***Original Article***

DOI: 10.5582/ddt.2025.01034

# Discovery of SARS-CoV-2 papain-like protease inhibitors through machine learning and molecular simulation approaches

**Li Li[1], Jinyang Li[1], Quanling Zhang[1], Yifei Huang[2], Yongxin Bao[1,3,*], Xiaowen Tang[1,4,*]**

[1] School of Pharmacy, Qingdao University, Qingdao, Shandong, China;
[2] College of Life Sciences, Qingdao University, Qingdao, Shandong, China;
[3] Qingdao Women and Children's Hospital Affiliated to Qingdao University, Qingdao, Shandong, China;
[4] Shandong Provincial Key Laboratory of Pathogenesis and Prevention of Brain Diseases, Qingdao University, Qingdao, Shandong, China.

**SUMMARY**: The papain-like protease (PLpro), a cysteine protease found in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), plays a crucial role in viral replication by cleaving the viral polyproteins and interfering with the host's innate immune response through deubiquitination and deISGylation activities. Consequently, targeting PLpro has emerged as an appealing therapeutic strategy against SARS-CoV-2 infection. Despite considerable efforts in the development of PLpro inhibitors, there is currently no drug available on the market that specifically targets PLpro. Improving drug screening strategies and identifying additional candidate compounds could significantly contribute to the advancement of antiviral agents targeting PLpro. To address this pressing issue, our present study has developed a highly efficient compound screening strategy based on a supervised machine learning approach. Integrated with further molecular simulation approaches such as molecular docking, molecular dynamics simulations, and quantum chemical calculations, we have identified seven compounds with potent inhibitory activity against PLpro. Notably, two of these compounds exhibited superior activity compared to Jun12682, which is currently considered the best-performing inhibitor against PLpro. Furthermore, some crucial residues in SARS-CoV-2 PLpro were recognized as favorable contributors to the binding with inhibitor, which would provide valuable insights for the development of more potent and highly selective SARS-CoV-2 PLpro inhibitors. The compound screening strategy and potential PLpro inhibitor candidates revealed in the present study would hold promise for advancing the development of antiviral drugs targeting SARS-CoV-2 and its variants.

***Keywords***: SARS-CoV-2, papain-like protease inhibitor, machine learning, virtual screening, molecular simulation

## 1. Introduction

The global Coronavirus Disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has posed a serious threat to public health security worldwide. Although various approved COVID-19 vaccines have played a critical role in controlling the pandemic (*1*), the continuous emergence of SARS-CoV-2 variants, such as the recently identified JN.1 (BA.2.86.1.1) and KP.2 (JN.1.11.1.2), threatens the efficacy of current vaccines (*2*). Moreover, vaccines are primarily used to prevent COVID-19, but for patients already infected with the virus, effective treatment options are still necessary. Therefore, the development of specific antiviral drugs targeting SARS-CoV-2 remains an essential measure in addressing this ongoing threat.

Papain-like protease (PLpro), a viral cysteine protease essential for SARS-CoV-2 replication, cleaves polyproteins pp1a and pp1ab to generate non-structural proteins (nsp). In addition to processing viral proteins, PLpro also targets host proteins such as ubiquitin and interferon-stimulated gene 15 (ISG15), performing deubiquitinating and deISGylating activities that suppress the host's innate immune response (*3*). Consequently, PLpro inhibition holds promise for suppressing viral propagation and restoring host's immune function (*4*), making it a key target for antiviral drug development.

Since the outbreak of the COVID-19 pandemic, numerous inhibitors targeting PLpro have been discovered through structure-based drug design, virtual screening, and high-throughput screening methods (*5*). These inhibitors can be categorized into non-covalent and covalent inhibitors based on their binding modes as summarized in Figure 1. GRL0617, the first non-covalent PLpro inhibitor (*6*), exhibited relatively low activity against SARS-CoV-2 in cell-based assays, despite the moderate enzymatic activity ($IC_{50}$~1.39

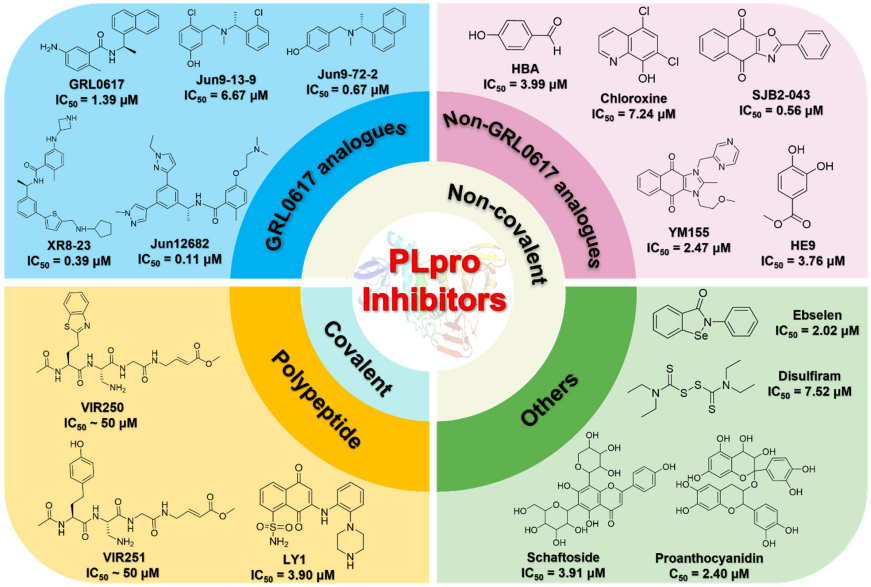


**Figure 1. Representation and classification of SARS-CoV-2 PLpro inhibitors.**

μM). Subsequently, Structural optimization led to more potent analogs such as XR8-23 ($IC_{50}$~0.39 μM) (*7*) and Jun12682 ($IC_{50}$~106.8 ± 5.0 nM) (*8*), the latter showing strong activity against multiple variants. In addition, some non-GRL0617 analogs (such as chloroxine (*9*), SJB2-043 (*10*), HE9 (*11*), HBA (*11*) and YM155 (*12*)) and several non-specific inhibitors, (including ebselen (*13*), disulfiram (*14*), schaftoside (*15*) and proanthocyanidi (*16*)) also exhibit promising antiviral activity. By contrast, covalent inhibitors targeting PLpro like LY1 (*17*) and peptide-based VIR250 and VIR251 (*18*) have also been reported. However, most candidates face limitations such as insufficient antiviral activity, poor pharmacokinetics, or inadequate target selectivity, preventing clinical translation. So far, only HL-21 has entered Phase I trials, and there are currently no FDA-approved drugs targeting PLpro. These challenges underscore the need for improved screening strategies and novel chemical scaffolds to accelerate PLpro-targeted drug discovery.

In recent years, artificial intelligence (AI) has become an increasingly powerful tool in drug discovery (*19*). Notably, the 2024 Nobel Prizes in Physics and Chemistry has underscored the pivotal role of AI in advancing scientific research. Recently, some potent SARS-CoV-2 Mpro inhibitors with strong cellular activity were discovered by a using machine learning approach (*20*), shortly thereafter, a lead compound (PF-07957472) targeting SARS-CoV-2 PLpro that showed high efficacy in mouse models was also identified by using AI technique (*21*). For the purpose of improving the drug screening efficiency and providing more candidate compounds to assist the development of anti-COVID-19 drugs, an integrated screening strategy that combined machine learning and molecular simulation approaches (*22,23*) was developed and further utilized to perform the screening of SARS-CoV-2 PLpro inhibitor. As a result, the current study identified seven compounds (Cpd-1~4, Cpd-6, Cpd-8, and Cpd-14) that exhibited higher binding affinity on PLpro compared to GRL0617. Among them, two compounds, Cpd-1 and Cpd-3, showed more potent inhibitory activity than the currently best-performing compound, Jun12682. These compounds hold promise for advancing the development of a new generation of inhibitors targeting SARS-CoV-2 and its variants.

## 2. Methods

The screening strategy of SARS-CoV-2 papain-like protease inhibitors in the current study refers to machine learning-based classification, molecular simulation (molecular docking, molecular dynamics simulation, and quantum chemical calculation) based screening and verification. The workflow of the screening strategy is exhibited in Figure 2.

### 2.1. Data preparation

The initial database was constructed based on the inhibitory activity data for 3,935 FDA-approved drugs and clinical trial candidate compounds against SARS-CoV-2 PLpro (*24*). To ensure data quality, the following preprocessing steps were performed before training the machine learning models: (1) Eliminating the compounds with invalid information, such as lacking structural information, containing ambiguous or even non-numeric data. Only compounds with structural information (in SMILES format) and inhibition rates against SARS-CoV-2 PLpro were retained. (2) Establishing an activity threshold of 60%, wherein

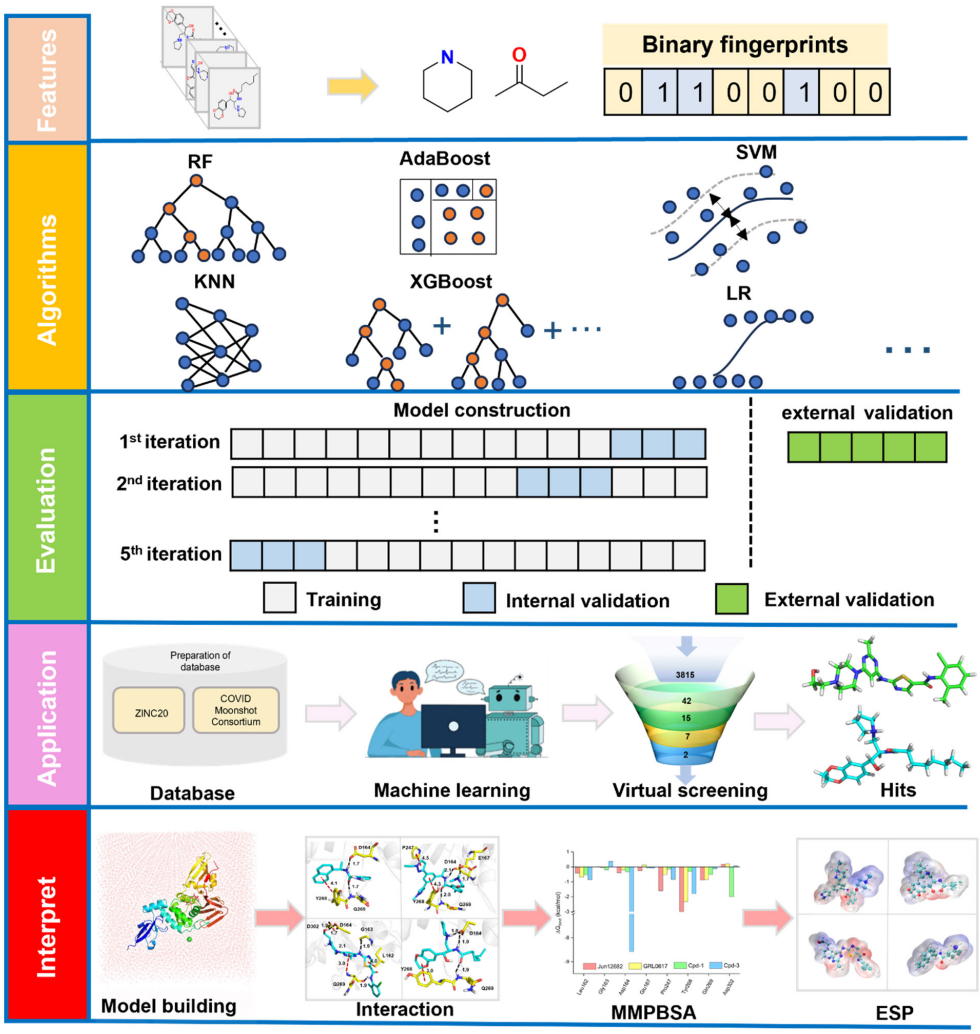


**Figure 2. Workflow of the screening strategy in the present study.**

compounds with an inhibition rate of 60% or higher were defined as active, while those lower than 60% were labeled as inactive. As a result, a binary dataset of 3,428 compounds consisting of 249 active and 3,179 inactive data was obtained (Supplementary Materials 1). (3) In view of the imbalance between active and inactive data, 1,000 compounds randomly selected from the inactive data, together with all the active data, were collected to construct a combined dataset of 1,249 compounds with a 4:1 ratio of inactive to active data (Supplementary Materials 2). (4) Further splitting the combined dataset into an 8:2 ratio for machine learning model training and internal validation. To achieve more realistic model performance, a 5-fold cross-validation strategy was applied. Additionally, 79 active compounds obtained from SARS-CoV-2 PLpro patent literature and 341 compounds randomly selected from the inactive data were collected for external validation (Supplementary Materials 3).

FDA-approved drugs (*https://www.fda.gov/*) are increasingly favored for new drug development because their well-documented toxicity profiles and human pharmacokinetics significantly reduce both the development time and costs. Moreover, surveys by the National Cancer Institute have shown that three-quarters of all drugs used globally over the past half-century to treat various human diseases are derived from natural resources. Therefore, for the final screening of potential SARS-CoV-2 PLpro inhibitors, 3,815 compounds from FDA-approved drugs (previously unevaluated) and natural products curated from the ZINC database, as well as the COVID Moonshot platform were compiled as prediction dataset (Supplementary Materials 4).

## 2.2. Machine learning

Six molecular descriptors and fingerprints, including Morgan fingerprint (MorganFP), MACCS fingerprint (MACCSFP), E-state fingerprint (E-stateFP), Avalon fingerprint (AvalonFP), Atom-pairs fingerprint (Atom-PairFP) and RDKit descriptors (RDKit-Des), were employed to describe the molecular structures when performing the machine learning. Further details regarding the molecular fingerprints and descriptors

can be found in Table S1 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*). Meanwhile, 14 algorithms, including decision trees, random forests, extreme gradient boosting (XGBoost), support vector machines (SVM), gradient boosting decision trees, gradient boosting machines (GBM), logistic regression, K-nearest neighbors (KNN), linear discriminant analysis (LDA), stochastic gradient descent, adaptive boosting, bootstrap aggregating, voting classifier, and multilayer perceptron classifier were adopted for machine learning model construction. Eventually, a total of 84 machine learning models were generated and further evaluated to establish the SARS-CoV-2 PLpro inhibitors screening models. In addition, RDKit was employed for the generation of all molecular features and fingerprints, and scikit-learn was employed to implement the model construction.

To evaluate the predictive ability and robustness of the constructed machine learning models, the following evaluation metrics were used in this study: area under the receiver operating characteristic curve (AUC), F1 score (F1), precision (Pre), sensitivity (Se), and specificity (Sp). The calculation methods of them are listed as follows:

$$Pre = TP / (TP + FP)$$
$$Se = TP / (TP + FN)$$
$$Sp = TN / (TN + FP)$$
$$F1 = 2 \times (Pre \times Se) / (Sp + Se)$$

Where, true positive (TP) refers to instances correctly classified as positive, while true negative (TN) denotes those correctly classified as negative. False positive (FP) represents instances incorrectly classified as positive, and false negative (FN) denotes those incorrectly classified as negative. The F1 score is a metric that provides a comprehensive evaluation of the model by considering both precision (Pre) and sensitivity (Se). A higher F1 score (closer to 1) indicates a stronger generalization ability of the mode. AUC is a key indicator for evaluating the performance of classification models. The model performance will be better if the AUC value is closer to 1 (*25*).

2.3. Molecular docking

The initial receptor model was constructed based on the crystal structure of the SARS-CoV-2 PLpro in complex with GRL0617 (PDB ID: 7CJM). Protein and ligand parts were extracted and processed for subsequent molecular docking experiments. Compounds screened by the machine learning model underwent further optimization using the OPLS4 force field. Protonated states of ionizable groups were defined at pH 7.0 ± 0.2, which simulated the slightly fluctuating pH conditions in the physiological environment. The protonated states of titratable residues in receptor structure were also calculated at the same pH for ligand preparation. Molecular docking analysis utilized AutoDock Vina (*26*), where the centroid of the ligand (GRL0617) was defined as the center of the docking grid, and the size of the grid was set to 25 × 25 × 25 $Å^3$. Finally, flexible molecular docking based on induced fit theory was executed, and results (binding mode and docking score) with the best docking score were recorded.

2.4 Classical molecular dynamics simulation

The ligand-receptor complex models were obtained from molecular docking. All molecular dynamics (MD) simulations were performed using the *pmemd* module in the AMBER18 molecular simulation package. The Amber ff14SB force field (*27*) was employed for the protein, and the TIP3P model (*28*) was used for the solvent water molecules. The force field parameter of the ligand was generated from the general AMBER force field (GAFF), and the partial atomic charge was defined by the restrained electrostatic potential (RESP) (*29*) charge based on HF/6-31G* calculation with the Gaussian09 package.

The initial coordinates and topology files were generated by the *tleap* program with neutralization and solvation. The subsequent classical MD simulations were carried out by using the periodic boundary condition with the cubic model. After a series of energy minimization, programmed heating (0 to 300 K, NVT, 100 ps), density equilibrium (300 K, 1.0 atm, NPT, 100 ps), and preequilibrium (300 K, 1.0 atm, NPT, 100 ps), a final 100 ns MD simulation with a 2 fs time step was performed under the NVT ensemble to generate trajectories. During the MD simulation, the high-frequency stretching vibration of all hydrogen-containing bonds was constrained by using the SHAKE algorithm (*30*), and a 12 Å cutoff was applied to van der Waals (LJ-12 potential) and electrostatic interactions (PME strategy). Finally, *cpptraj* was used for trajectories analysis and PyMOL was used for visualization.

The binding free energy was calculated using the MM/GBSA method (*31*) *via* the *MMPBSA.py* module, based on 100 snapshots extracted from the stable phase of the MD trajectory. All energies were expressed in kcal/mol. The calculation method for binding free energy is listed as follows:

$$\Delta G_{bind} = G_{complex} - G_{receptor} + G_{ligand}$$

Where, $\Delta G_{bind}$ represents the total binding free energy between PLpro and its inhibitor. $G_{complex}$ denotes the energy of the protein-inhibitor complex, while $G_{receptor}$ and $G_{ligand}$ refer to the individual energies of the PLpro and the inhibitor, respectively. The free energy components in the MM/GBSA approach were determined as follows:

$$\Delta G_{bind} = \Delta G_{gas} + \Delta G_{solv} - T\Delta S$$
$$= \Delta E_{vdw} + \Delta E_{ele} + \Delta G_{polar} + \Delta G_{nonpola} - T\Delta S$$
$$= \Delta E_{vdw} + \Delta E_{ele} + \Delta E_{GB} + \Delta E_{SA} - T\Delta S$$

Herein, $\Delta G_{gas}$ and $\Delta G_{solv}$ denote the gas-phase and solvation energy components of the total free energy ($\Delta G_{bind}$), respectively. $\Delta G_{gas}$ consists of van der Waals ($\Delta E_{vdw}$) and electrostatic ($\Delta E_{ele}$) contributions. $\Delta G_{polar}$ and $\Delta G_{nonpolar}$ refer to the polar and nonpolar components of the solvation free energy, respectively. The terms $\Delta E_{GB}$ and $\Delta E_{SA}$ represent the polar and nonpolar contributions, respectively. The absolute temperature of the system is denoted by $T$, and the entropy related to the system is denoted as $\Delta S$. The term $T\Delta S$ represents the entropy contribution.

### 2.5 Electrostatic potential calculation

The Gaussian09 program (Revision D.01) was utilized to calculate the electrostatic potential surface of the screened molecules. The calculation was performed based on density functional theory (DFT) at the B3LYP/6-311+G(2d,p) level. The restrained electrostatic potential (RESP) charges were computed using Multiwfn. Finally, Visual Molecular Dynamics (VMD, version 1.9.4a53) was used to visualize the molecular surface electrostatic potential maps, providing a clear graphical representation of the charge distribution across the molecules.

## 3. Results

### 3.1. Reliability analysis of datasets

Reliable datasets that refer to the training and validating set are an important guarantee for the accurate construction of the machine learning models. Herein, we applied t-distributed stochastic neighbor embedding (t-SNE) analysis by using a Euclidean distance metric to evaluate the reliability of the datasets for machine learning. Figure 3 illustrates the chemical space distributions of the collected compounds in training, validation, and prediction datasets, as visualized by t-SNE. Results showed that the distribution of training and validating datasets overlapped sufficiently, indicating that the construction and evaluation of the machine learning models are reliable. Furthermore, the distribution of datasets for model application (prediction dataset) and model construction (training and validation datasets) also presented a rough overlap, demonstrating the reliability of the subsequent machine learning-based SARS-CoV-2 PLpro inhibitor screening.

### 3.2. Machine learning-based model construction and application

To construct an accurate SARS-CoV-2 PLpro inhibitor prediction strategy, we developed 84 classification models based on 14 machine learning algorithms combined with 6 molecular fingerprints. Figure 4 shows the performance of the constructed classification models, which were evaluated through AUC and F1 score.

The performance of these models exhibited significant variations. For the machine learning algorithms, the average AUC values for the Random Forest, XGBoost, SVM, and GBDT models were dramatically higher than those of other models, as summarized on the right sidebar of Figure 4A, demonstrating the excellent performance of these four algorithms. Among them, models constructed with the Random Forest algorithm exhibited the best performance (with most models' AUC values closer to 1). For molecular fingerprints and

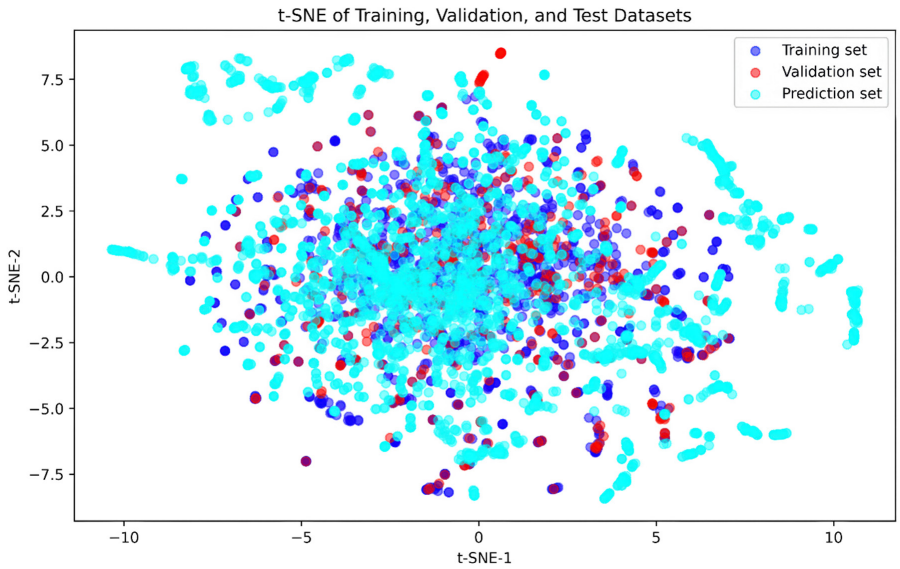


**Figure 3. t-SNE of training, validation, and prediction datasets.**

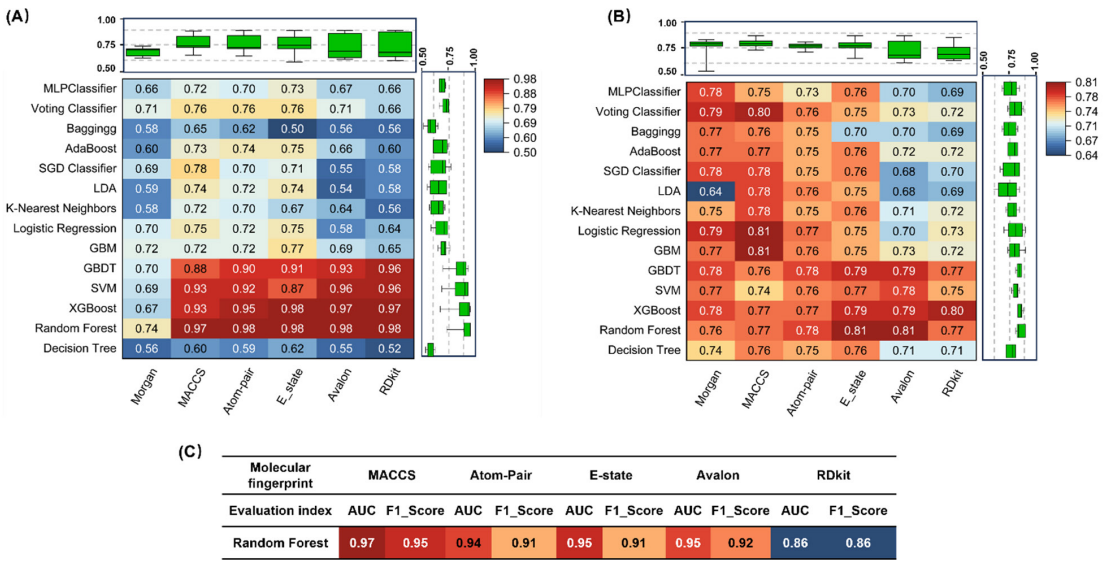


**Figure 4. Performance evaluation of the constructed machine learning models.** Heatmap of AUC **(A)** and F1 score **(B)** in internal validation. **(C)** The AUC values and F1 scores in external validation of models constructed by Random Forest algorithm with MACCSFP, E-stateFP, AtomPairFP, AvalonFP, and RDKit-Des.

descriptors, models employing MACCSFP, E-stateFP, AvalonFP, Atom-PairFP, and RDKit-Des achieved higher AUC values, especially for models with the four superior algorithms (Random Forest, XGBoost, SVM, and GBDT, where the AUC values exceeded 0.90 for most models). In contrast, models using MorganFP showed a lower performance (below 0.75). Additional assessment criteria (F1 score listed in Figure 4B) also highlighted the superior performance of models with the algorithms and descriptors aforementioned. Overall, models constructed by Random Forest algorithm with the five descriptors (MACCSFP, E-stateFP, AvalonFP, Atom-PairFP, and RDKit-Des) were level pegging in the internal validation (Figure 4A~B and Table S2, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*). Furthermore, an external validation dataset containing 79 active compounds and 341 inactive compounds was introduced to evaluate the generalization ability of the five models as shown in Figure 4C and Table S3 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*). Apparently, four models presented superior performance with AUC and F1 score about 0.90, whereas the Random Forest-RDKit model was slightly inferior compared with other models, with both assessment criteria being lower than 0.90. Consequently, models based on Random Forest models with four molecular fingerprints (MACCSFP, E-stateFP, Atom-PairFP, and AvalonFP) were selected for the subsequent screening of compounds with potential SARS-CoV-2 PLpro inhibition activity.

In order to obtain the SARS-CoV-2 PLpro inhibitor more efficiently, we adopted a strategy of using multiple models to present the intersection of results to perform the machine learning-based compound screening. Herein, the Venn diagram obtained from the online data visualization tool Venn (*http://www.ehbio.com/test/venn*) was employed. As shown in Figure 5, each of the four machine learning models was able to screen out approximately 100 compounds from the prediction database, among which 42 compounds listed in Table S4 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*) were obtained as the intersection of the four models eventually. In general, the 42 compounds that have been identified simultaneously by the four different classification models tend to have a higher probability of being active against SARS-CoV-2 PLpro. The current strategy that taking the intersection of multiple model predictions can reduce the probability of false positives, and further improve the efficiency of drug discovery.

### 3.3. Molecular simulation-based compound assessment

After screening out the compounds with potential inhibitory activity, molecular docking-based binding affinity evaluation was employed to obtain the compounds with high inhibitory activity targeting SARS-CoV-2 PLpro. Reliability evaluation of the molecular docking protocol used in the present study was performed in the first instance. As displayed in Figure S1 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*), the redocked binding conformations of the two ligands (GRL0617 and Jun12682) were consistent with their original conformations in the co-crystal structures, namely, the molecular docking protocol adopted in the present study can describe the ligand-protein interactions precisely. Subsequently, binding affinities of the 42 compounds screened from

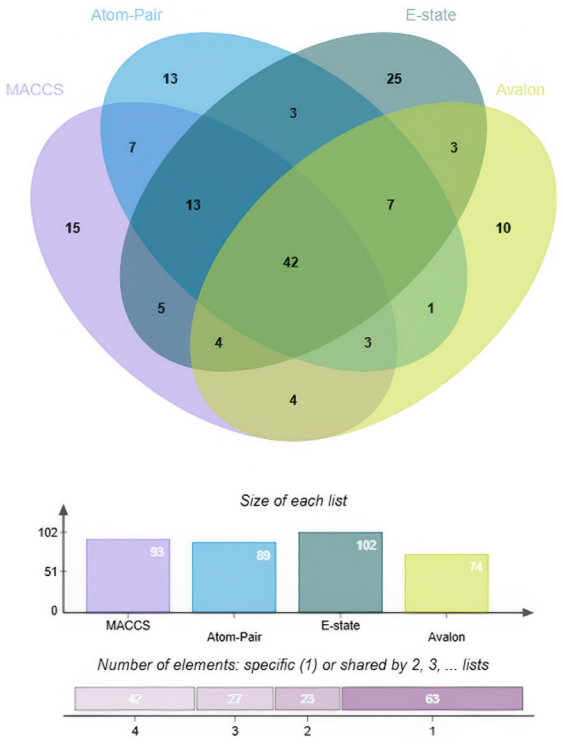


**Figure 5. Interactive Venn diagrams for the intersection of multiple model predictions.**

the machined learning-based classification models were calculated through molecular docking. Fifteen compounds were identified with high inhibitory activity compared with GRL0617, however, no compound was found to be more potent than Jun12682 (binding affinity data were listed in Table S4 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*), and the fifteen compounds were renamed as Cpd-1~15 for convenience in Figure S1 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*).

Considering that the target binding affinity of some compounds (such as Cpd-1, -9.38 kcal/mol) was very close to that of Jun12682 (-9.57 kcal/mol), we further employed a dynamic evaluation method based on molecular dynamics (MD) simulation to provide more accurate assessments on the potential inhibitory activity of the fifteen compounds. A total of 18 systems that contained Cpd-1~15, the two positive compounds GRL0617 and Jun12682, as well as the apo form of the target protein, were simulated through the MD simulation. Confirmed by the root mean square deviation (RMSD), all systems reached the equilibrium state within 100 ns MD simulation (Figure S2, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*). The target binding free energies of Cpd-1~15, GRL0617, and Jun12682 were calculated based on the MM/GBSA method and demonstrated in Figure S3 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*). Under the more accurate evaluation method, only 7 compounds (Cpd-1~4, Cpd-6, Cpd-8, Cpd-14) revealed superior target binding ability than GRL0617, which was significantly different from the result with a molecular docking-based evaluation approach. Surprisingly, Cpd-1 and Cpd-3 exhibited dramatically high target binding ability, suggesting their potential as more potent inhibitors than the current best-performing SARS-CoV-2 PLpro inhibitor Jun12682.

### 3.4. Analysis of ligand-receptor interactions

Figure 6A displays the decomposition of the binding free energy of the seven highly active compounds and the two positive compounds to the target. Apparently, the gas-phase energy component ($\Delta G_{gas}$) is the prime contributor to binding free energy ($\Delta G_{bind}$). Compounds with more tight binding to the target (Cpd-1~3 and Jun12682 with lower $\Delta G_{bind}$) possess significantly low $\Delta G_{gas}$ (about -150 kcal/mol for Cpd-1~3 *vs.* about -70 kcal/mol for others). According to the computational principle, $\Delta G_{gas}$ consists of van der Waals ($\Delta E_{vdw}$) and electrostatic ($\Delta E_{ele}$) terms. Values of the two terms for these compounds are both correlated with the trend of final $\Delta G_{bind}$. Nevertheless, the differences of $\Delta E_{ele}$ term for all the 9 compounds are apparently higher than those of $\Delta E_{vdw}$ (-16.66~-117.34 kcal/mol for $\Delta E_{ele}$ and -31.84~-54.60 kcal/mol for $\Delta E_{vdw}$), indicating that the electrostatic interactions are critical for the ligand binding to the target.

Further ligand-protein interaction analysis was performed to present more detailed descriptions of the binding pattern of the screened compounds with SARS-CoV-2 PLpro. As shown in Figure 6B, all the nine compounds bind to the binding site through some polar interactions such as hydrogen bonds and π–π interactions. Especially, these interactions are extremely abundant in the binding pattern of highly active compounds to the target, which could be a reasonable explanation on the critical role of electrostatic interactions to the ligand binding. For the binding site in target protein, the aromatic side chain of Tyr268 provides CH–π or π–π interactions with most compounds, and Asp164 and Gln269 are conserved in the hydrogen bond interactions of almost all compounds. Additional hydrogen bond occupancy analysis also suggested the pivotal role of these polar residues in ligand binding (Table S5, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*). In particular, Asp302 shows the same binding area as Asp164 as revealed in Figure 6B, contributes a hydrogen bond occupancy of 104.98% in the binding pattern of Cpd-1, and Asp164 donates as high as 179.52% in the binding pattern of Cpd-3. The total hydrogen bond occupancy related to Cpd-1 and Cpd-3 was significantly higher than that of Jun12682 and GRL0617, demonstrating again the superior target binding ability of Cpd-1 and Cpd-3.

### 3.5 Identification of key residues on ligand binding

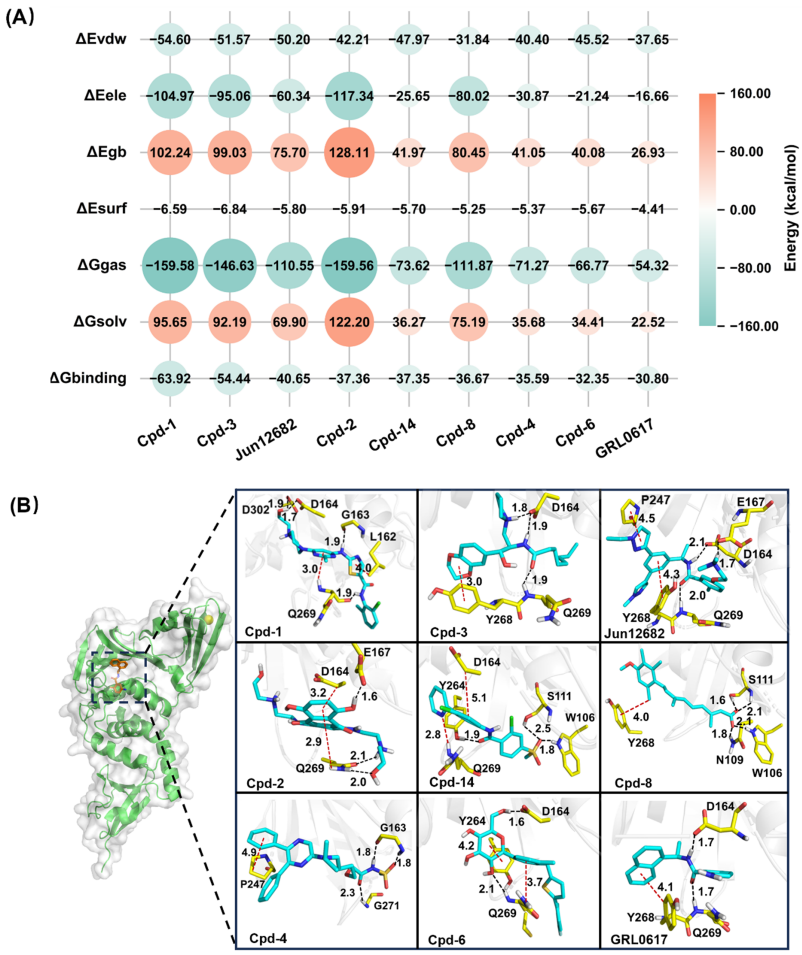


**Figure 6. (A)** Decomposition of binding free energy (kcal/mol) for the nine compounds. **(B)** Binding modes of the nine compounds in the binding site of SARS-CoV-2 PLpro. Ligands and key residues are shown with cyan and yellow stick models, respectively. Hydrogen bonds are represented by black dashed lines, π–π and CH–π interactions are represented by red dashed lines.

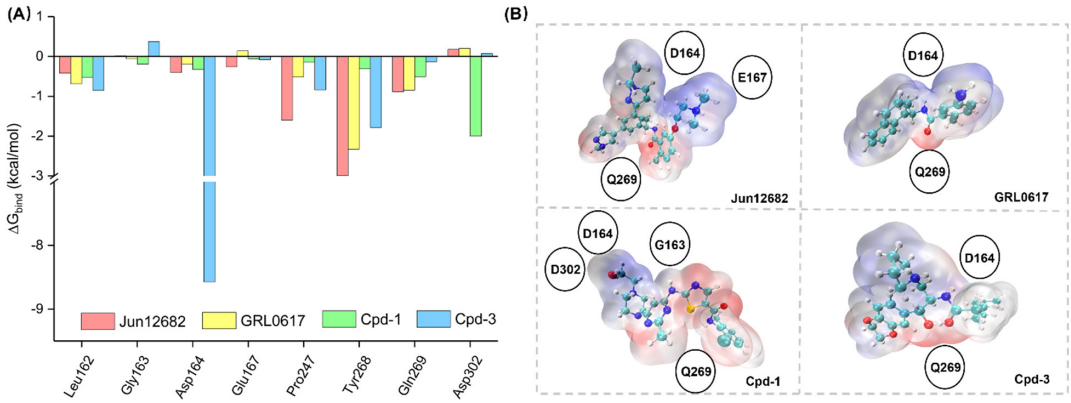


**Figure 7. (A)** Binding free energy contributions of some key residues in the four ligand-receptor binding systems. **(B)** Surface electrostatic potential maps of the four compounds in binding conformations. Electron-deficient and electron-rich regions are colored in blue and red, respectively. Some key residues around them are highlighted in circles.

For a more detailed presentation, binding free energy contributions of some crucial residues in the four ligand-receptor binding systems were calculated and displayed in Figure 7A. Results indicate that most of these residues make favorable contributions to the ligand binding, among which Tyr268 makes significant and conserved contributions to the four compounds. Notably, the contribution of Asp164 on the binding of Cpd-3 is dramatically high among all residues, and Asp302 provides a remarkable contribution to the binding of Cpd-1. The residue binding free energy contributions are consistent with the distributions of hydrogen bond

occupancy aforementioned, which highlights the significance of these polar residues in ligand binding, and also provides reasonable explanations for the high target binding free energy of Cpd-1 and Cpd-3.

Figure 7B illustrates the surface electrostatic potentials of the four compounds in the specific conformation when binding to SARS-CoV-2 PLpro. Apparently, some electronegative residues, like Asp164, Glu167, and Asp302, are situated around the electron-deficient region of the ligand, and Gln269 is close to the electron-rich region. Such an electrical matching mode can disperse the charge and provide a favorable ligand-protein binding pattern. In summary, the favorable electrostatic potential contributions of the key residues would provide valuable insights for the development of more potent and highly selective SARS-CoV-2 PLpro inhibitors.

## 4. Discussion

The current study highlights the significance of targeting the protease PLpro in the development of COVID-19 therapeutics, given the rapid mutation and widespread transmission of SARS-CoV-2. While some candidates, such as GRL0617 and its analogs, have shown weak to moderate *in vitro* potency, they often suffer from limitations that hinder clinical translation, such as insufficient antiviral activity and metabolic stability *in vitro* and *in vivo* (*32*), poor pharmacokinetic performance (*33*), limited selectivity (*34*), or toxicity concerns (*35*). To overcome these limitations, we employed an integrated screening strategy combining machine learning and molecular simulation approaches, which led to the identification of seven promising PLpro inhibitors (Cpd-1~4, Cpd-6, Cpd-8, and Cpd-14). Among them, Cpd-1 and Cpd-3 exhibited the strongest binding affinities and inhibitory potential against PLpro, making them prime candidates for further experimental validation. In addition, we identified several key residues critical for ligand binding, which may inform future optimization efforts aimed at enhancing potency and selectivity.

Although the present findings are promising, they merely represent initial steps towards drug development. To verify the reliability and therapeutic potential of the identified compounds, extensive experimental validation is still required. In this study, we employed a series of molecular simulation techniques—including molecular docking and molecular dynamics simulations—to assess the binding stability of candidate inhibitors with SARS-CoV-2 PLpro. While such computational strategies are highly valuable for identifying promising drug candidates (*36*), their outcomes must be substantiated by experimental data. Therefore, future research will focus on enzymatic assays and cell-based antiviral evaluations to confirm the inhibitory activity and antiviral efficacy of the screened compounds, thereby facilitating their further advancement toward clinical application.

In conclusion, our study contributes to overcoming the limitations that hinder clinical translation in two key ways. First, the integration of machine learning and molecular simulations offers an efficient framework for identifying structurally novel and potentially more drug-like inhibitors. Second, our residue-level interaction analysis provides mechanistic insights that may guide further lead optimization to improve target specificity and binding stability. While experimental validation is still required, our findings offer a solid foundation for the rational development of next-generation PLpro inhibitors. The candidate compounds and structural insights reported here may help accelerate the development of effective antivirals targeting SARS-CoV-2 and its evolving variants.

## Supporting Information

Detailed information on molecular fingerprints (Table S1), performance of classification models in internal validation (Table S2) and external validation (Table S3). Molecular docking scores of the 42 compounds from intersection of the four optimal machine learning models (Table S4), hydrogen bond occupancy analysis of the polar residues involved in ligand binding (Table S5). Redock analysis and docking scores of the screened compounds (Figure S1), RMSD of the 18 MD simulated systems (Figure S2), and binding free energy of the top 15 compounds (Figure S3). (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=261*)

Lists of the initial database (Supplementary Materials 1), training set (Supplementary Materials 2), external validation set (Supplementary Materials 3), and prediction dataset (Supplementary Materials 4) used in the machine learning models. (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=262*)

*Funding*: This work was supported by research grants of Natural Science Foundation of Shandong Province (ZR2024QH448), the Key Program of Brain Science and Technology (STI2030-Major Projects 2021ZD0202900) and Key Project by Qingdao Bureau of Sciences and Technology (22-3-3-HYGG-25-HY).

*Conflict of Interest*: The authors have no conflicts of interest to disclose.

## References

1. Zhu C, Pang S, Liu J, Duan Q. Current Progress, Challenges and prospects in the development of COVID-19 vaccines. Drugs. 2024; 84:403-423.
2. Kaku Y, Uriu K, Kosugi Y, Okumura K, Yamasoba D, Uwamino Y, Kuramochi J, Sadamasu K, Yoshimura K, Asakura H, Nagashima M, Ito J, Sato K. Virological characteristics of the SARS-CoV-2 KP.2 variant. Lancet Infect Dis. 2024; 24:e416.

3. Yang H, Rao Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat Rev Microbiol. 2021; 19:685-700.
4. Shin D, Mukherjee R, Grewe D, *et al*. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature. 2020; 587:657-662.
5. Protić S, Crnoglavac Popović M, Kaličanin N, Prodanović O, Senćanski M, Milićević J, Stevanović K, Perović V, Paessler S, Prodanović R, Glišić S. SARS-CoV-2 PLpro inhibition: evaluating *in silico* repurposed fidaxomicin's antiviral activity through *in vitro* assessment. ChemistryOpen. 2024; 13:e202400091.
6. Freitas BT, Durie IA, Murray J, Longo JE, Miller HC, Crich D, Hogan RJ, Tripp RA, Pegan SD. Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. ACS Infect Dis. 2020; 6:2099-2109.
7. Shen Z, Ratia K, Cooper L, Kong D, Lee H, Kwon Y, Li Y, Alqarni S, Huang F, Dubrovskyi O, Rong L, Thatcher G, Xiong R. Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity. J Med Chem. 2022; 65:2940-2955.
8. Tan B, Zhang X, Ansari A, Jadhav P, Tan H, Li K, Chopra A, Ford A, Chi X, Ruiz FX, Arnold E, Deng X, Wang J. Design of a SARS-CoV-2 papain-like protease inhibitor with antiviral efficacy in a mouse model. Science. 2024; 383:1434-1440.
9. Xu Y, Chen K, Pan J, Lei Y, Zhang D, Fang L, Tang J, Chen X, Ma Y, Zheng Y, Zhang B, Zhou Y, Zhan J, Xu W. Repurposing clinically approved drugs for COVID-19 treatment targeting SARS-CoV-2 papain-like protease. Int J Biol Macromol. 2021; 188:137-146.
10. Cho CC, Li SG, Lalonde TJ, Yang KS, Yu G, Qiao Y, Xu S, Ray Liu W. Drug repurposing for the SARS-CoV-2 papain-like protease. ChemMedChem. 2022; 17:e202100455.
11. Srinivasan V, Brognaro H, Prabhu PR, *et al*. Antiviral activity of natural phenolic compounds in complex at an allosteric site of SARS-CoV-2 papain-like protease. Commun Biol. 2022; 5:805.
12. Zhao Y, Du X, Duan Y, *et al.* High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. Protein Cell. 2021; 12:877-888.
13. Weglarz-Tomczak E, Tomczak JM, Talma M, Burda-Grabowska M, Giurg M, Brul S. Identification of ebselen and its analogues as potent covalent inhibitors of papain-like protease from SARS-CoV-2. Sci Rep. 2021; 11:3640.
14. Sargsyan K, Lin CC, Chen T, Grauffel C, Chen YP, Yang WZ, Yuan HS, Lim C. Multi-targeting of functional cysteines in multiple conserved SARS-CoV-2 domains by clinically safe Zn-ejectors. Chem Sci. 2020; 11:9904-9909.
15. Yi Y, Zhang M, Xue H, *et al.* Schaftoside inhibits 3CLpro and PLpro of SARS-CoV-2 virus and regulates immune response and inflammation of host cells for the treatment of COVID-19. Acta Pharm Sin B. 2022; 12:4154-4164.
16. Kuo CJ, Chao TL, Kao HC, Tsai YM, Liu YK, Wang LH, Hsieh MC, Chang SY, Liang PH. Kinetic characterization and inhibitor screening for the proteases leading to identification of drugs against SARS-CoV-2. Antimicrob Agents Chemother. 2021; 65:e02577-20.
17. Yu W, Zhao Y, Ye H, Wu N, Liao Y, Chen N, Li Z, Wan N, Hao H, Yan H, Xiao Y, Lai M. Structure-Based Design of a Dual-Targeted Covalent Inhibitor Against Papain-like and Main Proteases of SARS-CoV-2. J Med Chem. 2022; 65:16252-16267.
18. Rut W, Lv Z, Zmudzinski M, Patchett S, Nayak D, Snipas SJ, El Oualid F, Huang TT, Bekes M, Drag M, Olsen SK. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: a framework for anti-COVID-19 drug design. Sci Adv. 2020; 6:eabd4596.
19. Li Y, Li L, Wang S, Tang X. EQUIBIND: A geometric deep learning-based protein-ligand binding prediction method. Drug Discov Ther. 2023; 17:363-364.
20. Boby ML, Fearon D, Ferla M, *et al*. Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. Science. 2023; 382:eabo7201.
21. Garnsey MR, Robinson MC, Nguyen LT, *et al.* Discovery of SARS-CoV-2 papain-like protease (PLpro) inhibitors with efficacy in a murine infection model. Sci Adv. 2024; 10:eado4288.
22. Zhao J, Shi X, Wang Z, Xiong S, Lin Y, Wei X, Li Y, Tang X. Hepatotoxicity assessment investigations on PFASs targeting L-FABP using binding affinity data and machine learning-based QSAR model. Ecotoxicol Environ Saf. 2023; 262:115310.
23. Wei X, Liu N, Feng Y, Wang H, Han W, Zhuang M, Zhang H, Gao W, Lin Y, Tang X, Zheng Y. Competitive-like binding between carbon black and CTNNB1 to ΔNp63 interpreting the abnormal respiratory epithelial repair after injury. Sci Total Environ. 2024; 929:172652.
24. Hu H, Wang Q, Su H, Shao Q, Zhao W, Chen G, Li M, Xu Y. Identification of cysteine 270 as a novel site for allosteric modulators of SARS-CoV-2 papain-like protease. Angew Chem. 2022; 61:e202212378.
25. Li Y, Wang Z, Ma S, Tang X, Zhang H. Chemical space exploration and machine learning-based screening of PDE7A inhibitors. Pharmaceuticals. 2025; 18:444
26. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455-461.
27. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem. 2003; 24:1999-2012.
28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926-935.
29. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem. 1993; 97:10269-10280.
30. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977; 23:327-341.
31. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015; 10:449-461.
32. Velma GR, Shen Z, Holberg C, *et al.* Non-covalent inhibitors of SARS-CoV-2 papain-like protease (PLpro): *in vitro* and *in vivo* antiviral activity. J Med Chem. 2024; 67:13681-13702.
33. Garnsey MR, Robinson MC, Nguyen LT, *et al.* Discovery of SARS-CoV-2 papain-like protease (PLpro) inhibitors with efficacy in a murine infection model. Sci Adv. 2024; 10:eado4288.

34. Zmudzinski M, Rut W, Olech K, *et al.* Ebselen derivatives inhibit SARS-CoV-2 replication by inhibition of its essential proteins: PLpro and Mpro proteases, and nsp14 guanine N7-methyltransferase. Sci Rep. 2023; 13:9161.
35. Hu H, Wang Q, Su H, Shao Q, Zhao W, Chen G, Li M, Xu Y. Identification of cysteine 270 as a novel site for allosteric modulators of SARS-CoV-2 papain-like protease. Angew Chem Int Ed Engl. 2022; 61:e202212378.
36. Li Z, Li X, Huang YY, *et al.* Identify potent SARS-CoV-2 main protease inhibitors *via* accelerated free energy perturbation-based virtual screening of existing drugs. Proc Natl Acad Sci U S A. 2020; 17:27381-27387.

Received April 16, 2025; Revised May 21, 2025; Accepted June 22, 2025.

**Address correspondence to:*
Yongxin Bao, Qingdao Women and Children's Hospital Affiliated to Qingdao University, Qingdao 266034, China.
E-mail: byx285788575@163.com

Xiaowen Tang, School of Pharmacy, and Shandong Provincial Key Laboratory of Pathogenesis and Prevention of Brain Diseases, Qingdao University, Qingdao 266071, China.
E-mail: xwtang1219@qdu.edu.cn

Released online in J-STAGE as advance publication June 27, 2025.