REVISITING THE SARS-COV-2 MAIN PROTEASE: A 2023 IN SILICO ODYSSEY IN SEARCH OF POTENTIAL INHIBITORS

Chenyue Fan[1,2], Ayesha Abdul Qadir Memon[3], Prajit Adhikari[4], Muhammad Osama[3,5], Calvin R. Wei[2*]

[1]College of Pharmacy, University of Arizona, Tucson, Arizona, USA
[2*]Department of Research and Development, Shing Huei Group, Taipei, Taiwan
[3]Department of Pharmacy Practice, University of Karachi, Karachi, Pakistan
[4]Institute of Engineering, Tribhuvan University, Kathmandu, Nepal
[5]Department of Pharmacology, University of Karachi, Karachi, Pakistan

*Corresponding Author: Calvin R. Wei
*Department of Research and Development, Shing Huei Group, Taipei, Taiwan
Email: wei.calvin@shinghueigroup.com

Abstract

The novel coronavirus disease or Covid-19 is a global pandemic caused by the SARS-CoV-2 virus originated from Wuhan, China in December 2019. A rapidly spreading, contagious virus that caused more than 6.7 million deaths worldwide. The main protease of SARS-CoV-2 is believed to play a vital role in mediating viral replication and transcription, making it a potential target of interest against Covid-19. In this study, virtual drug screening methods were conducted against a current M^{pro} structure (Protein Data Bank ID: 8SXR) with 868 ligands from the NIH Clinical Collection of clinical trial molecules. Multiple possible hit compounds were identified with compound **1** and **2** outperforming the other compounds in binding conformation and binding free energy. Toxicity and ADMET properties of the top 5 compounds were further investigated computationally. To further validate the results, molecular dynamic simulations of the top 2 complexes were performed. The two complexes displayed stable affinity in respect to the root mean square distance (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA) and hydrogen bond.

Keywords: SARS-CoV-2; COVID-19; Mpro; Main Protease; Molecular Docking; Molecular Dynamics; ADMET; Toxicity; Drug Repurposing; Drug Discovery; Virtual Screening; Clinical Trials

Introduction

From the start of December 2019, the world experienced an acute respiratory disease outbreak known as SARS-CoV-2. To this day, the SARS-CoV-2 pandemic continues to devastate the world, with 663,640,386 total infections and 6,713,093 deaths as of 21, January 2023 [1]. In the early phases of the outbreak, Zhou and co-workers collected full-length genomic sequences of SARS-CoV-2 from five different patients [2]. Their analysis revealed 79.6% homology with SARS-CoV and 96%

resemblance to bat coronavirus at the genome level. The International Committee of Taxonomy of Viruses (ICTV) classified the virus as "severe acute respiratory syndrome coronavirus 2" or SARS-CoV-2, while the World Health Organization (WHO) pronounced this new disease as COVID-19 [3-4].

Although the virus is highly contagious, most cases of the coronavirus result in mild to moderate symptoms, with some cases resulting in severe symptoms and mortality. Currently, the WHO has classified five variants of SARS-CoV-2 across the world: Alpha, Beta, Gamma, Delta, and Omicron [5]. From a genomic standpoint, SARS-CoV-2 displays similarities to other RNA viruses, with the susceptibility of random mutations affecting both structural and non-structural genes. These genetic mutations not only influence transmissibility and viral phenotype but also contribute to resistant adaption towards anti-viral therapies and vaccines.
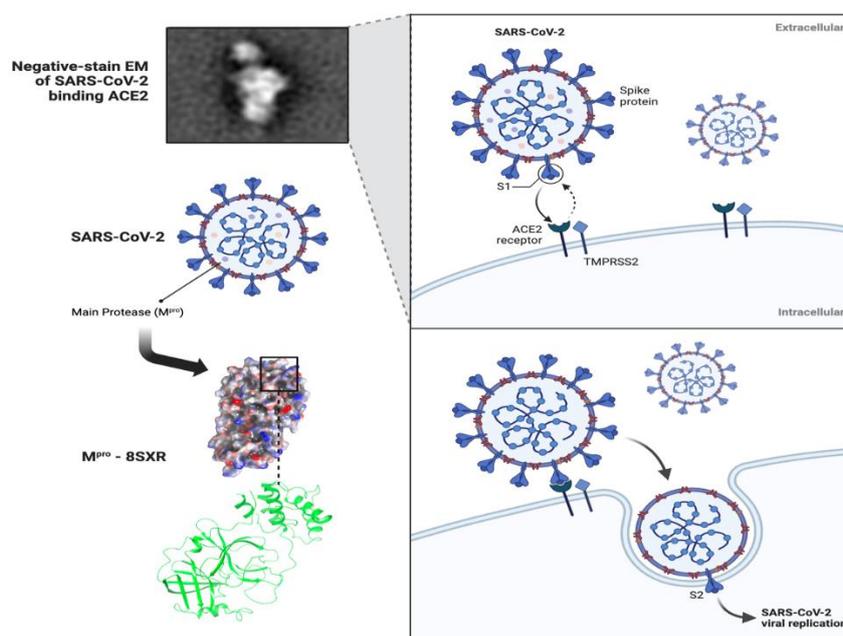

Figure 1: Diagram of SARS-CoV-2 in Relation to Main Protease

SARS-CoV-2 is made up of single stranded positive RNA genome that specify structural and non-structural proteins. Of interest to the non-structural protein are the RNA-dependent RNA polymerase (RdRp), the coronavirus main protease (M^{pro} or $3CL^{pro}$), and the papain-like protease (PL^{pro}) [6]. Upon entering the host cell, the viral genome is translated into polyprotein by the host cell's protein translation mechanism. Proteases such as M^{pro} and PL^{pro} then cleave the polyprotein into effector proteins [7-9]. It is thought M^{pro} can cleave polyproteins at no less than 11 conserved sites allowing it to play a pivotal role in viral replication [10], thus making it an attractive therapeutic target (Figure 1).

Due to how SARS-CoV-2 still plays a big part in society as well as being a major public concern in 2023, there is a need for the discovery of new therapeutics [11]. While researching and developing new therapeutics is time-consuming and necessitate the need for diverse financial responsibilities, the repositioning of existing or known novel therapeutics should be considered. A cost-effective and rapid method for identifying potential drug candidates would be the usage of computer-assisted virtual screening (VS) contrary to high-throughput screening (HTS) [12]. The VS methodology optimizes the selection of potential drug candidates or hits through in silico screening of library collections of different molecules against a target of interest. Recently, VS has played a pivotal role in discovery of small molecule inhibitors of therapeutic targets.

In this study, we explored a dataset from the National Institutes of Health (NIH) Clinical Collection library of small molecules that have an extensive history in human clinical trials. The dataset is composed of 868 ligands which were used in a docking-based virtual screening approach against the Mpro inhibitor. An array of screening strategies, such as molecular docking, molecular dynamics simulation, and ADMET prediction were carried out. A current 2023 Mpro crystalline structure was obtained and analyzed to provide a current theoretical insight. The purpose of this study was to provide a theoretical insight on possible drug hits for future biological experiments.

Materials and Methods
Analysis of Crystalline Structure
The COVID-19 Main Protease X-ray crystallography structure (PDB ID: 8SXR) with resolution 2.11 Å was obtained from the Protein Data Bank at the RCSB site (rscb.org) [13]. The structure was then evaluated on the PROCHECK, ERRAT, and Verify3D webserver to conduct chirality, dihedral angles, planarity, disulphide bonds, covalent geometry non-bonded interactions, stereochemical parameters, main-chain hydrogen bonds, parameter comparisons, and residue-by-residue analysis [14]. Ramachandran plot was then generated to predict Phi and Psi angles of the backbone conformation. Further quality assessment of the structure was performed on ProSA, a web-based analysis interactive to identify possible errors within the 3D structure [15].

Crystalline Structure Preparation
The Main Protease structure, PDB: 8SXR was then prepared in Maestro (Schrödinger Maestro version 2023-3) using the Protein Preparation Wizard within the suite to eliminate possible defects involving missing hydrogens, inaccurate bond orders, charges, alignment issues, and missing side chains using default parameters with the force field OPLS_4 (Optimized Potential for Liquid Simulations) for optimization and minimization [16-18]. Furthermore, a restrained energy minimization was used to remove steric hinderance and strained bonds or confirmation. Lastly, restrained minimization of the heavy atoms was converged to a root mean square deviation or RMSD of 0.30 Å for optimal docking simulation.
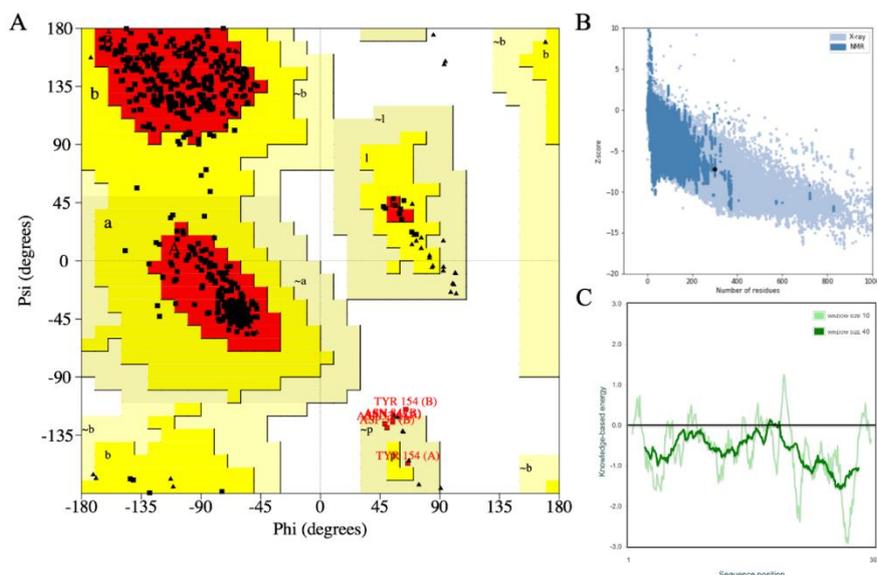

Figure 2. Ramachandran plot of φ−ψ distribution of 8SXR generated by PROCHECK (A) Protein structural analysis validation from ProSA webserver (B)(C)

Ligand Data Set Preparation for Virtual Screening
A collection of 868 2D structures were downloaded from the NIH Clinical Collection as a SDF file. The structure files were then loaded into BioLuminate workspace (BioLuminate version 2023-3) to identify possible ligand duplicates or errors [19-22]. LigPrep (LigPrep version 2023-3) interface

within the Schrödinger suite was then used to prepare the 2D structures into 3D structures as well as neutralizing charges and stereoisomer generation using the OPLS_4 force field at an ionization state of pH 7.0 ± 2.0 through Epik generating 35 poses per ligand [23].

Virtual Screening by Using Glide for Molecular Docking

The receptor grid creation panel in the Glide (Glide version 2023-3) module of Schrödinger suite was used to create a grid box around the co-crystallized ligand within the receptor protein [24-27]. The boxes were centered using the co-crystallized ligand as reference within the complex while scaling factor of van der Waal's radius was established at 1.0 Å and partial charge cutoff retained at 0.25 without constraints. Through this, a cubic box with coordinates of -23.72 in the *x* direction, 7.05 in the *y* direction, and 26.45 in the *z* direction was generated from the basis of chain A. Molecular docking was then performed with the ligand data set using the docking function in Glide for Standard Precision (SP).
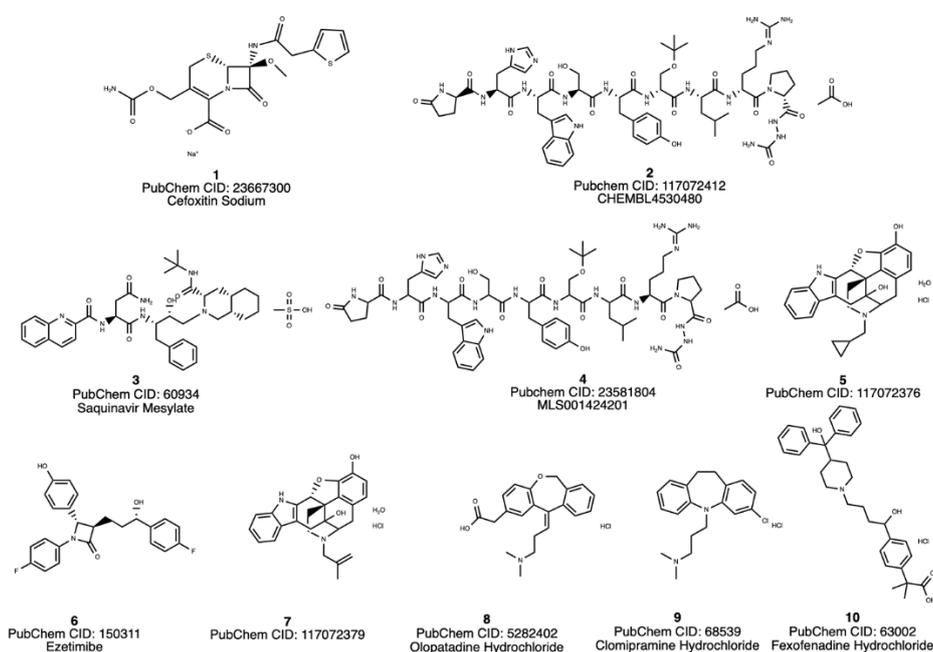

Figure 3. 2D Structure of the Top 10 Compounds of Interest from Docking Results

ADMET, Drug-Likeness Properties, and Toxicity Prediction

ADMET and drug-likeness properties of the top five top scoring compounds were analyzed with the webserver ADMETlab 2.0 [28]. This is to evaluate druggable effects and acceptability of screened compounds. The ProTox-II platform was utilized to predict possible toxicity [29-30].

Molecular Dynamic Simulation

The stability of the top 2 compounds docked in the active site in complex with the Main Protease was studied under physiological conditions using SiBioLead LLP (https://sibiolead.com), a GROMACS based molecular dynamic server [31-35]. The complexes were solvated using a triclinic box with water molecules as Simple Point Charges (SPC) and NaCl as counter ions. Physiological conditions were maintained with the addition of 0.15 M NaCl to the system. The CHARM27 force field was used for both simulations. The system was equilibrated for 100 ps using the NVT/NPT protocol with temperatures set at 300k and pressure at 1 bar. Leap Frog simulation integrator at 5000 frames was then carried out.

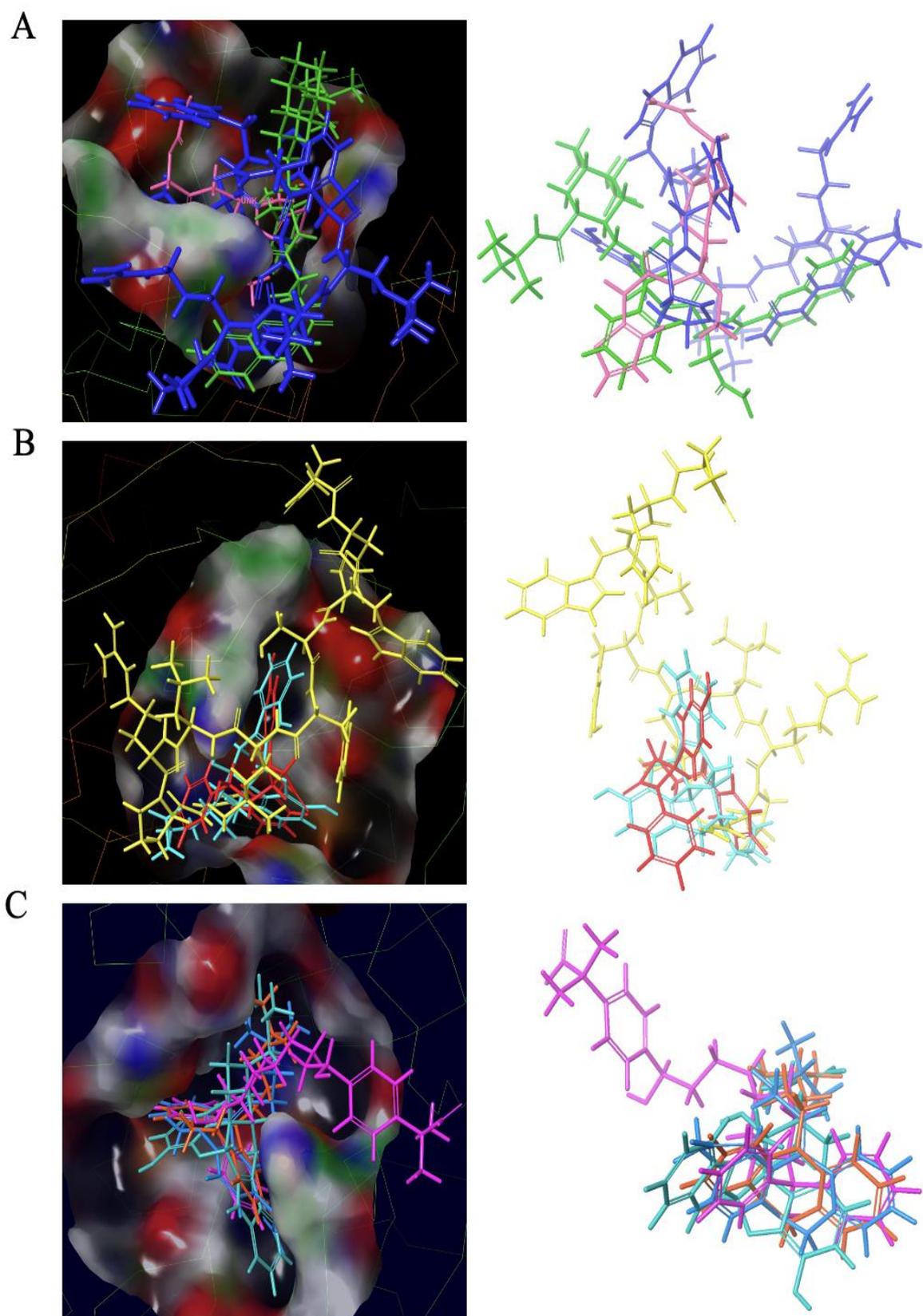**Figure 4**. Overlap of Compound **1** (Faded Salmon), **2** (Blue), and **3** (Green) (A) Overlap of Compound **4** (Yellow), **5** (Faded Teal), and **6** (Red) (B) Compound **7** (Teal), **8** (Azure), **9** (Red-Orange), and **10** (Magenta) (C)
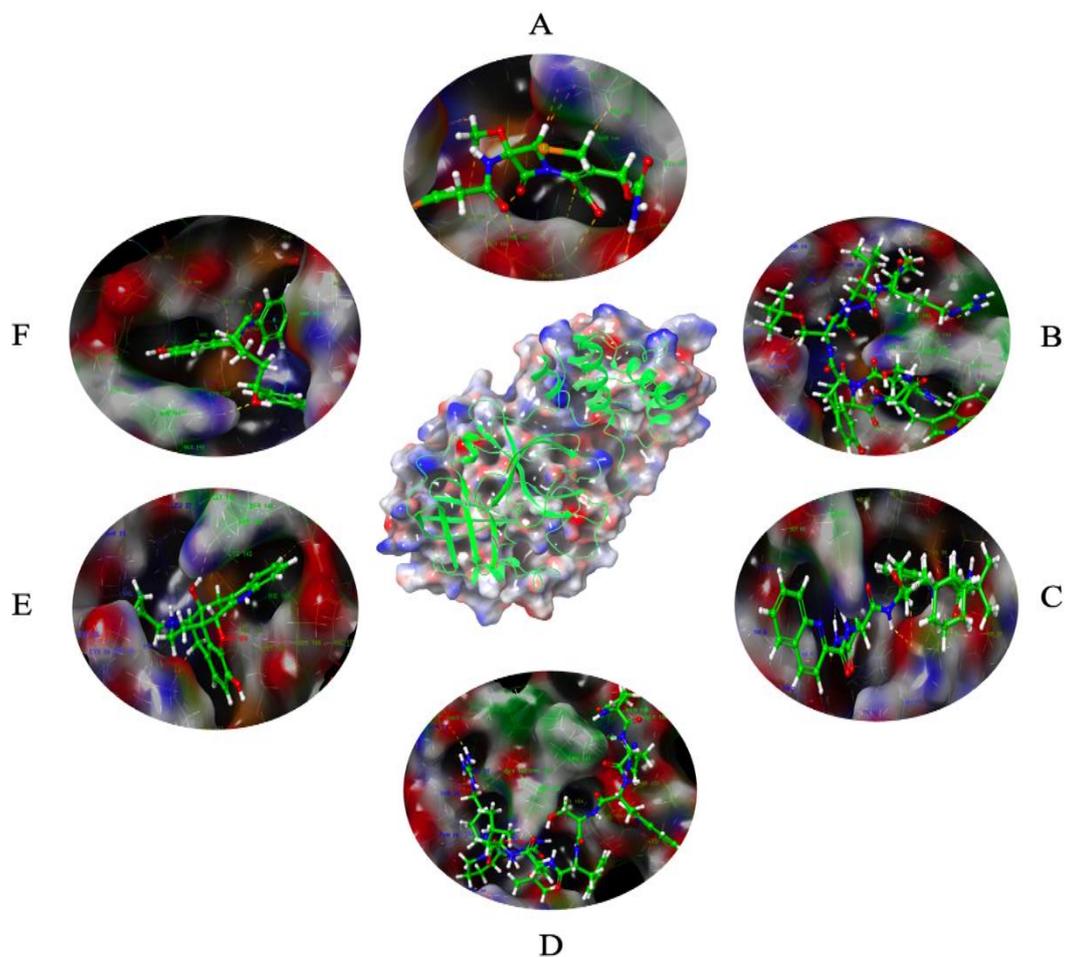
Figure 5. Center Main Protease Mpro 8SXR; Visualization of Compounds **1** (A), **2** (B), **3** (C), **4** (D), **5** (E), **6** (F) Docked Inside Binding Pocket

Results and Discussion

Compound Number	PubChem CID	Molecular Formula	Molecular Weight (g/mol)	Binding Energy (kcal/mol)
1	23667300	$C_{16}H_{16}N_3NaO_7S_2$	449.4	-8.933
2	117072412	$C_{61}H_{88}N_{18}O_{16}$	1329.5	-8.661
3	60934	$C_{39}H_{54}N_6O_8S$	766.9	-8.571
4	23581804	$C_{61}H_{88}N_{18}O_{16}$	1329.5	-8.253
5	117072376	$C_{26}H_{29}ClN_2O_4$	469	-8.110
6	150311	$C_{24}H_{21}F_2NO_3$	409.4	-8.096
7	117072379	$C_{26}H_{29}ClN_2O4$	469	-7.991
8	5282402	$C_{21}H_{24}ClNO_3$	373.9	-7.940
9	68539	$C_{19}H_{24}Cl_2N_2$	351.3	-7.910
10	63002	$C_{32}H_{40}ClNO_4$	538.1	-7.907

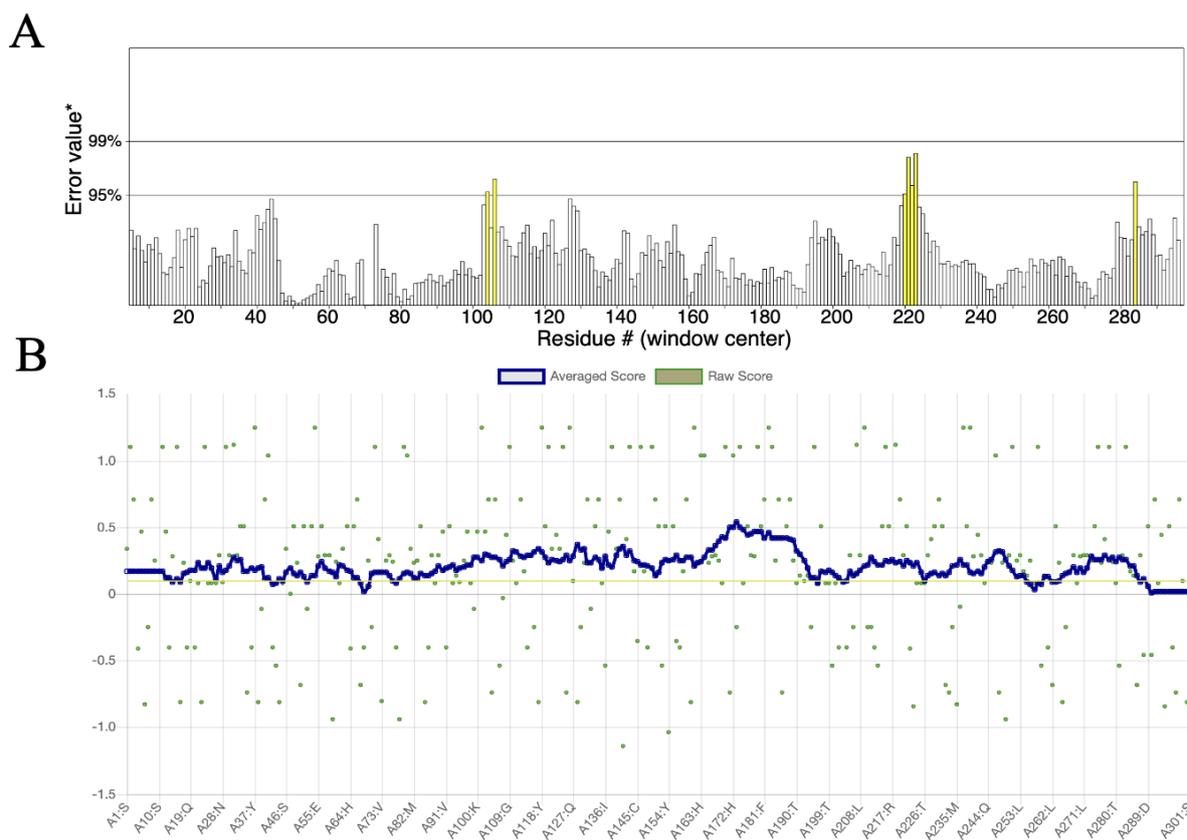**Table 1.** Top 10 Compounds from Docking Results

Figure 6. ERRAT Plot (A) and Verify 3D Plot (B)

The Ramachandran plot generated from Procheck for the Main Protease crystalline structure of 8SXR revealed phi/psi angles distribution of 93.5% residues were in the most favored regions, 5.3% residues were situated in the additional allowed regions, 1.0% residues fell in the generously allowed regions, while 0.2% were in the disallowed conformations (Figure 2A). From this stereochemical quality estimation by Procheck it is revealed the structure is reliable for molecular docking. The Z-score was analyzed by PROSA web tool to measure the deviation of the total energy of the complex to the energy distribution to the random conformations. The complex revealed an overall model quality score of -7.24 revealing it be of decent quality (Figure 2B and 2C).

Furthermore, ERRAT a protein structure analysis method used to detect incorrect regions within a protein structure through examining errors leading to random atoms distributions vs distinguishing correct distributions revealed an overall quality of 97.931% [36] (Figure 6A). Verify 3D, an analysis used to visualize the compatibility of the structure's atomic model in correlation with its amino acid sequence was simulated and revealed an above average score when comparing to other structures (Figure 6B).

During the docking simulation, 35 different conformations were generated for each ligand and docked into the prepared complex. From the docking outcomes, the conformation with the lowest docked energy was designated as the best conformation. Bases on the dataset used for the study, all compounds had a binding energy between -8.933 and -2.101 kcal/mol, suggesting most of the interactions are exothermic. According to the ranking of energy values, the top 10 compounds from the molecular docking analysis are shown in **Table 1** (Figure 3). Additionally, the visualization of the top 6 compounds overlaps and singular binding inside the pocket of the complex is shown (Figure 4 and 5).

The protein-ligand interactions provide visual insights into the interaction of the top compounds within the binding pocket (Figure 6). From an analysis standpoint, all docked complexes display nonbonding interactions such as hydrogen bonding, hydrophobic interactions, pi-pi stacking, polar

and non-polar interactions, and charged interactions. These interactions play a crucial role in helping to understand the binding ability of a compound to the active site.

An in-depth examination of the ligand interaction diagram display dominates polar and hydrophobic residues for most of the interactions. Compound **1** displayed hydrogen bond at residues ASN 142 and GLU 166. Compound **2** engaged hydrogen bonding at residues ASN 142, HIE 163, GLU 166, HIS 164, THR 45, and THR 26. Compound **3** showed hydrogen bonding at residues GLN 189 and formed a salt bridge at HIE 4. Compound **4** formed hydrogen bonds at residues HIE 41, ASN 142, GLU 166, THR 26, and ASN 119. Compound **5** involved hydrogen bonds at residue ASN 142 and formed a Pi-cation at residue HIE 41. Compound **6** involved hydrogen bond interactions at residues ASN 142 and HIE 163 while forming a π- stacking at residue HIE 41. Compound **7** formed a hydrogen bond at residue ASN 142 and a Pi-cation at residue HIE 41. Compound **8** produced a hydrogen bond at residue HIE 41. Compound **9** generated a hydrogen bond at residue HIE 41 and a halogen bond was overserved at residue HIE 163. Finally, compound **10** produced a hydrogen bond at residues ASN 119 and GLN 189.

Hydrogen bonding interactions are an important pillar for the binding stability of a complex. We've noticed most hydrogen bonding involves around the residues ASN 142, GLU 166, HIE 41, and HIE 163 for most of the interactions in the docking simulations.
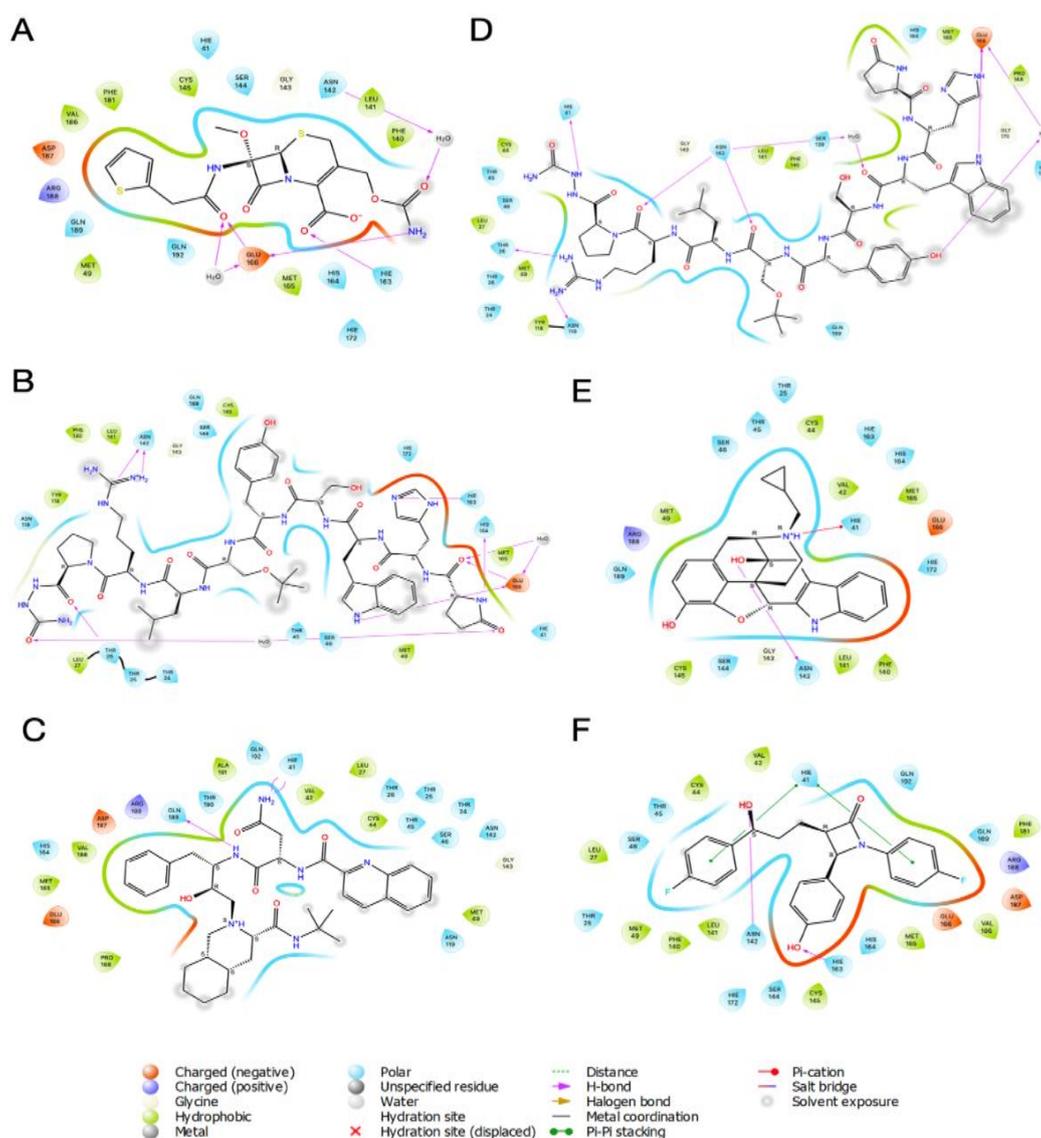

Figure 6. Ligand Interaction Diagram of the Top 6 Compound **1** (A), **2** (B), **3** (C), **4** (D), **5** (E), and **6** (F)
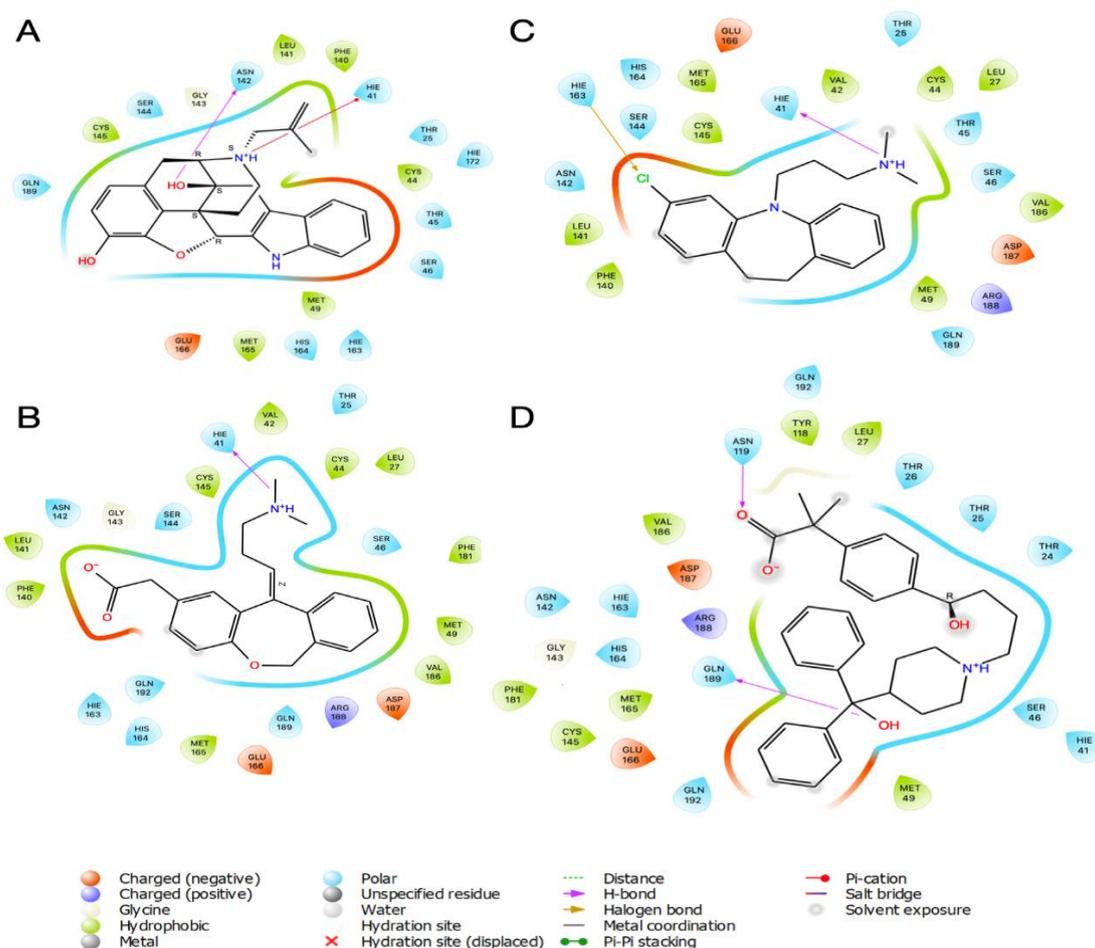
Figure 7. Ligand Interaction Diagram of the Top 6 Compound **7** (A), **8** (B), **9** (C), **10** (D)

Compound	1	2	3	4	5
logP	-0.675	-0.746	2.411	-0.746	4.054
logS	-1.863	-2.718	-2.846	-2.718	-4.095
logD	-0.616	-0.755	2.723	-0.755	3.520
TPSA	151.090	533.190	221.120	533.190	100.220
nHA	10	34	14	34	6
nHD	3	21	7	21	5
nRing	3	6	5	6	8
nRot	9	43	16	43	2

Table 2. In-Silico Values of Physicochemical Properties Obtained with ADMETlab 2.0

In addition to the wealth of experimental data available for the top 5 molecules of interested, we employed an in silico ADMET webserver, ADMETlab 2.0 to generate additional insights into their physicochemical, metabolic, and toxicological properties and to complement the extensive clinical data available for this computational study. Simulated physicochemical properties (**Table 2**) reveals the top five compounds are within acceptable values or within an acceptable range. The data further reveals, LogP and LogS, theories related to lipophilicity of a molecule were within acceptable range which one may deduce high possibilities of optimal oral bioavailability suggesting development of oral administration. **Table 3** displays simulated pharmacological properties where various parameters were calculated (Figure 8). It is believed the closer or positive the NP values are, the higher the probability that the molecule has a greater similarity to a natural product. Suggesting compound **5** might be closely related to a natural product while compounds **2**, **3**, **4** have a slight chance. Compound **5** generated a Chelator alert implying it might generate a reaction with metal present in the catalytic site of the enzyme.
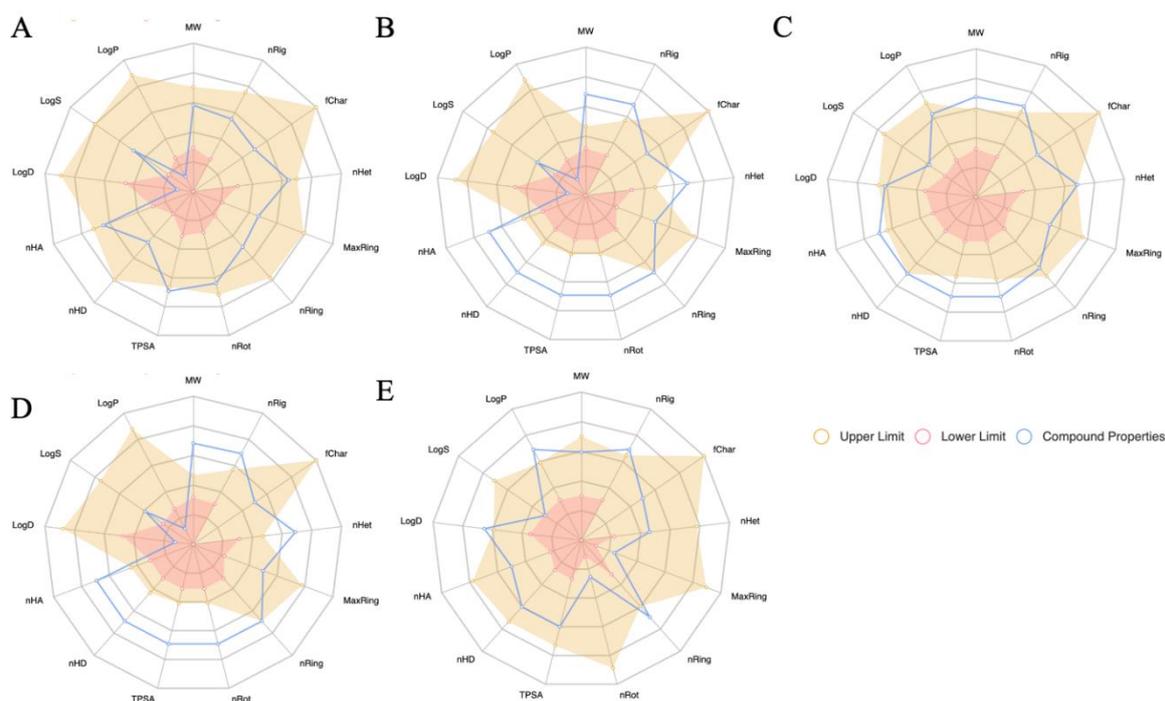
Figure 8. Predicted Physicochemical Properties Obtained with ADMETlab 2.0 Compound **1** (A), **2** (B), **3** (C), **4** (D), **5** (E)

Compound	1	2	3	4	5
NP Score	-0.275	-0.033	-0.080	-0.033	1.212
PAINS	0 Alerts	0 Alerts	0 Alerts	0 Alerts	0 Alerts
Lipinski's Rule	Accepted	Rejected	Rejected	Rejected	Accepted
Pfizer Rule	Accepted	Accepted	Accepted	Accepted	Accepted
GSK Rule	Rejected	Rejected	Rejected	Rejected	Rejected
Golden Triangle	Accepted	Rejected	Rejected	Rejected	Accepted
Chelator Rule	0 Alerts	0 Alerts	0 Alerts	0 Alerts	1 Alert

Table 3. In Silico Values of Drug Likeness Properties Obtained with ADMETlab 2.0

Compound	1	2	3	4	5
HIA	+++	+++	++	+++	--
Caco-2 Permeability	-5.357	-7.107	-5.724	-7.107	-5.917
MDCK Permeability	2E-05	5.4E-06	1.6E-05	5.4E-06	7.8E-06
Pgp-Inhibitor	---	---	---	---	---
Pgp-Substrate	-	+++	---	+++	--
F20%	+++	+++	+++	+++	+++
F30%	+++	+++	++	+++	-

Table 4. In Silico Values of Absorption Parameters Obtained with ADMETlab 2.0

Human Intestinal Absorption (HIA) values are shown in **Table 4**. Compound **5** displays the ability to be passively absorbed in the intestine, while the other four compounds may present difficulty in permeating passively into the gastrointestinal system. Caco-2 or the human colon epithelial cancer cell line is often predicted due to its ability to determine whether a substance is suitable for oral administration, intestinal permeability, and drug efflux. The optimal range for possible Caco-2 permeability are values greater than -5.15. Of the five compounds analyzed none are permeable. However, due to the statistical proximity of compound **1** and **5** to the border value once error is considered, these two compounds can be considered as possibilities. The MDCK cell line, known for drug transport and permeability was predicted. All compounds displayed low to medium probability of permeability. P-glycoprotein known as a transmembrane protein, acts as a pump that actively transfers drugs outside of the cell. Thus, the drug can act as a substrate, reducing and increasing

bioavailability, through interactions at allosteric sites causing conformational changes that blocks the channel. While all compounds show positive signs as possible PGP inhibitors favoring bioavailability, not all compounds show positive signs for as possible substrates for PGP, which disfavor the bioavailability of those compounds.

Compound	1	2	3	4	5
PPB	51.988%	16.476%	96.618%	16.476%	95.534%
VD	0.249	0.292	0.371	0.292	3.039
BBB Penetration	---	---	---	---	+++
Fu	63.202%	41.621%	1.612%	41.621%	2.450%

Table 5. In Silico Values of Distribution Profile Obtained with ADMETlab 2.0

Simulated results of the distribution and metabolism of the five molecules are shown in **Table 5 and 6**. It is assumed through studies that drugs with high plasma protein binding or PPB may have a lower therapeutic index. Compounds 3 and 5 display the highest value in this regard. All molecules displayed optimal range for volume of distribution (VD), while only compound 5 displayed the probability of penetrating the blood brain barrier (BBB). Fu parameter or fraction not bound in plasma indicates possible interactions with target proteins such as receptors, channels, and enzymes. While also displaying ability to diffuse between plasma and tissues. Compound **1**, **2**, and **4** displayed high probability while **3** and **5** have lower probability. Metabolic profiling of the molecules was analyzed through determining the substrate and inhibitor values of various cytochrome-450 (CYP) family. The CYP enzyme families usually involves CYP1, CYP2, CYP3, and CYP4. Cytochrome enzymes play an important role during drug metabolism; therefore, is of interest when studying molecule profiles. This is due to participation as either a substrate or inhibitor contributes to drug action. From the data, compound 2 and 3 shows inhibitory or substrate potential. However, compound 1, 3, and 5 shows possible interactions for various CYP.

Compound	1	2	3	4	5
CYP1A2 inhibitor	---	---	---	---	---
CYP1A2 substrate	--	---	---	---	++
CYP2C19 inhibitor	---	---	---	---	--
CYP2C19 substrate	--	---	---	---	+
CYP2C9 inhibitor	---	---	+	---	--
CYP2C9 substrate	--	-	++	-	--
CYP2D6 inhibitor	---	---	---	---	+++
CYP2D6 substrate	---	---	--	---	+++
CYP3A4 inhibitor	---	---	+++	---	--
CYP3A4 substrate	+++	---	+	---	++

Table 6. In Silico Values of Metabolism Profile Obtained with ADMETlab 2.0

Oral toxicity was simulated using the ProTox II web server that bases its prediction based on 2D similarity and recognition of toxic fragments expressed as LD_{50} (mg/kg). The platform is classified into five different categories: (1) acute toxicity using oral toxicity models with six different toxicity classes; (2) organ toxicity; (3) toxicological endpoints; (4) toxicological pathways; and (5) toxicological targets. The ProTox II web server incorporates 33 different models for the prediction of various toxicity endpoints by identifying molecular similarities, fragment propensities, fragment similarity-based CLUSTER cross validation, and machine-learning. Toxicity endpoints and organ toxicity prediction involves simulating hepatotoxicity, cytotoxicity, carcinogenicity, mutagenicity, and immunotoxicity. Furthermore, it simulates based on data from in vitro and in vivo assays. The ProTox II web server also simulates toxicological pathways based on nuclear receptor signaling pathways and stress response pathways. This methodology is based on how a compound can activate or inhibit a receptor or an enzyme. When interacted, it may result in perturbation of diverse biological pathways, thus disrupting cellular processes and causing cell death.

The rat oral toxicity prediction LD_{50} (mg/kg) and its corresponding toxicity classes (I-VI) for the top 5 compounds are shown in Table 8 (Figure 8). Among the 5 simulated compounds, compound **1** displayed the least toxicity at (LD_{50} = 1000 mg/kg) as a class VI with above average prediction accuracy. This classification designates compound 1 as a possible non-toxic class as its LD_{50} is greater than 5000. Compounds **2**, **4**, and **5** both fell under the class IV classification defining it as harmful after swallowing as its LD_{50} fell between 300 and 2000. Compound **3** is classified as a class III toxicity implying toxic after swallowing as its LD_{50} is between 50 and 300.
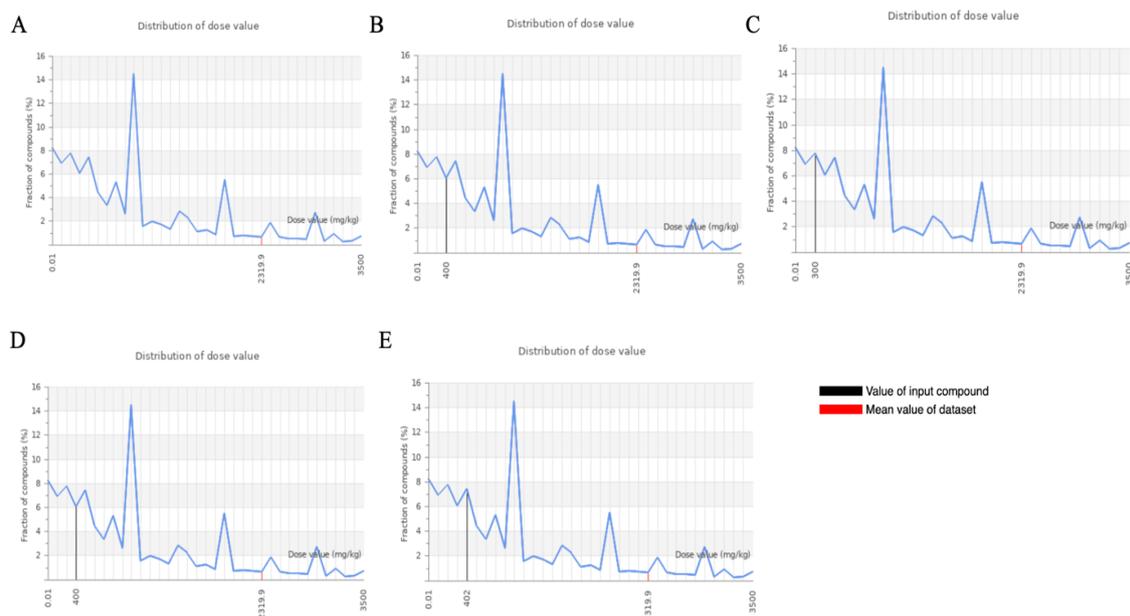

Figure 8. Graphical Representation of Predicted Dose Value Distribution for Compound **1** (A), **2** (B), **3** (C), **4** (D), and **5** (E)

Compound	Predicted LD50 (mg/kg)	Predicted Class	Toxicity Average (%)	Similarity Prediction (%)	Accuracy
1	10000	6	99.57	72.9	
2	400	4	60.75	68.07	
3	300	3	54.96	67.38	
4	400	4	60.75	68.07	
5	402	4	55.59	67.38	

Table 8. Acute Oral Toxicity Prediction Using ProTox II Web Server

Compound	Hepatotoxicity (%)	Carcinogenicity (%)	Immunotoxicity (%)	Mutagenicity (%)	Cytotoxicity (%)
1	Inactive (74)	Inactive (63)	Inactive (99)	Inactive (64)	Active (50)
2	Inactive (80)	Inactive (56)	Inactive (97)	Inactive (58)	Inactive (63)
3	Inactive (72)	Inactive (71)	Inactive (95)	Inactive (69)	Inactive (64)
4	Inactive (80)	Inactive (56)	Inactive (97)	Inactive (58)	Inactive (63)
5	Inactive (83)	Inactive (66)	Active (97)	Inactive (61)	Inactive (53)

Table 9. Organ Toxicity and Toxicological Endpoints Predicted Activity Using ProTox II Web Server

Predicted organ toxicity with emphasis to liver toxicity or hepatotoxicity is shown in Table 9. All compounds simulated for this study revealed above average probability in being hepatotoxicity, carcinogenicity, and mutagenicity inactive. While compound **5** displayed a high possibility of being immunotoxicity active and compound **1** display a low probability of being cytotoxic. All other compounds simulated showed inactivity for immunotoxicity and cytotoxicity.

Compound	1	2	3	4	5
Aryl Hydrocarbon Receptor (AhR) (%)	Inactive (98)	Inactive (98)	Inactive (97)	Inactive (98)	Inactive (90)
Androgen Receptor (AR) (%)	Inactive (99)	Inactive (99)	Inactive (98)	Inactive (99)	Inactive (99)
Androgen Receptor Ligand Binding Domain (AR-LBD) (%)	Inactive (99)	Inactive (99)	Inactive (98)	Inactive (99)	Inactive (99)
Aromatase (%)	Inactive (99)	Inactive (98)	Inactive (87)	Inactive (98)	Inactive (95)
Estrogen Receptor Alpha (ER) (%)	Inactive (98)	Inactive (96)	Inactive (91)	Inactive (96)	Inactive (93)
Estrogen Receptor Ligand Binding Domain (ER-LBD) (%)	Inactive (99)	Inactive (99)	Inactive (99)	Inactive (99)	Inactive (99)
Peroxisome Proliferator Activated Receptor Gamma (PPAR-Gamma) (%)	Inactive (99)	Inactive (99)	Inactive (98)	Inactive (99)	Inactive (100)

Table 10. Toxicological Pathways: Nuclear Receptor Signaling Pathways Prediction Using ProTox II Web Server

Nuclear receptor signaling pathways were simulated and all 5 compounds are given a high probability score of being inactive for AhR, AR, AR-LBD, ER, ER-LBD, and PPAR Gamma (**Table 10**). The results revealed that these compounds might exhibit weak estrogenic activities as well as antiestrogenic, antiandrogenic, and anti-TH activities. It is thought that nuclear receptor signaling maintains development, cellular growth, inflammation, and metabolism. While ligand distribution within the nuclear receptor varies with few receptors found predominantly in the nucleus (pregnane X receptor and peroxisome proliferator-activated receptor gamma). In some cases, some receptors are in both the nucleus and cytoplasm (vitamin D receptor and mineralocorticoid receptor), while others mostly in the cytoplasm (glucocorticoid receptor and androgen receptor) [37]. Based on the presented results, all five compounds indicate possible inactivity to carcinogenesis.

Table 11, stress response pathways such as nrf2/ARE, HSE, MMP, p53, and ATAD5 were predicted for the five compounds. In theory, adaptive stress response pathways are signal transduction pathways that are activated in response to cellular stress. These pathways ultimately lead to the transcriptional activation of cytoprotective genes, which are genes that encode proteins that protect the cell [38-40]. The simulation revealed all compounds were inactive for each category. However, compound 3 displayed high probability of Mitochondria Membrane Potential activity.

Compound	1	2	3	4	4
Nuclear Factor (Erythroid-Derived 2-Like 2/Antioxidant Responsive Element) (nrf2/ARE) (%)	Inactive (93)	Inactive (97)	Inactive (80)	Inactive (97)	Inactive (93)
Heat Shock Factor Response Element (HSE) (%)	Inactive (93)	Inactive (97)	Inactive (80)	Inactive (97)	Inactive (93)
Mitochondrial Membrane Potential (MMP) (%)	Inactive (99)	Inactive (93)	Active (98)	Inactive (93)	Inactive (89)
Phosphoprotein (Tumor Suppressor) p53 (%)	Inactive (97)	Inactive (97)	Inactive (90)	Inactive (97)	Inactive (97)
ATPase Family AAA Domain Containing Protein 5 (ATAD5) (%)	Inactive (99)	Inactive (98)	Inactive (96)	Inactive (98)	Inactive (92)

Table 11. Toxicological Pathways: Stress Response Pathways Predicted Using ProTox II Web Server

Mitochondria are double membrane organelles that are essential for cellular energy production and apoptosis prevention [41]. In response to mitochondrial stress, yeast cells have evolved a retrograde response pathway that signals from the mitochondria to the nucleus to induce the transcription of nuclear encoded mitochondrial genes [42]. This pathway helps to alleviate mitochondrial stress and promote cell survival. It is further believed that mitochondrial stress caused by toxins can lead to a

variety of diseases [43]. Richter and colleagues have shown that toxins can inhibit mitochondrial protein synthesis and block the stress response pathway, further exacerbating mitochondrial dysfunction [44]. Interestingly, p53 (Phosphoprotein tumor suppressor) and ATAD5 (ATPase family AAA domain-containing protein 5) were observed to be inactive for all five compounds. The p53 gene is a tumor suppressor that plays a critical role in cell cycle regulation, DNA damage response, apoptosis, and other cellular processes. All five compounds displayed inactivity in the prediction imply the probability not being carcinogenesis. ATAD5 is another important regulator involved in DNA damage response. It plays a role in RAD9A related damage checkpoint, which determines whether DNA damage can be repaired or if cells should undergo apoptosis. Since all five compounds displayed inactivity, this may suggest possible DNA damage repair may occur as there were no stress response to ATAD5.

To evaluate the stability of the complexes formed from molecular docking and obtain other insights into the different interactions that make up its stability, a 100 ns molecular dynamic simulation was performed on compounds **1** and **2**. The results show that compound **1** reached equilibrium after 40 ns of simulation at a Root Mean Square Deviation (RMSD) value of around 0.25 nm while compound **2** equilibrated between 45 ns and 65 ns before fluctuations were noted (Figure 9). The results suggest that the complexes does not experience significant conformational changes when interacting with the ligand, suggesting the probability of positive interaction with inhibition. From the RMSD of the ligands, values were below 0.7 nm, which supports the reliability of the docking analysis. Lower RMSD values also suggest that the complex presented stability over time, further supporting inhibition possibility. Both compounds displayed almost similar RMSF, but compound **1** displayed slightly higher RMSF indicating possible greater flexibility in the binding sites. Compound **1** also displayed higher solvent accessible surface area compared to **2**. Signifying possible larger interface with solvent during the simulation and influencing the interactions.
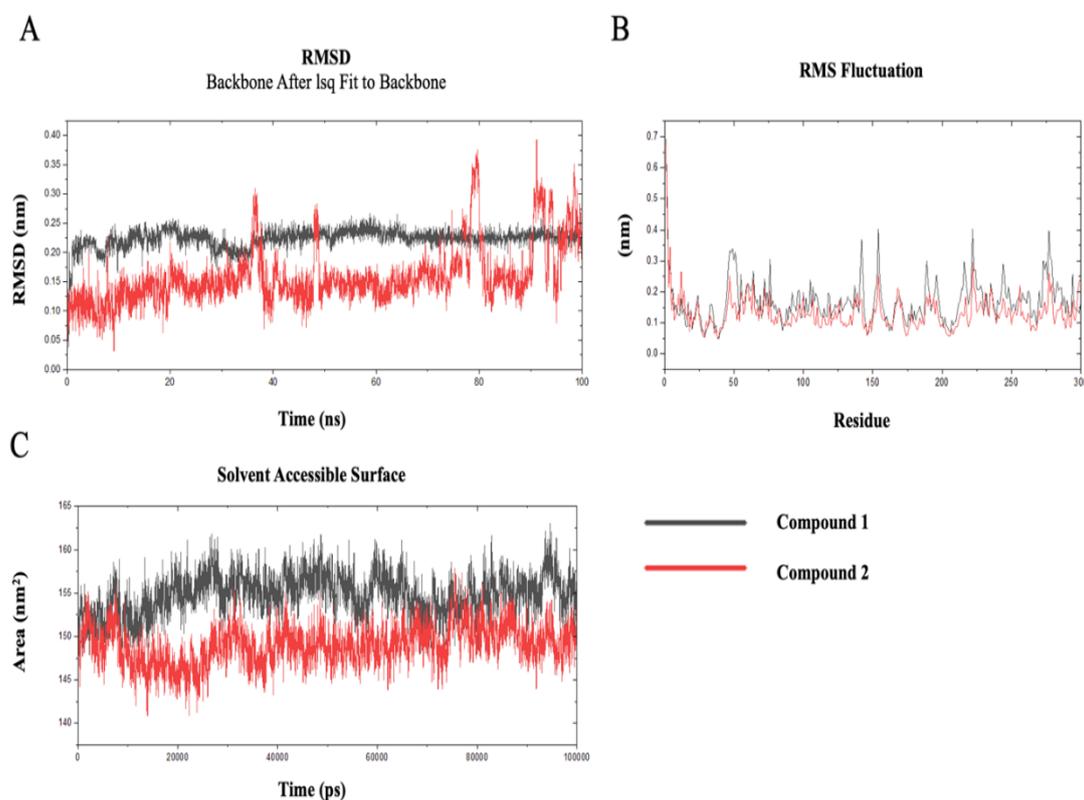

Figure 9. RMSD (A), RMSF (B), and Solvent Accessible Surface (C)
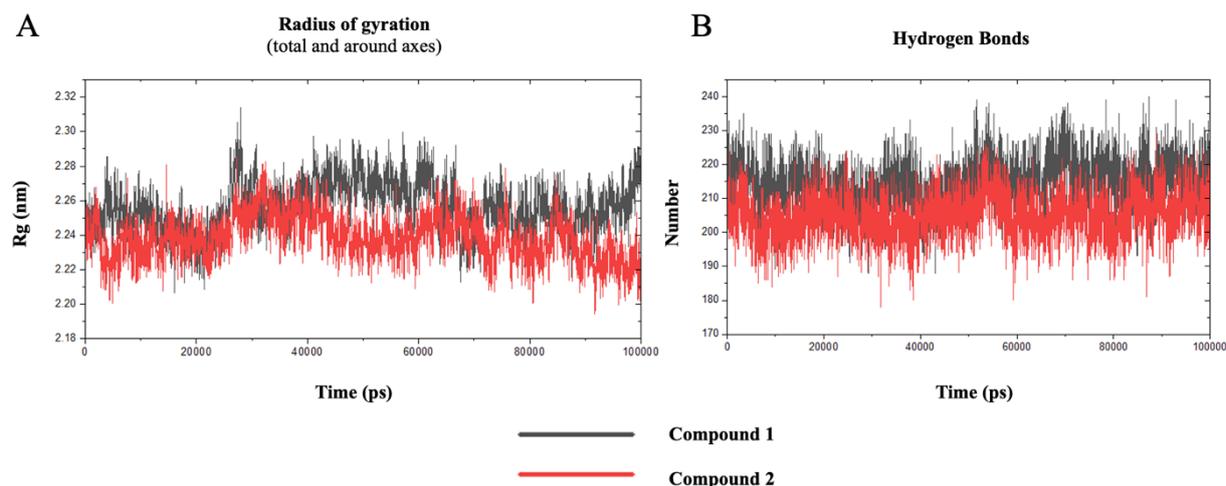
Figure 10. Radius of Gyration (A) and Hydrogen Bonds (B)

To observe the protein's compactness, the radius of gyration was analyzed. On average both simulations remained below the 2.28 Å threshold, suggesting the protein maintained its structural integrity throughout the simulation (Figure 10). Compound 1 displayed a minor increase above 2.30 Å around 30 nanoseconds implying a possible disruption of the protein's compactness. The number of hydrogen bonds formed remained relatively stable between 200 to 240 without any fluctuation. This suggests the complex interaction with the compounds upheld a steady interaction with the protein through the simulation.

Conclusion
Molecular docking and molecular dynamics simulations have been widely utilized in drug screening and drug design. In this study, we presented several interesting findings about SARS-CoV-2 M^{pro}. The compounds analyzed in this study would be interesting to study further as potential inhibitors of SARS-CoV-2 M^{pro}. Structural optimization and clinical trials should be further investigated after biological experiments. However, high-throughput molecular docking and molecular dynamic simulations has revealed the possibility that these compounds can form stable conformational structures with M^{pro} and potentially inhibit SARS-CoV-2. While at the same time, this study provides helpful insights for possible future studies.

Data Availability Statement
Publicly available datasets were used in this study. The NIH Clinical Collection can be found on the NIH website. The SARS-CoV-2 M^{pro} crystalline structure PDB: 8SXR can be found on the Protein Data Bank database.

Author Contributions
C.R.W. and M.O. contributed to the conception and study design. C.R.W., C.F., P.A., and A.A.Q.M. contributed to the collection, simulation, and analysis of the data. C.R.W., M.O., and C.F. wrote and revised the first draft of the manuscript. All authors contributed to the manuscript revision and approved the submitted version.

Funding

Conflict of Interest
The authors confirm and declare that the research was conducted without any commercial or financial relations that could be construed as potential conflict of interest.

References

1. Polatoğlu, I., Oncu-Oner, T., Dalman, I. and Ozdogan, S., 2023. COVID-19 in early 2023: Structure, replication mechanism, variants of SARS-CoV-2, diagnostic tests, and vaccine & drug development studies. *MedComm*, *4*(2), p.e228.

2. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L. and Chen, H.D., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, *579*(7798), pp.270-273.

3. Pal, M., Berhanu, G., Desalegn, C. and Kandi, V., 2020. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): an update. *Cureus*, *12*(3).

4. Sadeghi Dousari, A., Taati Moghadam, M. and Satarzadeh, N., 2020. COVID-19 (Coronavirus disease 2019): a new coronavirus disease. *Infection and drug resistance*, pp.2819-2828.

5. Drożdżal, S., Rosik, J., Lechowicz, K., Machaj, F., Szostak, B., Przybyciński, J., Lorzadeh, S., Kotfis, K., Ghavami, S. and Łos, M.J., 2021. An update on drugs with therapeutic potential for SARS-CoV-2 (COVID-19) treatment. *Drug Resistance Updates*, *59*, p.100794.

6. Wei, C.R., & Lang'at, G.C. (2023). In Silico Drug Discovery: The Next Frontier in the Fight Against SARS-CoV-2. Preprints. https://doi.org/10.20944/preprints202308.0887.v1

7. Zhou, Y.W., Xie, Y., Tang, L.S., Pu, D., Zhu, Y.J., Liu, J.Y. and Ma, X.L., 2021. Therapeutic targets and interventional strategies in COVID-19: mechanisms and clinical studies. *Signal transduction and targeted therapy*, *6*(1), p.317.

8. Moustaqil, M., Ollivier, E., Chiu, H.P., Van Tol, S., Rudolffi-Soto, P., Stevens, C., Bhumkar, A., Hunter, D.J., Freiberg, A.N., Jacques, D. and Lee, B., 2021. SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerging microbes & infections*, *10*(1), pp.178-195.

9. Jiang, H., Yang, P. and Zhang, J., 2022. Potential inhibitors targeting papain-like protease of SARS-CoV-2: two birds with one stone. *Frontiers in chemistry*, *10*, p.822785.

10. Yang, H. and Rao, Z., 2021. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature Reviews Microbiology*, *19*(11), pp.685-700.

11. Toussi, S.S., Hammond, J.L., Gerstenberger, B.S. and Anderson, A.S., 2023. Therapeutics for COVID-19. *Nature Microbiology*, *8*(5), pp.771-786.

12. Ashraf, F.B., Akter, S., Mumu, S.H., Islam, M.U. and Uddin, J., 2023. Bio-activity prediction of drug candidate compounds targeting SARS-Cov-2 using machine learning approaches. *Plos one*, *18*(9), p.e0288053.

13. Pérez-Vargas, Jimena et al. A novel class of broad-spectrum active-site-directed 3C-like protease inhibitors with nanomolar antiviral activity against highly immune-evasive SARS-CoV-2 Omicron subvariants." *Emerging microbes & infections* vol. 12,2 (2023): 2246594. doi:10.1080/ 22221751. 2023.2246594

14. Laskoswki, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. Main-chain bond lengths and bond angles in protein structures. *Journal of Molecular Biology*, 231(4), pp.1049-67.

15. Wiederstein, M.; Sippl, M.J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410.

16. Schrödinger Release 2023-3: Maestro, Schrödinger, LLC, New York, NY, USA

17. Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aid. Mol. Des.*, **2013**, *27(3)*, 221-234

18. Schrödinger **Release 2023-3**: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2023; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2023.

19. **Schrödinger Release 2023-3**: BioLuminate, Schrödinger, LLC, New York, NY, 2023.

20. Beard, H.; Cholleti, A.; Pearlman, D.; Sherman, W.; Loving, K.A., "Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes," *PLoS ONE*, **2013**, 8(12), e82849

21. Salam, N.K.; Adzhigirey, M.; Sherman, W.; Pearlman, D.A., "Structure-based approach to the prediction of disulfide bonds in proteins," *Protein Eng Des Sel*, **2014**, 27(10), 365-74

22. Zhu, K.; Day, T.; Warshaviak, D.; Murrett, C.; Friesner, R.; Pearlman, D., "Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction," *Proteins*, **2014**, 82(8), 1646-1655

23. Schrödinger Release 2023-3: LigPrep, Schrödinger, LLC, New York, NY, USA

24. Schrödinger Release 2023-3: Glide, Schrödinger, LLC, New York, NY, 2023.

25. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren,T. A.; Sanschagrin, P. C.; Mainz, D. T., "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes," *J. Med. Chem.*, **2006**, *49*, 6177–6196

26. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening," *J. Med. Chem.*, **2004**, *47*, 1750–1759

27. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S., "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, **2004**, *47*, 1739–1749

28. Xiong, Guoli et al. "ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties." *Nucleic acids research* vol. 49,W1 (2021): W5-W14. doi:10.1093/nar/gkab255

29. Priyanka Banerjee, Andreas O Eckert, Anna K Schrey, Robert Preissner, ProTox-II: a webserver for the prediction of toxicity of chemicals, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W257–W263

30. Malgorzata N. Drwal, Priyanka Banerjee, Mathias Dunkel, Martin R. Wettig, Robert Preissner, ProTox: a web server for the *in silico* prediction of rodent oral toxicity, *Nucleic Acids Research*, Volume 42, Issue W1, 1 July 2014, Pages W53–W58

31. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005 Dec;26(16):1701-18. doi: 10.1002/jcc.20291. PMID: 16211538.

32. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, G. Giambaşu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, Q. Zhu, and P.A. Kollman (2023), Amber 2023, University of California, San Francisco.

33. Sousa da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes. 2012 Jul 23;5:367. doi: 10.1186/1756-0500-5-367.

34. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. J Chem Theory Comput. 2021 Oct 12;17(10):6281-6291. doi: 10.1021/acs.jctc.1c00645.

35. Abohassan M, Alshahrani M, Alshahrani MY, Rajagopalan P. Insilco and Invitro approaches identify novel dual PI3K/AKT pathway inhibitors to control acute myeloid leukemia cell proliferations. Med Oncol. 2022 Oct 8;39(12):249. doi: 10.1007/s12032-022-01846-1.

36. [Colovos, C, and T O Yeates. "Verification of protein structures: patterns of nonbonded atomic interactions." *Protein science : a publication of the Protein Society* vol. 2,9 (1993): 1511-9. doi:10.1002/pro.5560020916

37. Kolodkin AN, Bruggeman FJ, Plant N, Moné MJ, Bakker BM, Campbell MJ, van Leeuwen JP, Carlberg C, Snoep JL, Westerhoff HV, Design principles of nuclear receptor signaling: how complex networking improves signal transduction. Mol Syst Biol. 6 (2010) 446 doi: 10.1038/msb.2010.102

38. Simmons SO, Fan C-Y, Ramabhadran R, Cellular stress response pathway system as a sentinel ensemble in toxicological screening. Toxicological Sciences 111(2) (2009) 202- 225

39. Kang KW, Lee SJ, Kim SG, Molecular mechanism of nrf2 activation by oxidative stress. Antioxid Redox Signal. 7 (2005) 1664-1673

40. Kensler TW, Wakabayashi N, Biswal S, Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu Rev Pharmacol Toxicol. 47 (2007) 89- 116

41. Hill S, Sataranatarajan K, Remmen HV, Role of signaling molecules in mitochondrial stress response. Front Genet. 9 (2018) 225 doi:10.3389/fgene.2018.00225

42. Parikh VS, Morgan MM, Scott R, Clements LS, Butow RA, The mitochondrial genotype can influence nuclear gene expression in yeast. Science 235(4788) (1987) 576- 580

43. Meyer JN, Hartman JH, Mello DF, Mitochondrial toxicity. Toxicological Sciences 162(1) (2018) 15-23

44. Richter U, Ng KY, Suomi F, Marttinen P, Turunen T, Jackson C, Suomalainen A, Vihinen H, Jokitalo E, Nyman TA, Isokallio MA, Stewart JB, Mancini C, Brusco A, Seneca S, Lombès A, Taylor RW, Battersby BJ, Mitochondrial stress response triggered by defects in protein synthesis quality control. Life Sci Alliance 2(1) (2019) e201800219 doi:10.26508/ lsa.201800219