



Research Article

The Integration of Pharmacophore-Based 3D-QSAR Modeling and Virtual Screening in Identification of Natural Product Inhibitors against SARS-CoV-2

Samira Norouzi¹, Maryam Farahani¹, Samad Nejad Ebrahimi^{1*}

¹Department of Phytochemistry, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Evin, Tehran, Iran.

Article Info

Article History:

Received: 6 August 2020
Accepted: 1 December 2020
ePublished: 3 December 2020

Keywords:

-Coronavirus
-SARS-CoV-2
-COVID-19
-SARS-CoV
-Pharmacophore
-3D-QSAR

Abstract

Background: The current outbreak of Coronavirus Disease 2019 (SARS-CoV-2) led to public health emergencies all over the world and made it a global concern. Also, the lack of an effective treatment to combat this virus is another concern that has appeared. Today, increasing knowledge of biological structures like increasing computer power brings about a chance to use computational methods efficiently in different phases of drug discovery and development for helping solve this new global problem.

Methods: In this study, 3D pharmacophores were generated based on thirty-one structures with functional affinity inhibition (antiviral drugs used for SARS and MERS) with $IC_{50} < 250 \mu M$ from the literature data. A 3D-QSAR model has been developed and validated to be utilized in virtual screening.

Results: The best pharmacophore models have been utilized as 3D queries for virtual screening to gain promising inhibitors from a data set of thousands of natural compounds retrieved from PubChem. The hit compounds were subsequently used for molecular docking studies to investigate their affinity to the 3D structure of the SARS-CoV-2 receptors. The ADMET properties calculate for the hits with high binding affinity.

Conclusion: The study outcomes can help understand the molecular characteristics and mechanisms of the binding of hit compounds to SARS-CoV-2 receptors and promising identification inhibitors that are likely to be evolved into drugs.

Introduction

The world was dismayed by the outbreak of a fatal infection at the beginning of 2020. The new severe viral cute respiratory syndrome coronavirus 2 (SARS-CoV-2) rapidly spreads from birthplace in Wuhan, China, to countries and continents boundaries.¹ The spread of this virus is mainly by close contact with other people, direct contact with infectious materials, and the respiratory droplets of a person infected with the virus when they cough or sneeze.² The fatality rate of this new CoV is much less than SARS and MERS, and it is around %2 in China. Rapid and active transmission of the virus from human to human has led to its spread worldwide.³⁻⁵ SARS-CoV-2 belongs to the beta coronavirus family.⁶ The SARS-CoV-2 showed the genetic similarity of about 79.6% compared to SARS-CoV and 50% sequence identity to MERS-CoV. Still, it is considered distinct from two of these high pathogenic viruses.⁷ Angiotensin-converting enzyme 2 (ACE2) is the target receptor of 2019-nCoV. By entering the virus into the host cell, the positive genomic RNA connects directly to

the host ribosome to translate two abundant, conterminal polyproteins. They are processed by proteolysis into components for packaging new virions.⁸ The papain-like protease (PLpro) and the coronavirus main proteinase (3CLpro) plays an essential role in the proteolysis process.^{9,10} when spike protein (S) binds the host cell-surface molecules, the replication of CoV starts inside the host cell.¹¹ The S protein contains two practical subunits that are in charge of linking to the host cell receptor (S1 subunit) and blending the viral and cellular membranes (S2 subunit). This protein is an ideal target for developing a vaccine and antiviral drugs due to its pivotal roles in receptor linkage and membrane fusion.

Recent trial agents reported some potential therapy for SARS-CoV-2, such as remdesivir¹² and combined protease inhibitor lopinavir-ritonavir,¹³ that are used to treat against SARS and MERS-CoV, but their efficacy is still unclear and needs further evaluation. Previously, native ligand peptide-like inhibitor PRD_002214 (N3) inhibitor was identified

*Corresponding Author: Samad Nejad Ebrahimi, E-mail: s_ebrahimi@sbu.ac.ir

©2021 The Author(s). This is an open access article and applies the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited.

and reported in the protein databank (PDB) with Mpro of the SARS-CoV-2 in the type of crystal structure complex (PDB ID, 6LU7). The native ligand (PRD_002214) was selected for ligand screening in the ZINC database using a pharmacophore-based online tool (ZINC Pharmer). Although many clinical trials are in progress, scientists have yet to find approved drugs or vaccines for SARS-CoV-2, and finding a cure can still be a great aid to the international community.¹⁴

The development of novel drugs is a time-consuming process, and generally, several years of work are required for clinical approval. Hence, computational methods can be utilized for the design and engineering of drugs.¹⁵⁻¹⁷ The short time requirements of computational methods are conducive for high throughput screening of available medications to identify potential drugs for new diseases and predict the adverse effects of novel drugs.¹⁸⁻²⁰

The pharmacophore-based ligands screening, 3D-QSAR modeling, and molecular docking were applied to discover new inhibitors acting against the SARS-CoV-2 essential receptors (main proteinase, spike ectodomain structure, and receptor-binding domain). In this regard, ligand-based pharmacophore models were built based on the structures with functional affinity inhibition displaying in vitro with IC_{50} value $<250 \mu\text{M}$. The validation of the ligand-based pharmacophores was implemented by enrichment analysis; then, the validated models were used for the HTVS and docking analysis. Finally, the fifteen structures were proposed as lead compounds based upon the validated pharmacophore models, scored by molecular docking and 3D-QSAR predictions.

Materials and Methods

Computational details

The PHASE 3.4 module performed the generation of Pharmacophore features and 3D QSAR models in the Maestro 9.3 (Schrödinger, LLC). Virtual screening based on pharmacophore models and docking studies was executed utilizing the PHASE 3.4 module and Glide 5.8 module (Maestro 9.3). The prediction of ADME properties was carried out by QikProp 3.5 in Maestro 9.3

Preparation of protein structures

Three crystal structures of SARS-CoV-2 were obtained from the protein data bank (PDB). For the SARS-CoV-2 main protease (Mpro), the selected crystal structure was 6LU7, whereas 6VYB and 6VW1 were considered for spike ectodomain structure and receptor-binding domain, respectively. The preparation of all protein structures was performed by the protein preparation wizard (Maestro 9.3). The protein grid box was created around the active pocket for using the protein grid generation module of the Glide program. The native ligand peptide-like inhibitor PRD_002214 (N_3) was settled as the core of the grid box of the 6LU7; for the 6VYB and 6VW1, the sitemap module was used for a generation of the active site. We used the default setting for the van der Waals scaling as well as

partial charges. The size of all structures binding pocket was settled to be proper for ligands with a dimension of 20 \AA .

Chemical structure preparation

Conversion of 2D to 3D structures was performed by LigPrep 2.5 through the Schrödinger suite package. The hydrogens added to the structures, and salts have been removed; the possible stereoisomers generated using Epik ($\text{pH } 7 \pm 2$) and energetically minimized. The generation of 3D conformers was performed based on ≤ 10 per rotatable bond and ≤ 100 per ligand, and the RMSD cut-off value of 1.0 \AA removed the redundant conformers. The OPLS3 force field performed the energy minimization of structures, and then to exclude high energy structures, filtration of conformers was done based upon a relevant energy window of 10.0 kcal/mol .

Creating pharmacophore-based models

Thirty-one structures with good affinity inhibition (antiviral drugs used for SARS and MERS) with IC_{50} value and diverse structures were collected from previously published data^{11,21-24} and prepared as said (The structures are reported in Table S1 in supplementary data). Generation of Pharmacophore hypothesis and statistical analyses carried out by PAHSE to form the 3D-QSAR model. After conformational development, the pharmacophore hypothesis was created on each ligand structure. The following formula shows the conversion of IC_{50} value into pIC_{50} to obtain a linear association in the QSAR equation as follow:

$$pIC_{50} = -\log IC_{50} \quad (1)$$

The IC_{50} is the half-maximal response inhibition. The structures and IC_{50} value of these compounds showed in supporting information. We set an activity threshold range of pIC_{50} for actives ≥ 5.5 and inactive < 5.0 . The evaluation and identification of best pharmacophores obtained by the scoring of the generated pharmacophore hypothesis. That was done by the adjustment of the site points, vector alignment, and volume overlap.²⁵

Enrichment calculations

The enrichment calculation was used to investigate and validation of the quality of the pharmacophore hypothesis. $EF(X \%)$ is the number of actives gained after the random screening of $X\%$ of the decoy set.²⁶ The validation set was composed of 1089 molecules containing 1000 decoy set (retrieved from Schrödinger database), and 89 known inhibitors were used to calculate enrichment factors. The significance of pharmacophore models was ensured based on the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC),²⁷ Enrichment Factor (EF), Goodness of Hit (GH) and Robust Initial Enhancement (RIE). The EF and GH values will be determined by using equation 4 and 5.²⁸

$$EF(\text{manual}) = \frac{H_a \times D}{H_t \times A} \quad (2)$$

$$GH = \left[\left(\frac{H_a}{4H_t A} \right) \times (3A + H_t) \right] \times \left[1 - \left(\frac{H_t - H_a}{D - A} \right) \right] \quad (3)$$

H_t is the sum of compounds, and H_a is the sum number of actives in the hit list. D the sum number of molecules and A the total number of actives in the decoy set. These two parameters (EF manual and GH metrics) are utilized to calculate the entire dataset.

Pharmacophore-based 3D QSAR modeling

Atom-based partial least square regression (PLS) method was utilized for the generation of 3D QSAR models. The default parameters were used for PLS. For three top-scored pharmacophore hypothesis, QSAR models were created by training compounds aligned to the pharmacophores at least on three sites with four features and sections for pharmacophores, where five components showed best alignments.²⁵ A rectangular grid was set into cubes of 1 Å to involve occupied space by the training set active ligands. Cubes were started as filled in a hypothesis pharmacophore site within the relevant sphere's radius. A compound can be described by a string of zeros and ones based upon the cube's occupancy variation and the various types of sites that lodge in these cubes. It led to binary values as 3D descriptors. PLS generated 3D-QSAR models to the pool of binary-valued variables.²⁵ The validation of QSAR models was performed by predicting activities of test set compounds.²⁹

External validation and PLS analysis of QSAR models

For the generated 3D-QSAR models, we used the statistical method of partial least-squares analysis. With the aid of the "Leave One Out" (LOO) method, the predictive ability of the models and cross-validation analysis was evaluated. R²_{cv} (cross-validated coefficient) was calculated as the following equation (4):

$$r_{cv}^2 = 1 - \frac{\sum(Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum(Y_{\text{observed}} - Y_{\text{mean}})^2} \quad (4)$$

Where Y_{predicted}, Y_{observed}, and Y_{mean} are the predicted, observed, and mean values of the target property (pIC₅₀) of the training set, respectively.³⁰

The predictive correlation coefficient (r_{pred}²), based on molecules of the test set, was calculated as:³¹

$$r_{\text{pred}}^2 = 1 - \frac{\sum(Y_{\text{pred}(\text{test})} - Y_{\text{test}})^2}{\sum(Y_{\text{test}} - \bar{Y}_{\text{training}})^2} \quad (5)$$

Y_{pred (test)} and Y_(test) show predicted as well as observed activity values respectively of the test set and $\bar{Y}_{\text{training}}$ shows the mean activity value of the training set compounds.

The following formula gained a high correlation coefficient R (or r² value):

$$R = \frac{\sum(Y_i - \bar{Y}_o)(\tilde{Y}_i - \bar{Y}_p)}{\sqrt{\sum(Y_i - \bar{Y}_o)^2 \sum(\tilde{Y}_i - \bar{Y}_p)^2}} \quad (6)$$

Y_i and \tilde{Y}_i indicate the observed and predicted activities of the test set compounds, respectively, and \bar{Y}_o and \bar{Y}_p are the mean values of the observed and anticipated activities of the test set molecules.

Utilizing regressions of Y_i against \tilde{Y}_i , or \tilde{Y}_i against Y_i through the origin, Y^{ro} = k \tilde{Y} and \tilde{Y}^{ro} = k' Y_i , respectively,³² the slopes k and k' were calculated as:

$$K = \frac{\sum Y_i \tilde{Y}_i}{\sum \tilde{Y}_i^2} \quad (7a)$$

$$K' = \frac{\sum Y_i \tilde{Y}_i}{\sum Y_i^2} \quad (7b)$$

R₀² and R'₀² define the correlation coefficients for the regression lines through the origin (Y^{ro}) is described by R₀² and R'₀² that were calculated as follows:³³

$$R_0^2 = 1 - \frac{\sum(\tilde{Y}_i - Y_i^{\text{ro}})^2}{\sum(\tilde{Y}_i - \bar{Y}_p)^2} \quad (8a)$$

$$R_0'^2 = 1 - \frac{\sum(Y_i - \tilde{Y}_i^{\text{ro}})^2}{\sum(Y_i - \bar{Y}_p)^2} \quad (8b)$$

The modified r² parameter that is utilized for the whole set compounds, r_m² value is described by the following equation:³³

$$r_m^2 = r^2 - \left(1 - \sqrt{r^2 - R_0^2} \right) \quad (9)$$

r² is the non-cross-validated correlation coefficient gained from the PLS process, and R₀² is gained from eq8a. According to literature,^{32,33} our 3D-QSAR models are valid and reliable if: r² > 0.5, r²_{cv} (Q²) > 0.5, the value of R₀² and R'₀² should be close to r², [(r² - R₀²)/r²] < 0.1, [(r² - R'₀²)/r²] < 0.1, the values of k and k' should be 0.85 ≤ k ≤ 1.15 or 0.85 ≤ k' ≤ 1.15, and r_m² > 0.5. Finally, the 3D-QSAR models build with the high prediction ability and the related pharmacophores to find the best models and new lead compounds.

Pharmacophore screening of 3D databases and 3D QSAR model

A 3D structure of two thousand natural compounds was energy minimized, and conformations were generated for making a PHASE library. Searching the natural compounds database with the best pharmacophores similarity was

carried out utilizing the PHASE module, and the tolerance distance set to 2.0 Å in the hypothesis. In pharmacophore screening, compounds must match 3 or 4 pharmacophoric sites on a hypothesis with four features and match four sites of the hypothesis with five or more sites. The hits derived from virtual screening ranked based on fitness scores.²⁵ This score shows the alignment of hits on the hypothesis and is based on on-site matching, RMSD, volume terms, and vector alignment. The fitness score range was 0 to 3, where the score of 3 indicates that our ligands have the maximum alignment with the hypothesis.

Molecular docking

A 3D database of hits with the best fitness scoring was used for docking study against different types of SARS-CoV-2 receptors. The docking study of these compounds against the active site of the protein structures was performed by Glide 5.7 (Glide, version 5.7, Schrödinger, LLC, New York, NY, 2011) in three steps: HTVS, standard precision (SP), and extra precision (XP). RMSD values calculation between the native and docked conformations were carried out utilizing core pattern comparison in Glide. The default settings were being used for both the grid generation and the docking.

ADME properties prediction

The QikProp (version 4.5, Schrodinger, LLC, and New York, NY) was utilized to study The ADME properties of the hits obtained from the virtual pharmacophore screening that showed high XP docking score. QikProp is being used for analyzing the related pharmaceutical futures, such as octanol/water log Ps, log S, predicted brain/blood partition coefficient (QPlogBB), human oral absorption, central nervous system activity, Caco-2, and Lipinski rule of five.³⁴

Result and Discussion

Protein preparation

Three crystal structures of SARS-CoV-2 with low resolution were downloaded from PDB. Protein preparation wizard prepared these proteins, and grid box generation was carried out by the Glide program. The active pocket of the SARS-CoV-2 main protease (PDB: 6lu7) was created

around the ligand N₃, and for indicating the active site of spike ectodomain structure (PDB: 6vyb) and receptor-binding domain (PDB: 6vw1), site map generation was used. Finally, five site maps were generated for each receptor, and sites with the highest site score were used for grid generation. Table 1 indicates the grid generated for each protein.

Ligand-based pharmacophore models

We generated the ligand-based pharmacophores based on antiviral inhibitors acting against SARS and MERS by the PHASE module (Schrödinger). These compounds were divided into active ligand and inactive ligands. Finally, 14 pharmacophore hypothesis was generated, and then three pharmacophore hypothesis were chosen based upon their excellent survival activity, site scores, best active adjustment, and the number of matches (Table 2). Hypothesis 1 (AADDR) indicated two hydrogen-bond acceptors, two hydrogen-bond donors, and one aromatic ring. Hypothesis 2 (AAADR) proved three hydrogen-bond acceptors, one hydrogen-bond donor and one aromatic ring, and hypothesis 3 (AAHR) showed two hydrogen bond acceptors, one hydrophobic and one aromatic ring as features (Figure 1). Table 3 shows the distances between the features of pharmacophore models.

Validation of the pharmacophores for virtual screening was needed. Hence, the enrichment factor was carried out for validation of the generated pharmacophores. The results in Table 4 show that hypothesis 1 (AADR) with the best survival score also indicated the highest EF1%, BEDROC, and GH values compared to other hypothesis. This could be inferred that the prediction ability of hypothesis AADDR is much more than the other hypothesis. The validation of these three hypothesis was carried out for QSAR predictability and PLS analysis. Figure 2 indicates the active and inactive ligands alignment based upon the best pharmacophore model (AADDR). Figure 2A clearly shows that all the most active ligands match entirely over all the pharmacophoric features – AADDR, whereas in Figure 2B all the least active ligands match the pharmacophore less precisely.

Table 1. Grid Information of the receptors with Their PDB IDs used for Docking Studies

PDB	COVID-19	X-Center	Y-Center	Z-Center
6LU7	Main protease	-26.10	12.57	59.10
6VYB	Spike ectodomain	212.49	210.07	185.54
6VW1	Receptor-binding domain	105.82	60.03	111.29

Table 2. Ligand-based pharmacophore hypothesis with their scores

Hypothesis	Survival score	Survival-inactive score	Vector score	Volume score	Score site	Activity
AADDR	3.37	2.79	0.91	0.74	0.69	10.15
AAADR	3.30	2.51	0.86	0.68	0.60	10.09
AAHR	2.46	1.55	0.85	0.50	0.46	10.15

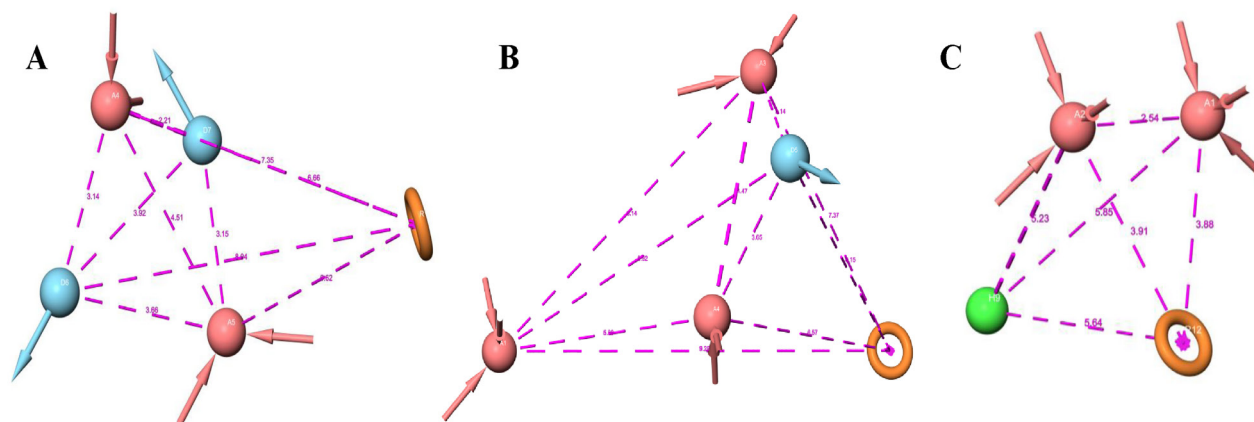


Figure 1. The final selected best pharmacophore hypothesis and distances between the features. **A:** Five-feature pharmacophore hypothesis AADDR with two acceptor features (A), two donor features (D), and one aromatic ring feature(R). **B:** Five-feature pharmacophore hypothesis AAADR with three acceptor features (A), one donor feature (D), and one aromatic ring feature(R). **C:** Four-feature pharmacophore hypothesis AARH with two acceptor features (A), one aromatic ring feature (R), and one hydrophobic feature (H). Pink sphere with arrows, hydrogen-bond acceptor (A); yellow open circle, aromatic ring (R); blue sphere with the arrow, hydrogen-bond donor (D); green area, hydrophobic (H).

Table 3. Distance between the Features of Ligand-Based Pharmacophores.

Hypothesis	Distance(Å)															
	A-A	A-A	A-A	A-R	A-R	A-R	A-D	A-D	A-D	A-D	D-D	D-R	D-R	A-H	A-H	H-R
AADDR	4.51	-	-	5.69	4.28	-	3.13	2.21	3.66	3.15	3.91	5.83	4.27	-	-	-
AAADR	4.70	7.98	4.82	5.24	3.62	5.95	9.04	4.74	3.15	-	-	5.61	-	-	-	-
AAHR	2.59	-	-	3.90	3.88	-	-	-	-	-	-	-	-	6.29	5.01	5.64

Table 4. Validation of Ligand-Based Hypothesis features: EF 1%^a, RIE^b, ROC^c, BEDROC^d (α -160.9), EF^e, GH^f.

No.	Hypothesis	EF 1%	RIF	BEDROC (α -160.9)	ROC	BEDROC (α -20.0)	ROC	EF	GH
1	AADDR	3.36	1.30	0.36	0.14	0.43	0.43	2.03	0.14
2	AAADR	3.36	1.30	0.36	0.14	0.43	0.43	2.03	0.14
3	AAHR	2.41	0.54	0.12	0.01	0.50	0.50	1.70	0.16

^a EF: Enrichment factor at 1% of the decoy data set. ^b RIE: Robust initial enhancement. ^c ROC: Receiver operating characteristic curve value. ^d BEDROC: Boltzmann-enhanced discrimination of receiver operating characteristic. ^e EF: Overall enrichment factor. ^f GH: Goodness of fit

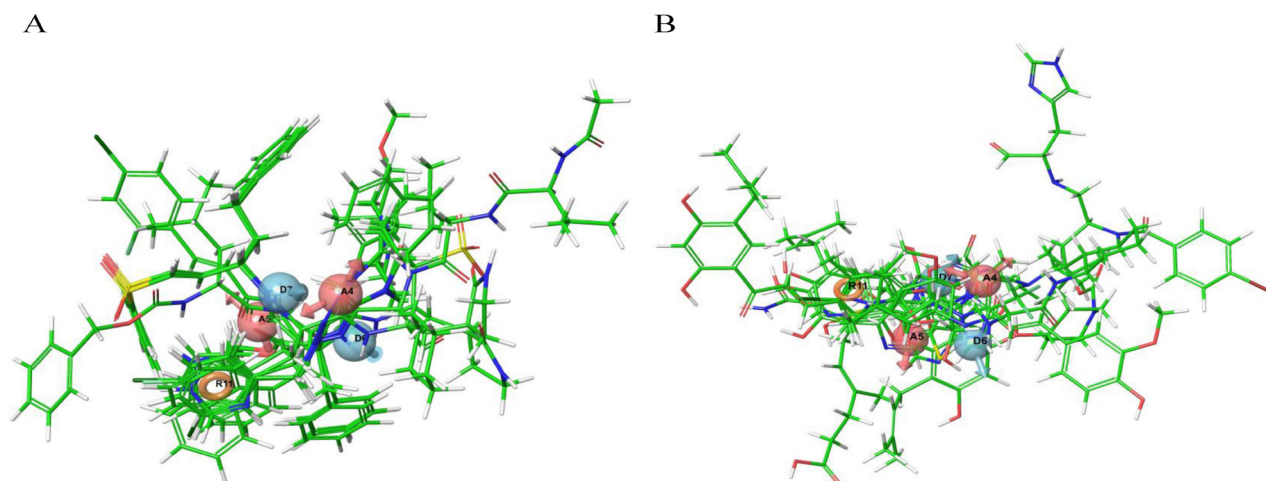


Figure 2. The overlap of (a) actives and (b) inactive ligands over the top-ranked pharmacophore hypothesis.

The generation of 3D-QSAR models and PLS analysis

The internal and external statistical validation is necessary for developing excellent 3D-QSAR models, which could show more reliable predictions. The more reliable estimation obtains from the models that can accomplish statistical validation parameter boundaries. Randomly, nineteen chosen for the training and twelve compounds in the test set were selected for a generation of the 3D-QSAR model. Based on the LOO method, significant parameters were gained and validated using PLS analysis. Based on PLS analysis, hypothesis 1 (AADDR) presented good external predictive ability compared to others. Hypothesis 1 showed an excellent R^2 value (0.9985) for the training set and a good predictive power Q^2 (0.5698) for the test sets, with an SD of 0.068 and F value of 3347.6. The r^2_{pred} predicted the unity of the model with a value of 0.5418 (Table 5). To attain functional predictive capacity in a QSAR model, R^2 (the accepted LOO-cross validated the value of training set) should show a value greater than 0.6, and Q^2 (the accepted LOO cross-validated value for test set) should be higher than 0.55. Also, a good model should

show a low value of Standard deviation (SD) (< 0.3), a minimum root-mean-square error (RMSE), and a high value of variance ratio (F).

zQ^2 is commonly utilized for assessing the prediction ability of a QSAR model. Although a high value of Q^2 is imperative, Q^2 alone is not an adequate condition for a QSAR model to possess a high predictive ability. A high value of correlation coefficient R (or r^2) of the observed and predicted external test set activities is also needed for a reliable model. In this study, the model with the best predictive ability was identified by the correlation coefficient $R^2 = 0.774$. The R_0^2 and $R_0'^2$ values were 0.988 and 0.986 respectively, which further utilized for calculating the association between r^2 , R_0^2 , and $R_0'^2$, $[(r^2 - R_0^2)/r^2]$ and $[(r^2 - R_0'^2)/r^2]$, which gave a value of -0.751 and -0.753 , respectively, within statistical limits (Table 6). Also, r_m^2 - a parameter of modified r^2 regarded for the better external predictive potential for the whole set of compounds- was calculated and found to be 0.864. Hence, the prediction ability of the 3D-QSAR model in foretelling the active compounds was confirmed based upon the external numerical parameters. Figure 3 shows

Table 5. PHASE 3D-QSAR and PLS statistic for internal validation of the data set.

	Training set	Test set	R^2	F	Q^2	SD	Pearson-R	RMSE	R	r^2_{pred}
AADDR	19	12	0.99	3347.6	0.56	0.06	0.7490	1.07	0.51	0.541
AAADR	19	12	0.94	596.1	0.53	0.14	0.68	1.04	0.43	0.523
AAHR	19	12	0.89	30.8	0.36	1.04	0.47	1.34	0.26	0.478

Notes: SD: Standard deviation of the regression; R^2 : regression co-efficient, F: the ratio of the observed activity variance to the model variance, P: Significant level of variance ratio, RMSE: root-mean-squared-error, and Q^2 : cross-validated correlation co-efficient of the test set.

Table 6. External Statistical Validation of 3D-QSAR model based on Hypothesis AADDR.

r_{cv}^2	R	r^2	k	k'	R_0^2	$R_0'^2$	$[(r^2 - R_0^2)/r^2]$	$[(r^2 - R_0'^2)/r^2]$	r_m^2
0.56	0.891	0.564	0.988	0.986	0.989	0.987	-0.751	-0.753	0.864

r_{cv}^2 (cross-validated coefficient) >0.5 , R (correlation coefficient between the actual and predicted activities) ~ 1 , slope values of regression lines k and k' $0.85 \leq k, k' \leq 1.15$, correlation coefficients for the regression lines through the origin R_0^2 or $R_0'^2$ close to r^2 , $[(r^2 - R_0^2)/r^2]$ and $[(r^2 - R_0'^2)/r^2] < 0.1$, $r_m^2(LOO) > 0.5$.

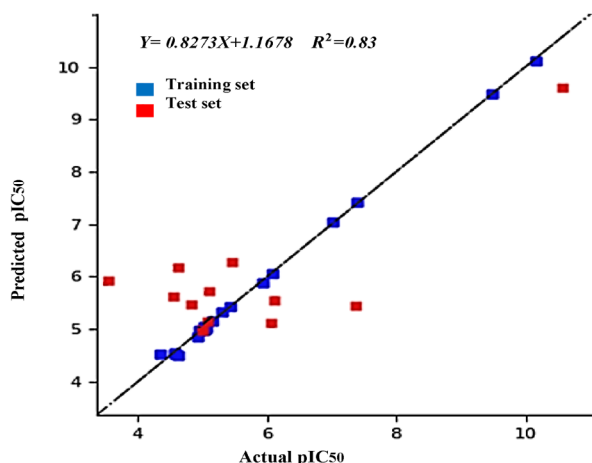


Figure 3. Scatter plot of actual and predicted biological activity of the training and the test set.

the scatter plot with the linear regression between the predicted activity values and the real activity values of the training and test set compounds with the R² value of 0.96 and 0.56, respectively.

Contour map analyses

To generate contour maps, we used the last validated

hypothesis AADDR gained from 3D-QSAR. With these maps aid, the substitutions place or atoms replacement can be identified to improve bioactivity. We can predict bioactivity by visualizing and realizing the Contour maps of the most active and least active ligands. It could lead to finding new scaffolds with valuable biological activity. The contour maps of the most and least active ligand are shown in Figure 4A and B indicate contour maps of the hydrogen-bond donor effect for the most productive and the least active compound with their most favorable zone, blue cubes, and unfavorable zone, red cubes. Hydrogen-bond donor mapping displayed that promising zones are situated near the nitrogen atoms amid functional groups of the most active ligand; this suggests that hydrogen bond donor groups at this location could increase the activity. The presence of hydrogen donor and acceptor groups could establish polar contact with coronavirus receptors. Also, the located blue parts around the hydroxyl group of the least active compound indicate that the hydrogen donor group's presence with cyclic rings in the scaffold of hits might be required for bioactivity.

Figure 4C and D show contour maps of the hydrophobic effect for the most and the least active compound with their favorable region, yellow cubes, and unfavorable regions, purple cubes. The favored yellow part around aromatic

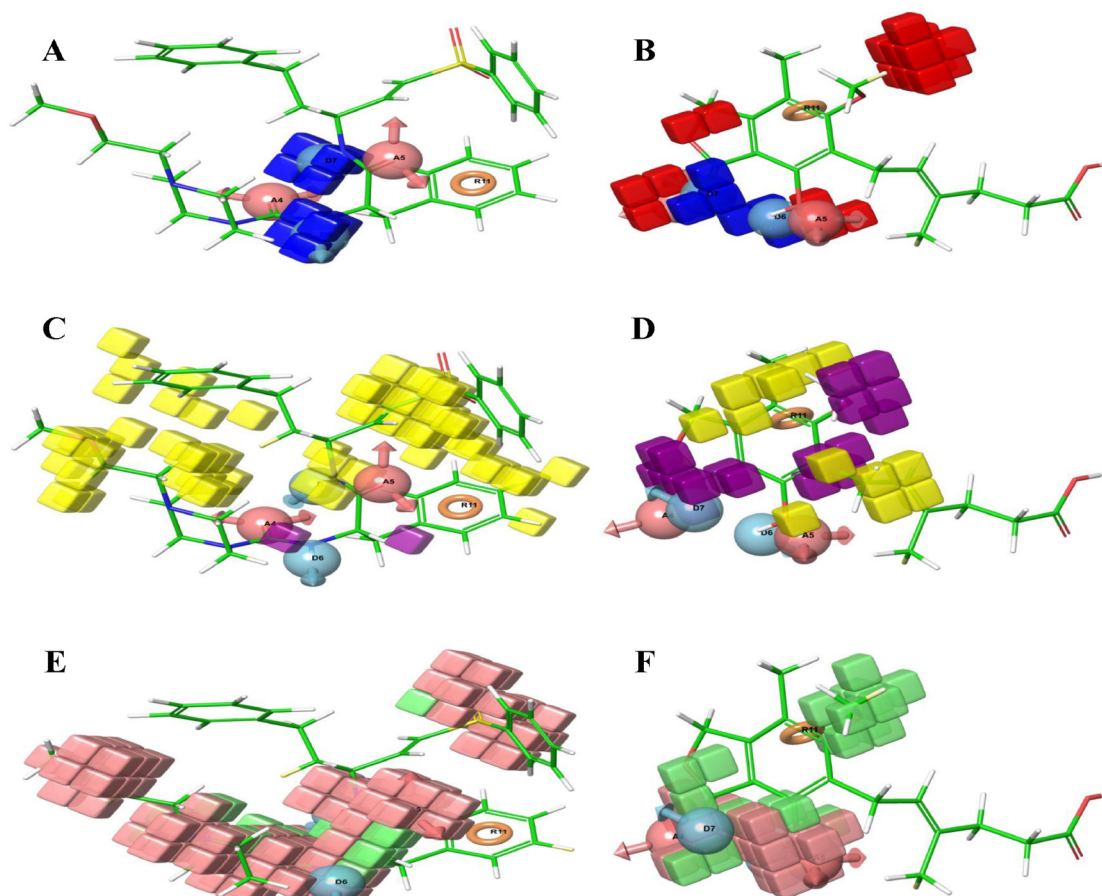


Figure 4. (A) H-bond donor effect, most active; (B) least active (blue, favorable; red, unfavorable); (C) hydrophobic effect, most active; (D) least active (green, favorable; yellow, unfavorable); (E) electron-withdrawing effect, most active; and (F) least active (red, favorable; blue, unfavorable)

rings indicates that the presence of terminal hydrophobic rings could increase the activity of hit compounds, and the unfavorable purple cubs around ring hydrogen moieties suggest that increasing the length of a carbon chain could lead to Figure 4E and F show contour maps of the electron-withdrawing effect with the most favorable region and unfavorable regions, pale pink cubs, and green cubs, respectively. The favored areas were located nearby hydrogen-bond acceptors and acceptor features of most active ligand, which showed that hydrogen-bond acceptor moieties were very much needed for bioactivity, and these groups should be considered in lead adjustment. As shown in Figure 4F the unfavorable area enclosed the ring moiety that exposed the compound's inactivity in the least active compound.

Pharmacophore models-based molecular modeling

In this study, virtual screening of 2000 natural compounds was carried out based on 3D-QSAR and its relative pharmacophore models to help find the lead compounds. By applying a fitness score of more than 1, the HTVS data filtered and yielded 1174 compounds. The fitness score is a value of how well a compound matches the pharmacophore hypothesis. Therefore, the hit compounds with a great fitness score are likely very active. The obtained ligands were docked to the receptors utilizing the default grid, and finally, 926 compounds with the docking score above than -6 kcal/mol, fitness score > 1.2, and the number of H-bonds > 2 were selected as hits. The HTVS filtered ligands were subjected to Glide SP (standard precision) docking program, and 654 ligands were chosen based upon their docking score > -7 kcal/mol. Eventually, a docking analysis of hits was carried out using the Glide XP simulation. The best 15 ligand molecules with their docking score, RMSD value, fitness score, and predicted pIC_{50} are listed in tables 7-9. In this study, we used the RMSD value to validate

docking results. In fact, the docking accuracy is assessed by the low value of RMSD (RMSD < 1.5 Å is wise to consider), and most hit ligands showed a RMSD value of less than 1 Å.

Neoeriocitrin, with the docking score, -11 kcal/mol showed a more binding affinity than other compounds against spike ectodomain structure (PBD: 6vyb) (Table 7). Docking results showed that this compound interacts by hydrogen bonding, pi-pi and pi-cation interactions with active pocket residues of 6VYB (Figure 5). The analysis of the binding affinity of five top-ranked compounds against 6VYB shows that the significant binding includes amino acid residues Asp 428, Asp 228, Lys 222, Lys 41, and His 519. Summarizing the between the five best ligands and the amino acid involved in the binding pocket of the crystal structure of the main protease (PBD: 6LU7) and five best ligands in the co-crystal structure, Gln127, Lys 5, Lys 102, Lys 137, Asp 289, and Glu290 were found to be the most frequently bound with the ligands (Table 8). Figure 6 depicts the interactions of Eriocitrin with the binding site of 6lu7. The 6lu7 binding pocket mainly consists of a hydrogen binding area formed by ASP A:228- LYS A:202- LYS A:41- HIS C:519- ASP A:40- PHE C:515-SRE C:514- ASP A:198.

Docking results of the five best compounds against the 3D structure of the receptor-binding domain (PBD: 6vw1) depict that Eriocitrin with docking score -10 has the best complements to the protein binding site (Table 9). Hydrogen bonds are the primary interaction between these compounds and active site residues of 6vw1, and amino acid residues Asn370, Thr 372, Trp436, and Val 367 play a significant role in docking interactions (Figure 7).

In this study, we used native ligand peptide-like inhibitor PRD_002214 (N3) as a positive control. Against 3D structure of spike ectodomain (PBD: 6vyb), this ligand showed docking score -4.86 kcal/mol, and established

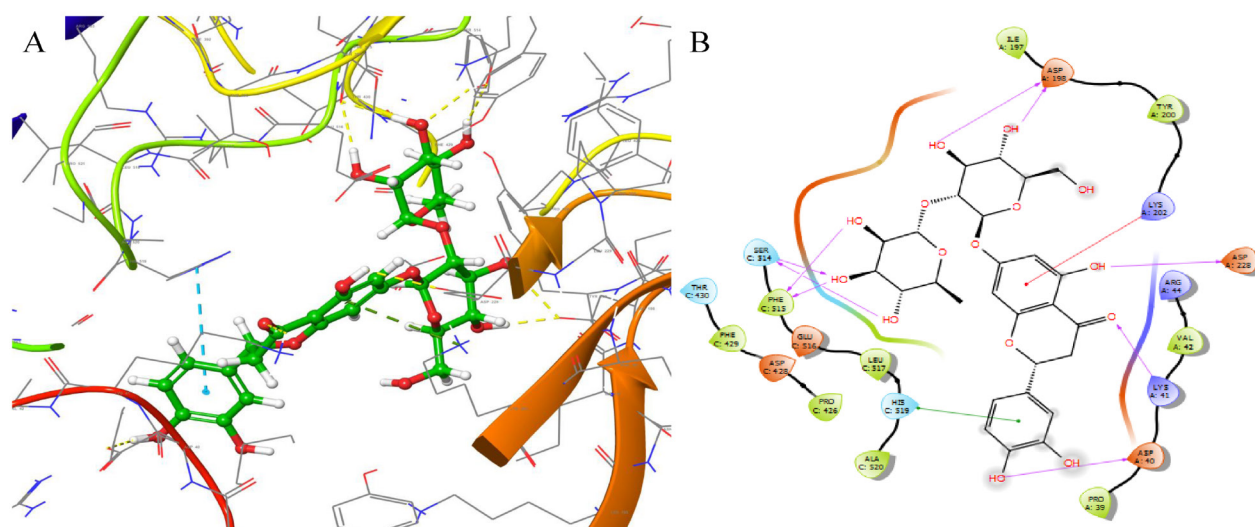


Figure 5. The binding pose of Spike ectodomain-Neoeriocitrin complex. A: 3D interaction, B: 2D interaction. The 2D interaction map indicates the important involved amino acid in the binding pocket.

Table 7. Glide docking energy, RMSD, fitness score, and predicted activity of top five compounds against the SARS-CoV-2 Spike ectodomain receptor.

Structure	Code	Name	Important AA	Fitness score	Dock score (kcal/mol)	RMSD (Å)	pIC ₅₀ (predicted)
	114627	Neoperiocitrin	ASP A:228- LYS A:202- LYS A:41- HIS C:519- ASP A:40- PHE C:515-SRE C:514- ASP A:198	1.58	-11.201	0.444	5.98
	128229	Glycyrrhizin	PRO C:579- LYS A:41- GLU C:516- LEU A:226- ASP A:228- ASN C:394- ARG C:357	1.38	-9.906	0.609	7.00
	5280805	Rutin	LYS A:41- THR C:430- ASP C:428- GLU C:516- LYS A:202- TYR A:200- SER C:514	1.22	-9.832	0.493	6.03
	442431	Narirutin	GLU C:516- HIS C:519- LYS A:41- LYS A:202- TYR A:200- ASP A:428- PHE C:515- THR C:430- ILE A:973	1.45	-9.777	0.019	6.14
	6475724	Licuroside	GLU C:516- HIS C:519- PRO A:39- LYS A:195- ASP A:53- ILE A:973- ARG A:983-	1.27	-9.646	0.461	5.74

Table 8. Glide docking energy, RMSD, fitness score, and predicted activity of top five compounds against the SARS-CoV-2 main protease receptor.

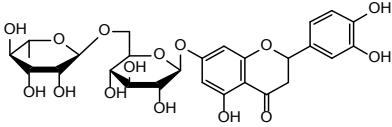
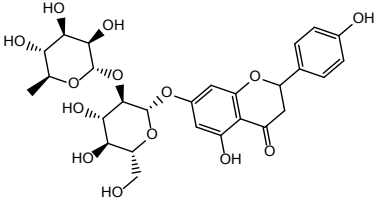
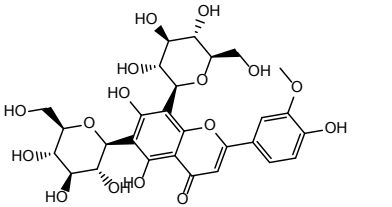
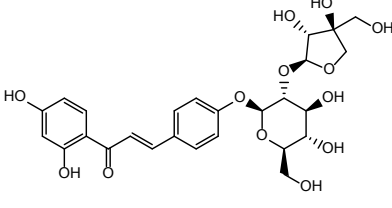
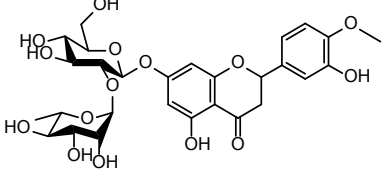
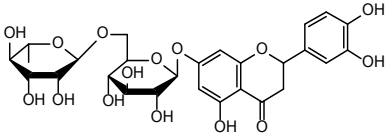
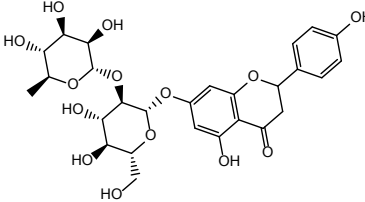
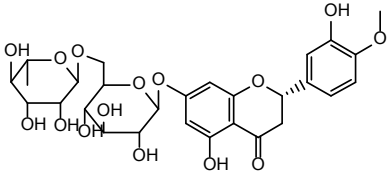
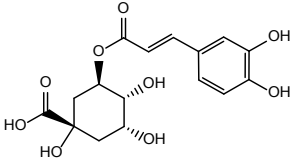
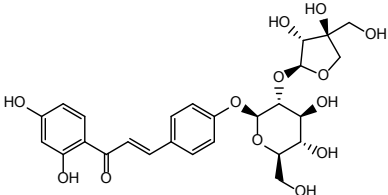
Structure	Code	Name	Important AA	Fitness score	Dock score (kcal/mol)	RMSD (Å)	pIC ₅₀ (predicted)
	3564542	Eriocitrin	GLN 127- ALA 7- LYS 5- GLU 228- ASP 289- LYS 137- GLU 290	1.57	-9.135	0.021	6.08
	74787988	Naringin	TYR 118- LYS 5- GLU 290- LYS 137- GLN 127-	1.71	-9.068	0.081	6.15
	44258167	Chrysoeriol 6,8-di- C-gluc (Stellarin-2)	LYS 102- ASP 153- SER 158- GLN 127- ARG 105	1.34	-8.903	0.534	5.08
	6442433	Isoliquiritin Apioside	ASP 153- ARG 298- THR 292- THR 111-	1.41	-8.526	0.310	5.38
	122173139	Neohesperidin	GLN 110- THR 111- SER 158- LYS 102- ASP 153- ASP 295- ARG 105-	1.40	-8.460	0.347	5.97

Table 9. Glide docking energy, RMSD, fitness score, and predicted activity of top five compounds against the SARS-CoV-2 receptor-binding domain.

Structure	Code	Name	Important AA	Fitness score	Dock score (kcal/mol)	RMSD (Å)	pIC ₅₀ (predicted)
	3564542	Eriocitrin	ASN 370- THR 372- NAG 601 -NAG 602- TRP 436- VAL 367	1.57	-10.614	0.068	6.08
	74787988	Naringin	ASN 370- ASN 440- ALA 344- ASN 343- PHE 342	1.71	-10.396	0.086	6.15
	10621	Hesperidin	THR 372- VAL 367- TRP 436- PHE 342- ASN 370	1.47	-9.928	0.181	5.56
	1794427	Chlorogenic Acid (5z-Caffeoylquinic Acid)	VAL 367- ASN 343- PHE 342- ARG 509- TRP 436	1.32	-9.700	1.901	5.89
	6442433	Isoliquiritin Apioside	ALA 344- PHE 342- ARG509- TRP436-SER 371- VAL 367	1.41	-9.155	0.242	5.38

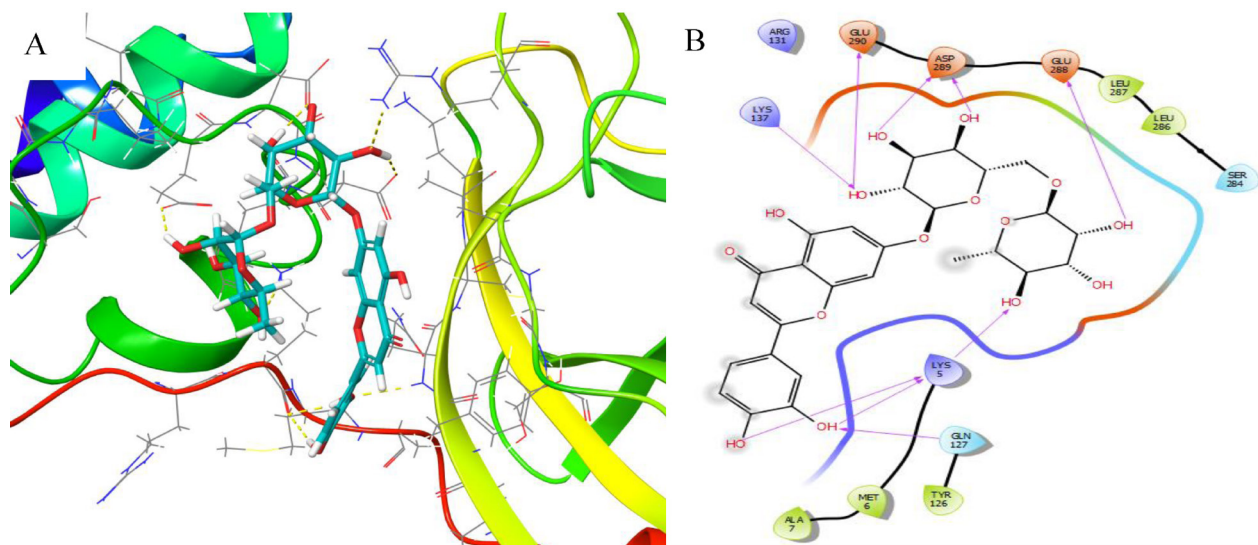


Figure 6. The binding pose of the main protease-Eriocitrin complex. A: 3D interaction, B: 2D interaction. The 2D interaction map indicates the essential involved amino acid in the binding pocket.

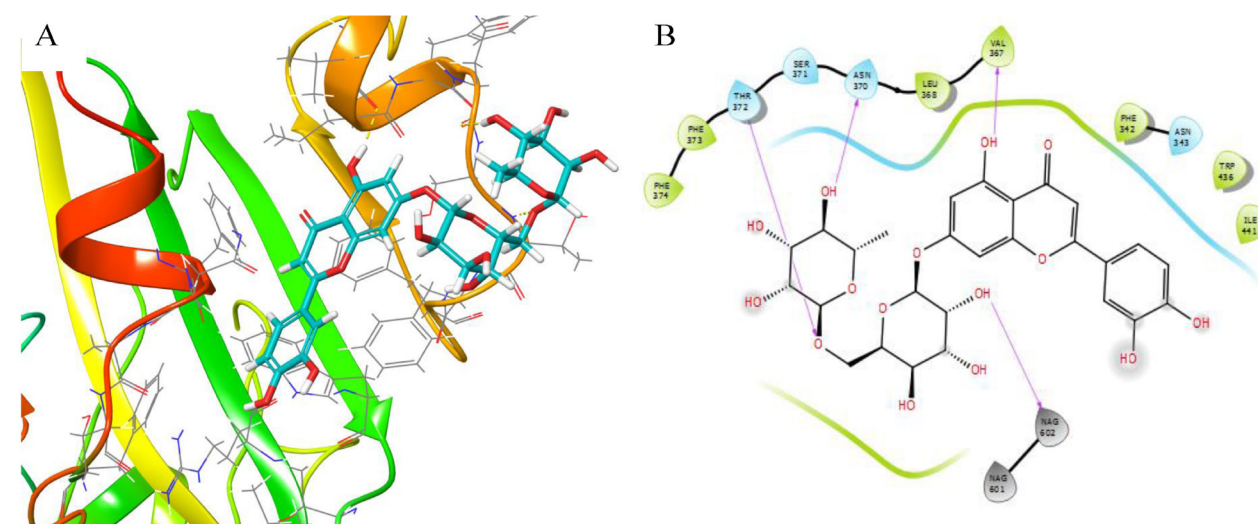


Figure 7. The binding pose of the receptor-binding domain - Eriocitrin complex. A: 3D interaction, B: 2D interaction. The 2D interaction map indicates the important involved amino acid in the binding pocket.

hydrogen binds, pi-pi and pi-cation interactions with amino acid residues Asp 428, Arg 983, Lys 202, Lys 41, and His519 through the active site of 6vyb. Against 3D structure of the main protease (PBD: 6LU7), this native ligand indicated docking affinity -4.76 kcal/mol through hydrogen interactions with amino acid residues Gln127, Lys 137, and Asp 289. Against 3D structure of the receptor-binding domain (PBD: 6vw1), N_3 depicted docking score -4.01 kcal/mol, and established hydrogen bond with amino acid residues Asn 370, Asn 343, and Trp436 through the active site of 6vw1 (Figure S1-S3). The comparison of N_3 against SARS-CoV-2 receptors and the interactions of hit ligands showed that our hit ligands interacted with the active pocket of SARS-CoV-2 receptors the same as N_3 did. The docking results of hit compounds were significantly correlated with the contour map analysis results and the number of hydrogen bond interactions between

ligands and the amino acid residues. Among these hits, Glycyrrhizin, narirutin, and naringin indicated superior predicted activity based upon the generated 3D-QSAR model. Figure S4-S6 shows the binding mode of these hit compounds.

ADME predictions

The pharmaceutical properties of the 15 lead compounds were evaluated to study drug resemblance and predict the drug's pharmacokinetics, involving ADME (absorption, distribution, metabolism, and excretion). We used the QikProp module of the Schrodinger software for assessing the drug-likeness of the 15 lead compounds by analyzing pharmacokinetic restrictions needed for ADME (Table 10). The factor partition coefficient (QPlogPo/w) is a crucial parameter to understand the compound's absorption and distribution, ranging from -2.770 to 2.221

Table 10. QikProp Properties of the lead compounds.

Name	MW ^a (<500)	DonorHB (<5)	Accept HB (<10)	QPlogPo/w ^b (-2 to 6.5)	Rule of Five (<4)	QPPCaco ^c (<25 poor, >500 great)	QplogBB ^d (-3.0-1.2)	QplogS ^e (-6.5-0.5)	PSA ^f	HOA ^g
Cholorogenic acid	354.31	6.00	9.6	-0.359	1	2.561	-3.002	-2.128	173.56	1
Chrysoeriol 6,8-di-C-gluc (Stellarin-2)	624.55	10.0	21.5	-2.517	3	12.405	-3.172	-1.902	215.71	1
Eriocitrin	594.52	8.0	19.8	-2.194	3	38.329	-4.165	-4.664	256.66	1
glycyrrhizin	822.94	6.0	21.3	2.221	3	0.146	-3.087	-3.616	255.04	1
Hesperidin	610.56	7.0	20.0	-1.585	3	16.530	-2.661	-1.348	210.39	1
Isoliquiritin apioside	550.51	7.0	17.6	-1.112	3	10.950	-3.715	-1.878	200.50	1
Licuroside	550.51	7.0	17.6	-1.065	3	7.683	-4.072	-2.291	210.27	1
Naringin	580.54	7.0	19.3	-1.608	3	3.838	-4.066	-2.943	225.68	1
Narirutin	580.54	7.0	19.3	-1.621	3	5.742	-3.479	-2.261	223.18	1
Neoeriocitrin	596.54	8.0	20.0	-2.000	3	4.378	-4.046	-2.615	233.22	1
Neohesperidin	610.56	7.0	20.0	-1.329	3	11.979	-3.638	-2.913	219.60	1
Rutin(Quercetin 3-O-rut)	610.52	9.0	20.5	-2.770	3	1.458	-3.776	-1.429	242.68	1

^a MW – molecular weight; ^b logPo/w – predicted octanol/water partition coefficient; ^c predicted apparent Caco-2 cell permeability in nm/s; ^d logBB – predicted brain/blood partition coefficient; ^e logS – predicted aqueous solubility; ^f PSA – polar surface area; ^g HOA – predicted qualitative human oral absorption: 1, 2, or 3 for low, medium, or high

for the lead compounds. Most of the lead compounds indicated an excellent partition coefficient (QPlogPo/w). The factors like MW, H-bond donors, acceptors, blood-brain barrier, permeability (QPPCaco), predicted aqueous solubility (QplogS), and human oral absorption was also calculated for drug-like behavior of the lead compounds.

Conclusion

Three crystal structures of SARS-CoV-2 and thirty-one antiviral drugs with high inhibitory potential against SARS and MERS were utilized for pharmacophore analysis integrated with HTVS as well as docking studies to retrieve natural lead compounds. This study aimed to generate the pharmacophore hypothesis and 3D-QSAR models and utilize

the models on virtual screening to investigate and identify new scaffolds, and predict the approximate hit molecules activity based on the best 3D-QSAR model. In this work, 3D pharmacophore models were developed from 31 antiviral compounds and the three best pharmacophore models were performed as 3D queries for virtual screening. Then, the hit ligands were submitted to molecular docking studies and RMSD calculations. Finally, 15 compounds were chosen as hit compounds based on the scoring function. These hit compounds were evaluated for ADME properties. From the overall analyses, it could be concluded that the best three pharmacophore models reflect the characteristics of SARS-CoV-2 inhibitors and these pharmacophore models could be utilized as a fast model to help in the identification and discovery of SARS-CoV-2 inhibitors. Also, both the 3D-QSAR

visualization and docking result of hit ligands against the target receptors presented considerable information about the relationship between structure and activity.

Acknowledgments

The research council of Shahid Beheshti University for financial support of this project much acknowledge.

Author Contributions

SN and MF did the molecular modeling part and preparing the draft of the manuscript. SNE supervised the project and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The authors did not have any conflict of interest.

Supplementary Data

The binding pose of hit compounds and IC₅₀ values (Table S1 and Figure S1-S6) are available on the journal's web site along with the published article.

References

- Macchiagodena M, Pagliai M, Procacci P. Inhibition of the main protease 3cl-pro of the coronavirus disease 19 via structure-based ligand design and molecular modeling. ArXiv Preprint ArXiv. 2020;arXiv:2002.09937.
- Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the outbreak of coronavirus disease 2019 (covid-19) in Wuhan, China with individual reaction and governmental action. Int J Infect Dis. 2020;93:211-6. doi:10.1016/j.ijid.2020.02.058
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in china, 2019. New N Engl J Med. 20;382(8):727-733. doi:10.1056/NEJMoa2001017
- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3. doi:10.1038/s41586-020-2012-7
- World Health Organization. Coronavirus disease 2019 (covid-19): Situation report, 72. 2020. <https://apps.who.int/iris/handle/10665/331685>
- Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A novel coronavirus emerging in china—key questions for impact assessment. N Engl J Med. 2020;382(8):692-4. doi:10.1056/NEJMp2000929
- Hosseini FS, Amanlou M. Simeprevir, Potential candidate to repurpose for coronavirus infection: Virtual screening and molecular docking study. Preprints. 2020. doi:10.20944/preprints202002.0438.v1
- Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT. Programmed ribosomal frameshifting in decoding the sars-cov genome. Virology. 2005;332(2):498-510. doi:10.1016/j.virol.2004.11.038
- Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the nidovirales. J Gen Virol. 2000;81(Pt 4):853-79. doi:10.1099/0022-1317-81-4-853
- Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. Molecular model of SARS coronavirus polymerase: Implications for biochemical functions and drug design. Nucleic Acids Res. 2003;31(24):7117-30. doi:10.1093/nar/gkg916
- Pillaiyar T, Meenakshisundaram S, Manickam M. Recent discovery and development of inhibitors targeting coronaviruses. Drug Discov Today. 2020;25(4):668-88. doi:10.1016/j.drudis.2020.01.015
- Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020;30(3):269-71. doi:10.1038/s41422-020-0282-0
- Lim J, Jeon S, Shin H-Y, Kim MJ, Seong YM, Lee WJ, et al. Case of the index patient who caused tertiary transmission of COVID-19 infection in Korea: The application of lopinavir/ritonavir for the treatment of COVID-19 infected pneumonia monitored by quantitative RT-PCR. J Korean Med Sci. 2020;35(6):e79. doi:10.3346/jkms.2020.35.e79
- Lu H. Drug treatment options for the 2019-new coronavirus (2019-nCoV). Biosci Trends. 2020;14(1):69-71. doi: 10.5582/bst.2020.01020
- Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 2016;12(1):2694-718. doi: 10.3762/bjoc.12.267
- Kuntz ID. Structure-based strategies for drug design and discovery. Science. 1992;257(5073):1078-82. doi: 10.1126/science.257.5073.1078
- Macalino SJY, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. Arch Pharm Res. 2015;38(9):1686-701. doi:10.1007/s12272-015-0640-5
- Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997;267(3):727-48. doi:10.1006/jmbi.1996.0897
- Irwin JJ, Shoichet BK. Docking screens for novel ligands conferring new biology: Miniperspective. J Med Chem. 2016;59(9):4103-20. doi:10.1021/acs.jmedchem.5b02008
- Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Brief Bioinform. 2016;17(1):2-12. doi:10.1093/bib/bbv020
- Barnard DL, Hubbard VD, Burton J, Smee DF, Morrey JD, Otto MJ, et al. Inhibition of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) by calpain inhibitors and β -D-N⁴-hydroxycytidine. Antivir Chem Chemother. 2004;15(1):15-22. doi:10.1177/095632020401500102
- Ghosh AK, Takayama J, Rao KV, Ratia K, Chaudhuri R,

- Mulhearn DC, et al. Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: Design, synthesis, protein–ligand x-ray structure and biological evaluation. *J Med Chem.* 2010;53(13):4968-79. doi:10.1021/jm1004489
23. Hsieh L-E, Lin C-N, Su B-L, Jan T-R, Chen C-M, Wang C-H, et al. Synergistic antiviral effect of *Galanthus nivalis* agglutinin and nelfinavir against feline coronavirus. *Antiviral Res.* 2010;88(1):25-30. doi:10.1016/j.antiviral.2010.06.010
24. Keyaerts E, Vijgen L, Maes P, Neyts J, Van Ranst M. In vitro inhibition of severe acute respiratory syndrome coronavirus by chloroquine. *Biochem Biophys Res Commun.* 2004;323(1):264-8. doi:10.1016/j.bbrc.2004.08.085
25. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des.* 2006;20(10-11):647-71. doi:10.1007/s10822-006-9087-6
26. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem.* 2004;47(7):1750-9. doi:10.1021/jm030644s
27. Truchon J-F, Bayly CI. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. *J Chem Inf Model.* 2007;47(2):488-508. doi:10.1021/ci600426e
28. Sakkiah S, Meganathan C, Sohn Y-S, Namadevan S, Lee KW. Identification of important chemical features of 11 β -hydroxysteroid dehydrogenase type1 inhibitors: Application of ligand based virtual screening and density functional theory. *Int J Mol Sci.* 2012;13(4):5138-62. doi:10.3390/ijms13045138
29. Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q2 parameter for QSAR validation. *J Chem Inf Model.* 2009;49(7):1669-78. doi:10.1021/ci900115y
30. Pulla VK, Sriram DS, Viswanadha S, Sriram D, Yogeewari P. Energy-based pharmacophore and three-dimensional quantitative structure-activity relationship (3D-QSAR) modeling combined with virtual screening to identify novel small-molecule inhibitors of silent mating-type information regulation 2 homologue 1 (SIRT1). *J Chem Inf Model.* 2016;56(1):173-87. doi:10.1021/acs.jcim.5b00220
31. Pratim Roy P, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. *Molecules.* 2009;14(5):1660-701. doi:10.3390/molecules14051660
32. Kirubakaran P, Karthikeyan M. Pharmacophore modeling, 3D-QSAR and DFT studies of IWR small-molecule inhibitors of Wnt response. *J Recept Signal Transduct Res.* 2013;33(5):276-85. doi:10.3109/10799893.2013.822888
33. Golbraikh A, Tropsha A. Beware of q2! *J Mol Graph Model.* 2002;20(4):269-76. doi:10.1016/s1093-3263(01)00123-1
34. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods.* 2000;44(1):235-49. doi:10.1016/s1056-8719(00)00107-6