



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

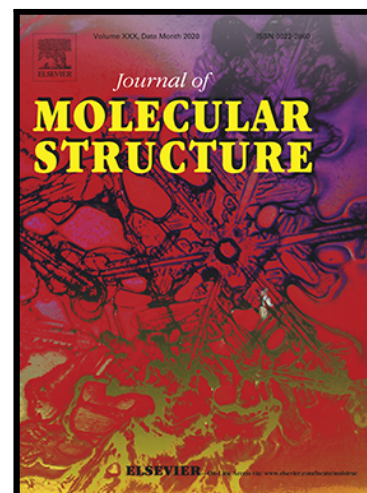
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Journal Pre-proof

Computational strategies towards developing novel SARS-CoV-2
M^{PRO} inhibitors against COVID-19

Ding Luo , Jian-Bo Tong , Xing Zhang , Xue-Chun Xiao ,
Shuai Bian Supervision

PII: S0022-2860(21)01506-4
DOI: <https://doi.org/10.1016/j.molstruc.2021.131378>
Reference: MOLSTR 131378



To appear in: *Journal of Molecular Structure*

Received date: 6 July 2021
Revised date: 15 August 2021
Accepted date: 17 August 2021

Please cite this article as: Ding Luo , Jian-Bo Tong , Xing Zhang , Xue-Chun Xiao ,
Shuai Bian Supervision , Computational strategies towards developing novel SARS-
CoV-2 M^{PRO} inhibitors against COVID-19, *Journal of Molecular Structure* (2021), doi:
<https://doi.org/10.1016/j.molstruc.2021.131378>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Elsevier B.V. All rights reserved.

Computational strategies towards developing novel SARS-CoV-2 M^{pro} inhibitors against COVID-19

Ding Luo^{1,2}, Jian-Bo Tong^{*1,2}, Xing Zhang^{1,2}, Xue-Chun Xiao^{1,2}, Shuai Bian^{1,2}

¹ College of Chemistry and Chemical Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China

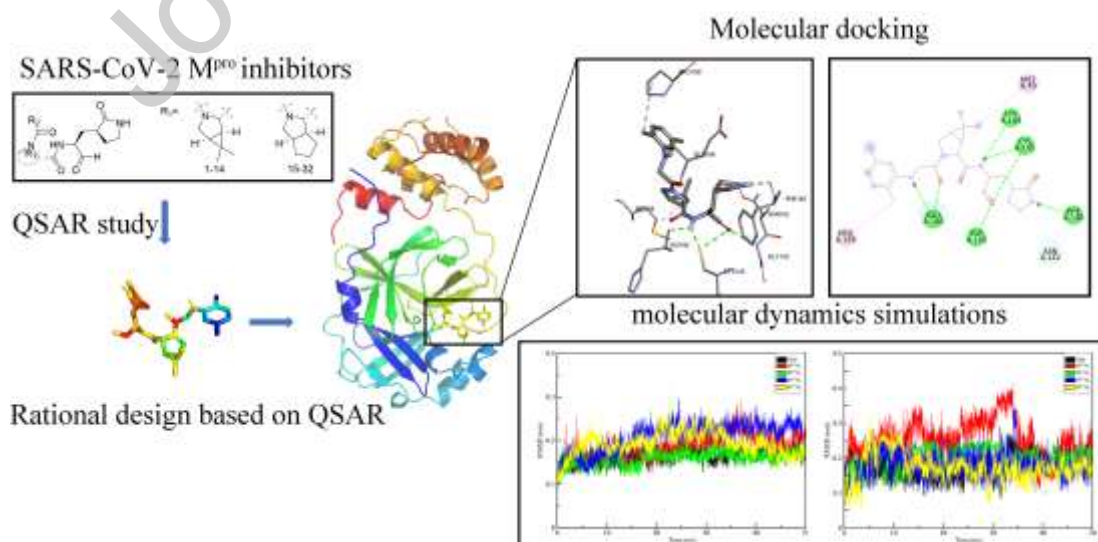
² Shaanxi Key Laboratory of Chemical Additives for Industry, Xi'an 710021, China

*Corresponding author: Jian-Bo Tong^{*1,2}; E-mail: jianbotong@aliyun.com

Highlights

- Four QSAR models are used to predict enzyme inhibition of SARS-CoV-2 M^{pro}.
- Key amino acids influencing the receptor-ligand interactions have been identified.
- The binding affinity is more dependent on hydrophobic interactions.
- Six novel SARS-CoV-2 M^{pro} inhibitors were designed.

Graphical Abstract



Abstract: The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remains to be a serious threat due to the lack of a specific therapeutic agent. Computational methods are particularly suitable for rapidly fight against SARS-CoV-2. This present research aims to systematically explore the interaction mechanism of a series of novel bicycloproline-containing SARS-CoV-2 M^{pro} inhibitors through integrated computational approaches. We designed six structurally modified novel SARS-CoV-2 M^{pro} inhibitors based on the QSAR study. The four designed compounds with higher docking scores were further explored through molecular docking, molecular dynamics (MD) simulations, free energy calculations, and residual energy contributions estimated by the MM-PBSA approach, with comparison to compound 23(PDB entry 7D3I). This research not only provides robust QSAR models as valuable screening tools for the development of anti-COVID-19 drugs, but also proposes the newly designed SARS-CoV-2 M^{pro} inhibitors with nanomolar activities that can be potentially used for further characterization to treat SARS-CoV-2 virus.

Keywords: QSAR • SARS-CoV-2 • Molecular docking • Molecular dynamics simulations • MM/PBSA

1. INTRODUCTION

The unprecedented coronavirus disease (COVID-19) that spreads rapidly around the globe has led to a severe health crisis [1]. The COVID-19 pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which shares approximately 82% sequence identity with SARS-CoV [2]. The surprising speed of SARS-CoV-2 is

attributed to a large number of mild or asymptomatic infections [3]. Remdesivir, which demonstrates pan-coronavirus suppression potential, and has obtained FDA approval for treating COVID-19 [4, 5]. More data is still needed to prove its effectiveness. Although the strategy of using other existing drugs for the treatment of this disease has been exploited, effective treatment options are still limited [6, 7]. It is imminent to provide effective and safe antiviral drugs against SARS-CoV-2.

For coronavirus, the cleavage of two polyproteins (pp1a/pp1ab) into a single non-structural protein is critical for the assembly of the viral replication–transcription [8], which is processed by main protease (M^{pro} , 3CL pro) and papain-like protease (PLpro), thus are essential in the normal expression of SARS-CoV-2 virus [9]. Especially, M^{pro} was identified as an attractive target for antiviral drug design on account of its vital role and no human homolog [10, 11]. Compared with complex drug synthesis process, the computational approaches such as QSAR (quantitative structure-activity relationship), molecular docking, pharmacophore modeling [12-14], etc., are effective for finding new drug targets and reusing existing drugs [15], in the development of the anti-SARS-CoV-2 drugs. Furthermore, despite the SARS-CoV and SARS-CoV-2 have approximately 82% sequence similarity of M^{pro} , most of the computational strategy is to leverage drugs and biomolecules previously reported to have activity against related coronaviruses [16-20], and few reports on data specific to SARS-CoV-2 M^{pro} . In addition, previous literature on inhibitors of SARS-CoV M^{pro} has not included infection data in an animal model. The synthesized compounds tested in the animal model have more significance to accelerate the development of novel therapies against COVID-19.

Amin et al. [21] performed the first structure-activity relationship analysis on the structure-activity data of recently reported diverse SARS-CoV-2 M^{pro} inhibitors to understand the structural requirements for higher inhibitory activity. However, the activity deviation between different data sets is not considered and it is unavoidable. Herein, we chose a series of 32 new SARS-CoV-2 M^{pro} inhibitors synthesized by Qiao et.al [22] to investigate the structural requirements of such inhibitors against SARS-CoV-2 M^{pro} using computational methods. Those SARS-CoV-2 M^{pro} inhibitors were based on the reported crystal structures of SARS-CoV-2 M^{pro} and cocrystal structures of SARS-CoV-2 M^{pro} in complex with the approved antivirals against hepatitis C virus infection, boceprevir (PDB entry 7COM) and telaprevir (PDB entry 7C7P), and two compounds were tested for antiviral activity in mice. The cocrystal structure of M^{pro} in complex with the synthesized compound 23, which is one of the most potent compounds, was released to revealed its interaction modes (PDB entry 7D3I). This provides new information for the rational design of new type SARS-CoV-2 M^{pro} inhibitors based on its good inhibitory effect and binding mode revealed in detail.

Molecular modeling in combination to molecular dynamics simulation has been widely used in computer-aided drug design [23, 24]. QSAR is one of the effective methods for predicting the activity of compounds when data and appropriate experimental facilities are lacking [25]. Whereas molecular docking is a very fast and efficient tool for predicting how a small molecule will have a conformational geometry on the active site of the target protein [26]. These computational methods are particularly suitable for rapidly fighting against SARS-CoV-2. Therefore, the present

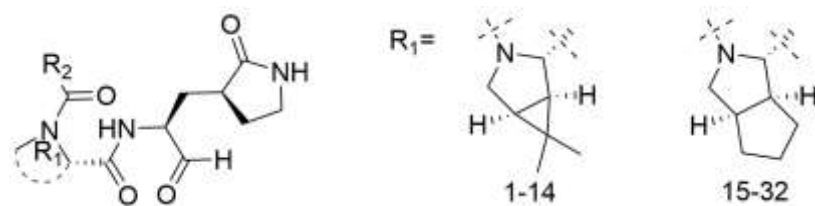
study was conducted using QSAR coupled with molecular docking approach, MD simulation, and free binding energy MM/PBSA analysis, to investigate the structural requirements of SARS-CoV-2 M^{pro} inhibitors and to design novel compounds that can be helpful against SARS-CoV-2 infection. The outcomes could provide insightful direction to design more active peptide-type compounds which can be further synthesized and screened for their potential utilization in the SARS-CoV-2 M^{pro} inhibition.

2. MATERIALS AND METHODS

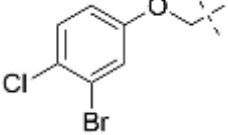
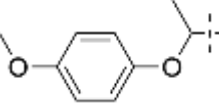
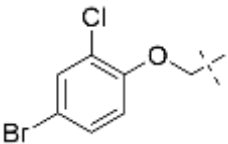
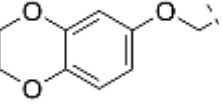
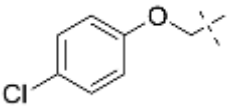
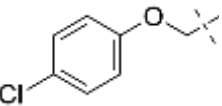
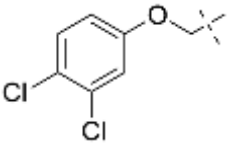
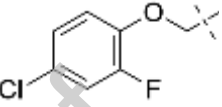
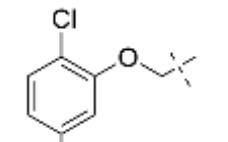
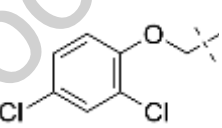
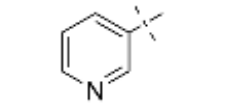
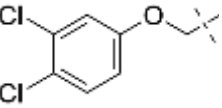
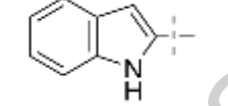
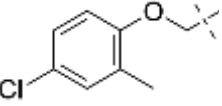
2.1 Data set

The experimental activity data to perform the molecular modeling of 32 new bicycloproline-containing M^{pro} inhibitors derived from either boceprevir or telaprevir were taken from the recent literature [22]. The corresponding half-maximal inhibitory concentration (IC₅₀) nM value was converted into pIC₅₀ (-log IC₅₀) value, as the dependent variable in the analysis of QSAR models. 32 compounds were divided into a training set composed of 24 compounds (75%) for establishing QSAR models and a test set composed of 8 compounds (25%) for model validation (Fig. 1). The test and training sets were chosen in such a way that the chemical diversity and a good range of activity were maintained throughout both sets. Outliers are generally removed as they result in ambiguity in statistical models. The range of the biological activities spans 1.99 log units, from 6.13 to 8.12. The structures of all compounds and their activity values (pIC₅₀) are listed in Table 1.

Table 1: Structures and activities of SARS-CoV-2 M^{pro} inhibitors.



Com p.	R ₂	IC ₅₀ (n M)	pIC ₅₀	Com p.	R ₂	IC ₅₀ (nM)	pIC ₅₀
1		453	6.34	17		298.8	6.53
2		52.1	7.28	18		525.9	6.28
3*		16.5	7.78	19*		195.6	6.71
4		18.5	7.73	20		375.0	6.43
5		13.2	7.88	21		7.6	8.12
6		14.5	7.84	22*		17.4	7.76
7*		43.3	7.36	23		7.6	8.12
8		37.2	7.43	24*		378.2	6.42
9		15.2	7.82	25		36.2	7.44

10		50.8	7.29	26		69.1	7.16
11*		13.3	7.88	27		93.5	7.21
12		19.0	7.72	28*		9.2	8.04
13		12.4	7.91	29		34.7	7.46
14		13.0	7.89	30		17.2	7.76
15*		748.5	6.13	31		30.0	7.52
16		153.1	6.82	32		19.7	7.71

*: selected as the test set.

2.2 Molecular optimization

The 3D conformation of each compound was generated by the SketchTool module in the SYBYL software package. Partial atomic charges were calculated by Gasteiger–Huckel method. In addition, under the Tripos force field the Powell conjugate gradient minimization algorithm minimized the potential energies of the compounds. The energy convergence criterion was 0.005 kcal/mol Å with the maximum number of iterations at 1000 [27]. The IC₅₀ values were all converted to pIC₅₀ (-LogIC₅₀) and added to the spreadsheet as compound variables. Other parameters adopt the default values of

SYBYL 2.0

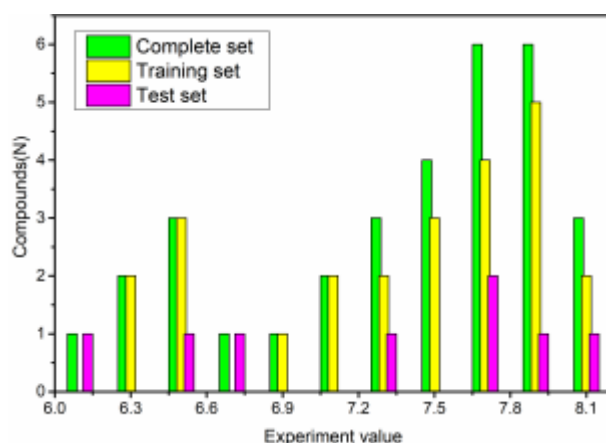


Fig. 1: The distribution of experimental inhibitory activity (pIC_{50}) of the training and test set compounds in QSAR models.

2.3 Molecular alignment

The alignment was carried out by the distill module implemented in SYBYL 2.0 in which a common core scaffold was identified among all the 32 molecules and was shown in Fig. 2(a). The lowest conformation of compound 23 was selected as a template because the most active compound of the data set. The remaining compounds of the dataset were aligned on this highest active compound based upon the common substructure similarity (Fig. 2(b)).

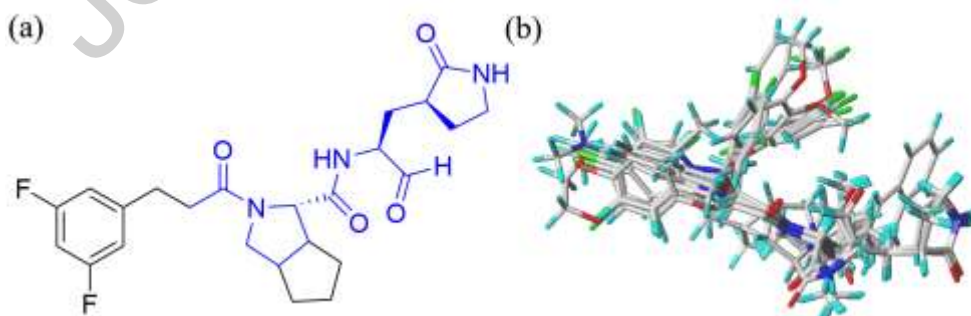


Fig. 2: Molecular alignment mode. (a): 2D structure of compound 23 (blue represents the common skeleton); (b): Molecular alignment based on compound 23.

2.4 CoMFA and CoMSIA descriptors

The CoMFA (Comparative molecular field analysis) and CoMSIA (Comparative molecular similarity index analysis) studies have been performed based on the molecular alignment using SYBYL software to explore the contributions of diverse interactions. CoMFA and CoMSIA fields were calculated for each compound in a dataset at every grid point using an sp^3 -hybridized carbon probe with a charge of +1.0 [28]. The stereo field action is calculated by Lennard-Jones Eq. (1), and the electrostatic field calculation adopts the Coulomb Eq. (2).

$$E_{vdw} = \sum_{i=1}^n (A_j r_{ij}^{-12} - C_j r_{ij}^{-6}) \quad (1)$$

$$E_C = \sum_{i=1}^n \frac{q_i q_j}{D r_{ij}} \quad (2)$$

Where E_{vdw} represents the space potential energy of the compound, A_j and C_j represent the corresponding atomic van der Waals radius constants, n is the total number of atoms in the molecule, E_C is the electrostatic field energy of the compound, q_i is the net charge of the atom calculated by the Gasteiger-Huckel method, q_j is the probe charge of the atom, and D is the dielectric constant.

In the CoMSIA analysis, descriptors such as spatial, electrostatic, hydrophobic, hydrogen bond donor, and acceptor fields were obtained [29]. Compared to CoMFA, CoMSIA uses a Gaussian function and the similarity index of the Gaussian distribution to avoid mutations in the grid-based probe-atom interaction [30].

2.5 Topomer CoMFA descriptors

The Topomer CoMFA modeling is a combination of the spatial, electrostatic, and

topological properties of molecules. Topomer CoMFA modeling eliminates the steps that affect the prediction results, such as molecular overlap, and defines the activity value of the compound as the contribution value represented by each fragment through molecule cutting [31]. Topomer is similar to CoMFA in creating a spatial contour map produced by a stereocenter and position, with the Tripos force field method of +1 atomic charge. Importantly, the robustness of the generated QSAR model depends on the appropriate cutting method. Herein, we chose the highest M^{pro} inhibitors (compound 23) as the template for cutting. According to the structural characteristics of the inhibitors, the 32 compounds are divided into fragments by the same cutting method (Fig. S1).

2.6 HQSAR descriptors

HQSAR (Holographic QSAR) converts the molecular structure into characteristic molecular fingerprints according to the types of molecular fragments and marks with numbers (53-401) [32]. These marked numbers are used as QSAR descriptors to establish the corresponding structure-activity relationship to predict the biological activity of the compounds. Different fragment descriptors that including Atoms (A), Bonds(B), Connections (C), Chirality (Ch), Hydrogen atoms(H), Donor and Acceptor (DA) were utilized in combination with adjusted dimensions and holographic lengths to obtain an HQSAR model with good predictive ability.

2.7 Partial least square (PLS) analysis

The Partial least square method [33] is applied to generate QSAR models as an extension of multiple regression analysis. The generated CoMFA, CoMSIA, Topomer

CoMFA, and HQSAR descriptors were set as independent variables and M^{pro} inhibitory activity (pIC_{50}) as the dependent variable. Cross-validation correlation coefficient q^2 (Eq. (3)) and the optimal number of components (ONC) was executed by Leave-one-out (LOO) cross-validation analysis. In order to further verify the model internally, the non-cross-validated and bootstrap PLS analysis was performed using ONC obtained from the LOO method. The non-cross-validation correlation coefficient (r^2) (Eq. (4)) value, the standard error estimate (SEE), the F-test value (F), and contributions of each field were determined to better evaluate the accuracy and robustness of the developed models. Usually, high q^2 and r^2 ($q^2 > 0.5$, $r^2 > 0.6$) values are considered to be the proof that the high predictive ability of the QSAR model is in line with expectations [34].

$$q^2 = 1 - \frac{\sum_{n=1}^{\infty} (\hat{y}_i - y_i)^2}{\sum_{n=1}^{\infty} (y_i - \bar{y})^2} \quad (3)$$

$$r^2 = \frac{[\sum (y_i - \bar{y}_i) (\hat{y}_i - \hat{y})]^2}{\sum (y_i - \bar{y}_i)^2 \times \sum (\hat{y}_i - \hat{y})^2} \quad (4)$$

whereas, \hat{y}_i and y_i are predicted, observed activity values, and \bar{y} and \hat{y} are observed and predicted mean activity values of the training set, respectively.

2.8 Validation of the QSAR models

Validation of the QSAR model is essential for understanding the predictive power of the established QSAR models. Critical evaluation of developed models involves internationally recognized internal and external validation parameters to test the robustness of the model in terms of fitness, stability, classical fitness measurement, and predictability. External verification parameters r_{pred}^2 was calculated by Eq. (5) to judge the quality of external verification capabilities.

$$r_{pred}^2 = \left(\frac{SD - PRESS}{SD} \right) \quad (5)$$

SD represents the sum of square deviations between the biological activity of the molecules in the test set and the average activity of the molecules in the training set, and PRESS is the sum of the square derivatives of the predicted and actual activities of the molecules in the test set. The established QSAR model considered to have the desired predictive ability with $r_{pred}^2 > 0.6$ [35].

Other external parameters including Q_{F1}^2 , Q_{F2}^2 , concordance correlation coefficient (CCC), Δr_m^2 metrics, the root means square error (RMSE), mean absolute error (MAE) [35] (RMSE and MAE are close to zero) for both training and test set compounds were calculated by Eq. (6)-(12):

$$Q_{F1}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_{TR})^2} \quad (6)$$

$$Q_{F2}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_{EXT})^2} \quad (7)$$

$$r_m^2 = r^2 \left(1 - \sqrt{|r^2 - r_o^2|} \right) \quad (8)$$

$$\Delta r_m^2 = |r_m^2 - r_m^{2'}| \quad (9)$$

$$CCC = \frac{2 \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \hat{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - \hat{y})^2 + n(\bar{y} - \hat{y})^2} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (11)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (12)$$

Where \bar{y}_{TR} and \bar{y}_{EXT} indicate the response means of the training set and the external test set, respectively. The $r_m^2 > 0.5$, $\Delta r_m^2 < 0.2$, Q_{F1}^2 , Q_{F2}^2 values more than 0.7 usually represent that the model has ideal external predictability [36].

The Golbraikh-Tropsha method [37] was used to calculate other verification parameters of the QSAR model, where conditions are satisfied: $R^2 > 0.6$, $(R^2 - R_0^2/R^2) < 0.1$ and $0.85 \ll k \ll 1.15$ or $((R^2 - R_0'^2)/R^2 < 0.1)$ and $0.85 \ll k' \ll 1.15$, $|R_0^2 - R_0'^2| < 0.1$. $k, k', R_0^2, R_0'^2$ was calculated by Eq. (13) - (16).

$$k = \frac{\sum(y_i \times \hat{y}_i)}{\sum(\hat{y}_i)^2} \quad (13)$$

$$k' = \frac{\sum(y_i \times \hat{y}_i)}{\sum(y_i)^2} \quad (14)$$

$$R_0^2 = 1 - \frac{\sum(y_i - k \times \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (15)$$

$$R_0'^2 = 1 - \frac{\sum(\hat{y}_i - k' \times y_i)^2}{\sum(\hat{y}_i - \hat{y})^2} \quad (16)$$

R^2 represents the correlation between the experimental and predicted activities of the test set; k and k' represent the regression slope of the experimental activity and the predicted activity or the regression slope of the predicted activity and the observed activity through the origin. R^2 and R_0^2 represent the correlation coefficients between the observed and predicted activity of the test set without intercept.

2.9 Molecular docking

We have implemented molecular docking studies to explore the interaction pattern of compounds with their relevant enzyme (M^{pro}). Choosing an appropriate docking method is a crucial factor in determining whether the molecular docking simulation is realistic. We attempt to select the most suitable one from the following docking protocols to evaluate the binding mode between protein and ligand:

(1) Surflex-Dock: Surflex-Dock is a Protomol-based method with fast and accurate molecular docking mean, which uses a unique empirical scoring function and a patented search engine (based on the molecular similarity search engine) to dock ligand in the

binding site of the target protein [38].

(2) Autodock 4.2: AutoDock4.2 relies on many approximate values to predict the conformation and binding free energy during docking simulation. Ligands are considered to be flexible. But different from traditional molecular mechanics methods, only torsion degrees of freedom is discussed to keep the bond angle and bond length constant [39]. The empirical free energy force field is based on a molecular mechanics force field, which includes typical terms for dispersion/repulsion, hydrogen bonding, electrostatics, desolvation, and torsional entropy.

(3) Autodock Vina: Vina improves the average accuracy of combined mode prediction, and speeds up the search by using a simpler scoring function [40]. Vina's scoring function is completely empirical, including Gaussian spatial interaction, repulsion, hydrogen bonding, hydrophobic, and torsion terms. Furthermore, Vina has parallel calculation functions and is very easy to use [41].

The energies of the prepared proteins and ligands are minimized and unnecessary water molecules are removed. Under the Kollman united force field with the non-bonding cut-off value set to 8.0 and the dielectric constant set to 1.0, the protein performs 1000 energy minimization iterations according to the gradient of the Powell method. The grids were chosen to be sufficiently large to include not only the active site but also a significant portion of the surrounding surface at receptor protein with grid points $60 \times 60 \times 60$, along with a grid spacing of 0.375 Å.

2.10 Molecular dynamics simulation

The MD simulation was performed in Gromacs 5.15 [42] with CHARMM36 (Feb-

2021) force field [43] to investigate the interaction between the receptor and ligand, and the time step is set to 2 fs. The topology files of the protein and ligand were obtained by executing the `pdb2gmx` command and employing the CHARMM General Force Field [44], respectively. The system is immersed in a periodic boundary condition (PBC) 12-sided box composed of Tip3p water model, followed by ionization and neutralization in Na^+ and Cl^- ions. The energy minimization was performed for 50,000 steps using the steepest descent algorithm. Two equilibration phases for NVT and NPT are in sequence to balance the temperature of 300 K and the atmospheric pressure of 1 atm. The simulation of each selected complex was performed for 50 ns.

Molecular dynamics trajectories were analyzed using tools in the GROMACS suite, including root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent-accessible surface area (SASA), number of hydrogen bonds, and Radius of gyration (Rg). PCA/essential dynamics statistical tools are used to identify similarities and differences in specific data sets, and to detect patterns in specific data sets. The principal component analysis (PCA) method in Gromacs was utilized to analyze the eigenvectors and eigenvalues to obtain the projection of the first two principal components (PC1 and PC2) [45]. The trajectories on eigenvectors were projected by `gmx anaeig` module, and covariance matrix was constructed by `Gmx covar` module.

The molecular mechanics/Poisson-Boltzmann surface area (MM-PBSA) method was performed to estimate the binding free energy (ΔG_{bind}) of the SARS-CoV-2 M^{pro}-ligand complex, which is a widely used endpoint method in free energy calculations [46, 47]. The binding free energies of the complex in liquid/solution were determined with the

g_mmpbsa tool [48] by Eq. (17):

$$\begin{aligned}\Delta G_{\text{bind}} &= \Delta G_{\text{complex}} - (\Delta G_{\text{protein}} + \Delta G_{\text{ligand}}) \quad (17) \\ &= \Delta E_{\text{MM}} + \Delta E_{\text{polar}} + \Delta E_{\text{SASA}} - T\Delta S \\ &= \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} + \Delta E_{\text{polar}} + \Delta E_{\text{SASA}} - T\Delta S\end{aligned}$$

ΔG_{bind} represents the net free binding energy, and ΔE_{MM} indicates molecular mechanics potential energy, which consists of van der Waals (ΔE_{vdw}) and electrostatic energy (ΔE_{elec}). ΔE_{polar} and ΔE_{SASA} indicate polar solvation and nonpolar solvation contributions, respectively. $T\Delta S$ refers to conformational entropy contribution at temperature T. The contribution of $T\Delta S$ term to the binding energy can be ignored when calculating relative binding affinity energy [49]. To further confirm the key residues involved in the binding modes, the binding free energy decomposition was calculated, which utilized the python script `MmPbSaDecomp.py` (available at http://rashmikumari.github.io/g_mmpbsa) after the calculation of every term of binding energy contribution. The amino acid residues whose binding energy contribution is less than -2 KJ/mol and greater than 2 KJ/mol are selected to reveal their favorable and unfavorable contributions in the simulation process.

2.11 In silico ADMET prediction

ADMET (drug absorption, distribution, metabolism, excretion, and toxicity) pharmacokinetic property is a very essential means in contemporary drug design and drug screening. The early ADMET property evaluation can effectively solve the problem of species differences, significantly improve the success rate of drug development, and reduce drug development costs, drug toxicity, and side effects, The

ADMET properties of newly designed compounds were predicted using the Online web server ADMETlab2.0 [50], followed SwissADME online tool [51] for the prediction of their drug-likeness and synthetic difficulty.

3. RESULTS AND DISCUSSION

3.1 Statistical results of CoMFA and Topomer CoMFA models

The statistical values of CoMFA and Topomer CoMFA models constructed by steric and electrostatic descriptors are listed in Table 2. The CoMFA model possessed the cross-validated q^2 value of 0.590 with nine components, a non-cross-validated correlation coefficient r^2 value of 0.966, a lower SEE value of 0.044, and a significant F value of 414.196, which indicates the good statistical correlation of the models. The contributions of the steric and electrostatic fields are 0.664 and 0.336, respectively. Furthermore, high bootstrapped r^2 (r_{bs}^2) value of 0.992 and SEE_{bs} of 0.013 indicate a high confidence level in the evaluation for the established CoMFA Model. The Topomer CoMFA model results in a q^2 value of 0.522 with components of 5, high r^2 of 0.952 with a low SEE value of 0.139, and F values of 74.147. These values initially recommend that the generated CoMFA and Topomer CoMFA models are supposed to have competent predictive ability.

Table 2: Statistical parameters of QSAR models.

Parameters	CoMFA	Topomer COMFA	CoMSIA	HQSAR 2-10
q^2	0.590	0.522	0.571	0.827
r^2	0.966	0.952	0.989	0.945
SEE	0.044	0.139	0.074	0.145
N	9	5	8	4
F	414.196	71.147	173.352	-
r_{bs}^2 (100 run)	0.992		0.998	
SEE_{bs}	0.013		0.003	

Field contributions				
Steric	0.664	-	0.321	-
Electrostatic	0.336	-	-	-
Hydrophobic	-	-	0.679	-

3.2 Statistical results of CoMSIA models

The CoMSIA model has been generated by considering all possible combinations of CoMSIA field descriptors, including steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A) using Gasteiger-Hückel partial atomic charges. Different combinations of CoMSIA field descriptors were carried out to determine and generate an efficient CoMSIA model (Table 2 and Table S1). The combination of steric and hydrophobic field descriptors showed that best statistical results, which yielded q^2 value of 0.571 with the component of 8, r^2 value of 0.989 with a SEE value of 0.072. The contributions of the steric and hydrophobic fields are 0.321 and 0.679, respectively. It seems that only considering the steric field descriptors gives a higher q^2 , but its low F value makes it not more robust than the combination of steric and hydrophobic field descriptors. Additionally, results of bootstrap analysis for 100 runs shown in Table 2 demonstrate that the CoMSIA model along with low values of the standard error estimations and the absence of systematic errors within the training data set ($r_{bs}^2 = 0.998$, $SEE_{bs} = 0.003$).

3.3 Statistical results of HQSAR

Eighteen HQSAR models were established by combining different fragment parameter descriptors (Table S2). The optimal HQSAR model is obtained while combining fragment descriptors Atoms (A), Bonds (B), Connections (C), Chirality (Ch), and followed the default fragment lengths 4-7, which gives q^2 to be 0.827 and r^2 0.945,

with a hologram length of 61 and a component of 4. Moreover, the best HQSAR model 9 was selected for further improvement of the quality. For this purpose, the fragment sizes of 1–10 were explored, with the higher q^2 as a chosen criterion. It was found in Table S3 that the best model represents the results the same as HQSAR model 9, which achieved default fragment lengths of 4-7.

3.4 Validation of QSAR model

High value of q^2 does not imply that the model has high predictive power, and the independent test set was chosen for further evaluation of the QSAR model to ensure its robustness and stability. Before using the test set as the external verification, it is crucial to eliminate possible systematic errors. The experimental and predicted pIC_{50} with residual values of CoMFA, CoMSIA, Topomer CoMFA, and HQSAR are shown in Table 3, and graphically depicted in Fig. 3. Compound 24 has large residual values in all QSAR models that are reflected in both Table 3 and Fig. 3, but it does not influence the quality of the QSAR model established by the training set. So we treat compound 24 as an outlier and remove it during calculated external verification. As shown in Table 4, the QSAR models for the test set without compound 24 results in R^2 of 0.8471 (CoMFA), 0.8226 (Topomer CoMFA), 0.8179 (CoMSIA), 0.7508 (HQSAR) and high slope regression lines with k and k' values of 0.9975 and 1.0019 (CoMFA), 0.9939 and 1.0049 (Topomer CoMFA), 0.9942 and 1.0047 (CoMSIA), 0.9967 and 1.0022 (HQSAR), respectively. R_0^2 and $R_0'^2$ values of 0.8392 and 0.8078 (CoMFA), 0.8150 and 0.8129 (Topomer CoMFA), 0.8168 and 0.7854 (CoMSIA), 0.7503 and 0.7255 (HQSAR), respectively. $(R^2 - R_0^2)/R^2$ and $(R^2 - R_0'^2)/R^2$ are calculated by R_0^2

and $R_0'^2$ to be 0.0734 and 0.0793 (CoMFA), 0.0051 and 0.0052 (Topomer CoMFA), 0.0754 and 0.0816 (CoMSIA), 0.0650 and 0.0695 (HQSAR), respectively. r_m^2 and Δr_m^2 values are 0.6828 and 0.1550 (CoMFA), 0.7572 and 0.0585 (Topomer CoMFA), 0.6841 and 0.1659 (CoMSIA), 0.6576 and 0.1532 (HQSAR), respectively.

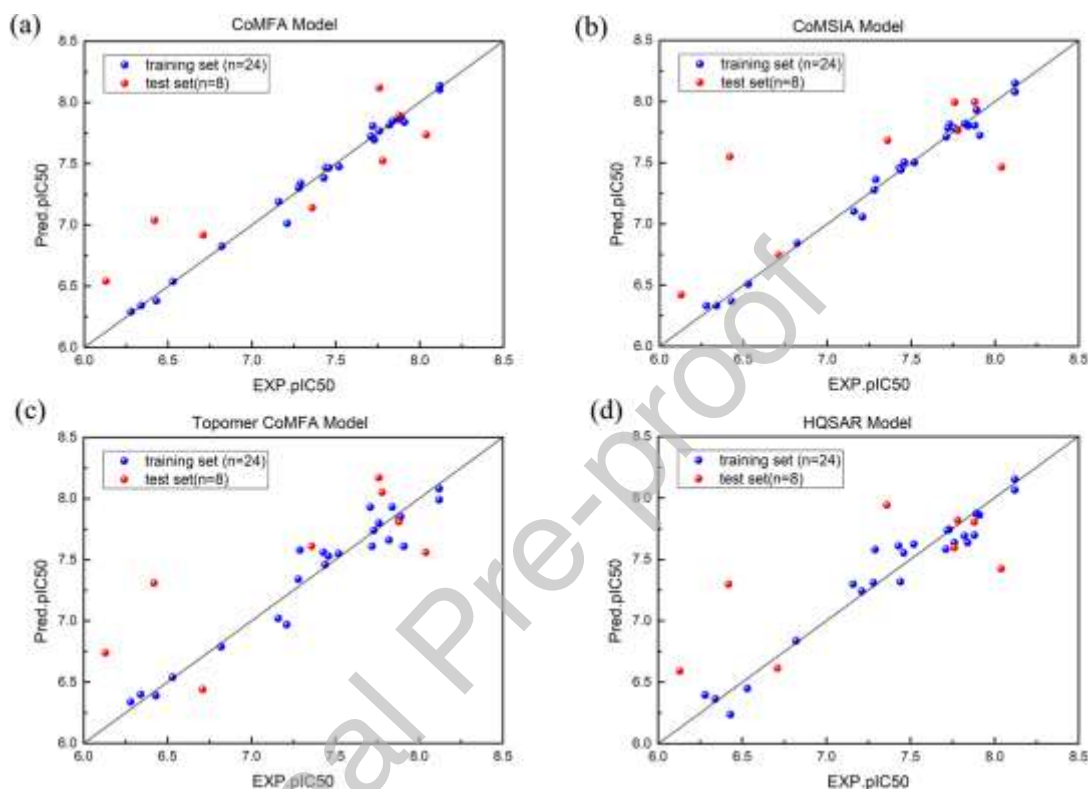


Fig. 3: The plots of experimental and predicted activities based on training and test sets.

Table 3: Actual and predicted activities of training and test set compounds.

Comp.	Exp. pIC ₅₀	CoMFA		CoMSIA		Topomer CoMFA		Contribution		HQSAR10-4	
		Predicted	Residual	Predicted	Residual	Predicted	Residual	R _a	R _b	Predicted	Residual
1	6.34	6.342	-0.002	6.334	0.006	6.40	-0.06	0.02	-0.64	6.364	0.024
2	7.28	7.303	-0.023	7.278	0.002	7.34	-0.06	0.02	0.31	7.311	0.031

3*	7.7 8	7.523	0.25 7	7.765	0.01 5	8.05	-0.27	0.0 2	1.0 2	7.819	- 0.03 9
4	7.7 3	7.696	0.03 4	7.814	- 0.08 4	7.74	-0.01	0.0 2	0.7 0	7.743	- 0.01 3
5	7.8 8	7.862	0.01 8	7.805	0.07 5	7.83	0.05	0.0 2	0.8 0	7.697	0.18 3
6	7.8 4	7.849	-0.09	7.801	0.03 9	7.93	-0.09	0.0 2	0.9 0	7.637	0.20 3
7*	7.3 6	7.141	- 0.05 1	7.682	- 0.32 2	7.61	-0.25	0.0 2	0.5 7	7.943	- 0.58 3
8	7.4 3	7.384	0.04 6	7.459	- 0.02 9	7.56	-0.13	0.0 2	0.5 2	7.610	-0.18
9	7.8 2	7.817	0.00 3	7.821	- 0.00 1	7.66	0.16	7.6 6	0.0 2	7.691	0.12 9
10	7.2 9	7.340	- 0.05 0	7.364	- 0.07 4	7.58	-0.29	0.0 2	0.5 5	7.580	-0.29
11*	7.8 8	7.893	- 0.01 3	7.998	- 0.11 8	7.81	0.07	0.0 2	0.7 8	7.805	0.07 5
12	7.7 2	7.807	- 0.08 7	7.784	- 0.06 4	7.61	0.11	0.0 2	0.5 8	7.738	- 0.01 8
13	7.9 1	7.839	0.07 1	7.724	0.18 6	7.61	0.3	0.0 2	0.5 8	7.861	0.04 9
14	7.8 9	7.878	0.01 2	7.933	- 0.04 3	7.85	0.04	0.0 2	0.8 2	7.874	0.01 6
15*	6.1 3	6.541	- 1.01 1	6.421	- 0.29 1	6.74	-0.61	- 0.0 4	- 0.2 4	6.591	- 0.46 1
16	6.8 2	6.827	- 0.00 7	6.843	- 0.02 3	6.79	0.03	- 0.0 4	- 0.1 9	6.836	- 0.01 6
17	6.5 3	6.538	- 0.01 4	6.509	0.02 1	6.54	-0.01	- 0.0 4	- 0.4 4	6.450	0.08

18	6.2 8	6.293	- 0.01 3	6.332	- 0.05 2	6.34	-0.06	0.0 4	- 0.6 3	6.397	- 0.11 7
19*	6.7 1	6.918	- 0.20 8	6.747	- 0.03 7	6.44	0.27	0.0 4	- 0.5 3	6.614	0.09 6
20	6.4 3	6.380	0.05 0	6.368	0.06 2	6.39	0.04	0.0 4	- 0.5 9	6.238	0.19 2
21	8.1 2	8.105	0.01 5	8.079	0.04 1	7.99	0.13	0.0 4	- 1.0 2	8.065	0.05 5
22*	7.7 6	8.120	- 0.36 0	7.995	- 0.23 5	8.17	-0.41	0.0 4	- 1.1 9	7.596	0.16 4
23	8.1 2	8.137	- 0.01 7	8.151	- 0.03 1	8.08	0.04	0.0 4	- 1.1 1	8.152	- 0.03 2
24*	6.4 2	7.038	- 1.01 8	7.550	- -1.13	7.31	-0.89	0.0 4	- 0.3 3	7.298	- 0.87 8
25	7.4 4	7.469	- 0.02 9	7.442	- 0.00 2	7.46	-0.02	0.0 4	- 0.4 8	7.318	- 0.12 2
26	7.1 6	7.189	- 0.02 9	7.103	0.05 7	7.02	0.14	0.0 4	- 0.0 4	7.299	- 0.13 9
27	7.2 1	7.013	0.01 7	7.058	0.15 2	6.97	0.24	0.0 4	- 0.0 0	7.241	- 0.03 1
28*	8.0 4	7.736	0.30 4	7.465	0.57 5	7.56	0.48	0.0 4	- 0.5 8	7.424	0.61 6
29	7.4 6	7.466	- 0.00 6	7.504	- 0.04 4	7.53	-0.07	0.0 4	- 0.5 6	7.553	- 0.09 3
30	7.7 6	7.766	- 0.00 6	7.781	- 0.02 1	7.80	-0.04	0.0 4	- 0.8 2	7.637	- 0.12 3
31	7.5 2	7.477	0.04 3	7.502	0.01 8	7.55	-0.03	0.0 4	- 0.5 8	7.624	- 0.10 4
32	7.7 1	7.724	- 0.01	7.710	0.00 0	7.93	-0.22	- 0.0	0.9 5	7.583	0.12 7

The models gave low RMSE, MAE, and high CCC of 0.2351, 0.1657 and 0.9033 (CoMFA), 0.3165, 0.2573 and 0.9726 (Topomer CoMFA), 0.2648, 0.1788 and 0.8908 (CoMSIA), 0.3187, 0.2383 and 0.8607 (HQSAR) for training and test set, respectively. In addition, Q_{F1}^2 and Q_{F2}^2 values of QSAR models represent 0.8388 and 0.8354 (CoMFA), 0.8123 and 0.8088 (Topomer CoMFA), 0.8082 and 0.8054 (CoMSIA), 0.7507 and 0.7503 (HQSAR), respectively. In summary, four optimal models produced ideal internal and external verification parameters, which shows that the QSAR models have good evaluation stability and good predictive ability.

Table 4: Verification parameter data statistics results.

Parameters	Criterion	CoMFA	Topomer COMFA	CoMSIA	HQSAR 10-4
External validation	$r_{pred}^2 > 0.6$	0.832	0.674	0.805	0.678
Golbraikh–Tropsha method	$R^2 > 0.6$	0.8471	0.8226	0.8179	0.7508
	R_0^2	0.8392	0.8150	0.8168	0.7503
	$R_0'^2$	0.8078	0.8129	0.7854	0.7255
$\frac{(R^2 - R_0^2)}{R^2}$	$\frac{(R^2 - R_0^2)}{R^2} < 0.1$	0.0734	0.0051	0.0754	0.0650
$\frac{(R^2 - R_0'^2)}{R^2}$	$\frac{(R^2 - R_0'^2)}{R^2} < 0.1$	0.0793	0.0052	0.0816	0.0695
k	$0.85 \leq k \leq 1.15$	0.9975	0.9939	0.9942	0.9967
k'	$0.85 \leq k' \leq 1.15$	1.0019	1.0049	1.0047	1.0022
r_m^2	$r_m^2 > 0.5$	0.6828	0.7572	0.6841	0.6576
Δr_m^2	$\Delta r_m^2 < 0.2$	0.1550	0.0585	0.1659	0.1532
Q_{F1}^2	$Q_{F1}^2 > 0.7$	0.8388	0.8123	0.8082	0.7507

Q_{F2}^2	$Q_{F2}^2 > 0.7$	0.8354	0.8088	0.8054	0.7503
RMSE	Close to 0	0.2351	0.3165	0.2648	0.3187
MAE		0.1657	0.2573	0.1788	0.2383
CCC	$CCC \geq 0.85$	0.9033	0.9726	0.8908	0.8607

3.5 Y-randomization test and applicability domain validation

The Y-randomization test was performed on the QSAR models to evaluate the robustness of the model and to avoid accidental correlations that may appear [52]. 10 random trials were run for the PLS model (Table S4), and all the results show small average values of q^2 and r^2 ($q^2 < 0.5$, $r^2 < 0.5$). This proves that the good results of our original model are not due to chance correlation or structural dependency of the training set. The applicability domain (AD) is the theoretical region in the chemical space defined by the model descriptor and the model response in which the prediction is reliable [53]. The robust and predictable QSAR model that can be used to predict new compounds should be followed by an applicability domain [54]. The applicability domain test was performed using the standardized method developed by Roy (<http://dtclab.webs.com/software-tools>). Regarding the applicability domain QSAR models, there is no compound located outside the AD for the QSAR models, which is consistent with the characteristics of this type of SARS-CoV-2 inhibitor. No compound has a divergence from other compounds.

3.6 CoMFA and Topomer CoMFA contour map analysis

The contour map of the CoMFA/Topomer CoMFA/CoMSIA model is a tool to explain the important structural aspects through the favorable and unfavorable regions in different fields. In steric contour maps, the green contours indicate that the large substituent has higher activity, while yellow contours stand for the large substituent

with lower activity. In electrostatic contour maps, red contours represent electronegative charge favored region and blue contour map represents the electropositive charge favored region. Regarding the CoMFA steric maps in Fig. 4(a), big green contours covered the 2nd and 3rd positions at a phenyl moiety of compound 23, which suggests that the bulky group was favorable in the region. This agrees with the fact that the activity of compound 19 ($pIC_{50} = 6.293$) is higher than compound 18 ($pIC_{50} = 6.918$) and compound 20 ($pIC_{50} = 6.380$). The green contours that appear at the introduction of the phenyl ring indicate that bulky group favorable in the region, such as compound 2 with O atom has higher activity than compound 1 with a C atom. Furthermore, a big yellow contour around the R₁ substituent suggests that the less-bulky group here is beneficial to increase the inhibitory activity, such as the activity of compound 21 ($pIC_{50} = 8.12$) is higher than compound 3 ($pIC_{50} = 7.78$).

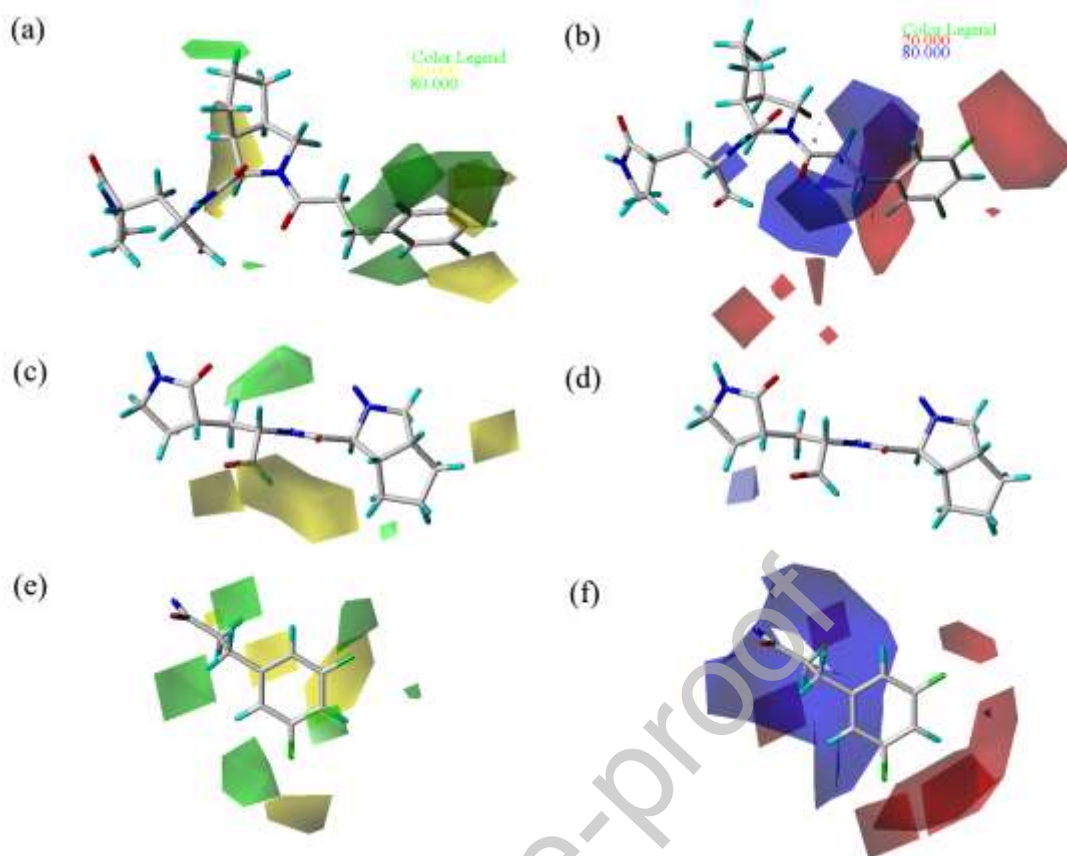


Fig. 4: The 3D contour maps of the steric and electrostatic fields of the CoMFA and Topomer CoMFA models with compound 23 as the template. (a): The steric contour map of the CoMFA model. (b): the electrostatic contour map of the CoMFA model; (c): The steric contour maps of the Topomer CoMFA model for R_a fragment of compound 23. (d): The electrostatic contour maps of the Topomer CoMFA model for R_b fragment of compound 23. (e): The steric contour maps of the Topomer CoMFA model for R_b fragment of compound 23. (f): The electrostatic contour maps of the Topomer CoMFA model for R_b fragment of compound 23.

Fig. 4(b) is the contour map of the electrostatic field of the compound 23. A large red area wrapped around the 2nd and 3rd positions at a phenyl moiety of compound 23 indicates that the electron-withdrawing (electronegative) substituent linked here will enhance the biological activity of compounds. This might be the reason that the pIC_{50}

of compound 30 ($pIC_{50} = 7.76$, 2nd = -F) is higher than compound 29 ($pIC_{50} = 7.46$, 2nd = -Cl). The blue contours appeared at the introduction of phenyl ring suggests that electron-donating (electropositive) substituents are favorable in the region, which validates biological activities of compound 16 ($pIC_{50} = 6.82$, R_2 = indole substituent) > 17 ($pIC_{50} = 6.52$, R_2 = pyrazolo[1,5-a]pyridine substituent) > 18 ($pIC_{50} = 6.28$, vinylbenzene substituent).

Fig. 4(c) and Fig. 4(d) show the steric contour maps and electrostatic contour maps of the Topomer CoMFA model for the R_a fragment, respectively. Since the structural substitution of the R_a fragments for 32 inhibitors is not obvious, the R_1 groups of compound 1-14 and compound 15-32 are the same, which leads to their contour maps being similar. Similar to CoMFA steric contour results, the big yellow contour is around the R_1 substituent suggests that a less bulky region here is favored. This is consistent with the Topomer CoMFA contribution results that the R_1 substituent of compound 1-14 is higher than the R_1 substituent of compound 15-32 (Table 3).

Fig. 4(e) represents the steric contour map of Topomer CoMFA for the R_b fragment. The region where the R_2 substituent is connected to the amide group is wrapped in green contours, which indicates that the bulky group is favorable in the region. This explains that the activity of compound 20 ($pIC_{50} = 6.43$) is higher than compound 24 ($pIC_{50} = 6.42$). In addition, the green contours around the 3rd position at a phenyl moiety of compound 23 suggest that a bulky group here is beneficial to increase the SARS-CoV-2 inhibitory activity. An example is that the activity of compound 21 ($pIC_{50} = 8.12$, 3rd of phenyl = -F) is higher than compound 24 ($pIC_{50} = 6.24$, 3rd of phenyl = -H, 4th of

phenyl = -N(CH₃)₂). The yellow contours around the 4th position at a phenyl moiety of compound 23 also suggest a less bulky group here is beneficial to increase the SARS-CoV-2 inhibitory activity. For instance, the activity of compound 28 (pIC₅₀ = 8.04, 4th of phenyl = Cl) > compound 25 (pIC₅₀ = 7.44, 4th of phenyl = -OCH₃) > compound 27 (pIC₅₀ = 7.21, 3rd and 4th of phenyl = 1,4-dioxane).

Fig. 4(f) represents the electrostatic contour map of Topomer CoMFA for the R_b fragment. The red contours near the 2nd, 3rd, and 4th positions at a phenyl ring of compound 23 indicate that the introduction of electronegative groups will increase the inhibitory activity. This is reflected in the inhibitory activity that compound 1 (pIC₅₀ = 6.34) and compound 15 (pIC₅₀ = 6.13) without an electronegative group at a phenyl ring result in a relative lower inhibitory activity. The blue contours around the introduction positions at the phenyl ring of compound 23 indicate electropositive region groups here is favored, which could explain that the activity of compound 2 (pIC₅₀ = 7.28) which have an electropositive group at the introduction positions of the phenyl ring is higher than compound 1 (pIC₅₀ = 6.34).

3.7 CoMSIA contour map analysis

The CoMSIA steric contour maps were both similar to the CoMFA/Topomer CoMFA contour maps discussed above (Fig. 5(a)). Thus, only hydrophobic fields of CoMSIA were analyzed. In the hydrophobic contour map (Fig. 5(b)), the cyan contours show favorable hydrophobic regions, while magenta contours represent unfavorable hydrophobic regions. For the hydrophobic map, the cyan favorable region could be found around 2nd and 3rd positions at a phenyl ring of R₂ substitutes, which indicates

that the addition of hydrophobic substituents in this region would lead to an increase in the activity. For instance, compound 6 ($pIC_{50} = 7.84$) with a $-OCH_3$ substituents at 3rd position of phenyl has higher activity than compound 8 ($pIC_{50} = 7.43$) with a less hydrophobic substituent (3rd of phenyl = -F). The magenta unfavorable region could be found around the 4th position at a phenyl ring of R_2 substitutes, which indicates that the addition of hydrophobic substituents in this region would lead to decreased activity. For instance, the activity of compound 21 (4th of phenyl = -H) is higher than compound 22 (4th of phenyl = -F). Also, the magenta unfavorable region could be found at R_1 substitutes, which reveals that R_1 substitutes should be less hydrophobic substituents. This is further verified by the Topomer CoMFA contribution results.

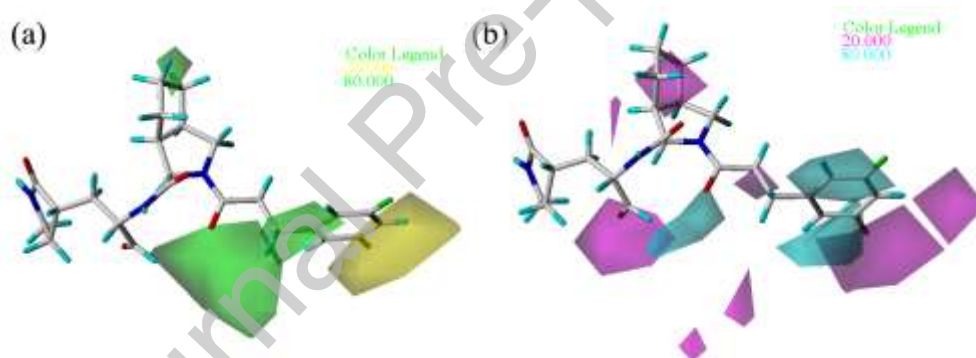


Fig. 5: The 3D contour maps of the steric and hydrophobic contour map fields of the CoMSIA model with compound 23 as the template.

3.8 HQSAR color code map analysis

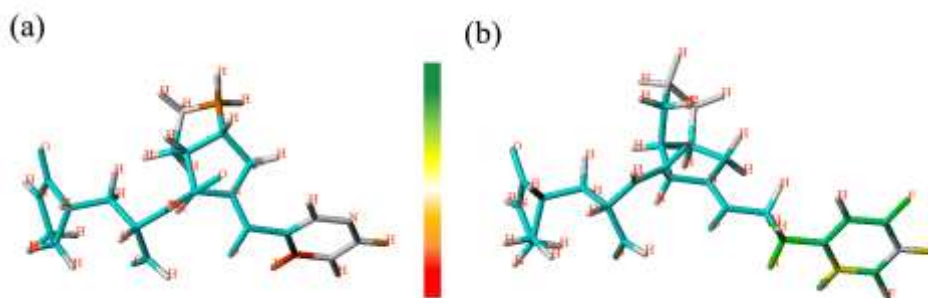


Fig. 6: Atomic contribution maps of compound 15 (a) and compound 23 (b). (Cyan

represents common backbone. The green or yellow on the contribution map shows a positive contribution, while orange and red indicate a negative contribution. Fragments of atoms show white as an intermediate contribution to biological activity).

The results of the HQSAR contribution map visualized the results of the HQSAR model by color-coding different atoms. The color of each atom reflects its contribution to the overall activity of each compound. The highest activity compound 21 ($pIC_{50} = 8.12$) and a moderate activity compound 15 ($pIC_{50} = 6.13$) were chosen to explore the relationship between the activity and atom contribution. As shown in Fig. 6(a), the orange color at the R_1 substitute revealed that the cyclopentyl provides a negative contribution, which is consistent with Topomer CoMFA results. In addition, the orange color appeared at 2nd position and 4th of the pyridine ring suggests that pyridine substituents here are not beneficial to increase the SARS-CoV-2 inhibitory activity, and the five-membered ring in the R_2 substituent could not be a pyridine ring. The color code map of the highest activity compound 21 is shown in Fig. 6(b). The green atom at the junction of the benzene ring group and the amide group indicates this methylene group is beneficial to improve the inhibitory activity, e.g., compound 16 ($pIC_{50} = 6.82$) and compound 17 ($pIC_{50} = 6.53$) have lower inhibitory activity compared with other compounds with a methylene group. Furthermore, the carbon atom on the phenyl ring represents a yellow or green positive contribution except for the C atom at the 2nd position. However, the hydrogen atoms at the 4th and 2nd positions show positive contributions, which suggests that the 4th and 2nd positions at the phenyl ring should be hydrogen atoms instead of others groups. This is reflected on the inhibitory activity that

compound 21 ($pIC_{50} = 8.21$, $4^{th} = -H$) is higher than compound 22 ($pIC_{50} = 8.21$, $4^{th} = -F$), and that compound 12 ($pIC_{50} = 7.72$, $4^{th} = -H$) is higher than compound 32 ($pIC_{50} = 7.5$, $2^{nd} = -CH_3$).

3.9 Molecular docking evaluations

Molecular docking analysis has been utilized to clarify the binding modes between bicycloproline-containing compounds and the SARS-CoV-2 M^{pro}. The Root means square deviation (RMSD) value between the homologous ligand and the re-docking ligand is a criterion to ensure docking reliability. Herein, six SARS-CoV-2 M^{pro}-ligand systems were used to perform the re-docking process for determining the most suitable docking protocol. As shown in Table S5, the RMSD values between the homologous ligand and the re-docking ligand of all six SARS-CoV-2 M^{pro} complex processed by Autodock Vina are within 2 Å. The minimum RMSD value between the homologous ligand and the re-docking ligand value is 1.0111 Å (7D3I), which indicates that this pair is most likely to reproduce the natural binding modes between the ligand and the protein. As shown in Fig. S2, the re-docking ligand (yellow) and homologous ligand (blue) are almost completely superimposed. Based on that, the Autodock Vina docking protocol was chosen to reveal the binding modes of SARS-CoV-2 M^{pro} with inhibitors.

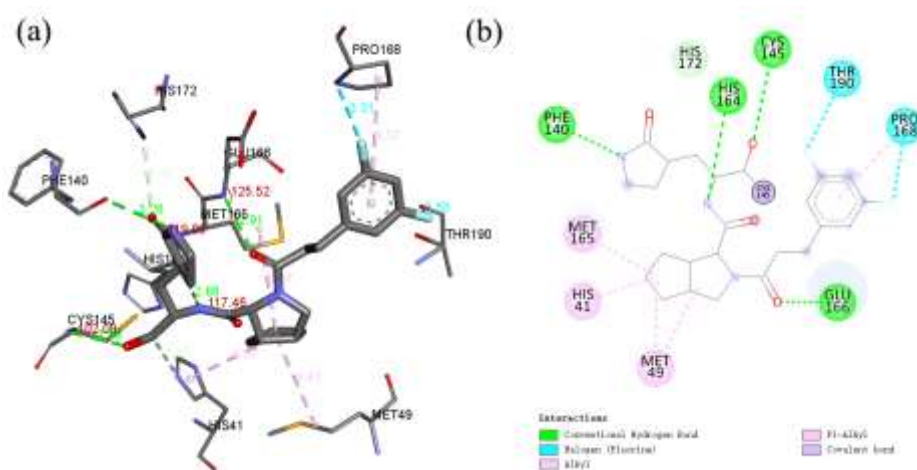


Fig. 7: (a) 3D and (b) 2D views of the binding conformations and ligand interactions of the most active compound 23 at the active site of SARS-CoV-2 M^{Pro} (7D3I).

(Visualized by Discovery Studio Visualizer 2021).

Fig. 7 illustrates the binding pocket of the receptor and compound 23 that interacts with the representative key residues. Amino acid residues Phe140, Cys145, His164 were observed to form the hydrogen bond with the N, and O atoms at the aligned common skeleton structure (Phe140-OH---N, 3.26 Å, 119.00 °; Cys145-N---HO, 2.89 Å, 102.9 °; His164-OH---N, 2.88 Å, 117.46 °). Multiple independent hydrogen bonds here could maintain the stability of the ligand at the active site. Hydrophobic cavities are formed among His41, Met49, Met165 amino acid residues at R₁ substituent. This is consistent with the CoMSIA hydrophobic field contour map results. Amino acid residue Glu166 was observed to form a hydrogen bond with the amide group among the R₁ substituent and R₂ substituent (Glu166-N---HO, 2.91 Å, 125.52 °). Importantly, amino acid residue Cys145 also formed a strong covalent bond with warhead aldehyde, which is essential for antiviral activity [22, 55]. In addition, certain interactions between the F atom at the benzyl ring of the R₂ substituent and the amino acid residues Pro168 and

Thr190 were observed.

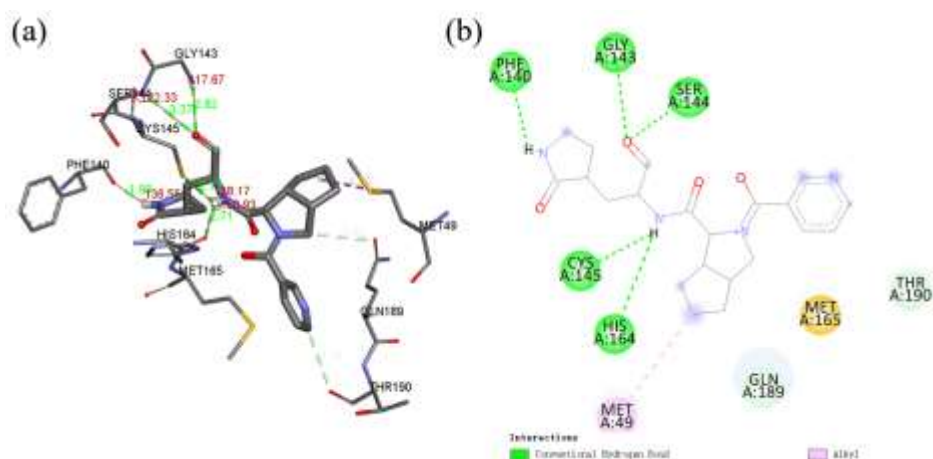


Fig. 8: (a) 3D and (b) 2D views of the binding conformations and ligand interactions of the moderate active compound 15 at the active site of SARS-CoV-2 M^{pro}.

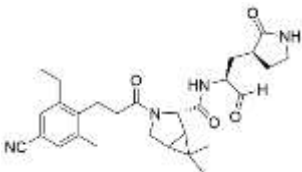
Fig. 8 shows the docking results of compound 15 with moderate activity. Similarly, amino acid residues Phe140, Gly143, Ser144, Cys145, and His164 were observed to form a hydrogen bond with the N, and O atoms at the aligned common skeleton of compound 15 (Phe140-O---HN, 1.99 Å, 136.55 °; Gly143-NH---O, 2.82 Å, 117.67 °; Ser144-NH---O, 3.37 Å, 122.33 °; Cys145-S---HN, 2.73 Å, 138.17 °; His164-O---HN, 2.71 Å, 150.93 °). The hydrophobic cavity at R₁ substituent shows hydrophobic interaction with amino acid residue Met49. For the structure of compound 15, there are only weak interactions on the R₂ substituents because of the piperidinyl group here. Moreover, the direct connection between the piperidine group and the amide group hinders the formation of a hydrogen bond. This may enhance the strong interactions on the R₂ substituents and would lead to a more stable bonding mode.

3.10 Prediction of new compounds with promising activity

The interpretation of QSAR models provide us more flexible design tips for

developing powerful SARS-CoV-2 M^{pro} inhibitors: (a): Bulky, electronegative groups at the 2nd, 3rd, and 4th position of benzene ring were useful with regards to the increasing of SARS-CoV-2 M^{pro} inhibitory activity; (b): the introduction place of phenyl ring surrounded by bulky, electron-donating groups will conduce improve activity; (c): The R₁ substitute should be a less bulky substitute for improving the inhibitory activity. Depending on the validated multi-QSAR modeling approaches, six new potential SARS-CoV-2 M^{pro} inhibitors were designed by modifying the R₂ substitute (Easily added bulky -CH₂CH₃, -CN, -F, and -Br group with relatively electronegative charge were added at the benzene ring, and the introduction place of phenyl ring replaced by the bulky, electron-donating groups that selected from the data set) and retaining a less bulky R₁ substitute (The R₁ group of compound 1-14). Table 5 represents the structures of designed compounds N₁-N₆ along with their predicted pIC₅₀ values for SARS-CoV-2 M^{pro}. All the newly designed inhibitors showed high predicted pIC₅₀ values compared to compound 23. This suggests that these compounds are potential candidates as SARS-CoV-2 inhibitors. However, these compounds still need to be synthesized combined with *in vitro* and *in vivo* evaluation to confirm their efficacy in inhibiting SARS-CoV-2 M^{pro}.

Table 5: Novel-designed compounds with their predicted activities.

Comp.	Structure	Pred. pIC ₅₀			
		CoMFA	CoMSIA	Topomer CoMFA	HQSAR
N ₁		7.659	8.012	8.06	8.269

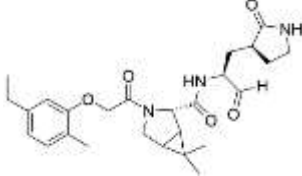
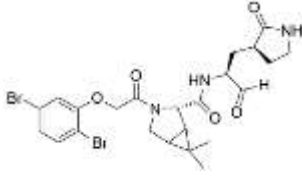
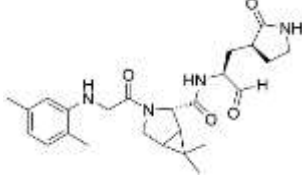
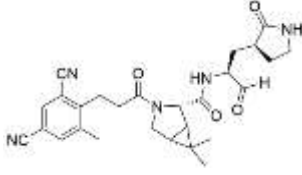
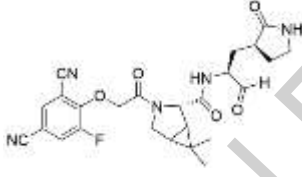
N ₂		7.747	7.772	7.86	7.827
N ₃		7.626	7.732	7.91	7.642
N ₄		7.638	7.708	8.01	7.665
N ₅		7.893	7.955	7.91	8.365
N ₆		7.980	7.611	7.84	7.615

Table 6 illustrates the Autodock vina docking scores and ligand interactions of newly designed compounds, and they show ideal docking scores from -7.1-8.0 Kcal/mol. The 2D and 3D views of the binding ligand interactions of the newly designed compounds are shown in Fig. S3 and Fig. S4, respectively. The docking poses of the newly designed compounds were similar with compound 23 and compound 15, in the active site of SARS-CoV-2 M^{pro}, which further verifies that these newly designed compounds have strong potential to produce similar inhibitory effects with the compounds in the data set. For compound N₁ with relatively low docking scores, the R₁ and R₂ substituents mainly have hydrophobic interactions with the amino acid residues Leu27 and Met49 at the

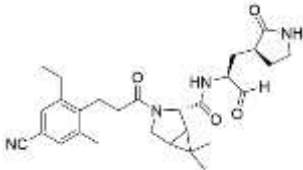
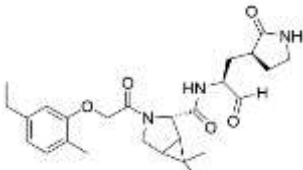
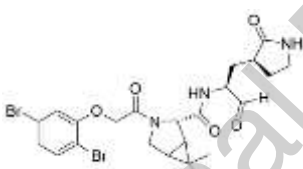
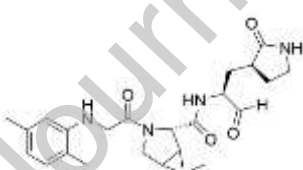
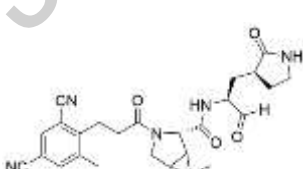
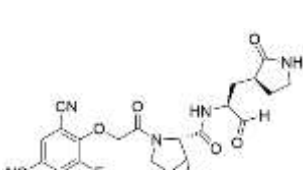
active site of the SARS-CoV-2 M^{pro}, and the remaining common skeleton position could be observed to form stable hydrogen bonds with amino acid residues Asn142, Ser144, His164 (Asn142-OH---O, 3.34 Å, 107.68 °; Ser144-OH---O, 2.98 Å, 166.46 °; His164-O---HN, 2.91 Å, 135.20 °). In addition, unlike other compounds, the pyrrolidine group at the common skeleton of compound N₁ does not produce hydrophobic or hydrogen bonding at the active site of SARS-CoV-2 M^{pro}. This explains that compound N₁ has a relatively lower docking score than other newly designed compounds. Regarding compound N₆ with a relatively higher docking score, the newly introduced -CN group at phenyl can form a stable hydrogen bond with the amino acid residue Gln192 (Gln192-NH---N, 3.23 Å). In addition, strong hydrogen bonds with amino acid residues Phe140, Gly143, Cys145, and His164 are also observed at the position of the aligned backbone (Phe140-O---HN, 2.09 Å, 162.19 °; Gly143-NH---O, 3.16 Å, 127.44 °; Cys145-SH---O, 3.68 Å, 107.5 °; Cys145-S---HN, 2.42 Å, 145.19 °; His164-O---HN, 2.52 Å, 130.39 °). The hydrophobic interaction with Met49 amino acid residue is also observed at the R₁ substituent. These hydrophobic interactions and hydrogen bonds stabilize the ligand at the active site of the protein, which also explains the high docking score of the compound N₆.

Taken together, the R₁ substituents in the active site mainly form on the hydrophobic interaction with amino acid residue Leu27, Met49 to maintain a stable conformation, and the R₂ substituents in the active site rely on the hydrophobic interaction with the amino acid residues Met165, Leu167, Pro168, and Thr190, and on the hydrogen bond with amino acid residue Gln192 to maintain a stable conformation. Several stable

hydrogen bonds can be formed at aligned backbone position with amino acid residues

Phe140, Gly143, Cys145, His164, and Glu166.

Table 6: Novel designed compounds with their Autodock Vina fitness scores and amino acids interactions.

Comp.	Structure	Binding energy score (Kcal/mol)	Hydrogen bond interacting residues	Hydrophobic interactions
N ₁		-7.1	Asn142, Ser144, His164	Leu27, Met49
N ₂		-7.5	Phe140, Cys145, His164	Met49, Leu167, Pro168
N ₃		-7.4	Phe140, Cys145, His164	Met49
N ₄		-7.9	Phe140, Gly143, Cys145, His164, Glu166,	Met49, Pro168
N ₅		-7.8	Phe140, Gly143, Cys145, His164, Gln192	Met49
N ₆		-8.0	Phe140, Gly143, Cys145, His164, Gln192	Met49

3.11 Molecular dynamics simulations

To validate the intrinsic atomic interactions and binding conformations of the newly designed compounds, 50 ns MD simulations were performed on the compounds with higher docking score (N₃, N₄, N₅, and N₆ with M^{pro}), and 7D3I (compound 23 with M^{pro}). The overall stability and convergence of the protein-ligand system are monitored by the RMSD values (relative to the simulation time of the initial structure) of the backbone atoms and ligand atoms. 7D3I protein and SARS-CoV-2 M^{pro}-N₄ system are stable after 5 ns of initial simulation, other SARS-CoV-2 M^{pro}-ligand systems are gradually stable after 20 ns (Fig. 9(a)). The backbone atoms of SARS-CoV-2 M^{pro} show fluctuations < 3 Å. Furthermore, 7D3I protein and SARS-CoV-2 M^{pro}-N₄ systems have lower fluctuations < 2 Å. This reveals that the fluctuation of the SARS-CoV-2 M^{pro} during the simulation is within an acceptable range. The RMSD values of the ligand fluctuation over time are shown in Fig. 9(b). Except that the N₃ ligand fluctuates greater during the initial simulation process (from 0.1 nm to 0.4 nm), and is stable after 35 ns, the RMSD fluctuations of other ligands are all stable between 0.1 nm - 0.2 nm during the whole simulation process.

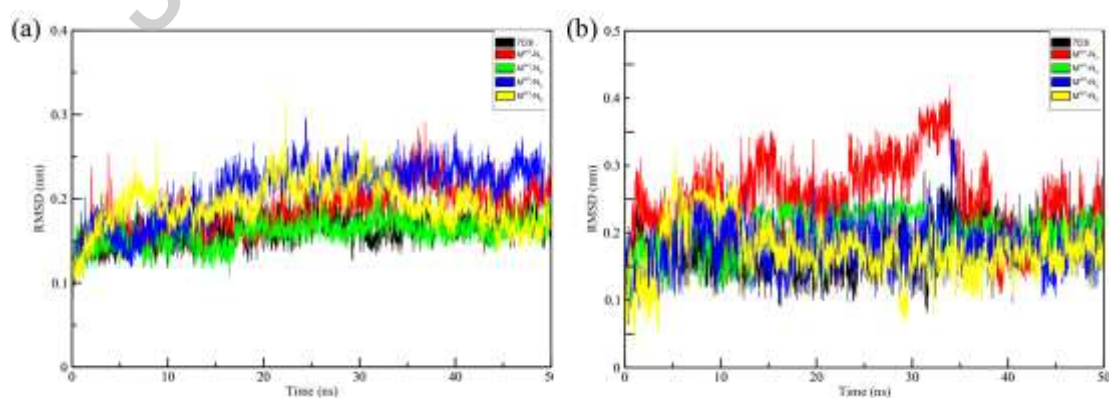


Fig. 9: RMSD values of protein backbone (a) and RMSD values of ligands (b) at the

active site of SARS-CoV-2 M^{pro} during 50 ns MD simulation.

The root means square fluctuation (RMSF) reflects the movement of each amino acid residue during the MD simulation. Analysis of RMSF values versus the residue numbers of the five simulated systems is depicted in Fig. 10(a). These five complexes have similar RMSF distributions, which indicates that the binding of these inhibitors has similar effects on protein residue fluctuation patterns. The smaller fluctuations in the region of 20-50, and 140-200 are caused by the relevant residues in the active site of SARS-CoV-2 M^{pro}, such as the residues Leu27 and Met49 in the residue region of 20-50, and Gly143, Cys145, Met165, Pro168, His164, and Glu192 in the residue region of 140-200. The hydrophobic or hydrogen interactions between the ligand and residues in these regions result in fewer fluctuations. This is consistent with the docking results.

The radius of gyration represents the change in the compactness of the protein after the ligand interaction. The respective average Rg values of 2.17, 2.21, 2.20, 2.19, and 2.21 nm are calculated for 7D3I, M^{pro}-N₃, M^{pro}-N₄, M^{pro}-N₅, and M^{pro}-N₆, respectively. The radius of gyration of M^{pro}-N₃ has relatively obvious fluctuations, then tends to balance around 30 ns (Fig. 10(b)). The Rg results show that all M^{pro}-ligand systems reach a level of compactness, which indicates that each compound can interact with SARS-CoV-2 M^{pro} without affecting its structural folding in the dynamic environment.

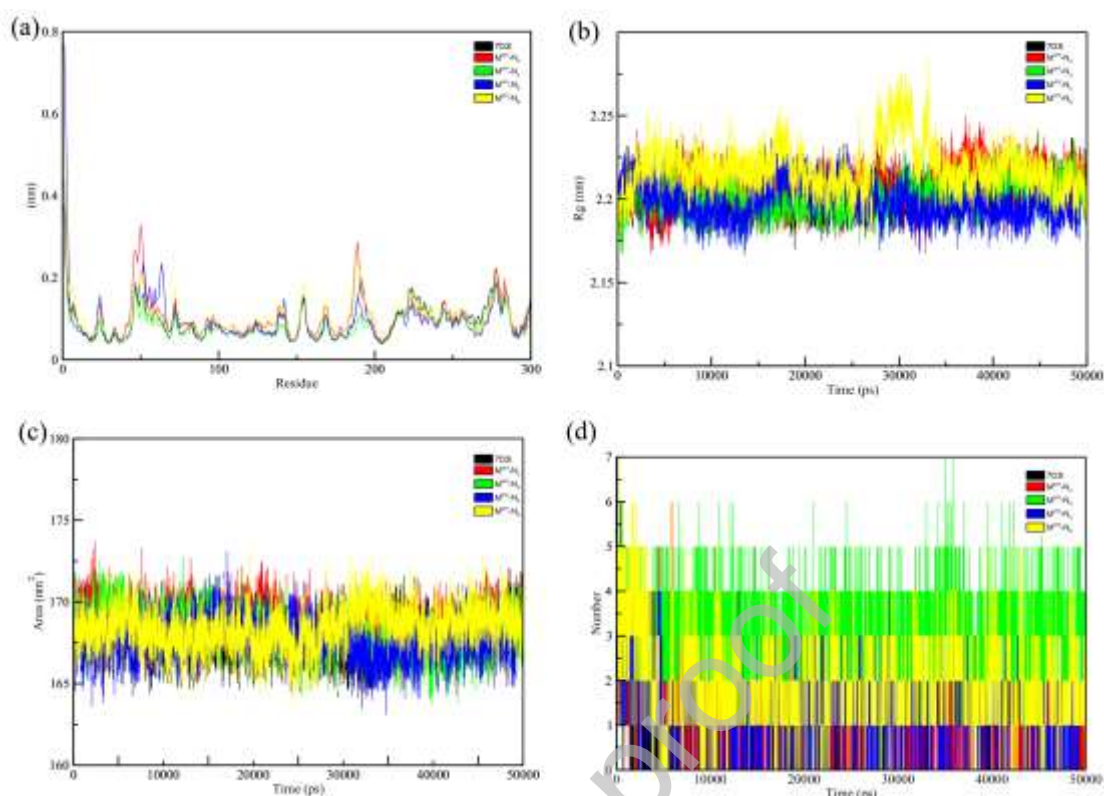


Fig. 10: Comparison and detail representation of (a) RMSF: Root mean square fluctuations, (b) Rg: Radius of gyration, (c) SASA: Solvent accessible surface area, and (d) H-bond numbers between protein and ligand over the time of 50 ns of 7D3I and all the selected complexes.

Solvent accessible surface area (SASA) is the surface area where ligands can interact with solvent molecules, which could examine interactions between the complex and the solvent during the simulation analysis. The average SASA values for the five complexes were calculated and plotted vs simulation time in Fig. 10(c), the ligand-protein complex 7D3I, M^{pro}-N₃, M^{pro}-N₄, M^{pro}-N₅, and M^{pro}-N₆, are observed to be 166, 170, 168, 167, and 166 nm², respectively. Those similar average SASA values reveal that no major differences in the available protein structure.

The hydrogen bonding between the protein and the ligand is a crucial factor in

maintaining the stability of the complex conformation. Fig. 10(d) shows the number of hydrogen bonds between the five ligands and protein complexes under 50 ns simulation. The number of hydrogen bonds varies for 7D3I (0-5), M^{pro}-N₃ (0-3), M^{pro}-N₄ (1-4), M^{pro}-N₅ (0-3), and M^{pro}-N₆ (1-5). Compared with other complexes, 2-3 number of hydrogen bonds of compound M^{pro}-N₄ system are more stable in the simulation process. This demonstrates that the N₄ ligand has better stability at the active site of SARS-CoV-2 M^{pro}.

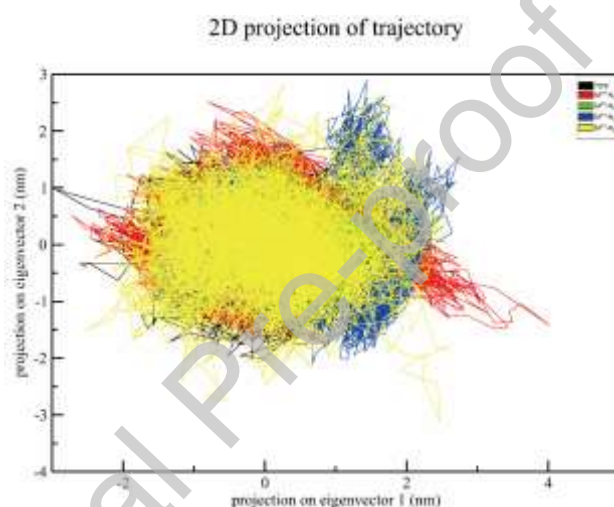


Fig. 11: Projection of the motion of the SARS-CoV-2 M^{pro} in phase space along the PC1 and PC2.

The principal component analysis could predict the relative movement of complex systems, and the overall movement in the protein is described by several key feature eigenvectors. The projections of two eigenvectors for SARS-CoV-2 M^{pro} are docked with selected compounds extracted from respective MD trajectories and are represented in clusters depicted in Fig. 11. Those M^{pro}-ligand systems have similar cluster patterns. It is known that a complex that occupies less phase space and shows a stable cluster would have higher stability, while a complex that occupies more space and shows an

unstable cluster indicates worse stability. The $M^{\text{pro}}\text{-N}_3$ and $M^{\text{pro}}\text{-N}_5$ systems occupy more space, followed by the $M^{\text{pro}}\text{-N}_6$ system, and show less stable clusters as compared to other complexes. The 7D3I and $M^{\text{pro}}\text{-N}_4$ complexes occupy less phase space as compared to the other complexes, and show a stable cluster. These observations suggest the stability of newly designed compounds is similar with 7D3I, and ultimately can restrict the protein essential motions required for exhibiting enzymatic activity, which is also supported by RMSD, RMSF, SASA, R_g , protein-ligand interaction profiling, and the substantial docking scores (Table 6, Fig. 9, Fig. 10).

3.12 MM/PBSA binding free energy calculation

The binding free energies of the five complexes were calculated using the MM-PBSA approach and the results are summarized in Table 7. The binding free energies (ΔG_{bind}) of the designed compounds N_3 (-81.862 ± 13.081 kJ/mol), N_4 (-75.300 ± 13.868 kJ/mol), and N_6 (-74.266 ± 18.405 kJ/mol) are bigger than the most active compound 23 (7D3I: -69.027 ± 11.484 kJ/mol). The van der Waals (ΔE_{vdw}) energy term is the main driving force for the molecular recognition of SARS-CoV-2 M^{pro} by the selected ligands, followed by electrostatic energy (ΔE_{elec}) term which plays a major contribution in strengthening the binding mode. Those results suggest the binding affinities of newly designed compounds (N_3 , N_4 , N_6) are bigger in comparison with the most active compound 23 with active residues in the active site of SARS-CoV-2 M^{pro} , and confirm that the structural modification information provided by QSAR models for designing novel potent SARS-CoV-2 M^{pro} inhibitors are credible as well.

Table 7: Average binding free energies (KJ/mol) of ligands obtained from MD

simulation with their energy terms.

Energy Terms (kJ/mol)	7D3I	M ^{pro} -N ₃	M ^{pro} -N ₄	M ^{pro} -N ₅	M ^{pro} -N ₆
	-	-	-	-	-
ΔE_{vdw}	137.326±13.	142.122±15.	155.720±17.	133.824±13.	173.253±11.
	519	282	691	860	491
	-	-	-	-	-
ΔE_{elec}	12.710±7.66	37.786±16.0	60.982±11.1	20.653±9.12	27.571±14.5
	9	39	20	32	23
	98.593±18.2	110.874±27.	160.117±15.	114.815±17.	145.169±22.
ΔE_{polar}	67	662	581	675	998
	-	-	-	-	-
ΔE_{SASA}	17.584±1.68	14.828±1.96	18.716±1.71	17.141±1.29	18.611±1.06
	-	1	5	9	1
	-	-	-	-	-
ΔG_{bind}	69.027±11.4	81.862±13.0	75.300±13.8	56.803±9.87	74.266±18.4
	84	81	68	0	05

The amino acid residues involved in the interactions between SARS-CoV-2 M^{pro} and ligands are observed by monitoring the binding free energy contribution (Fig. 12). The amino acid residues Cys27, Met49, Cys145, His164, Met165, and Pro168 have favorable contributions with more negative binding free energy to all five ligands, in which they exert the hydrophobic interaction, except for Cys145 and His164 with H-

bond interactions. This is common in almost all the studied SARS-CoV-2 M^{pro}-ligand systems. The amino acid residues Ala191 and Gln192 have favorable contributions with N₅ and N₆, which may be due to the new introduced -CN group at phenyl ring has H-bond interactions with acid residues Gln192 and hydrophobic interactions with Ala191. Furthermore, the amino acid residues of Met49, His164 and Met165 contribute more than others to the five complexes, which indicates their significant roles in ligand binding in five systems. Those findings confirm the docking results (Table 6). Interestingly, the important amino acid residue Cys145 seems not to have a higher binding free energy contribution in the simulation process, and the binding affinity between ligand and SARS-CoV-2 M^{pro} is more dependent on hydrophobic interactions.

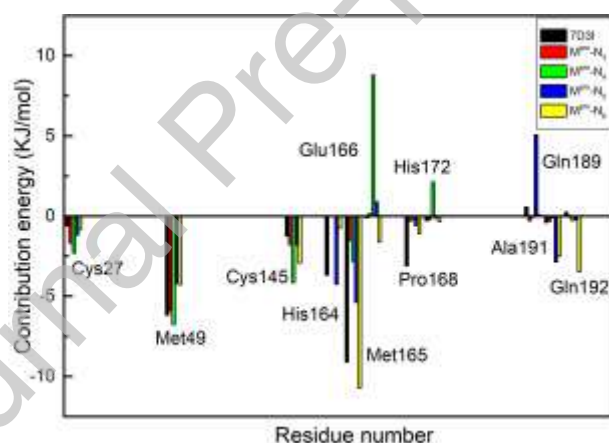


Fig. 12: Plots of binding free energy decomposition for the five complex systems.

3.13 ADMET and drug-likeness prediction

The bioavailability (absorption, distribution, metabolism, and excretion) of newly designed compounds were estimated to elucidate the drug-likeness of strong binders, which is an important criterion for approving any compound for clinical trials. ADMET scores for newly designed compounds are listed in Table 8. The designed compounds show good human intestinal absorption (HIA+: 0.053-0.31) except N₅, and the human

oral bioavailability 20% ($F_{20\%+}$: 0.023-0.079), which indicates that those designed compounds have good absorption and utilization in the human body. Madin-Darby Canine Kidney cells (MDCK) permeability is widely regarded as the gold standard for in-vitro evaluation of the absorption efficiency of chemical substances in the body. The designed compounds also give moderate MDCK Permeability values (3×10^{-6} - 1×10^{-5}) except compound N₃. The drugs with high protein-binding rates may have a low therapeutic index. The prediction results of PPB indicate that most compounds have a high plasma protein binding rate (PPB < 90%). The BBB+ prediction results demonstrate that newly designed compounds seem to be potential CYP3A4 protease substrates except for compound N₄. For the total clearance of newly designed compounds, they have moderate clearance rates except for compound N₃, which indicates that N₃ is not cleared from the body more slowly. Finally, the newly designed compounds were predicted to be AMES toxic.

Table 8: ADMET prediction of newly designed compounds.

Comp.	Absorption			Distribution		Metabolism		Excretion	Toxicity
	HI A+	F 20%+	MDCK Permeability	PP B (%)	BB B+	CYP2 D6 substrate	CYP3 A4 substrate	Total Clearance (mL/min/kg)	AMES toxicity
N ₁	0.249	0.127	3e-06	55.03	0.116	0.225	0.556	6.434	0.007
N ₂	0.053	0.038	3e-06	67.87	0.177	0.26	0.617	8.528	0.007
N ₃	0.574	0.023	1e-05	83.78	0.201	0.234	0.766	1.35	0.005
N ₄	0.13	0.075	2e-06	68.45	0.07	0.117	0.298	6.697	0.009

N ₅	0.3	0.0	3e-06	59.	0.0	0.207	0.515	6.422	0.007
	1	37		81	85				
N ₆	0.2	0.0	4e-06	59.	0.0	0.206	0.373	6.339	0.008
	93	79		67	69				

*Classification (HIA): Category 1: HIA+ (HIA < 30%); Category 0: HIA- (HIA < 30%);

The output value is the probability of being HIA+; (F_{20%}): Category 1: F_{20%}+

(bioavailability < 20%); Category 0: F_{20%}- (bioavailability ≥ 20%); The output value

is the probability of being F_{20%}+

MDCK Permeability: low permeability: < 2 × 10⁻⁶

cm/s. medium permeability: 2–20 × 10⁻⁶ cm/s high passive permeability: > 20 × 10⁻⁶

cm/s; PPB: Optimal: < 90%. Drugs with high protein-binding rate may have a low

therapeutic index; BBB: Category 1: BBB+; Category 0: BBB-; The output value is the

probability of being BBB+; Clearance: High: > 15 mL/min/kg; moderate: 5-15

mL/min/kg; low: < 5 mL/min/kg.

Drug-likeness prediction results and synthesis accessibility for all designed compounds are illustrated in Table S6. All designed compounds are acceptable considering the Lipinski Rule, Pfizer Rule, Golden Triangle and are rejected by GSK Rule. This is because the molecular weights of newly designed compounds, including compounds in the data set, are higher than 400 g/mol, which violates the GSK Rule. But the molecular weights of Remdesivir, which has been approved by the FDA for treating COVID-19, are also higher than 400 g/mol. Therefore, the GSK Rule is not an absolute condition for determining the drug-likeness properties. Furthermore, all the newly designed compounds are predicted to have convenient synthesis accessibility (4.37-4.71).

4. CONCLUSION

The quick and efficient development of active antiviral agents for therapeutic use is exceptionally challenging. In the current study, a series of bicycloproline-containing SARS-CoV-2 M^{pro} inhibitors are used for molecular modeling. The robust QSAR models were established to predict the enzyme inhibition of SARS-CoV-2 M^{pro}, with structural modification information to design six potent SARS-CoV-2 M^{pro} inhibitors. The binding modes of this kind of SARS-CoV-2 M^{pro} inhibitors reveal that Met 49, Pro168, Met165, Cys145, and His164 might be the key residues for maintaining a stable conformation at the active site of SARS-CoV-2 M^{pro}. The designed active compounds N₃, N₄, N₅, and N₆ with M^{pro} complexes that have higher activities are further evaluated in terms of intermolecular interactions, complex stability, and binding affinity against SARS-CoV-2 M^{pro}, with 7D3I as the reference complex. The evaluation results suggest all newly designed compounds with SARS-CoV-2 M^{pro} have a similar mechanism of activity and stability compared to 7D3I. Binding free energy calculations on the SARS-CoV-2 M^{pro}-ligand complexes also revealed that the compound N₃, N₄, and N₆ have bigger binding free energy than compound 23 (PDB entry 7D3I), which indicates their potential to become leading compounds for treating SARS-CoV-2 virus. Those results verify the prediction ability of the QSAR models. The Van der Waals interaction energies would dominate five compounds interactions over electrostatic contributions. Per-residue energy decomposition analysis demonstrates binding affinity between the ligands and SARS-CoV-2 M^{pro} is more dependent on hydrophobic interactions. Furthermore, all newly designed compounds are confirmed to have good ADMET properties through the webserver prediction. Also, compared to recent efforts to identify

promising attributes for previous coronavirus inhibitors, the present study deals with the existing SARS-CoV-2 M^{pro} inhibitors and avoids activity test deviation between different data sets [21]. Taken together, the results from this study provide a reference for the further rational design of novel SARS-CoV-2 M^{pro} inhibitors with high potency.

Author Statement

Jian-Bo Tong, Ding Luo: Conceptualization, Methodology, Software, Original draft preparation, Writing-Reviewing, and Editing, Software. Xue-Chun Xiao, Xing Zhang: Visualization, Investigation. Shuai Bian: Supervision.

Funding

This work was supported by the National Natural Science Foundation of China (21475081), the Natural Science Foundation of Shaanxi Province of China (2019JM-237), and the Graduate Innovation Fund of Shaanxi University of Science and Technology.

Declaration of competing interest

The authors declare that they have no conflict of interest

Acknowledgments

The authors would like to sincerely thank Prof. Ran Fang at Shaanxi University of Science and Technology for providing high-performance computer for research facilities.

References

[1] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen,

- X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579(7798) (2020) 270-273.
- [2] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E.C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *The Lancet* 395(10224) (2020) 565-574.
- [3] I.F.-N. Hung, V.C.-C. Cheng, X. Li, A.R. Tam, D.L.-L. Hung, K.H.-Y. Chiu, C.C.-Y. Yip, J.-P. Cai, D.T.-Y. Ho, S.-C. Wong, S.S.-M. Leung, M.-Y. Chu, M.O.-Y. Tang, J.H.-K. Chen, R.W.-S. Poon, A.Y.-F. Fung, R.R. Zhang, E.Y.-W. Yan, L.-L. Chen, C.Y.-K. Choi, K.-H. Leung, T.W.-H. Chung, S.H.-Y. Lam, T.P.-W. Lam, J.F.-W. Chan, K.-H. Chan, T.-C. Wu, P.-L. Ho, J.W.-M. Chan, C.-S. Lau, K.K.-W. To, K.-Y. Yuen, SARS-CoV-2 shedding and seroconversion among passengers quarantined after disembarking a cruise ship: a case series, *Lancet Infect Dis* 20(9) (2020) 1051-1060.
- [4] J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M.-d. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H.C. Lane, Remdesivir for the Treatment of Covid-19 — Final Report, *New England Journal of Medicine* 383(19) (2020) 1813-1826.
- [5] M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, G. Xiao, Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*, *Cell Res* 30(3) (2020) 269-271.
- [6] D.W. Kneller, S. Galanie, G. Phillips, H.M. O'Neill, L. Coates, A. Kovalevsky, Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals, *Structure* 28(12) (2020) 1313-1320.e3.
- [7] Y. Bai, F. Ye, Y. Feng, H. Liao, H. Song, J. Qi, G.F. Gao, W. Tan, L. Fu, Y. Shi, Structural basis for the inhibition of the SARS-CoV-2 main protease by the anti-HCV drug nardaparvir, *Signal Transduction and Targeted Therapy* 6(1) (2021) 51.
- [8] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, *J Gen Virol* 83(3) (2002) 595-599.
- [9] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582(7811) (2020) 289-293.
- [10] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H. Jung, An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy, *J Med Chem* 59(14) (2016) 6595-6628.

- [11] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases, *PLoS Biol* 3(10) (2005) e324.
- [12] S. Chtita, A. Belhassan, M. Bakhouch, A.I. Taourati, A. Aouidate, S. Belaidi, M. Moutaabbid, S. Belaouad, M. Bouachrine, T. Lakhlifi, QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods, *Chemometrics and Intelligent Laboratory Systems* 210 (2021) 104266.
- [13] R. Islam, M.R. Parves, A.S. Paul, N. Uddin, M.S. Rahman, A.A. Mamun, M.N. Hossain, M.A. Ali, M.A. Halim, A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2, *Journal of Biomolecular Structure and Dynamics* (2020) 1-12.
- [14] V.M. Alves, T. Bobrowski, C.C. Melo-Filho, D. Korn, S. Auerbach, C. Schmitt, E.N. Muratov, A. Tropsha, QSAR Modeling of SARS-CoV Mpro Inhibitors Identifies Sufugolix, Cenicriviroc, Proglumetacin, and other Drugs as Candidates for Repurposing against SARS-CoV-2, *Mol Inform* 40(1) (2021) 2000113.
- [15] S. Yuan, X. Yin, X. Meng, J.F.-W. Chan, Z.-W. Ye, L. Riva, L. Pache, C.C.-Y. Chan, P.-M. Lai, C.C.-S. Chan, V.K.-M. Poon, A.C.-Y. Lee, N. Matsunaga, Y. Pu, C.-K. Yuen, J. Cao, R. Liang, K. Tang, L. Sheng, Y. Du, W. Xu, C.-Y. Lau, K.-Y. Sit, W.-K. Au, R. Wang, Y.-Y. Zhang, Y.-D. Tang, T.M. Clausen, J. Pihl, J. Oh, K.-H. Sze, A.J. Zhang, H. Chu, K.-H. Kok, D. Wang, X.-H. Cai, J.D. Esko, I.F.-N. Hung, R.A. Li, H. Chen, H. Sun, D.-Y. Jin, R. Sun, S.K. Chanda, K.-Y. Yuen, Clofazimine broadly inhibits coronaviruses including SARS-CoV-2, *Nature* 593(7859) (2021) 418-423.
- [16] K. Pathan Mohsin, K. Vinay, R. Kunal, In Silico Modeling of Small Molecule Carboxamides as Inhibitors of SARS-CoV 3CL Protease: An Approach Towards Combating COVID-19, *Comb Chem High Throughput Screen* 23 (2020) 1-19.
- [17] K. Ghosh, S.A. Amin, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: Exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors, *J Mol Struct* 1224 (2021) 129026.
- [18] P. De, S. Bhayye, V. Kumar, K. Roy, In silico modeling for quick prediction of inhibitory activity against 3CLpro enzyme in SARS CoV diseases, *Journal of Biomolecular Structure and Dynamics* (2020) 1-27.
- [19] J.-B. Tong, D. Luo, H.-Y. Xu, S. Bian, X. Zhang, X.-C. Xiao, J. Wang, A computational approach for designing novel SARS-CoV-2 Mpro inhibitors: combined QSAR, molecular docking, and molecular dynamics simulation techniques, *New J Chem* 45(26) (2021) 11512-11529.
- [20] A.A.-A.A. Abu-Saleh, I.E. Awad, A. Yadav, R.A. Poirier, Discovery of potent inhibitors for SARS-CoV-2's main protease by ligand-based/structure-based virtual screening, MD simulations, and binding energy calculations, *Physical Chemistry Chemical Physics* 22(40) (2020) 23099-23106.
- [21] S.A. Amin, S. Banerjee, S. Singh, I.A. Qureshi, S. Gayen, T. Jha, First structure-activity relationship analysis of SARS-CoV-2 virus main protease (Mpro) inhibitors: an

- endeavor on COVID-19 drug discovery, *Mol Divers* (2021).
- [22] J. Qiao, Y.-S. Li, R. Zeng, F.-L. Liu, R.-H. Luo, C. Huang, Y.-F. Wang, J. Zhang, B. Quan, C. Shen, X. Mao, X. Liu, W. Sun, W. Yang, X. Ni, K. Wang, L. Xu, Z.-L. Duan, Q.-C. Zou, H.-L. Zhang, W. Qu, Y.-H.-P. Long, M.-H. Li, R.-C. Yang, X. Liu, J. You, Y. Zhou, R. Yao, W.-P. Li, J.-M. Liu, P. Chen, Y. Liu, G.-F. Lin, X. Yang, J. Zou, L. Li, Y. Hu, G.-W. Lu, W.-M. Li, Y.-Q. Wei, Y.-T. Zheng, J. Lei, S. Yang, SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model, *371(6536)* (2021) 1374-1378.
- [23] Z.T. Muhseen, A.R. Hameed, H.M.H. Al-Hasani, M. Tahir ul Qamar, G. Li, Promising terpenes as SARS-CoV-2 spike receptor-binding domain (RBD) attachment inhibitors to the human ACE2 receptor: Integrated computational approach, *J Mol Liq* *320* (2020) 114493.
- [24] Y.-L. Wang, F. Wang, X.-X. Shi, C.-Y. Jia, F.-X. Wu, G.-F. Hao, G.-F. Yang, Cloud 3D-QSAR: a web tool for the development of quantitative structure–activity relationship models in drug discovery, *Brief Bioinform* (2020).
- [25] S. Seetaha, B. Boonyarit, S. Tongshima, N. Songtawee, K. Choowongkamon, Potential tripeptides against the tyrosine kinase domain of human epidermal growth factor receptor (HER) 2 through computational and kinase assay approaches, *J Mol Graph Model* *97* (2020) 107564.
- [26] G. Weng, J. Gao, Z. Wang, E. Wang, X. Hu, X. Yao, D. Cao, T. Hou, Comprehensive Evaluation of Fourteen Docking Programs on Protein–Peptide Complexes, *J Chem Theory Comput* *16(6)* (2020) 3959-3969.
- [27] J.-B. Tong, D. Luo, X. Zhang, S. Bian, Design of novel SHP2 inhibitors using Topomer CoMFA, HQSAR analysis, and molecular docking, *Struct Chem* *32(3)* (2021) 1061-1076.
- [28] H. Razmazma, A. Ebrahimi, M. Hashemi, Structural insights for rational design of new PIM-1 kinase inhibitors based on 3,5-disubstituted indole derivatives: An integrative computational approach, *Comput Biol Med* *118* (2020) 103641.
- [29] M. Lorca, C. Morales-Verdejo, D. Vásquez-Velásquez, J. Andrades-Lagos, J. Campanini-Salinas, J. Soto-Delgado, G. Recabarren-Gajardo, J. Mella, Structure-Activity Relationships Based on 3D-QSAR CoMFA/CoMSIA and Design of Aryloxypropanol-Amine Agonists with Selectivity for the Human β 3-Adrenergic Receptor and Anti-Obesity and Anti-Diabetic Profiles, *Molecules* *23(5)* (2018).
- [30] J. Verma, V.M. Khedkar, E.C. Coutinho, 3D-QSAR in Drug Design - A Review, *Curr Top Med Chem* *10(1)* (2010) 95-115.
- [31] J.-B. Tong, D. Luo, S. Bian, X. Zhang, Structural investigation of tetrahydropteridin analogues as selective PLK1 inhibitors for treating cancer through combined QSAR techniques, molecular docking, and molecular dynamics simulations, *J Mol Liq* *335* (2021) 116235.
- [32] G.C. Veríssimo, E.F. Menezes Dutra, A.L. Teotonio Dias, P. de Oliveira Fernandes, T. Kronenberger, M.A. Gomes, V.G. Maltarollo, HQSAR and random forest-based QSAR models for anti-*T. vaginalis* activities of nitroimidazoles derivatives, *J Mol Graph Model* *90* (2019) 180-191.
- [33] T.S. David, QSAR and QSPR Model Interpretation Using Partial Least Squares

- (PLS) Analysis, *Curr Comput Aided Drug Des* 8(2) (2012) 107-127.
- [34] P. Gramatica, A. Sangion, A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology, *J Chem Inf Model* 56(6) (2016) 1127-1131.
- [35] K. Roy, R.N. Das, P. Ambure, R.B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, *Chemometrics and Intelligent Laboratory Systems* 152 (2016) 18-33.
- [36] V. Consonni, D. Ballabio, R. Todeschini, Comments on the Definition of the Q2 Parameter for QSAR Validation, *J Chem Inf Model* 49(7) (2009) 1669-1678.
- [37] A. Golbraikh, A. Tropsha, Beware of q²!, *J Mol Graph Model* 20(4) (2002) 269-276.
- [38] L. Meng, K. Feng, Y. Ren, Molecular modelling studies of tricyclic triazinone analogues as potential PKC- θ inhibitors through combined QSAR, molecular docking and molecular dynamics simulations techniques, *Journal of the Taiwan Institute of Chemical Engineers* 91 (2018) 155-175.
- [39] S. Cosconati, S. Forli, A.L. Perryman, R. Harris, D.S. Goodsell, A.J. Olson, Virtual screening with AutoDock: theory and practice, *Expert Opin Drug Discov* 5(6) (2010) 597-607.
- [40] N.T. Nguyen, T.H. Nguyen, T.N.H. Pham, N.T. Huy, M.V. Bay, M.Q. Pham, P.C. Nam, V.V. Vu, S.T. Ngo, Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity, *J Chem Inf Model* 60(1) (2020) 204-211.
- [41] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J Comput Chem* 31(2) (2010) 455-461.
- [42] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: Fast, flexible, and free, *J Comput Chem* 26(16) (2005) 1701-1718.
- [43] B.R. Brooks, C.L. Brooks Iii, A.D. Mackerell Jr, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoseck, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: The biomolecular simulation program, *J Comput Chem* 30(10) (2009) 1545-1614.
- [44] K. Vanommeslaeghe, A.D. MacKerell, Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing, *J Chem Inf Model* 52(12) (2012) 3144-3154.
- [45] U. N, K. S, Molecular dynamics simulation study on *Thermotoga maritima* EngA: GTP/GDP bound state of the second G-domain influences the domain-domain interface interactions, *Journal of Biomolecular Structure and Dynamics* (2020) 1-13.
- [46] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J.Z.H. Zhang, T. Hou, End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design, *Chem Rev* 119(16) (2019) 9478-9508.
- [47] H. Sun, Y. Li, M. Shen, S. Tian, L. Xu, P. Pan, Y. Guan, T. Hou, Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring,

Physical Chemistry Chemical Physics 16(40) (2014) 22035-22045.

- [48] R. Kumari, R. Kumar, A. Lynn, *g_mmpbsa*—A GROMACS Tool for High-Throughput MM-PBSA Calculations, *J Chem Inf Model* 54(7) (2014) 1951-1962.
- [49] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations, *J Chem Inf Model* 51(1) (2011) 69-82.
- [50] G. Xiong, Z. Wu, J. Yi, L. Fu, Z. Yang, C. Hsieh, M. Yin, X. Zeng, C. Wu, A. Lu, X. Chen, T. Hou, D. Cao, ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties, *Nucleic Acids Res* (2021).
- [51] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci Rep* 7(1) (2017) 42717.
- [52] D.L. Nossa González, J.A. Gómez Castaño, W.E. Rozo Núñez, P.R. Duchowicz, Antiprotozoal QSAR modelling for trypanosomiasis (Chagas disease) based on thiosemicarbazone and thiazole derivatives, *J Mol Graph Model* 103 (2021) 107821.
- [53] M. Chatterjee, K. Roy, Prediction of aquatic toxicity of chemical mixtures by the QSAR approach using 2D structural descriptors, *J Hazard Mater* 408 (2021) 124936.
- [54] N.-E.-H. Hammoudi, W. Sobhi, A. Attoui, T. Lemaoui, A. Erto, Y. Benguerba, In silico drug discovery of Acetylcholinesterase and Butyrylcholinesterase enzymes inhibitors based on Quantitative Structure-Activity Relationship (QSAR) and drug-likeness evaluation, *J Mol Struct* 1229 (2021) 129845.
- [55] W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang, H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* (80-) 368(6497) (2020) 1331-1335.