.* ChemMedChem, 2021. **16**(6): p. 942-948.
22. Duan, Y., et al., *Structural biology of SARS-CoV-2 M(pro) and drug discovery.* Curr Opin Struct Biol, 2023. **82**: p. 102667.
23. Duraisamy, N., et al., *Machine learning tools used for mapping some immunogenic epitopes within the major structural proteins of the bovine coronavirus (BCoV) and for the in silico design of the multiepitope-based vaccines.* Front Vet Sci, 2024. **11**: p. 1468890.
24. Khan, M.Y., et al., *Repurposing of Some Nucleoside Analogs Targeting Some Key Proteins of the Avian H5N1 Clade 2.3.4.4b to Combat the Circulating HPAI in Birds: An In Silico Approach.* Viruses, 2025. **17**(7).
25. Khan, M.Y., et al., *Leveraging Artificial Intelligence and Gene Expression Analysis to Identify Some Potential Bovine Coronavirus (BCoV) Receptors and Host Cell Enzymes Potentially Involved in the Viral Replication and Tissue Tropism.* Int J Mol Sci, 2025. **26**(3).
26. Chang, H.P., C.Y. Chou, and G.G. Chang, *Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride.* Biophys J, 2007. **92**(4): p. 1374-83.
27. Chou, C.Y., et al., *Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease.* Biochemistry, 2004. **43**(47): p. 14958-70.
28. Parmar, M., et al., *Structural differences in 3C-like protease (Mpro) from SARS-CoV and SARS-CoV-2: molecular insights revealed by Molecular Dynamics Simulations.* Struct Chem, 2022: p. 1-18.
29. Narayanan, A., et al., *Identification of SARS-CoV-2 inhibitors targeting Mpro and PLpro using in-cell-protease assay.* Commun Biol, 2022. **5**(1): p. 169.
30. Fu, L., et al., *Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease.* Nat Commun, 2020. **11**(1): p. 4417.
31. Fukumoto, Y., et al., *Development of a Biosafety Level 1 Cellular Assay for Identifying Small-Molecule Antivirals Targeting the Main Protease of SARS-CoV-2: Evaluation of Cellular Activity of GC376, Boceprevir, Carmofur, Ebselen, and Selenoneine.* Int J Mol Sci, 2024. **25**(11).
32. Kneller, D.W., et al., *Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room temperature X-ray crystallography.* Nat Commun, 2020. **11**(1): p. 3202.

33. Jin, Z., et al., *Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur.* Nat Struct Mol Biol, 2020. **27**(6): p. 529-532.
34. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.* Science, 2003. **300**(5626): p. 1763-7.
35. Shi, J., J. Sivaraman, and J. Song, *Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease.* J Virol, 2008. **82**(9): p. 4620-9.
36. Opo, F., et al., *Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein.* Sci Rep, 2021. **11**(1): p. 4049.
37. Jiang, S., et al., *The structure and function of membrane protein in coronavirus infection and its applications in the development of vaccines and therapeutic drugs.* Front Microbiol, 2026. **17**: p. 1762041.
38. Khan, M.Y., et al., *Identification of potential inhibitors of the main protease from feline infectious peritonitis virus using molecular docking and dynamic simulation approaches.* PeerJ, 2025. **13**: p. e19744.
39. Narwal, M., et al., *SARS-CoV-2 polyprotein substrate regulates the stepwise M(pro) cleavage reaction.* J Biol Chem, 2023. **299**(5): p. 104697.
40. Ullrich, S. and C. Nitsche, *The SARS-CoV-2 main protease as drug target.* Bioorg Med Chem Lett, 2020. **30**(17): p. 127377.
41. Hsu, C.K., et al., *Effect of sofosbuvir-based treatment on clinical outcomes of patients with COVID-19: a systematic review and meta-analysis of randomised controlled trials.* Int J Antimicrob Agents, 2022. **59**(3): p. 106545.
42. Motyan, J.A., et al., *Potential Resistance of SARS-CoV-2 Main Protease (Mpro) against Protease Inhibitors: Lessons Learned from HIV-1 Protease.* Int J Mol Sci, 2022. **23**(7).
43. Melano, I., Y.C. Lo, and W.C. Su, *Characterization of host substrates of SARS-CoV-2 main protease.* Front Microbiol, 2023. **14**: p. 1251705.
44. Temml, V., et al., *Discovery of the first dual inhibitor of the 5-lipoxygenase-activating protein and soluble epoxide hydrolase using pharmacophore-based virtual screening.* Sci Rep, 2017. **7**: p. 42751.
45. Gao, K., et al., *Perspectives on SARS-CoV-2 Main Protease Inhibitors.* J Med Chem, 2021. **64**(23): p. 16922-16955.
46. Gao, Q., L. Yang, and Y. Zhu, *Pharmacophore based drug design approach as a practical process in drug discovery.* Curr Comput Aided Drug Des, 2010. **6**(1): p. 37-49.
47. Jin, Z., et al., *Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors.* Nature, 2020. **582**(7811): p. 289-293.
48. Galasiti Kankanamalage, A.C., et al., *Structure-guided design of potent and permeable inhibitors of MERS coronavirus 3CL protease that utilize a piperidine moiety as a novel design element.* Eur J Med Chem, 2018. **150**: p. 334-346.
49. Umar, A.K., et al., *Antiviral phytocompounds "ellagic acid" and "(+)-sesamin" of Bridelia retusa identified as potential inhibitors of SARS-CoV-2 3CL pro using extensive molecular docking, molecular dynamics simulation studies, binding free energy calculations, and bioactivity prediction.* Struct Chem, 2022. **33**(5): p. 1445-1465.
50. Ahmad, P., et al., *Targeting SARS-CoV-2 main protease (Mpro) and human ACE-2: A virtual screening of carotenoids and polyphenols from tomato (Solanum lycopersicum L.) to combat Covid-19.* Intelligent Pharmacy, 2024. **2**(1): p. 51-68.
51. Ashraf, Z., et al., *Synthesis, Bioevaluation and Molecular Dynamic Simulation Studies of Dexibuprofen-Antioxidant Mutual Prodrugs.* Int J Mol Sci, 2016. **17**(12).
52. Mathpal, S., et al., *A dynamic simulation study of FDA drug from zinc database against COVID-19 main protease receptor.* J Biomol Struct Dyn, 2022. **40**(3): p. 1084-1100.
53. Rasyid, H., B. Purwono, and H.D. Pranowo, *Design of New Quinazoline Derivative as EGFR (Epidermal Growth Factor Receptor) Inhibitor through Molecular Docking and Dynamics Simulation.* Indonesian Journal of Chemistry; Vol 21, No 1 (2021)DO - 10.22146/ijc.57012, 2020.
54. Nada, H., A. Elkamhawy, and K. Lee, *Identification of 1H-purine-2,6-dione derivative as a potential SARS-CoV-2 main protease inhibitor: molecular docking, dynamic simulations, and energy calculations.* PeerJ, 2022. **10**: p. e14120.
55. Abramson, J., et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3.* Nature, 2024. **630**(8016): p. 493-500.

56. Halder, D., R.S. Jeyaprakash, and B. Ghosh, *A Structure-Based Design Strategy with Pyrazole-Pyridine Derivatives Targeting TNFalpha as Anti-Inflammatory Agents: E-Pharmacophore, Dynamic Simulation, Synthesis and In Vitro Evaluation.* Chem Biodivers, 2024. **21**(9): p. e202400778.
57. Maarfi, F., et al., *Blocking ASIP to Protect MC1R Signaling and Mitigate Melanoma Risk: An In Silico Study.* Pharmaceuticals (Basel), 2026. **19**(1).
58. James, N., V. Shanthi, and K. Ramanathan, *Drug Design for ALK-Positive NSCLC: an Integrated Pharmacophore-Based 3D QSAR and Virtual Screening Strategy.* Appl Biochem Biotechnol, 2018. **185**(1): p. 289-315.
59. Luo, L., et al., *Structure-Based Pharmacophore Modeling, Virtual Screening, Molecular Docking, ADMET, and Molecular Dynamics (MD) Simulation of Potential Inhibitors of PD-L1 from the Library of Marine Natural Products.* Mar Drugs, 2021. **20**(1).
60. Singh, A., et al., *Computational assessment of Withania somnifera phytomolecules as putative inhibitors of Mycobacterium tuberculosis CTP synthase PyrG.* J Biomol Struct Dyn, 2023. **41**(11): p. 4903-4916.
61. Lohani, M., et al., *Structure based in silico screening of natural Vitamin D analogs for targeted and safer treatment of resistant multiple sclerosis.* Sci Rep, 2025. **15**(1): p. 29321.
62. Hospital, A., et al., *Molecular dynamics simulations: advances and applications.* Adv Appl Bioinform Chem, 2015. **8**: p. 37-47.