Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated 3D-QSAR, Docking and Molecular Dynamics Reveal Novel Drug-Like Inhibitors of SARS-CoV-2 Main Protease

Soukaina Zahm

zahm.s@ucd.ac.ma

Chouaïb Doukkali University

Nabil Nor

Chouaïb Doukkali University

Soukaina El orchi

Chouaïb Doukkali University

Hassan Haddouchy

Moulay Ismail University

El Mostafa Erradi

ENSAM, Mohammed V University in Rabat

Abderrahmane El melouky

University Chouaib Doukkali

Ouassila Riouchi

Cadi Ayyad University

Nassima Riouchi

Mohammed Premier University

Elhouceine Benhadria

ENSAM, Mohammed V University in Rabat

Oussama Riouchi

Mohammed Premier University

Mohammed Salah

Chouaïb Doukkali University

Research Article

Keywords: SARS-CoV-2, Mpro, 3D-QSAR, docking, molecular dynamics, ADMET, Auto-QSAR

Posted Date: June 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9684307/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Chemical Papers on September 6th, 2026.
See the published version at https://doi.org/10.1007/s11696-026-05518-6.

Integrated 3D-QSAR, Docking and Molecular Dynamics Reveal Novel Drug-Like Inhibitors of SARS-CoV-2 Main Protease

Soukaina Zahm [1*], Nabil Nor [1], Soukaina El orchi [1], Hassan Haddouchy [2], El Mostafa Erradi [3a, 3b], Abderrahmane El melouky [4], Ouassila Riouchi [5], Nassima Riouchi [6], Elhouceine Benhadria [7], Oussama Riouchi [8], and Mohammed Salah [1]

[1] Molecular Modeling and Spectroscopy Research Team, Department of Chemistry, Faculty of Sciences, Chouaïb Doukkali University, El Jadida, Morocco. zahm.s@ucd.ac.ma; nor.n@ucd.ac.ma; elorchisoukaina2210@gmail.com; salah.m@ucd.ac.ma

[2] Physical Chemistry, Electrochemistry and Environment Team, Moulay Ismail University, Faculty of Sciences and Technology, B.P. 509, Boutalamine, Errachidia, Morocco. ha.haddouchy@gmail.com

[3a] Molecular Electrochemistry and Inorganic Materials, Sultan Moulay Slimane University, Faculty of Science and Technology, Beni Mellal, Morocco. elmostafa.erradi@usms.ma

[3b] Laboratory of Spectroscopy, Molecular Modeling, Materials, Nanomaterials, Water and Environment, Environmental Materials Research Team, ENSAM, Mohammed V University in Rabat, Morocco. elmostafa.erradi@usms.ma

[4] University Chouaib Doukkali, Laboratory Physics of Condensed Matter (LPCM), El Jadida, Morocco. elmlouky_abderrahmane@yahoo.fr

[5] Interdisciplinary Laboratory for Research in Environment, Management, Energy and Tourism (LIREMET), School of Technology Essaouira, Cadi Ayyad University, Essaouira, Morocco. Nassima.riouchi@ump.ac.ma

[6] LCM2E, Laboratory of Molecular Chemistry, Materials and Environment, Multidisciplinary Faculty of Nador (FPN), Mohammed Premier University, B.P. 300. 62700, Selouane, Nador, Morocco. nassima.riouchi@gmail.com

[7] LS3MN2E, CERNE2D, ENSAM, Mohamed V University in Rabat Morocco. Elhouceine.benhadria@um5r.ac.ma

[8] University Mohammed Premier, Faculty of Sciences, Laboratory of Environment and Applied Chemistry (LCAE), Oujda, Morocco. oussama.riouchi@ump.ac.ma

*Corresponding author: zahm.s@ucd.ac.ma

Abstract

Inhibition of the SARS-CoV-2 main protease (Mpro) remains a validated strategy for antiviral development. Here, we report an integrated in silico workflow combining Auto QSAR (KPLS), 3D QSAR (CoMFA/CoMSIA), reaction based enumeration, structure based docking, 100 ns molecular dynamics (MD), and ADMET/toxicity profiling to prioritize novel Mpro inhibitor candidates. A curated dataset of 80 aromatic ligands with experimentally determined Ki values (converted to pKi) was used to build predictive models. The top Auto QSAR model (kpls_linear_1) showed strong performance (R^2 = 0.8803, Q^2 = 0.8970, RMSE = 0.3083), while the CoMSIA model achieved robust fit with acceptable internal and external predictivity (R^2 = 0.9792; R^2CV = 0.6509; test Q^2 = 0.6618; RMSE = 0.47). Reaction based enumeration expanded accessible chemical space and yielded 12 designed ligands prioritized by predicted pKi and docking against Mpro (PDB: 7TLL). Docking protocol reliability was supported by redocking Nirmatrelvir (RMSD = 1.23 Å). Among designed candidates, C6 (predicted pKi = 8.9) achieved the best Glidescore (−7.851 kcal/mol) and formed stabilizing active site interactions. MD simulations indicated that C6 remained bound and maintained persistent polar contacts under explicit solvent conditions. In silico ADMET profiling suggested favorable oral absorption (HIA ≈ 91.6%) and acceptable drug likeness, while flagging liabilities that warrant follow up (e.g., predicted CYP3A4 and an ambiguous hERG signal). Overall, this reproducible computational pipeline identifies C6 as a promising lead for experimental validation and provides a transferable route for rapid antiviral lead discovery.

Keywords: SARS-CoV-2; Mpro; 3D-QSAR; docking; molecular dynamics; ADMET; Auto-QSAR.

1. Introduction

Coronaviruses are a diverse group of enveloped viruses characterized by a positive-sense single-stranded RNA genome. They belong to the family *Coronaviridae*, order *Nidovirales*, and realm *Riboviria*. These viruses are known to infect a variety of hosts, including mammals and birds, leading to respiratory, gastrointestinal, hepatic, and neurological diseases[1]. The coronavirus genome is notably large for an RNA virus, ranging from approximately 26 to 32 kilobases. This extensive genome encodes several structural proteins, including the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, which play crucial roles in virus entry, assembly, and replication [2]. Historically, coronaviruses have been associated with mild human illnesses, such as the common cold. However, the emergence of highly pathogenic strains like SARS-CoV in 2002, MERS-CoV in 2012, and SARS-CoV-2 in 2019 has underscored their potential to cause severe respiratory syndromes with significant morbidity and mortality [3]. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of the COVID-19 pandemic, has necessitated the rapid development of effective therapeutic interventions. Among the viral proteins essential for the virus's life cycle, the main protease (Mpro), also known as 3-chymotrypsin-like protease (3CLpro), plays a pivotal role in mediating viral replication and transcription. Due to its indispensable function and lack of a closely related homolog in humans, Mpro has emerged as an attractive and high-priority target for antiviral drug development [4].Mpro is a cysteine protease that cleaves the viral polyproteins pp1a and pp1ab at no fewer than 11 distinct sites, generating nonstructural proteins (nsps) necessary for the formation of the replication–transcription complex. Its recognition sequence is highly conserved, typically characterized by the pattern Leu-Gln↓(Ser, Ala, Gly), where the cleavage occurs at the site. This specificity is distinct from that of human proteases, thereby reducing the risk of off-target toxicity [5].Structurally, SARS-CoV-2 Mpro functions as a homodimer, with each monomer composed of three domains: domains I (residues 10–99) and II (residues 100–182) form an antiparallel β-barrel structure, while domain III (residues 201–303) is involved in dimerization through a cluster of five α-helices. The active site, situated in the cleft between domains I and II, contains a catalytic dyad composed of cysteine (Cys145) and histidine (His41) residues [6].Given its essential enzymatic role and structural distinctiveness, Mpro inhibition effectively halts viral replication, making it a centerpiece in antiviral drug discovery efforts. The successful development of Mpro inhibitors could thus lead to potent therapeutic agents with broad-spectrum activity against current and emerging coronaviruses. Several strategies, including

structure-based drug design, high-throughput virtual screening, and fragment-based approaches, have been employed to identify novel Mpro inhibitors [7].Recent progress, such as the development of Nirmatrelvir (PF-07321332), which is now part of the authorized antiviral therapy Paxlovid, further validates Mpro as a clinically relevant target [8].In silico approaches such as three-dimensional quantitative structure–activity relationship (3D-QSAR) modeling, molecular docking, molecular dynamics (MD) simulations, and ADMET predictions have become indispensable in accelerating the identification and optimization of Mpro inhibitors. These computational strategies not only provide insight into the key molecular interactions governing binding affinity but also predict pharmacokinetic and toxicity profiles, thereby enhancing the efficiency of lead optimization [9].Given the continuous emergence of SARS-CoV-2 variants, the ongoing search for novel and robust inhibitors targeting Mpro remains a strategic priority in global efforts to mitigate COVID-19 and future coronavirus-related threats.In this work, we present an integrated computational workflow that combines Auto-QSAR, CoMFA/CoMSIA-based 3D-QSAR, reaction-based chemical enumeration, molecular docking, molecular dynamics simulation, and ADMET profiling to identify novel inhibitors of SARS-CoV-2 Mpro. To our knowledge, few studies have unified all these complementary methodologies into a single coherent framework for coronavirus protease inhibitor discovery. The present study aims to generate robust predictive models, expand the chemical space around known inhibitors, and prioritize drug-like lead candidates with strong binding affinity, structural stability, and favorable pharmacokinetic profiles.

2. Material and methods

2.1. Dataset Collection and Preparation

A curated dataset of 80 aromatic ligands was retrieved from the BindingDB, based on experimental data obtained through FRET-based enzymatic assays. The assay employed recombinant Mpro (750 nM) and a fluorogenic peptide substrate, with fluorescence recorded at 365/445 nm. Recombinant Mpro (750 nM) was incubated with the test compounds in assay buffer composed of 20 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, and 2 mM dithiothreitol (DTT) for 10 min at room temperature. The enzymatic reaction was initiated by the addition of 25 μM fluorogenic substrate (MCA-AVLQSGFR-Lys(Dnp)-Lys-NH_2), and proteolytic activity was quantified by measuring fluorescence at 365/445 nm[10].Each compound was evaluated across nine serial concentrations, and IC50 values were determined by fitting the dose–response data to a four-parameter logistic model using GraphPad Prism.

All IC50 values represent the mean of three independent replicates, with DMSO-treated samples defining the upper and lower response limits. Reported IC50 values were subsequently converted to inhibition constants (K_i)using the Cheng–Prusoff equation to correct for substrate concentration and competitive binding kinetics. As an intrinsic measure of binding affinity, Ki reflects the equilibrium dissociation constant of the enzyme–inhibitor complex, with lower values indicating stronger inhibition.Only compounds with complete and unambiguous K_i data were retained for model development. For quantitative structure–activity relationship (QSAR) modeling, K_i values (nM) were transformed into their negative logarithmic form (pK_i).For classification, compounds with $K_i \leq$ 1190 nM were considered active, while those above this threshold were labeled inactive. A subset with $K_i \leq$ 7 nM was designated as highly active for further focused analysis[10].

Table 1 Classification of Compounds Based on Ki Ranges and Activity Labels

Class	**K_i Range (nM)**	**No. of Compounds**
Highly active	< 7	0
Active	7 – 1190	75
Inactive	> 1,190	5

These molecules are based on core scaffolds outlined in **Figure 1**, characterized by heterocyclic rings (X = O or S) and diverse R-group substitutions, including alkyl, aryl, heteroaryl, halogen, hydroxyl, cyano, amino, and carboxyl functionalities.Key structural features include:

- Five- or six-membered heterocycles (R6) bearing polar or lipophilic substituents.
- Linker groups at R3–R5 enabling flexible connectivity (e.g., alkylene, alkenylene, CO, NH, CONH).
- Variants across Formulas II–IV supporting broader structure–activity relationship (SAR) analysis.

•

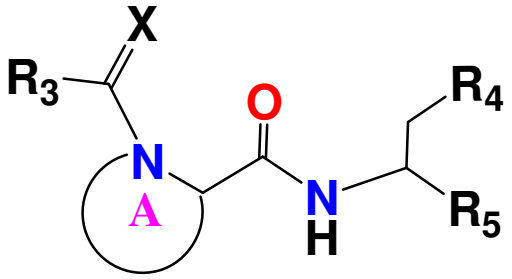


Figure 1 General chemical scaffold of SARS-CoV-2 Mproinhibitors, featuring a heterocyclic core (A) and variable substituents (R3–R5, X)

2.2. Rationale for MproTargetingand Protein Preparation

The SARS-CoV-2 main protease (Mpro) is a highly conserved viral enzyme that plays a pivotal role in viral replication by cleaving the viral polyproteins pp1a and pp1ab at specific recognition sites containing a P1 glutamine residue, a sequence motif absents in human proteases. This unique substrate specificity, together with the absence of close human homologs, significantly reduces the risk of host off-target effects and establishes Mpro as an attractive target for antiviral drug development[11].

The three-dimensional structure of SARS-CoV-2 Mpro was obtained from the Protein Data Bank and prepared for molecular modeling using the Protein Preparation Wizard implemented in Maestro (Schrödinger LLC). The protein structure was processed by adding missing hydrogen atoms, assigning correct bond orders, optimizing the protonation states of ionizable residues at physiological pH, correcting incomplete side chains and loops, and performing restrained energy minimization using the OPLS force field to remove steric clashes and optimize structural geometry[12].

2.3. Ligand Preparation

A set of 80 SARS-CoV-2 Mpro inhibitors were prepared using Schrödinger's LigPrep workflow to ensure high quality, lowenergy 3D structures suitable for subsequent docking and property prediction. LigPrep converts input 2D or 3D structures into minimized 3D conformers, systematically sampling protonation states, stereoisomers, tautomers, and ring conformations. We set a maximum ligand size of 200 atoms to exclude overly large molecules, and employed the OPLS4 force field for all geometry optimizations. Ionization states were generated for pH 7.0 ± 2.0 using the default "Ionizer" algorithm, adding or

removing protons to reflect physiological conditions. Counter-ions and extraneous waters were removed by the “Desalt” option, retaining only the largest molecular species in each file[13].

2.4. Molecular Descriptor Calculation

A total of 277 molecular descriptors were calculated for the 80 SARS-CoV-2 Mpro inhibitors to capture their structural, physicochemical, and electronic characteristics. Topological descriptors, including Wiener, Zagreb, and Hosoya indices, were computed using graph-theoretical definitions. Physicochemical parameters relevant to ADMET profilingsuchas predicted aqueous solubility (log S), partition coefficient, Caco-2 permeability, blood–brain barrier penetration (log BB), and human oral absorptionwere generated via Schrödinger’s QikProp tool. Quantum-chemical descriptors were also calculated such as dipole moment, HOMO-LUMO energies, and electrostatic potential. The combined descriptor matrix provided a comprehensive profile of each ligand and served as input for feature selection and QSAR modeling. Prior to model development, all descriptors were inspected for missing values, constant or near-constant variables, and excessive intercorrelation to minimize noise and reduce the risk of overfitting.

2.5. Model building

Auto-QSAR employing Kernel Partial Least Squares (KPLS) was utilized to achieve robust predictive accuracy through descriptor-based machine learning approaches. Additionally, three-dimensional QSAR (3D-QSAR) analyses using Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) were conducted to provide spatial and mechanistic insights into ligand–Mpro interactions. Field-based 3D-QSAR models were constructed using CoMFA and CoMSIA methodologies [14]. Eighty structurally aligned SARS-CoV-2 main protease (Mpro) inhibitors were positioned within a uniformly spaced three-dimensional grid. In CoMFA, steric and electrostatic interaction fields were computed, whereas CoMSIA extended this evaluation by incorporating hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor fields through Gaussian-type functions. To prevent singularities, grid points within 2 Å of ligand atoms were excluded, and all field values were auto scaled to normalize variance across different field types.Partial Least Squares (PLS) regression was used to correlate field data with experimental pK_ivalues, employing up to 10 components to prevent overfitting. In the Auto-QSAR workflow, a 70/30 training–test split was used. To obtain representative subsets, K-means clustering was first applied to the molecular descriptor matrix, grouping compounds according to structural and physicochemical similarity. Samples from each cluster were then

proportionally distributed between the training and test sets and ensuring balanced chemical diversity.

The applicability domain of the developed QSAR models was assessed using the leverage approach and visualized with Williams plots. In this method, the leverage value (h) for each compound was calculated to quantify its influence on the model, while standardized residuals were used to evaluate prediction reliability. Compounds with leverage values exceeding the critical threshold (h) or with standardized residuals outside ±3 standard deviations were considered outliers. To design new ligands based on the QSAR outcomes, an "Automated Reaction-Based Enumeration (ARBE)" strategy was used to expand the chemical space around known inhibitors. Pre-defined SMARTS rules were applied to generate synthetically accessible analogs based on existing lead scaffolds[15]. These virtual products were then filtered based on ADMET criteria, structural alerts, and medicinal chemistry rules. The top-ranked compounds were prioritized using predictions from the CoMFA/CoMSIA models and further evaluated by molecular docking against Mpro.

2.6 Docking, Molecular Dynamics and ADMET:

Molecular docking studies were conducted using the Glide module within Schrödinger's Maestro software suite (release 2024-1) [16]. Standard Precision (SP) docking mode was employed to assess the binding affinity of the designed SARS-CoV-2 main protease (Mpro) inhibitors against its crystal structure (PDB ID: 7TLL). The active site was identified using the SiteMap tool, which ranked the top five potential binding pockets based on site score and draggability parameters. The docking grid was subsequently centered on the most suitable site, encompassing key catalytic and substrate-recognition residues.

To ensure reliability of the docking protocol, validation was performed through redocking of the co-crystallized ligand, Nirmatrelvir, into the 7TLL structure. The resulting root-mean-square deviation (RMSD) between the redocked and crystallographic ligand conformations was calculated and confirmed to be below 2 Å, indicating high accuracy of the docking setup. Post-docking analyses included GlideScore evaluation, hydrogen bond mapping, hydrophobic interaction profiling, and visualization of ligand–protein contacts to elucidate binding modes and interaction patterns.

Molecular dynamics (MD) simulations were performed for the top-ranked ligand–protein complex (ligand C6 bound to SARS-CoV-2 Mpro) using GROMACS (version 2021.4) on a high-performance computing system (FUJITSU CELSIUS R970, Intel Xeon Silver 4210R CPU @ 2.40 GHz × 40 cores, accelerated by NVIDIA GPU). The simulation was conducted under explicit solvent conditions for 100 ns using the TIP3P water model and

the CHARMM36 force field, selected for its optimized performance in protein–ligand systems. Protein topology was generated using pdb2gmx, while ligand topology was prepared via SwissParam[17]. The complex was solvated in a triclinic unit cell, and NaCl ions were added to achieve a physiological concentration of 0.15 M, ensuring system neutrality. Equilibration was performed in two phases: 100 ps under NVT ensemble (canonical conditions) followed by 100 ps under NPT ensemble (isothermal–isobaric conditions) at 300 K and 1 bar, employing the Berendsen thermostat and Parrinello–Rahman barostat. The production MD simulation was executed for 100 ns with a 2fs time step, and trajectory frames were saved every 5 ps for subsequent analysis. Post-simulation analyses includedRoot-mean-square deviation (RMSD), Root-mean-square fluctuation (RMSF) to evaluate residue flexibility, Radius of gyration (Rg) for monitoring protein compactness and Hydrogen bond profiling.In silico ADMET profiling was performed using the PreADMET web platform [18] to comprehensively evaluate drug-likeness, pharmacokinetic behavior, and toxicity risks of the designed SARS-CoV-2 Mpro inhibitors. The analysis included assessment of Lipinski's rule of five for oral bioavailability, Caco-2 cell permeability as an indicator of intestinal absorption, and blood–brain barrier (BBB) penetration to predict central nervous system exposure. Additionally, potential cytochrome P450 (CYP450) enzyme interactions were examined to identify metabolic liabilities, and mutagenicity tests (Ames assay simulation) were conducted to estimate genotoxic risk. Further parameters such as human intestinal absorption (HIA), plasma protein binding (PPB), and skin permeability were also evaluated to provide a holistic pharmacokinetic profile.

3. Result and discussion:

3.1 Auto QSAR modelling: KPLS model

An Auto-QSAR workflow was applied to the full set of 80 SARS-CoV-2 Mpro inhibitors to generate predictive regression models based on kernel-based partial least squares (KPLS) analysis[19]. The dataset was partitioned into a training set (70%, 56 compounds) and a test set (30%, 24 compounds). Among seven generated models, **kpls_linear_1** demonstrated superior performance, with an R^2 of 0.8803 and a Q^2 of 0.8970, coupled with a low RMSE of 0.3083 (**Table 2**). These metrics indicate high internal consistency and strong predictivity. The correlation between experimental and predicted pK_i values for both the training and test sets is illustrated in **Figure 2**, with predictions closely following the ideal regression line (y = x). Model selection was further supported by cross-validation and comparison against the null hypothesis, with Q^2 MW values remaining acceptably low. The observed versus predicted activity using KPLS model are listed in **Table 2S** (supplementary Data)

Table 2 Performance metrics of the top-ranked Auto-QSAR models. The highlighted model (kpls_linear_1) was selected for further prediction based on its superior R^2, Q^2, and RMSE.

Model Code	Score	S.D.	R^2	RMSE	Q^2	Q^2 MW (Null Hypothesis)
Kpls_linear_1	0.8757	0.2999	0.8803	0.3083	0.8970	-0.1180
Kpls_molprint2D_47	0.8676	0.3551	0.8635	0.2676	0.8674	-0.1402
Kpls_molprint2D_8	0.8573	0.3620	0.8712	0.3256	0.6469	-0.2617
Kpls_molprint2D_46	0.8427	0.3591	0.8622	0.3397	0.7712	-0.0240
Kpls_linear_8	0.8257	0.4141	0.8283	0.3670	0.5514	-0.2617
Kpls_linear_49	0.8247	0.2828	0.8951	0.3333	0.8774	0.0221
Kpls_dendritic_49	0.7995	0.3056	0.8775	0.3582	0.8584	0.0221

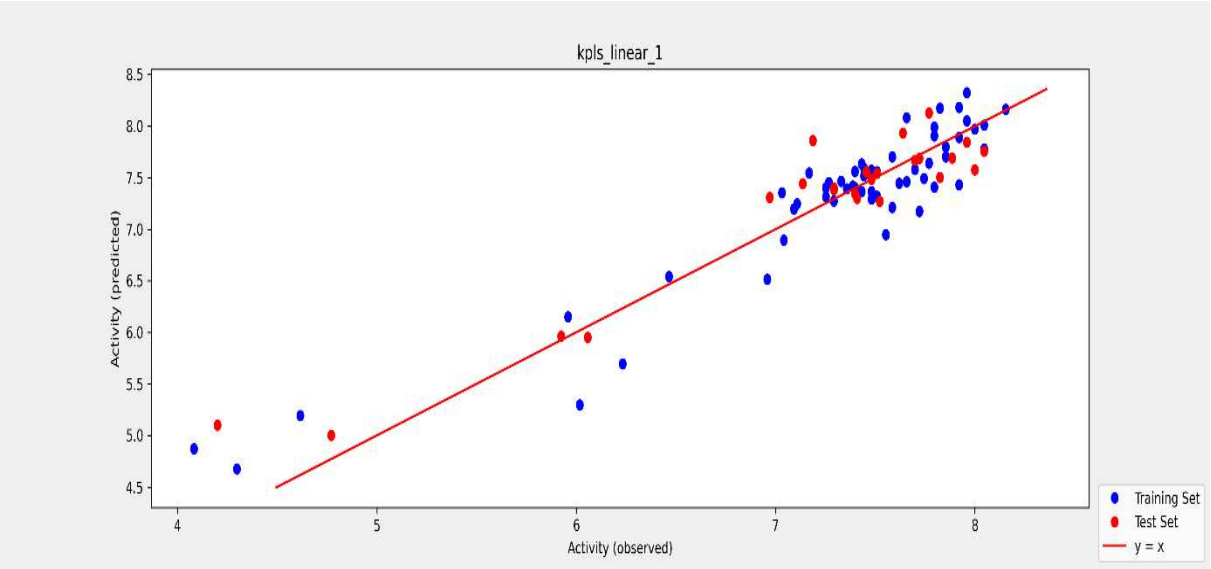


Figure 2 Correlation plot between observed and predicted pKi values using the kpls_linear_1 model. Blue dots represent training set compounds; red dots indicate the test set. The red line denotes the ideal correlation (y = x).

The applicability domain of the Auto-QSAR (KPLS) model was examined using a combined structural–response analysis. In the Williams plot (standardized residuals vs. leverage), the vast majority of compounds lie well within the ±3 residual bounds and to the left of the warning leverage (h*), indicating predictions made by interpolation inside the training descriptor space with acceptable error dispersion**(Figure 3)**. Two compounds are flagged as outside the AD: one (index 78) is a response outlier with a standardized residual close to −3 despite leverage near, but below, h*, and the other (index 73) is a structural outlier with high leverage (h > h*) but moderate residuals.

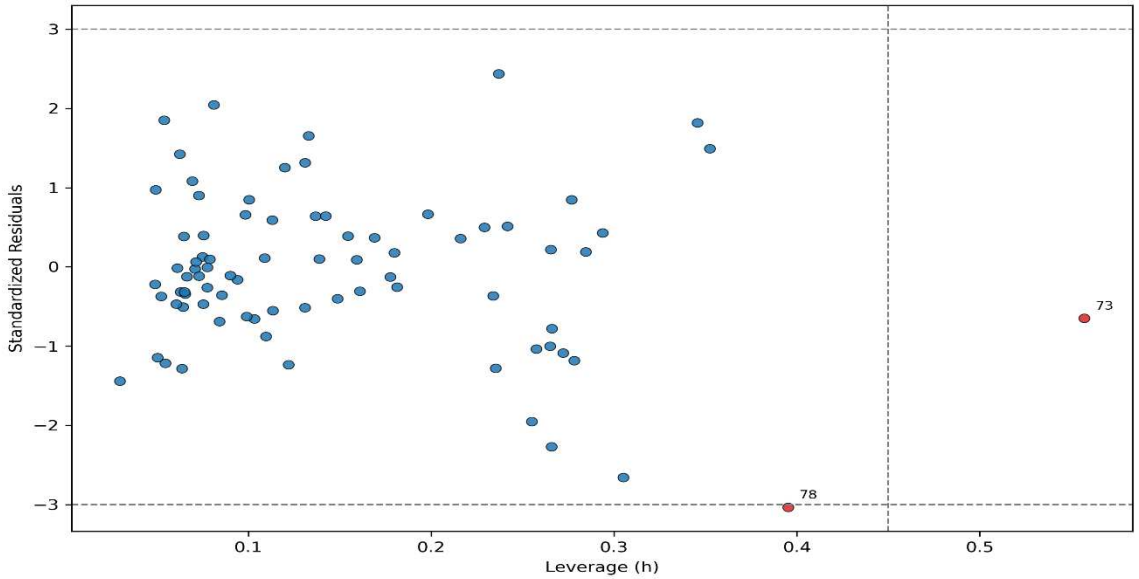


Figure 3 Williams Plot (StandardizedResiduals vs. Leverage, h) with ±3 ResidualBounds and Warning Leverage (h*)

Consistently, the Mahalanobis distance plot confirms these observations: both flagged compounds exhibitMD values above the empirical 97.5 thpercentile threshold, which confirm that they are chemically distant from the training domain (**Figure 4**). Taken together, these diagnostics show that the model is reliable across the dense core of the chemical space, where predictions are simultaneously low leverage and low error. However, ligands 73 and 78) carry higher uncertainty and should be either experimentally verified.

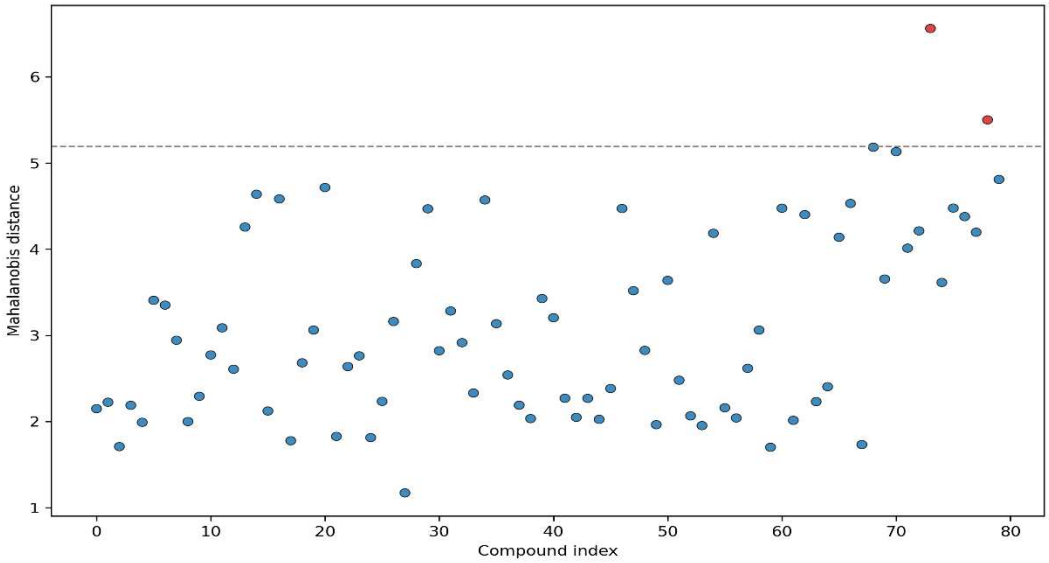


Figure 4 Mahalanobis Distance in Descriptor Space with Empirical 97.5 thPercentile Threshold

3.2QSAR based field:

To ensure balanced representation of activity and chemical space, the 80 structurally aligned ligands were partitioned into a training set (70%, 56 compounds) and a test set (30%, 24 compounds) using hierarchical clustering. This method preserved chemical diversity across both subsets, enabling reliable model development and external validation. CoMFA and CoMSIA models were built on the training set using up to 10 partial least squares (PLS) components. The number of components was optimized to maximize predictive performance while minimizing the risk of overfitting for the model. Both models captured relevant steric, electrostatic, and physicochemical contributions governing SARS-CoV-2 Mpro inhibition.

The CoMSIA model constructed on the 56-compound training set demonstrated a strong fit (factor 8 was chosen for best performance), with an R^2 of 0.9792, indicating that 97% of the variance in experimental pKi values was explained by the model. The root mean square error (SD) was 0.1433, reflecting good predictive agreement the training data. The cross-validated R^2 (R^2_CV) was limited to 0.627, suggesting acceptable internal predictive power. The scrambling validation result (R^2_Scramble = 0.7808) low than R^2 of 0.9792confirmed that the model is not due to chance (**Table 3**).

Table 3 Training Set QSAR Statistics (56 compounds)

Statistic	Value	Comment
# Factors	10	Number of PLS components
SD	0.1433	RMS error of fitted pK_i (n–m–1 DOF)
R^2	0.9792	Fraction of variance in observed activity explained
R^2_CV	0.6509	Leave-N-out cross-validated R^2
R^2_Scramble	0.7808	Average R^2 from models with randomized activities
Stability	0.18	Robustness to changes in training set (lower than R^2 suggests over-fitting)
F	276.6	Variance ratio (model vs. observed activity)
P (F-ratio)	7.7e-37	Significance level of F (highly significant)

External validation of the CoMSIA model was performed using the independent test set comprising 24 compounds, with results summarized in **Table 4**. The model achieved a root mean square error (RMSE) of 0.47, reflecting an average prediction error that is within the expected range for 3D-QSAR models. The predictive coefficient of determination (Q^2) reached 0.66, indicating that 66% of the variance in experimental test set activities was captured by the model. According to Tropsha and his collaborators, this level of predictive power is considered acceptable for virtual screening and prioritization of candidate inhibitors in early-stage drug discovery[20]. Furthermore, the Pearson correlation coefficient (r = 0.84) demonstrates a strong linear association between predicted and observed pKi values, supporting the model's reliability for ranking compound potency.

Table 4 Test Set QSAR Statistics (24 compounds)

Statistic	Value	Comment
RMSE	0.47	RMS error of predicted vs. observed pKi
Q^2	0.6618	Predictive R^2 on the test set
Pearson-r	0.8424	Correlation between predicted and observed pKi

Figures 5 and **6** provide graphical representations of the model's performance for the training and test sets, respectively. In **Figure 5**, the experimental and predicted pKi values for the training set are tightly clustered along the diagonal, reflecting the strong fit and high power observed during model development. **Figure 4**illustrates the test set predictions, where a broader scatter around the diagonal is evident, consistent with the RMSE and Q^2 values.

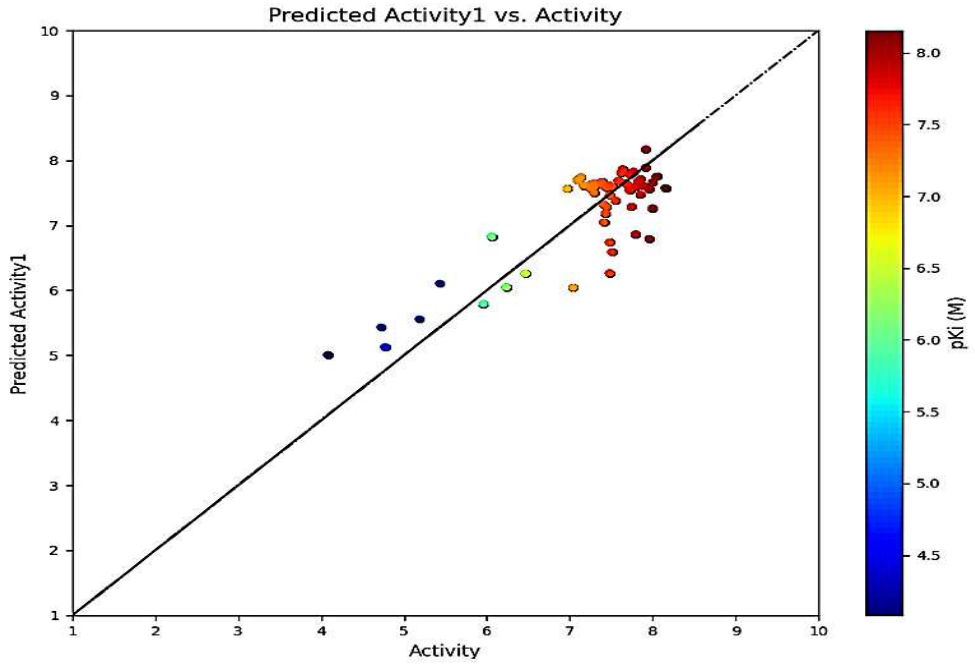


Figure 5 Correlation plot of predicted versus experimental pKi values for the training set.

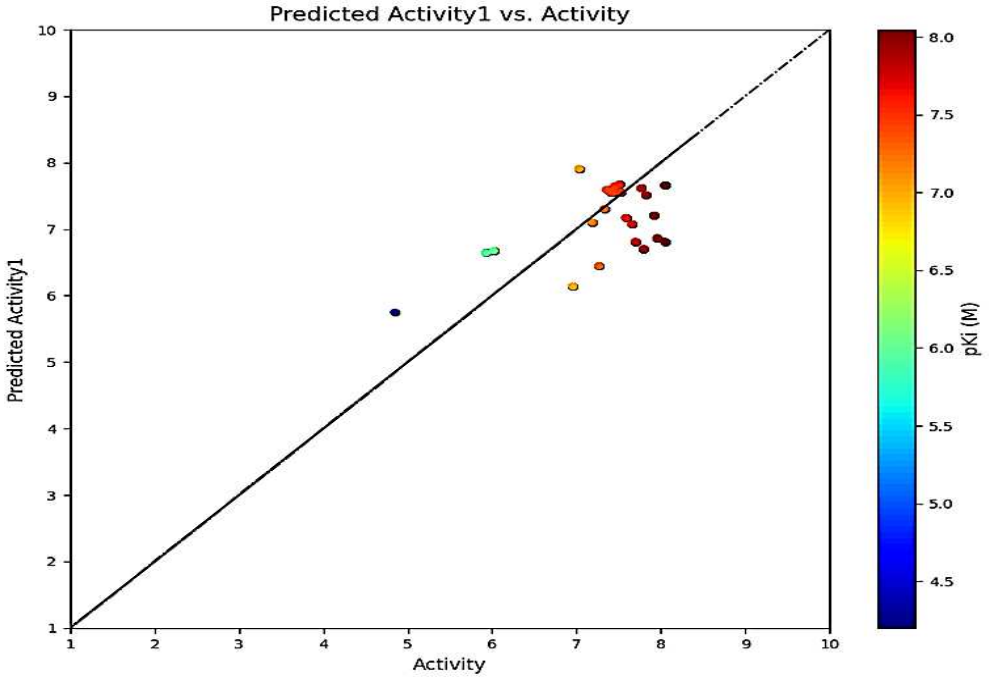


Figure 6 Correlation plot of predicted versus experimental pKi values for the test set.

3.2.1 Field Contribution Analysis

The CoMSIA model's field contribution analysis revealed that steric effects were the most influential (41%), followed by hydrophobic (24%), H-bond acceptor (13%), H-bond donor (12%), and electrostatic (8%) components (**Figure 7**). These results suggest that molecular

shape, lipophilicity, and the ability to accept hydrogen bonds are critical determinants of inhibitory activity against SARS-CoV-2 Mpro.

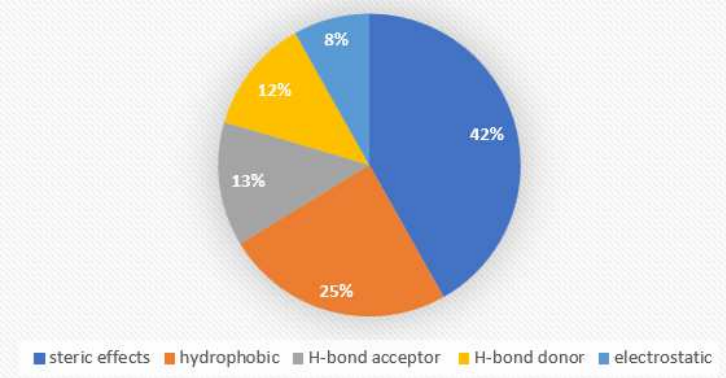
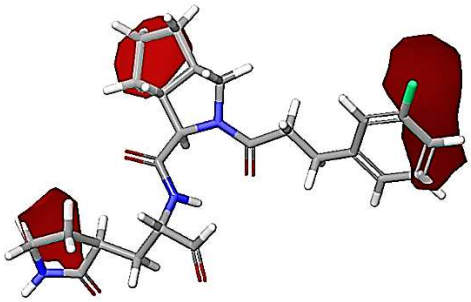


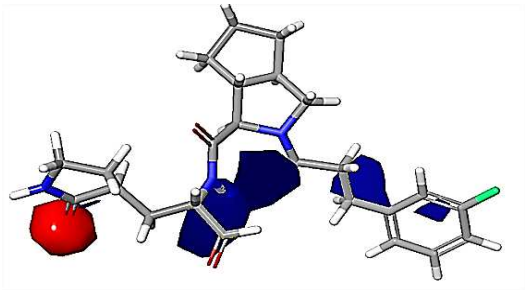
Figure 7 Pie chart representing the fractional contributions of CoMSIA fields in the final QSAR model.

Five individual Gaussian field maps were generated for the most active compound to visualize these spatial contributions (**Figure 6**).

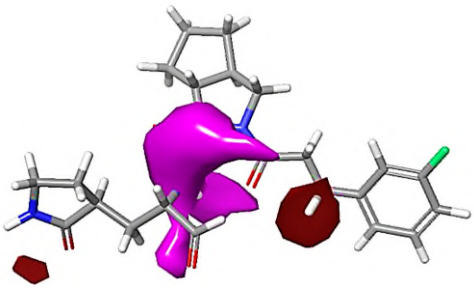
(a) Gaussian steric



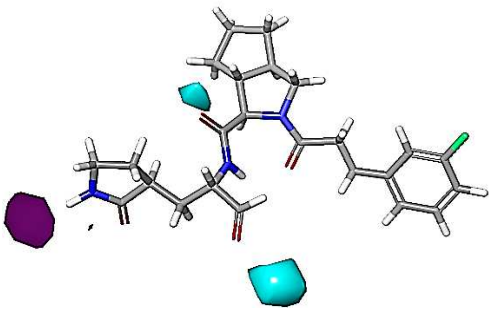
(b) Gaussian electrostatic



(c) Gaussian H-bond acceptor



(d) Gaussian H-bond donor



(e) Gaussian hydrophobic

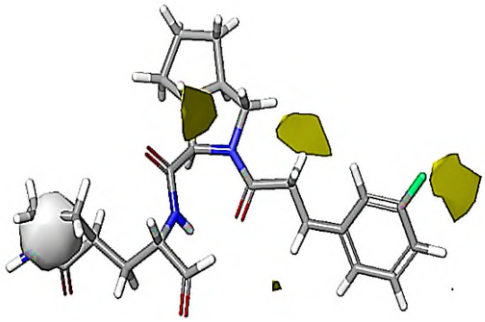


Figure 8 CoMSIA field contour maps around the most active SARS-CoV-2 Mpro inhibitor, color-coded by field contribution.

Overall, the CoMSIA model exhibited an excellent fit to the training data, as reflected by the high R^2 and F-statistic, confirming that the selected Gaussian fields effectively capture the underlying structure–activity relationships.

3.3. Model Prediction and Design of New Compounds

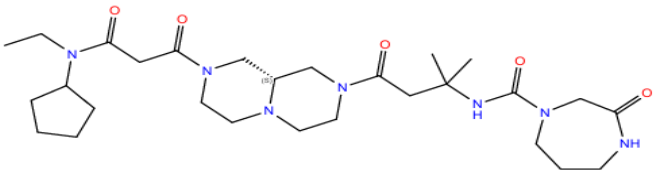
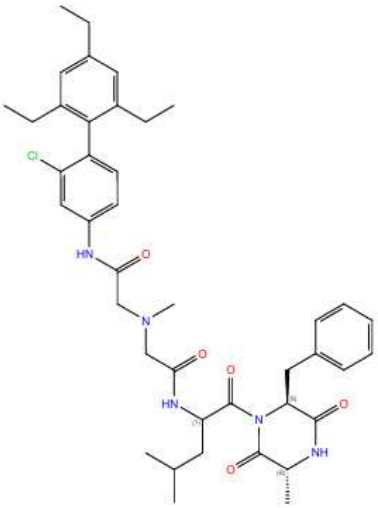
To expand the chemical space beyond known SARS-CoV-2 Mpro inhibitors, we employed an Automated Reaction-Based Enumeration strategy to generate novel drug-like analogs with potentially enhanced inhibitory activity[21]. Reference ligands were selected from the curated dataset of 80 compounds, serving as templates for reaction-based transformations. A total of 50 synthetic pathways were identified and applied to the reference scaffolds using predefined SMARTS rules and medicinal chemistry-relevant reaction schemes.

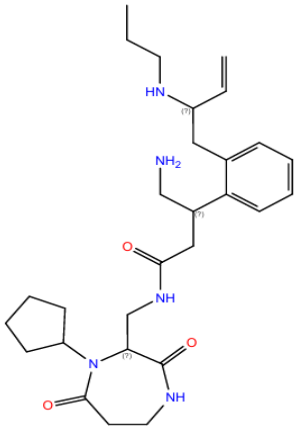
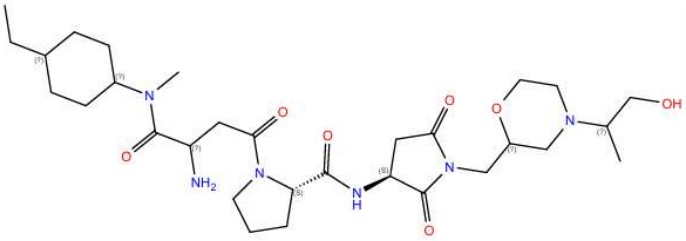
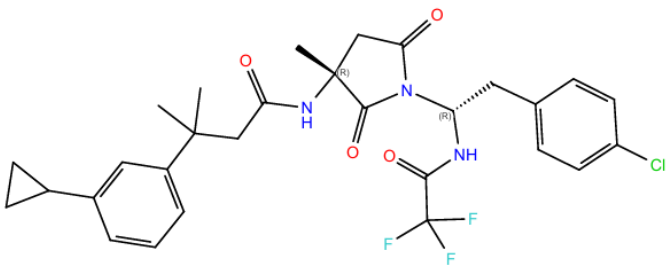
This enumeration generated 100 analogs, which were then filtered according to the predicted activities from the best-performing QSAR models (CoMFA and CoMSIA). From this set, 12 ligands exhibiting favorable drug-like properties were selected for further evaluation.

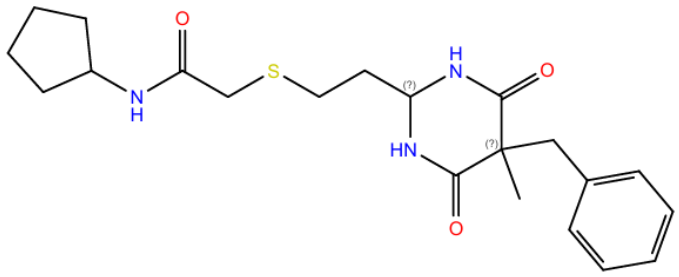
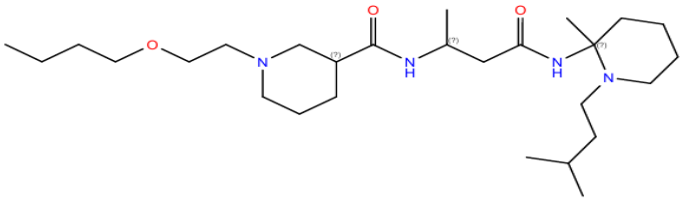
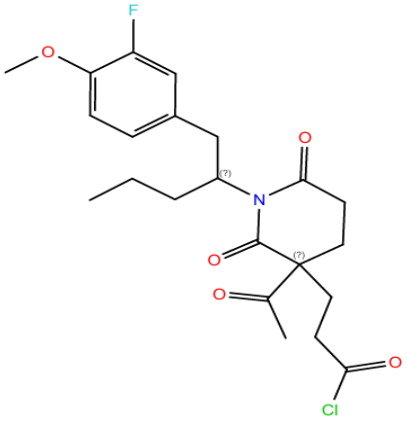
Activity predictions for the 12 newly designed ligands were performed using the validated CoMFA and CoMSIA models. The predicted pK_ivalues ranged from 8.5 to 8.8, indicating a potentially enhanced inhibitory effect compared to most compounds in the original training set and the 88 additional analogs generated by enumeration. Among the 12 designed ligands, activity prediction using the CoMFA and CoMSIA models identified C10, C8, C7, and C1 with high pKi values of 8.8. Notably, compound C6 exhibited the highest predicted activity (pKi = 9.0). To further evaluate their binding interactions and prioritize potential leads, all 12 ligands were subjected to molecular docking studies, providing structural insights into their compatibility with the protease active site.

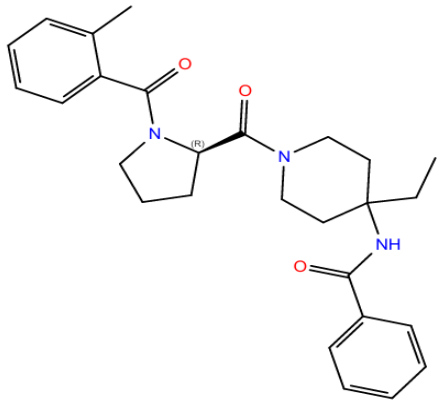
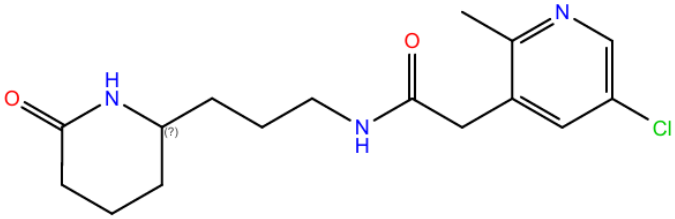
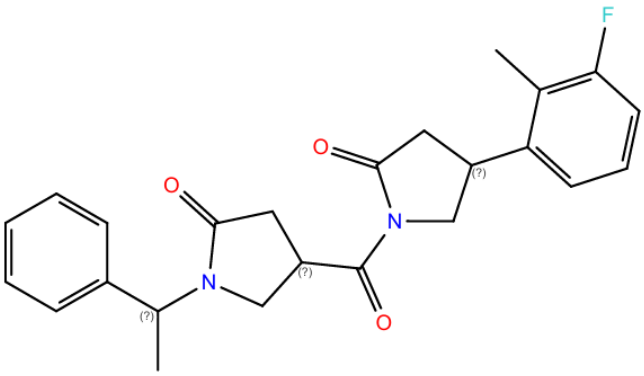
Table 5 summarizes the designed ligands, their synthetic pathways, and predicted pKi values. These virtual hits represent promising candidates for further optimization and biological validation.

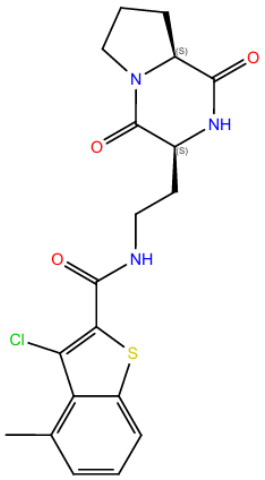
Table 5 Designed Ligands with Their Synthetic Pathways and Predicted pKi

ID designed ligand	**Reference reaction Pathway**	**Structure**	**Predicted pKi (M)**
C1	38855013 + 315355503 + 323970322 [1/1]		8.8
C2	301254474 + 313656976 + 37451093 + 38415995 [1/2]		8.6

C3	41349971 + 329968599 + 48579639 + 41479707 [3/3]		8.6
C4	314906524 + 39735836 + 107100887 [1/1]		8.7
C5	313848266 + 31261161 + 331233085 [1/2]		8.6

C6	5853600 37936209 [2/3]	+		8.9
C7	38341062 313627114 313564590 [1/1]	+ +		8.8
C8	313842787 38088428 [1/1]	+		8.8

C9	480478 261549818 [3/3]	+		8.7
C10	53991418 313481328 [2/2]	+		8.8
C11	29817551 313856690 [1/1]	+		8.6

C12	29818689 + 4798284 [2/6]		8.6

3.4. Molecular Docking Studies

Molecular docking studies were performed to further evaluate the binding affinity and interaction profiles of the 12 designed ligands against the SARS-CoV-2 main protease (Mpro), using the crystal structure 7TLL as the receptor model. The docking protocol employed Nirmatrelvir (PF-07321332), the active component of the approved drug Paxlovidas a reference ligand. The docking protocol was validated by redocking the co-crystallized ligand Nirmatrelvir into the active site of the target protein (PDB ID: 7TLL). The accuracy of the docking procedure was evaluated by calculating the root-mean-square deviation (RMSD) between the crystallographic pose and the redocked conformation using the Discovery Studio Visualizer V21. An RMSD value of 1.23 Å was obtained, indicating acceptable reliability of the docking protocol.

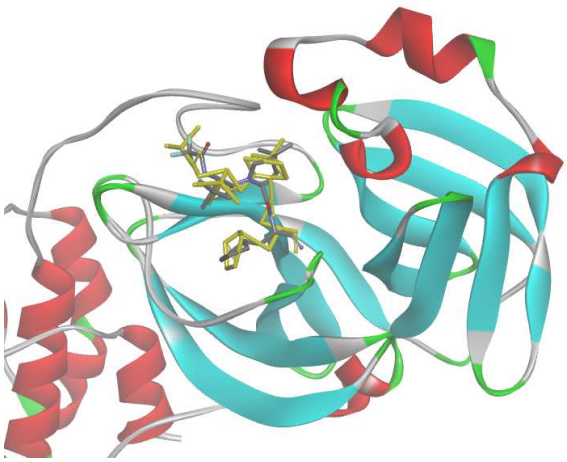
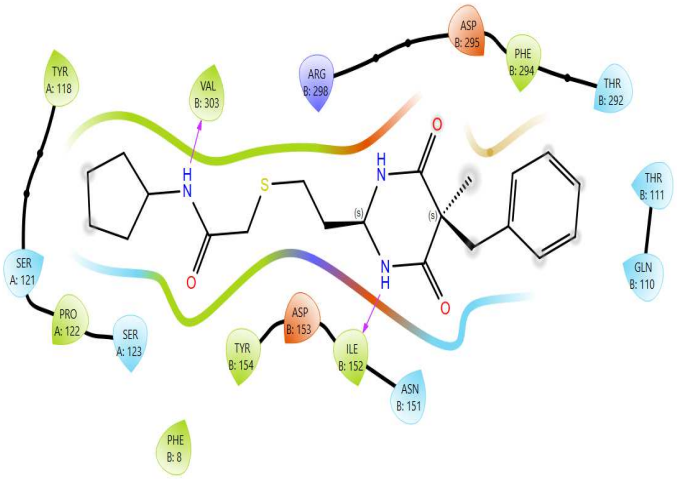
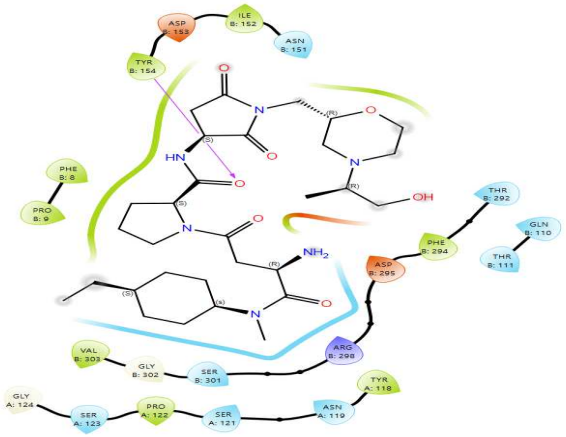
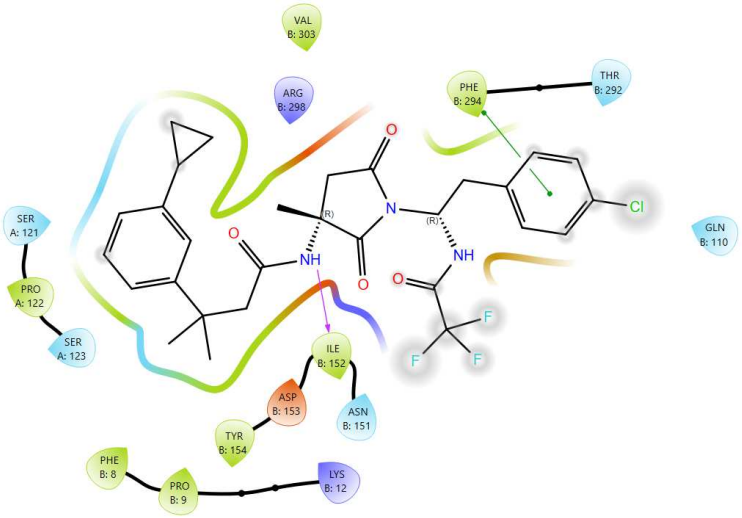
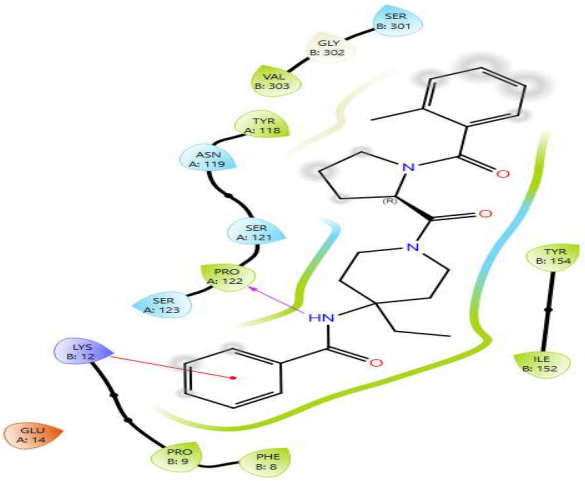
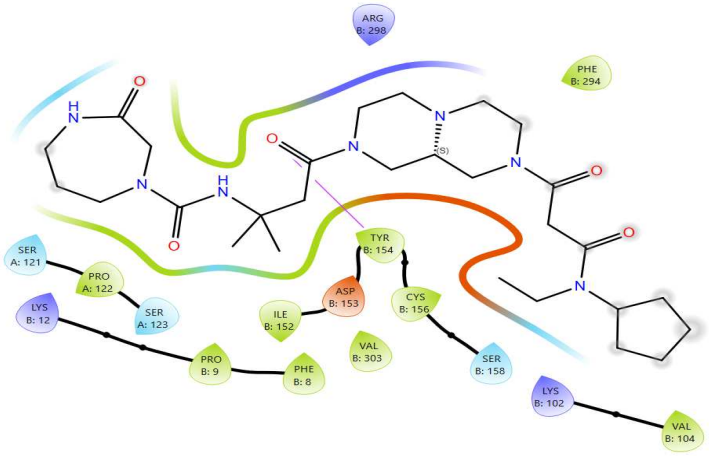
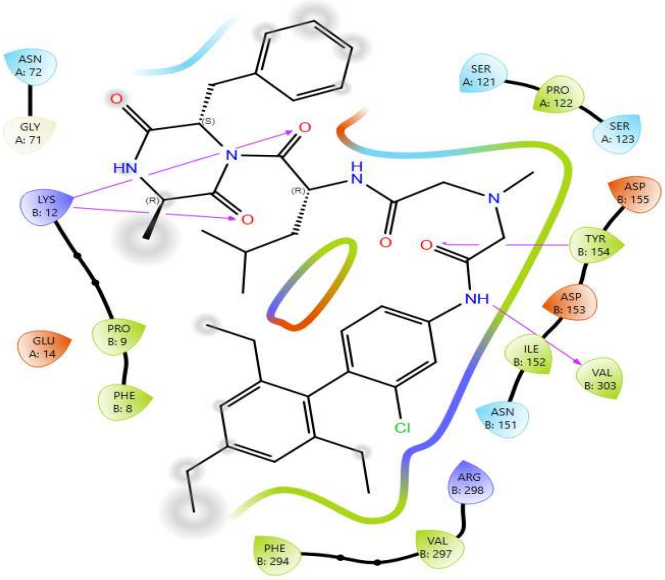


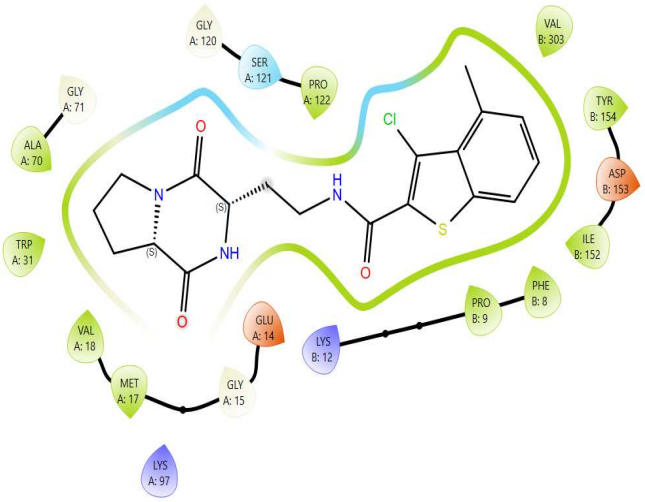
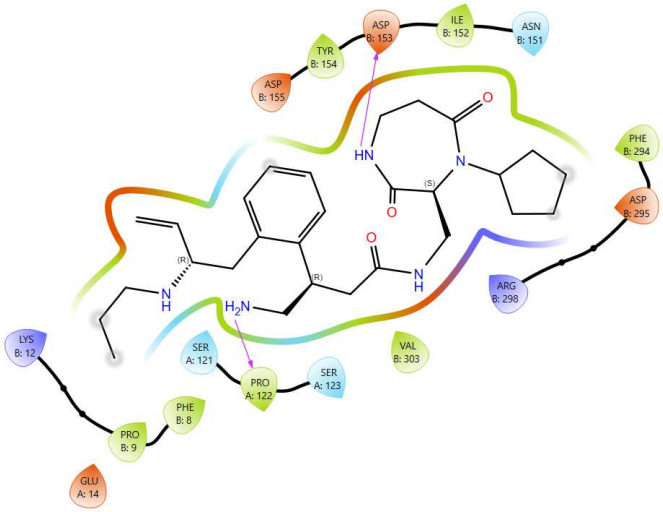
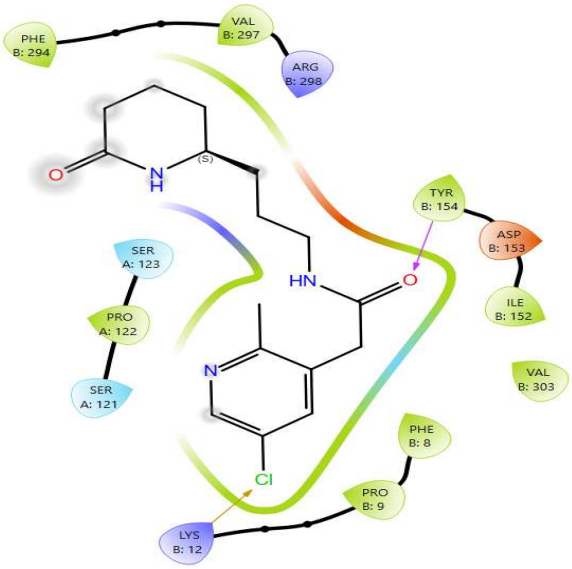
Figure 9 Re Docking of the co-crystallized Ligand Nirmatrelvir

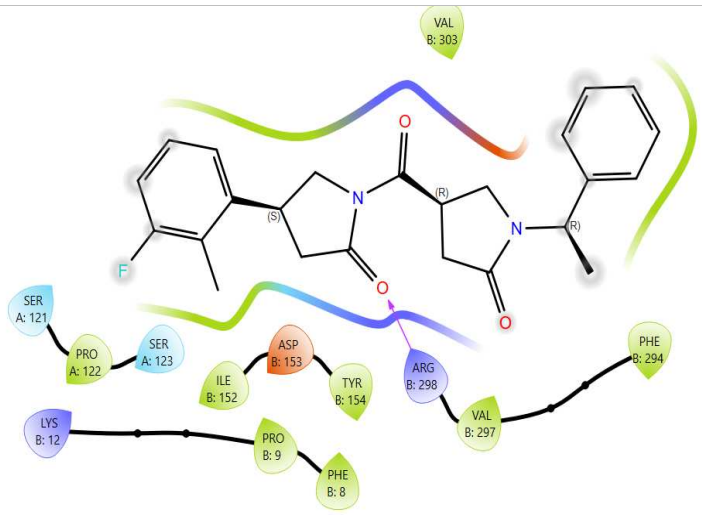
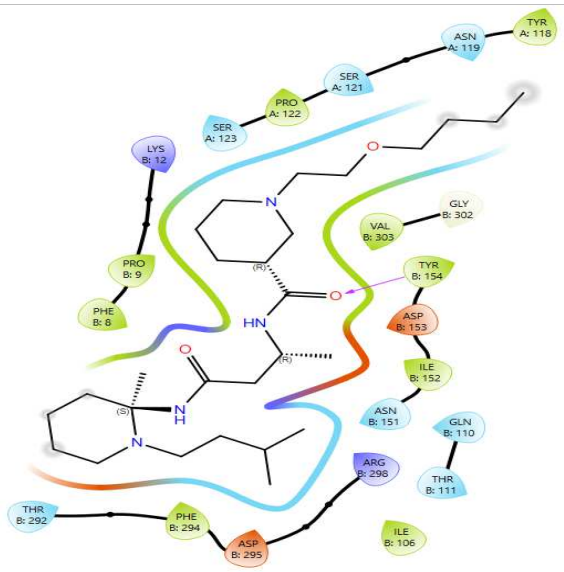
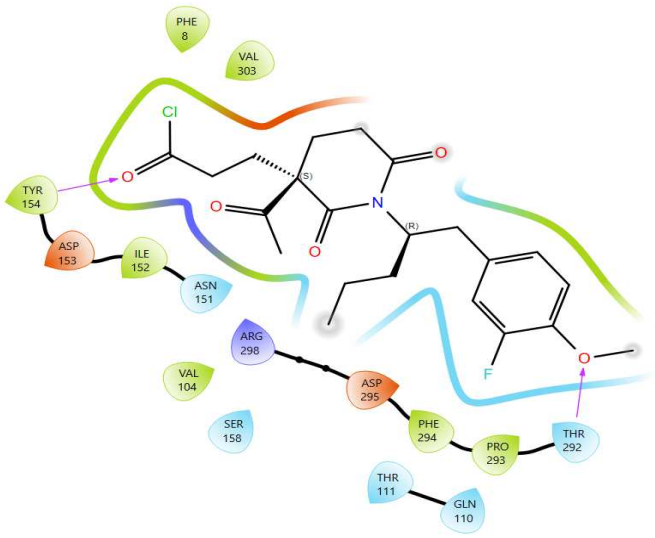
The designed ligands exhibited significantly improved binding affinity compared to the reference inhibitor Nirmatrelvir. Compound C6 showed the highest docking score (−7.851 kcal/mol), forming stabilizing interactions with key active-site residues VAL B:303 and ILE B:152 through hydrogen bonding, hydrophobic, and polar contacts, indicating strong and well-oriented binding. The second-ranked ligand, C4, also demonstrated favorable binding with a docking score of −7.232 kcal/mol, engaging residue TYR B:154 via hydrogen-bond interactions. In contrast, the reference ligand Nirmatrelvir showed a lower docking score (−4.32 kcal/mol) despite forming hydrogen-bond, hydrophobic, and polar interactions with ARG B:298 and ILE B:152.The superior docking score and interaction profile of C6 indicate enhanced binding stability within the protease active site, which is earlier supported by its highest predicted activity according to the CoMFA and CoMSIA QSAR models (pK_i=8.9).
Table 6 summarizes the docking scores, key interacting residues, and interaction types for all 12 ligands.

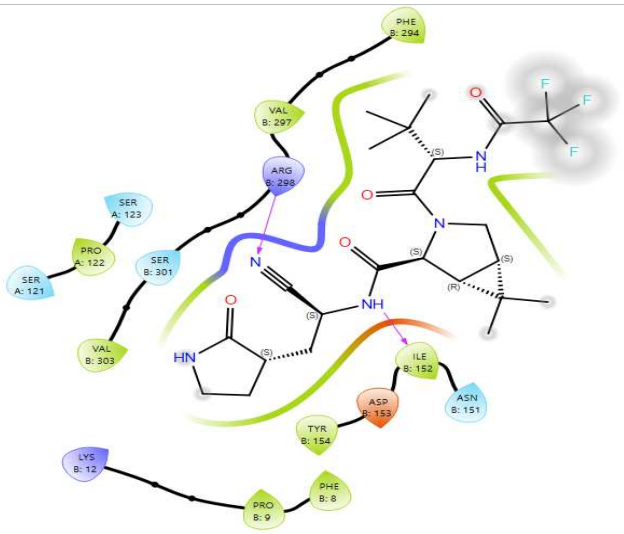
Table 6 Docking Scores and Key Protein–Ligand Interactions of the 12 Designed Compounds Against SARS-CoV-2 Mpro (PDB ID: 7TLL)

Complex	2D Diagram	Interactions
Protein 7TLL – Ligand C6		Amino acid: VALB:303, ILE B:152 Docking Score: -7.851 Intercation: Hydrogen Bonds Hydrophobic Polar
Protein 7TLL – Ligand C4		Amino acid: TYR B:154 Docking Score: -7.232 Intercation: Hydrogen Bonds
Protein 7TLL – Ligand C5		Amino acid: PHE B:294 and ILE B:152 Docking Score: -7.163 Intercation: Hydrogen Bonds Pi-Pi stacking Solvent exposure

Protein 7TLL – Ligand C9		Amino acid: LYSB:12, PRO A:122 Docking Score: -6.937 Intercation: Hydrogen Bonds Pi-Cation
Protein 7TLL – Ligand C1		Amino acid: TYR B:154 Docking Score: -6.821 Intercation: Hydrogen Bonds Hydrophobic
Protein 7TLL – Ligand C2		Amino acid: LYS B:12, TYR B:154, VAL B:303 Docking Score: -6.606 Intercation: Hydrogen Bonds Hydrophobic Polar

Protein 7TLL – Ligand C12		Amino acid: NA Docking Score: -6.369 Intercation: Hydrogen Bonds Hydrophobic Polar
Protein 7TLL – Ligand C3		Amino acid: ASPB:153 and PROA:122 Docking Score: -6.218 Intercation: Hydrogen Bonds Hydrophobic Polar
Protein 7TLL – Ligand C10		Amino acid: TYR B:154 Docking Score: -5.989 Intercation: Hydrogen Bonds Hydrophobic Polar

Protein 7TLL – Ligand C11		Amino acid: ARG B:298 Docking Score: -5.727 Intercation: Hydrogen Bonds Hydrophobic Polar
Protein 7TLL – Ligand C7		Amino acid: TYR B:154 Docking Score: -5.670 Intercation: Hydrogen Bonds Hydrophobic Polar
Protein 7TLL – Ligand C8		Amino acid: TYR 154, THR 292 Docking Score: -5.455 Intercation: Hydrogen Bonds Hydrophobic Polar

Protein 7TLL – Ligand Nirmatrelvir		Amino acid: ARG B:298, ILE B:152 Docking Score: -4.32 Intercation: Hydrogen Bonds Hydrophobic Polar

3.5. Molecular Dynamics Simulations

To benchmark the dynamic stability of the newly designed ligand C6, we performed parallel 100 ns MD simulations of C6–Mpro and Nirmatrelvir–Mpro under identical conditions. The trajectories were interrogated using ligand RMSD after least-squares fit to the protein (**Figure 9A**), RMSF (**Figure 9B**), hydrogen-bond time series (**Figure 10A**), and radius of gyration (Rg) (**Figure 10B**). Throughout, red traces denote C6 and black traces denote Nirmatrelvir.

3.5.1. Root Mean Square Deviation (RMSD)

The ligand RMSD quantifies the positional variability of the bound ligand relative to the Mpro. C6 exhibits a stable band centered near ~2.0 nm, punctuated by short-lived excursions but without any monotonic drift across 100 ns. Nirmatrelvir displays a lower RMSD baseline (~1.0 nm), consistent along the full trajectory. The higher yet bounded RMSD for C6 suggests that C6 remains confined and engaged while sampling internal conformers/microstates (e.g., rotameric adjustments) rather than escaping the binding site. The absence of long-timescale drift supports persistent binding under the simulated conditions. **Table 7**summarizes the RMSD statistics of C6 and Nirmatrelvir relative to the protein backbone over the 100-ns MD trajectory.

Table 7 Ligand RMSD Statistics Relative to the Protein Backbone (100-ns MD Simulation)

Ligand	Median RMSD (nm)	Mean RMSD (nm)	IQR (nm)	RMSD Range (nm)
C6	2.046	2.091	0.266	1.636 – 3.678
Nirmatrelvir	0.541	0.59	0.258	0.168 – 2.147

3.5.2. Root Mean Square Fluctuation (RMSF)

RMSF values are uniformly low for both complexes, with the majority of atoms/residues fluctuating below ~0.2 nm and sporadic peaks confined to loop/terminal segments that are naturally flexible. No systematic increase is observed for residues in the active-site region in the presence of C6 compared with Nirmatrelvir. This indicates that C6 does not introduce local destabilization and that the native dynamical profile of Mpro is preserved.

3.5.3. Hydrogen Bond Analysis

The hydrogen-bond time series shows near-continuous presence of ≥1 H-bond for both complexes, with intermittent formation of a second H-bond over the trajectory. For C6, the dense red bars indicating pairs within 0.35 nm corroborate sustained polar contacts that periodically strengthen the interaction network. The continuity of hydrogen bonding, combined with bounded ligand RMSD, supports a stable, electrostatically anchored binding pose for C6, comparable to the reference.

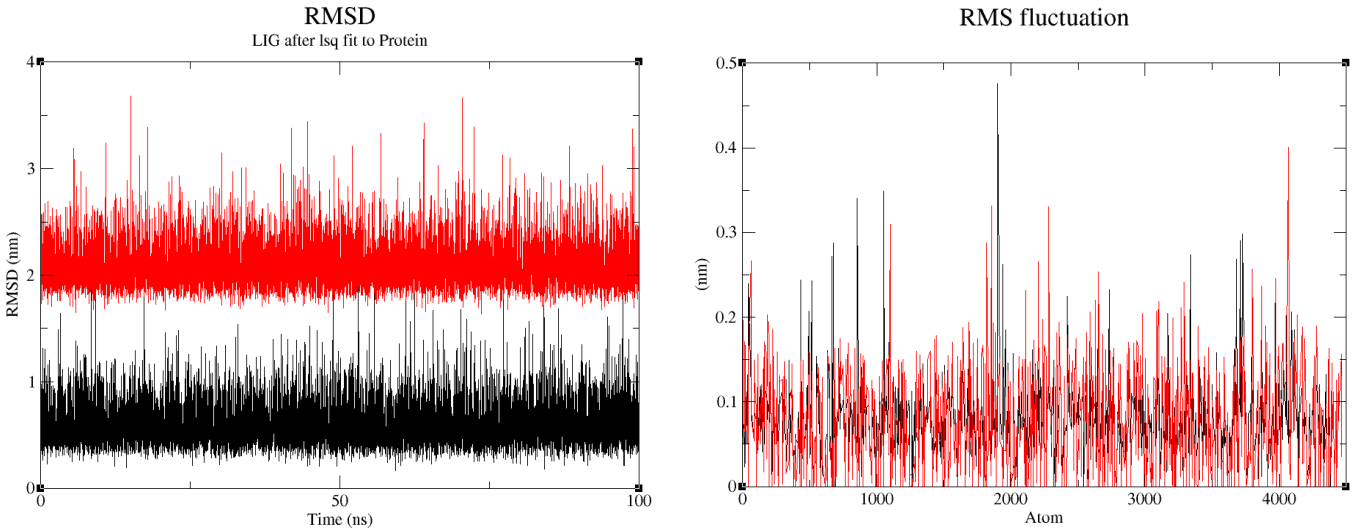


Figure 10 A: Compound C6 RMSD after least-squares fit to the protein (Red plot) andNirmatrelvirRMSD after least-squares fit to the protein (Black plot). B: Root Mean Square Fluctuation for Compound C6 (Red plot) and Root Mean Square Fluctuation for Nirmatrelvir (Black plot).

3.5.4. Radius of Gyration (Rg)

The Rg traces remain flat, indicating no global compaction/expansion events and no unfolding-like transitions for Mpro in either complex. The C6 compound (red) is centered lower than the Nirmatrelvir series (black), consistent with a slightly more compact assembly or differences in mass distribution under the same conditions. Importantly, there are no step changes or long-timescale trends, which confirms global structural stability of the protein during ligand binding. **Table 8** summarizes the Rg values of Mpro in complex with C6 and Nirmatrelvir, reflecting their global structural stability during the MD trajectories.

Table 8 Radius of gyrationstatistics

Ligand Complex	Mean Rg (nm)	IQR (nm)	SD (nm)
C6	4.19	0.198	0.129
Nirmatrelvir	5.388	0.512	0.291

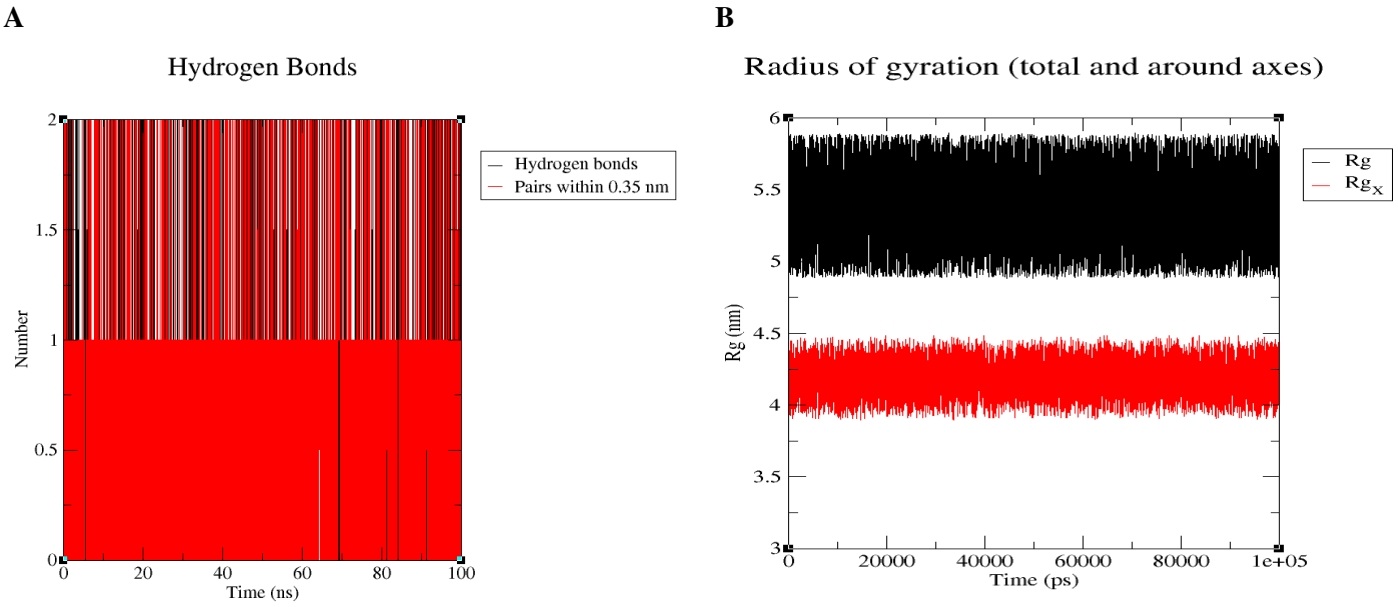


Figure 11 A: Hydrogen Bonds for Compound C6 (Red plot) and Nirmatrelvir drug reference (Black plot). B: Radius of gyration for Compound C6 (Red plot) and Nirmatrelvir drug reference(Black plot).

3.6. ADMET Analysis

In silico ADMET analysis of ligand C6 and reference drugNirmatrelvirwas carried out using the PreADMET platform to comprehensively assess its absorption, distribution, metabolism, excretion, and toxicity properties, as well as its druglikeness profile. These predictive evaluations are crucial in early-stage drug discovery to prioritize compounds with desirable pharmacokinetic and safety attributes while minimizing experimental costs and downstream attrition. The results are summarized in **Tables 9, 10 and 11**.

3.6.1 ADME Properties:

The Caco2 apparent permeability (Papp), an in vitro measure of transcellular intestinal permeability expressed in $\times 10^{-6}$ cm/s, is high for C6 (21.04), indicating efficient epithelial transport. The MDCK apparent permeability (Papp) which is a complementary epithelial model also reported in $\times 10^{-6}$ cm/sis likewise high for C6 (23.15), reinforcing strong permeability across cell systems. Consistently, the human intestinal absorption (HIA) prediction, expressed as % absorbed, is high for C6 (91.6%), supporting favorable oral absorption. In comparison, Nirmatrelvir shows similarly high Caco2 Papp (20.59) but low MDCK Papp (0.065), and a near high HIA (79.3%), which indicate that while the reference drug benefits from good Caco2 permeability, C6 demonstrates a stronger, system agnostic permeability profile with higher predicted intestinal absorption.

Aqueous buffer solubility (mg/L)is high for C6 (12,866 mg/L), facilitating pre-absorptive availability. Pure-water solubility (mg/L) is moderate for C6 (46.27 mg/L), which remains compatible with oral development when coupled with strong permeability. The SKlogP (or SKlogD) descriptor, an estimate of lipophilicity where moderate values (~2) typically balance permeability and solubilitygaves ≈ 2.04 for C6, aligning with its permeability driven absorption. In contrast, Nirmatrelvir exhibits extremely high buffer solubility (1.11×10^{6} mg/L) and higher pure water solubility (180.2 mg/L), but a very low SKlogP (~0.19), consistent with hydrophilic, solubility led absorption. Overall, C6 combines adequate solubility with optimally moderate lipophilicity, yielding a favorable absorption profile.

Plasma protein binding (PPB, % boundis moderate to high for C6 (75.6%), a range commonly compatible with small molecule therapeutics[22]. Blood–brain barrier penetration probability (BBB) is moderate low for C6 (0.141), appropriate for non CNS targets while allowing adequate systemic distribution. In comparison, Nirmatrelvir has moderate PPB (65.0%) and low BBB probability (0.036). Taken together, C6's distribution parameters (moderate PPB, moderate low BBB) confirm an acceptable systemic profile.

Table 9 ADME Properties for Compound C6 and Nirmatrelvir

ID	Value	
	Compound C6	**Nirmatrelvir**
Blood–Brain BarrierPermeability	0.140806	0.0358981
Solubility in Buffer (mg/L)	12865.7	1.10959e+006
Caco-2 CellPermeability	21.036	20.589
CYP2C19 Enzyme Inhibition	Non	Non
CYP2C9 Enzyme Inhibition	Non	Non
CYP2D6 Enzyme Inhibition	Non	Non
CYP2D6 SubstrateStatus	Non	Non
CYP3A4 Enzyme Inhibition	Inhibitor	Inhibitor
CYP3A4 Substrate Status	Weakly	Substrate
Human Intestinal Absorption (%)	91.5972	79.31649
MDCK Cell Permeability	23.1511	0.0649109
P-glycoprotein Inhibition	Inhibitor	Non
Plasma Protein Binding (%)	75.60172	65.03451
Solubility in Pure Water (mg/L)	46.2695	180.199
Skin Permeability Coefficient	-3.68072	-4.12093
Log D (Distribution Coefficient)	2.04019	0.189780
Log P (Partition Coefficient)	2.04019	0.189780
Log Solubility in Buffer	-1.49645	0.346600
Log Solubility in Pure Water	-3.94059	-3.44281

3.6.2 Drug likeness Evaluation

The Rule of Five is Lipinski's guideline for oral drug likeness (limits on molecular weight, logP, hydrogenbond donors/acceptors) and itis Suitable for C6 with 0 violations, and likewise Suitable for the reference with 0 violations, supporting oral ADME compatibility for both candidates[23]. The CMClike rule (similarity to the Comprehensive Medicinal Chemistry drug database) is Qualified for C6 (0 violations) whilenot qualified for the reference (1 violation: molecular weight), indicating that C6 aligns more closely with established drug space property ranges[24]. The MDDR like rule (similarity to the MDL Drug Data Report database) classifies both compounds as Drug like with 0 violations.

The Lead like rule which designed for early discovery leads and thus stricter than Ro5 (favoring smaller, less lipophilic scaffolds)is Violated by C6 with 1 violation (molecular weight) and by the reference with 2 violations (molecular weight and AlogP98)[25]. While

both breach lead like constraints, C6's single violation places it closer to lead like space than the reference. Importantly, lead like thresholds are intentionally conservative; C6's favorable ADME (high permeability, high HIA, adequate solubility) remains consistent with development suitability.

The WDI like rule (proximity to World Drug Index distributions of drug properties and topological descriptors) flags both compounds "Out of 90% cutoff", yet C6 shows only 1 violation (Kier_alpha_03) compared with 6 violations for the reference (e.g., Kier_alpha_01/03, VChi_00/02, VChi_03_cluster, 1st_Zagreb standard connectivity indices).

Table 10 Drug likeness evaluation for Compound C6 and Nirmatrelvir

ID	**Value**	
	Compound C6	**Nirmatrelvir**
CMC-like Rule	Qualified	Not qualified
CMC-like Rule Violation Fields		Molecular_weight
Number of violations for CMC-like rule	0	1
Lead-like Rule Violation Fields	Molecular_weight	Molecular_weight, AlopP98_value
Molecular Weight	Violated	Violated
Lead-like Rule	1	2
Lead-like Rule Violations	Drug-like	Drug-like
MDDR-like Rule	Suitable	Suitable
Rule of Five	Out of 90% cutoff	Out of 90% cutoff
WDI-like Rule	Kier_alpha_03	Kier_alpha_01, Kier_alpha_03, VChi_00, VChi_02, VChi_03_cluster, 1st_Zagreb
WDI-like Rule Violation Fields	1	6

3.6.3 Toxicity Profiling

a- Mutagenicity (Ames test; TA strains):

The Ames test (in vitro mutagenicity screen) is non mutagen for C6 and the reference, indicating no mutagenic signal in the overall call. The strain specific assays TA100 and TA1535, both with metabolic activation (10 RLI) and without activation (NA) are negative for C6 across all conditions, and negative for the reference as well which confirm the absence of point-mutation risk in these Salmonella strains. These results collectively support a

mutagenicity favorable profile for C6 relative to early discovery risk. (C6: Ames non mutagen; TA100/TA1535 10 RLI & NA negative. Reference: same negative outcomes.)

b- Carcinogenicity (rodent 2 year; rat & mouse) and hERG.

In rodent carcinogenicity (2 year) assays, C6 is negative in both rat and mouse, whereas the reference is positive in rat but negative in mouse. This distinction indicates a more favorable carcinogenicity profile for C6 in long term rodent testing. For cardiotoxicity risk, the hERG(human Ether à go-go Related Gene) inhibition prediction is ambiguous for both compounds an inconclusive signal that warrants follow-up experimental confirmation. (C6: Carcino Rat/Mouse negative; hERG ambiguous. Reference: Carcino Rat positive, Mouse negative; hERG ambiguous.)

c: Aquatic toxicity (algae, daphnia, medaka, minnow):

Model metrics for aquatic toxicityalgae_at, daphnia_at, medaka_at, and minnow_atare lower for C6 than for the reference across all four species, suggesting reduced predicted ecotoxicity for C6 relative to the comparator. Specifically, C6 vs. reference: algae (0.0159 vs. 0.0293), daphnia (0.0766 vs. 0.6212), medaka (0.0122 vs. 0.6204), minnow (0.0258 vs. 0.3272). While units are model specific, the consistent directionality (lower values for C6) supports a more favorable environmental toxicity profile in these in silico predictions.

Table 11 Toxicity assessment for Compound C6 and Nirmatrelvir

Test ID	**Value**	
	Compound C6	**Nirmatrelvir**
Algae Acute Toxicity	0.0158678	0.0292586
Ames Test (Mutagenicity)	non-mutagen	non-mutagen
Mouse Carcinogenicity	negative	negative
Rat Carcinogenicity	negative	positive
Daphnia Acute Toxicity	0.0766021	0.62124
hERG Channel Inhibition	ambiguous	ambiguous
Medaka Fish Acute Toxicity	0.0121812	0.620391
Minnow Acute Toxicity	0.0258106	0.327193
Salmonella strain TA100, 10RLI condition	negative	negative
Salmonella strain TA100, NA condition	negative	negative
Salmonella strain TA1535, 10RLI condition	negative	negative
Salmonella strain TA1535, NA condition	negative	negative

Conclusion

In this study, we developed and applied an integrated in silico workflow that couples Auto QSAR (KPLS), 3D-QSAR (CoMFA/CoMSIA), reaction based enumeration, structure based docking, 100 ns molecular dynamics, and ADMET/toxicity profiling to accelerate the identification of novel SARS-CoV2 Mpro inhibitors. Using a curated dataset of 80 experimentally characterized Compounds, the best-performing Auto QSAR model (kpls_linear_1) achieved strong predictive performance ($R^2 \approx 0.88$, $Q^2 \approx 0.90$, RMSE ≈ 0.31), while the CoMSIA model provided a high-quality structure–activity interpretation and acceptable external predictivity (test $Q^2 \approx 0.66$; RMSE ≈ 0.47). These models guided the exploration of synthetically tractable chemical space through automated reaction enumeration, enabling prioritization of 12 designed candidates with high predicted potency. Structure based assessment against SARS-CoV2 Mpro (PDB: 7TLL) further supported this prioritization: among the designed ligands, C6 emerged as the most promising hit, combining the highest predicted activity (pKi ≈ 8.9), favorable docking score, and conserved interaction patterns consistent with stable active site of the protein. Dynamic evaluation via 100 ns MD simulations confirmed that the C6–Mpro complex remains conformationally stable over time, with persistent polar contacts and no evidence of global protein destabilization relative to the reference complex. ADMET profiling further suggested that C6 has a favorable developability profile, including high predicted intestinal absorption and balanced physicochemical properties. Although the in-silico mutagenicity readout is overall favorable in the reported strain based outputs, the broader safety profileespecially cardiotoxicity and drug–drug interaction riskshould be validated using standardized experimental assays. Collectively, these results position C6 as a promising lead-like starting point for biochemical testing and iterative medicinal chemistry optimizationfor rapid antiviral lead discovery targeting coronavirus proteases.

References

[1] "Designing highly efficient heterocyclic inhibitors for SARS-CoV-2 3C-like proteinase: a comprehensive in silico study: Journal of Biomolecular Structure and Dynamics: Vol 0, No 0 - Get Access." Accessed: Aug. 05, 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/07391102.2025.2494842

[2] D. Cavanagh and P. Britton, "Coronaviruses: General Features," *Encyclopedia of Virology*, pp. 549–554, 2008, doi: 10.1016/B978-012374410-4.00370-8.

[3] "Origin and evolution of pathogenic coronaviruses | Nature Reviews Microbiology." Accessed: Apr. 28, 2025. [Online]. Available: https://www.nature.com/articles/s41579-018-0118-9

[4] Q. Hu *et al.*, "The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19," *MedComm (2020)*, vol. 3, no. 3, p. e151, Sept. 2022, doi: 10.1002/mco2.151.

[5] R. Arya *et al.*, "Structural insights into SARS-CoV-2 proteins," *J Mol Biol*, vol. 433, no. 2, p. 166725, Jan. 2021, doi: 10.1016/j.jmb.2020.11.024.

[6] G.-G. Chang, "Quaternary Structure of the SARS Coronavirus Main Protease," *Molecular Biology of the SARS-Coronavirus*, pp. 115–128, July 2009, doi: 10.1007/978-3-642-03683-5_8.

[7] G. U. Jeong, H. Song, G. Y. Yoon, D. Kim, and Y.-C. Kwon, "Therapeutic Strategies Against COVID-19 and Structural Characterization of SARS-CoV-2: A Review," *Front. Microbiol.*, vol. 11, July 2020, doi: 10.3389/fmicb.2020.01723.

[8] R. M. Yaghi, D. C. Wylie, C. L. Andrews, O. H. Dickert, A. Ram, and B. L. Iverson, "An Investigation of Nirmatrelvir (Paxlovid) Resistance in SARS-CoV-2 Mpro," *ACS Bio Med Chem Au*, vol. 4, no. 6, pp. 280–290, Dec. 2024, doi: 10.1021/acsbiomedchemau.4c00045.

[9] L. E. Khoury *et al.*, "Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation," *Chem. Sci.*, vol. 13, no. 13, pp. 3674–3687, 2022, doi: 10.1039/D1SC05892D.

[10] "Assay in Summary_ki." Accessed: Dec. 26, 2025. [Online]. Available: https://bdb99.ucsd.edu/rwd/jsp/dbsearch/assay.jsp?assayid=1&entryid=10640

[11] Y. Zhao *et al.*, "Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2," *Proceedings of the National Academy of Sciences*, vol. 119, no. 16, p. e2117142119, Apr. 2022, doi: 10.1073/pnas.2117142119.

[12] “Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments | Journal of Computer-Aided Molecular Design.” Accessed: Dec. 26, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s10822-013-9644-8

[13] *Schrödinger Release 2025-4: LigPrep, Schrödinger, LLC, New York, NY, 2025.*

[14] R. D. Cramer, D. E. Patterson, and J. D. Bunce, “Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins,” *J. Am. Chem. Soc.*, vol. 110, no. 18, pp. 5959–5967, Aug. 1988, doi: 10.1021/ja00226a005.

[15] T. Knehans, N. A. Boyles, and P. H. Bos, “Enumerable Libraries and Accessible Chemical Space in Drug Discovery,” in *Computational Drug Discovery*, John Wiley & Sons, Ltd, 2024, pp. 315–336. doi: 10.1002/9783527840748.ch14.

[16] *Schrödinger Release 2025-1: Maestro, Schrödinger, LLC, New York, NY, 2025.*

[17] “SwissParam: A fast force field generation tool for small organic molecules - Zoete - 2011 - Journal of Computational Chemistry - Wiley Online Library.” Accessed: Dec. 26, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/jcc.21816

[18] S. Lee, I. H. Lee, H. J. Kim, G. S. Chang, J. E. Chung, and K. T. No, “The PreADME Approach: Web-based program for rapid prediction of physico-chemical, drug absorption and drug-like properties,” *euro QSAR 2002 - Designing Drugs and Crop Protectants: Processes Problems and Solutions*, pp. 418–420, Jan. 2002.

[19] A. Tropsha, “Best Practices for QSAR Model Development, Validation, and Exploitation,” *Mol Inform*, vol. 29, no. 6–7, pp. 476–488, July 2010, doi: 10.1002/minf.201000061.

[20] A. Golbraikh and A. Tropsha, “Beware of q2!,” *J Mol Graph Model*, vol. 20, no. 4, pp. 269–276, Jan. 2002, doi: 10.1016/s1093-3263(01)00123-1.

[21] “Reaction-Based Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-Dependent Kinase 2 Inhibitors | Journal of Chemical Information and Modeling.” Accessed: Dec. 26, 2025. [Online]. Available: https://pubs.acs.org/doi/10.1021/acs.jcim.9b00367

[22] “10: Quantitative Biochemistry in the Pharmacological Evaluation of Drugs | Basicmedical Key.” Accessed: Dec. 26, 2025. [Online]. Available: https://basicmedicalkey.com/10-quantitative-biochemistry-in-the-pharmacological-evaluation-of-drugs/

[23] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1," *Advanced Drug Delivery Reviews*, vol. 46, no. 1, pp. 3–26, Mar. 2001, doi: 10.1016/S0169-409X(00)00129-0.

[24] A. K. Ghose, V. N. Viswanadhan, and J. J. Wendoloski, "A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases," *J. Comb. Chem.*, vol. 1, no. 1, pp. 55–68, Jan. 1999, doi: 10.1021/cc9800071.

[25] S. J. Teague, A. M. Davis, P. D. Leeson, and T. Oprea, "The Design of Leadlike Combinatorial Libraries," *Angew Chem Int Ed Engl*, vol. 38, no. 24, pp. 3743–3748, Dec. 1999, doi: 10.1002/(SICI)1521-3773(19991216)38:24%3C3743::AID-ANIE3743%3E3.0.CO;2-U.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.docx