



## VIRTUAL SCREENING FOR SARS-COV-2 ENTRY INHIBITORS BY DUAL TARGETING OF TMPRSS2 AND CTSL

Ian Lemuel Sigue Virtucio<sup>1</sup>, Jervée Malabanan Punzalan<sup>1</sup>, Junie Barotil Billones<sup>1\*</sup>

1. Department of Physical Sciences and Mathematics, College of Arts and Sciences, University of the Philippines Manila, Ermita, Manila, Philippines.

### ARTICLE INFO

#### Received:

05 Nov 2022

#### Received in revised form:

17 Jan 2023

#### Accepted:

24 Jan 2023

#### Available online:

28 Feb 2023

**Keywords:** SARS-CoV-2, TMPRSS2, CTSL, Virtual screening, Molecular docking

### ABSTRACT

The COVID-19 pandemic remains to be a global public health crisis due to the emergence of new variants of concern and the scarcity of drug treatments. The cell entry of SARS-CoV-2 requires activation of its spike protein by host proteases TMPRSS2 and CTSL, which triggers membrane fusion and facilitates the endocytic uptake mechanism, respectively. This study employed a structure-based virtual screening technique to identify drugs and natural products that simultaneously target TMPRSS2 and CTSL. Two pharmacophore models were generated from the binding sites of the proteins in complex with their co-crystallized ligands. Both structure-based pharmacophores were used to screen a ligand library composed of 41,775 compounds (10,849 drugs from the ChEMBL database and 30,926 natural products from the NPASS database). A total of 115 compounds (54 drugs and 61 natural products) that fit both TMPRSS2 and CTSL pharmacophore models were identified. The common hits were docked into both proteases to obtain a short list of compounds. Molecular docking filtered 17 compounds (5 drugs and 12 natural products) that have higher binding energy values than the co-crystallized ligands and known inhibitors of both proteins. The top hits were then subjected to ADMET, drug-likeness, and synthetic accessibility filters. Based on docking scores, pharmacokinetics, and drug-likeness, Silibinin was the most promising repurposed drug candidate as a treatment for SARS-CoV-2 infection via dual inhibition of TMPRSS2 and CTSL. Among the natural products, barettin was the best candidate for further development as a novel dual TMPRSS2 and CTSL inhibitor.

This is an **open-access** article distributed under the terms of the [Creative Commons Attribution-Non Commercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows others to remix, tweak, and build upon the work non commercially, as long as the author is credited and the new creations are licensed under the identical terms.

**To Cite This Article:** Virtucio ILS, Punzalan JM, Billones JB. Virtual Screening for SARS-COV-2 Entry Inhibitors by Dual Targeting of TMPRSS2 and CTSL. *Pharmacophore*. 2023;14(1):9-18. <https://doi.org/10.51847/6IMWqjwVPa>

### Introduction

Coronaviruses (CoVs) are known for disease outbreaks that received global attention due to their capability to cause mortality in humans. These include the 2002-2004 severe acute respiratory syndrome (SARS) and the 2012 Middle East respiratory syndrome (MERS) outbreaks. Last March 2020, the World Health Organization (WHO) declared that the world is facing a pandemic of coronavirus disease 2019 (COVID-19) due to the rapid spread of SARS-CoV-2. As of October 2022, over 617 million cases of COVID-19 have been recorded worldwide with more than 6 million confirmed deaths [1]. Along with mass vaccination and convalescent plasma transfusion, many hopes have been pinned on drug repurposing in the pursuit of developing safe and effective COVID-19 treatments. This line of scientific research has brought newfound interest into many approved and experimental drugs such as remdesivir, nirmatrelvir/ritonavir, molnupiravir, baricitinib, [2] and ivermectin [3] as potential therapeutic options for COVID-19-afflicted individuals. While a large number of drug candidates are under study, the search for effective therapeutic agents against SARS-CoV-2 continues to be a global necessity.

CoVs are a large family of enveloped, single-stranded positive-sense RNA viruses that cause respiratory tract infections in mammals and birds. A distinguishing feature of CoVs is the spike (S) protein on the surface of the viral envelope which plays a vital role in receptor recognition and entry of the viral genome to the host cell [4]. For SARS-CoV-1 and SARS-CoV-2, the S protein binds to the angiotensin-converting enzyme 2 (ACE2) receptor present in type II pneumocyte cells. The S protein consists of the S1 and S2 subunits — the S1 subunit contains the receptor-binding domain (RBD) which facilitates ACE2 recognition and binding, while the S2 subunit mediates viral fusion and entry [5]. After binding to the ACE2 receptor, there are two pathways for SARS-CoV-2 entry into the host cells, either by fusion of the viral membrane with the host plasma

**Corresponding Author:** Junie Barotil Billones; Department of Physical Sciences and Mathematics, College of Arts and Sciences, University of the Philippines Manila, Ermita, Manila, Philippines. E-mail: [jbillones@up.edu.ph](mailto:jbillones@up.edu.ph)

membrane or via endosomes by receptor-mediated endocytosis. It has been well-demonstrated that the virus hijacks host proteases to cleave and activate its S protein, such as transmembrane protease serine 2 (TMPRSS2) for membrane fusion, and cathepsin L (CTSL) for endosomal entry mechanism [6]. The simultaneous inhibition of both proteases to block both surface fusion and endosomal entry is considered to be a robust mechanism for prophylactic agents or early-stage treatments for SARS-CoV-2 infection.

The biological function of TMPRSS2 remains unknown but it is found to be overexpressed in prostate cancer and is also responsible for priming SARS-CoV-1 and MERS-CoV S proteins [7]. On the other hand, CTSL is a cysteine protease involved in cellular homeostasis with implications for many types of human cancer. It is also overexpressed during COVID-19 infection and is positively correlated with disease course and severity [8]. As TMPRSS2 and CTSL exert cell-type dependent viral entry mechanisms, large databases may be screened to discover potential drugs that work against both host proteases. A drug acting on two druggable enzymes rather than one is intuitively more potent and less prone to drug resistance. Additionally, therapeutics targeting host rather than viral proteins should be less susceptible to drug resistance arising from viral mutations [9].

In the modern drug discovery process, computational techniques including pharmacophore modeling and molecular docking are often employed as complementary screening tools for fast and efficient processing of finding drug leads. A pharmacophore is a group of geometrically-mapped chemical features that are required for interactions to occur between a receptor target and its partner molecule [10]. However, because pharmacophore-based screening only weighs the alignment between the ligand and pharmacophore features that do not necessarily reflect affinity, potential leads are often filtered further by molecular docking. Docking is a method that predicts the binding pose and affinity of a ligand to the receptor to form a stable complex. These two methods for virtual screening have become a popular starting approach for *in silico* drug discovery because they enable the prioritization and cost-effective classification of potential leads that inhibit drug targets, and ultimately complement high-throughput screening [11]. We have demonstrated the pharmacophore-based screening and molecular docking techniques in discovering potential leads against crucial *M. tuberculosis* targets [12-16].

This study aims to discover existing drugs and natural products that simultaneously target TMPRSS2 and CTSL. Two pharmacophore models were generated from the crystal structures of both proteases complexed with their co-crystallized ligands. Both structure-based pharmacophores were individually used to screen drugs included in the ChEMBL database and natural products in the NPASS database, then the common hits were docked into both proteases to obtain shortlisted candidates. The top hits were subjected to ADMET, drug-likeness, and synthetic accessibility filters to obtain potentially repurposable drugs and naturally occurring leads that act against both proteases.

## Materials and Methods

All computational screening methods were performed on Windows 11 operating system in a machine with an Intel® Core™ i5-10300H CPU @ 2.50 GHz (8 CPUs) processor, 8 GB RAM, and NVIDIA GeForce GTX 1650 Ti graphics card.

### Generation of Pharmacophore Models

The crystal structures of human TMPRSS2 with cleaved nafamostat or ligand GBS (PDB ID: 7MEQ; resolution 1.95 Å) and human CTSL with dipeptidyl nitrile inhibitor NOW (PDB ID: 3HHA; resolution 1.27 Å) were used to generate structure-based pharmacophore models. For 7MEQ, the ligand GBS was returned to its unhydrolyzed form as nafamostat complexed with TMPRSS2, while 3HHA was uploaded using the PDB search function in LigandScout 4.4 software [17]. The binding site for each receptor was selected by clicking the co-crystallized ligand automatically detected by the software, and the receptor-ligand complexes were optimized using the MMFF94 force field. The software then created the pharmacophore models based on the chemical features of the active sites bound by the co-crystallized ligands.

### Pharmacophore Model Validation

The pharmacophore models were validated for their capability to discriminate active compounds from decoys. Experimentally validated active compounds against TMPRSS2 and CTSL were identified based on literature and submitted in SMILES notation to the DUDE•E directory (<http://dude.docking.org/>). The active and decoy ligands were separately converted into a database file format (ldb) in LigandScout 4.4 software. The pharmacophore models were then validated by simultaneously screening both active and decoy libraries. The receiver operating characteristic (ROC) curves were generated to analyze the area under the curve (AUC) and enrichment factor (EF) for evaluating the accuracy of the pharmacophores.

### Database Building

A total of 41,775 structures, composed of 10,849 small-molecule drugs from ChEMBL (<https://www.ebi.ac.uk/chembl/>) and 30,926 natural products from NPASS (<http://bidd.group/NPASS/>), were obtained. The ligand structures were prepared and optimized in OpenBabel 2.3.1 [18] by conversion into 3D, the addition of polar hydrogens, assignment of MMFF94 partial charges, and removal of duplicates. After processing, the library contained a total of 40,039 unique structures (9,945 drugs and 30,094 natural products). The ligands were then uploaded into the LigandScout 4.4 platform for MMFF94 minimization and compilation into a library and converted into a database file format (ldb) while generating a maximum of 25 conformations for each structure for subsequent screening.

### Virtual Screening of Compound Libraries

The prepared library was screened against the two pharmacophore models at optimal parameters as determined by the validation step. For both models, the parameters *pharmacophore-fit* was chosen as the scoring function, *match all query features* as the screening mode, and *first matching confirmation* as the retrieval mode, where a maximum of one feature were omitted, and with exclusion volumes. Initially, the ligand library was screened against the TMRSS2 pharmacophore model, then rescreened against the CTSL pharmacophore model. The hits were sorted according to the average rank of their pharmacophore fit scores for each model, then exported to xls format for analysis. The overlapping hits in both models were manually extracted from the ligand libraries and then saved in mol2 file format.

### Molecular Docking

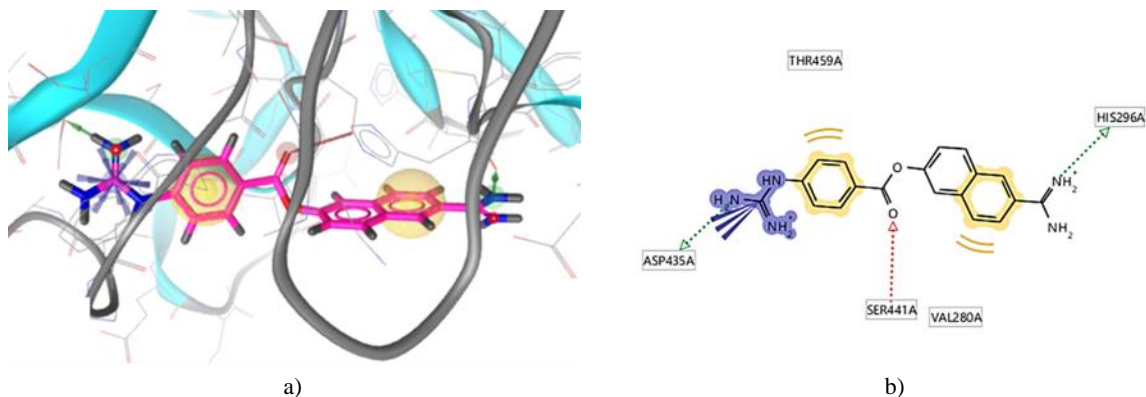
The structures of TMRSS2 (7MEQ) and CTSL (3HHA) were first prepared by insertion of missing residues via homology modeling using SWISS-MODEL and optimization in UCSF Chimera 1.17 via Dock Prep tool which removed solvent molecules and heteroatoms, inserted polar hydrogens, and assigned Gasteiger charges that were computed using Amber's Antechamber module. The protein structures were further minimized under the MMFF94 force field in OpenBabel 2.3.1. Before docking the common hits from virtual screening, the AutoDock Vina protocol [19] was validated by extracting the co-crystallized ligands GBS and NOW from the crystal structures, then re-docking them, after optimization, into the binding pockets of TMRSS2 and CTSL, respectively. For TMRSS2, the grid box was defined with a search space size of  $25 \text{ \AA} \times 25 \text{ \AA} \times 25 \text{ \AA}$  covering all essential residues centered at the coordinates (-8.55, -2.05, 16.07). For CTSL, the grid box was defined with a search space size of  $25 \text{ \AA} \times 25 \text{ \AA} \times 25 \text{ \AA}$  covering all essential residues centered at the coordinates (8.65, 8.06, -2.91). After confirming interaction profile reproducibility, the virtual screening hits were added in the UCSF Chimera 1.17 platform as mol2 files and optimized similarly. The ligands were docked into the active sites of TMRSS2 and CTSL under the validated protocol, and the binding energy for the best docking pose of each compound was recorded. Experimentally validated inhibitors of each receptor, camostat on TMRSS2 and CLIK-148 on CTSL, were also docked as controls in addition to the co-crystallized ligands. The receptor-ligand complexes were saved as single files in pdb format, then the 2D and 3D interaction diagrams were viewed in Discovery Studio Client 2021.

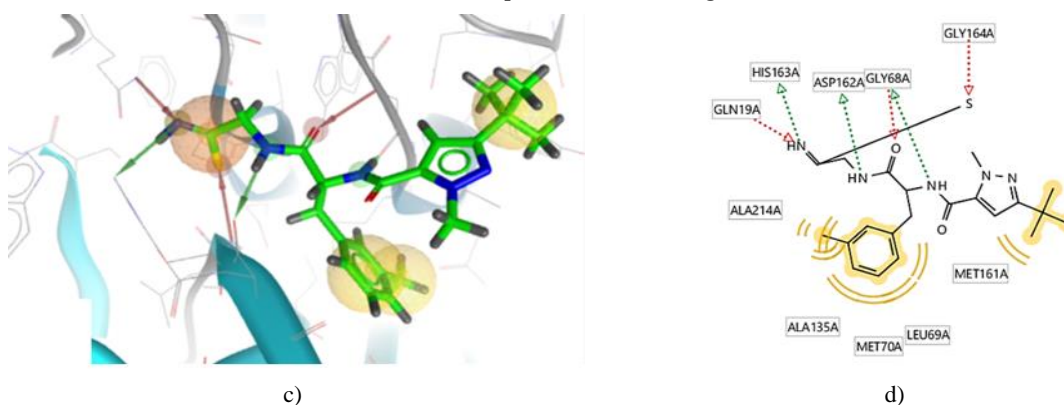
### ADMET Predictions

Computational prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of compounds was carried out using SwissADME (<http://www.swissadme.ch/>) and OSIRIS Property Explorer (<https://www.organic-chemistry.org/prog/peo/>). Parameters include water solubility, gastrointestinal (GI) absorption, blood-brain barrier (BBB) penetration, permeability glycoprotein (P-gp) binding, cytochrome P450 2D6 (CYP2D6) inhibition, and risks for mutagenic, tumorigenic, irritant, and reproductive effects. Drug- and lead-likeness were evaluated using the Lipinski rule and synthetic accessibility score (SAscore) in SwissADME.

### Results and Discussion

To identify novel scaffolds for compounds acting against both TMRSS2 and CTSL, two different pharmacophore models were generated using the available crystal structures of TMRSS2 and CTSL in complex with their respective ligands (**Figure 1**). Exclusion volume coats generated during the pharmacophore modeling process were also considered to maintain the steric characteristics and shape of the binding sites of both receptors. The TMRSS2 pharmacophore consisted of two hydrogen bond donors (HBD), one hydrogen bond acceptor (HBA), two hydrophobic interactions, one positive ionizable area (PI), and 18 exclusion volume coats. Meanwhile, the CTSL pharmacophore was composed of three HBD, three HBA, three hydrophobic interactions, one residue bonding point, and 20 exclusion volume coats.



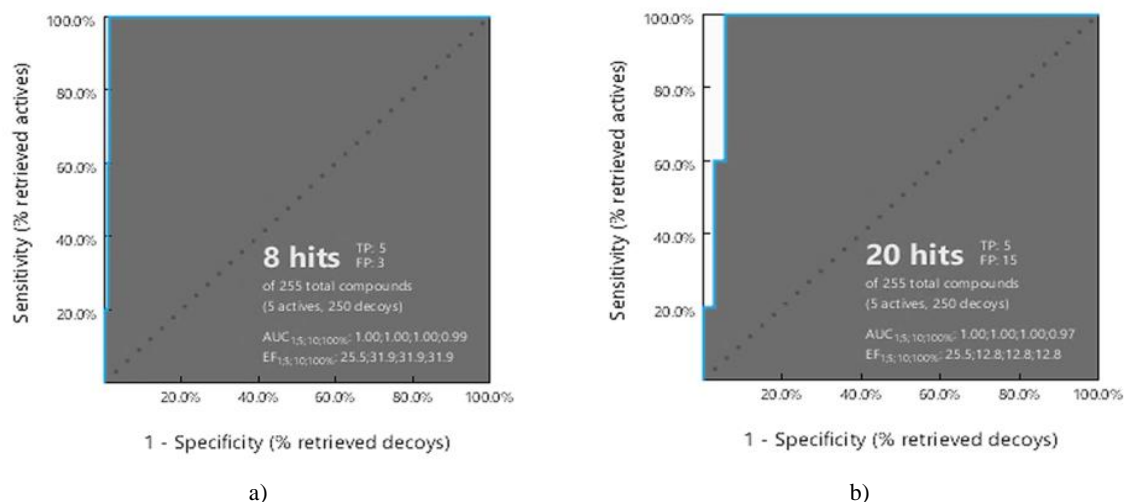


**Figure 1.** 3D and 2D pharmacophore models of a-b) TMPRSS2 and c-d) CTSL. The HBD features are shown in green vectors, HBA in red vectors, hydrophobic interactions in yellow spheres, PI area in blue star, residue bonding point in the orange sphere, and exclusion volume coats in grey spheres.

Validation of a pharmacophore model is necessary before screening large databases because it enables one to evaluate the quality and reliability of the model, especially after each selection of features and adjustment of tolerances. The structure-based pharmacophore models of TMPRSS2 and CTSL were validated for their capability to distinguish active compounds from decoy sets. For each target receptor, the test set consisted of five experimentally active compounds collected from the literature and 250 decoys generated by the DUDE•E directory corresponding to the active compounds. Decoys are based on similar physical properties (molecular weight,  $\log P$ , number of rotatable bonds, HBD, HBA, etc.) but different topologies with the active compounds to minimize the likelihood of actual binding [20].

To determine the performance of the pharmacophore models, the receiver-operating characteristic (ROC) curve generated during the virtual screening process of the test sets was analyzed. The ROC curve provides information regarding the accuracy of the pharmacophores by plotting the true positive percentage (sensitivity) against the false positive percentage (specificity). Classifiers that give curves closer to the top-left corner indicate better performance, while an unfavorable test gives a curve along the diagonal [21]. The overall summary of the model accuracy can be calculated from the Area Under the Curve (AUC) and Enrichment Factor (EF). The AUC represents the degree of discrimination ability and ranges between 0 and 1, wherein higher AUC values or values closer to 1 indicate better predictability of the model. The EF provides an idea about the number of active compounds found from a specific model compared to hypothetically active compounds found from a randomly screened model. The EF factor can range from 1 to >100, where 1 indicates the number of randomly sorted molecules and >100 indicates the least number of compounds needed to screen in vitro to find a large number of active compounds [22].

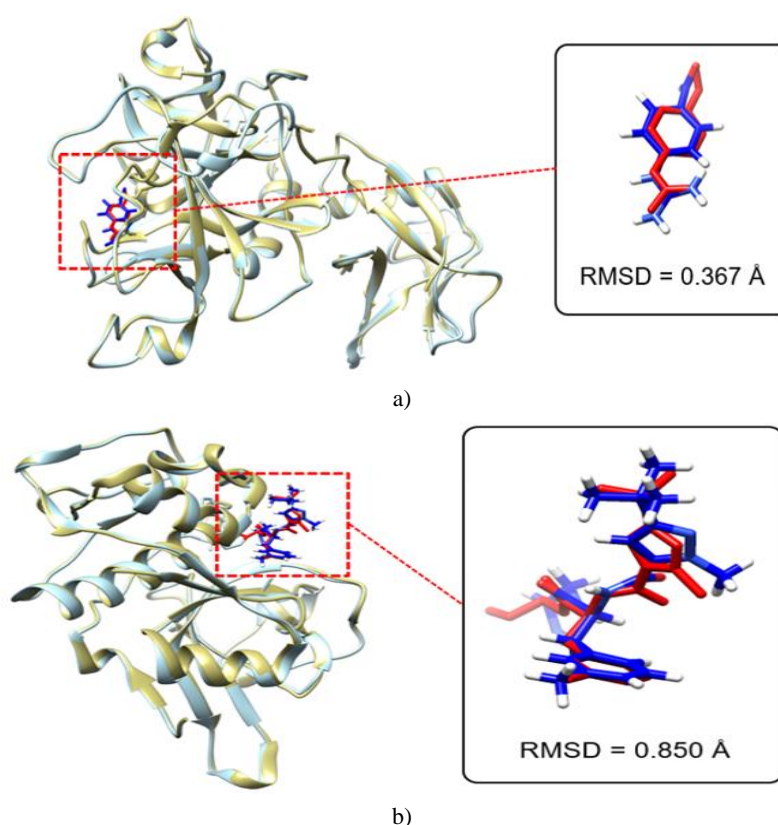
Both models were able to distinguish all five active molecules among the test set. Among 250 decoys, the TMPRSS2 model had three false positives, while the CTSL model had 15 false positives. The AUC and EF values at the 1% threshold found in both pharmacophore models were 1.00 and 25.5, respectively, indicating excellent discrimination ability and robustness of the models as shown in **Figure 2**. The enrichment factor of 25.5 essentially suggests that around 25 times more active compounds would be expectedly found in the top 1% of screening results as compared to their concentration throughout an entire database.



**Figure 2.** ROC curves (blue) were generated based on virtual screening of active and decoy set with pharmacophore models of a) TMPRSS2 and b) CTSL. Both models were validated using a test set of five active and 250 decoy compounds.

Pharmacophore-based virtual screening of the ligand library retrieved 115 compounds (54 drugs and 61 natural products) that overlapped in both TMRSS2 and CTSL pharmacophore screens. The resulting fit scores of overlapping hits ranged from 42.08 to 47.08 for TMRSS2 pharmacophore, and from 71.31 to 72.04 for CTSL pharmacophore. Since the fit scores weigh how well the ligand substructures match the query features of the structure-based pharmacophore models during the screening process, higher fit scores suggest better fitting of compounds against both target proteins.

Molecular docking predicts the binding pose and affinity of ligands with the receptor binding pocket to modulate its activity. Before docking the overlapping hits from pharmacophore-based virtual screening, the performance of the docking protocol was validated by the pose selection method. The bound ligands in TMRSS2 and CTSL crystal structures were re-docked to their respective active sites, and AutoDock Vina returned the poses of the ligands which were aligned with their co-crystallized conformation. Typically, the RMSD is used as the quantitative measure of the structural similarity between two superimposed structures, which is small if the structures are more similar to each other. An RMSD of 1.5 Å is usually set as the cut-off for docking accuracy [23]. **Figure 3** shows that the re-docked ligands closely resemble their co-crystallized conformation in the reference structures, with RMSD values of 0.367 Å for ligand GBS on TMRSS2, and 0.850 Å for ligand NOW on CTSL. The RMSD values were lower than 1.5 Å which indicates high structural similarity and thus high binding mode reproducibility for both protein targets.



**Figure 3.** Superimposed molecular pose of the reference protein structure (cyan) with co-crystallized ligand (red) against the prepared protein structure (gold) with re-docked ligand (blue) for a) TMRSS2 and b) CTSL. The structural similarity of the aligned ligands was calculated in terms of RMSD values.

The re-docked ligand GBS retained important interactions with TMRSS2, namely hydrogen bonding with substrate-binding residue ASP435 and catalytic residue SER441 (**Figure 1**). Since ligand GBS is only the phenylguanidino acyl moiety of nafamostat after it slowly hydrolyzed upon interacting with the enzyme, the remaining two important residues in the catalytic triad, HIS296, and ASP345, were not present in the resulting complex after re-docking. The ligand interactions of the re-docked ligand GBS also retained hydrogen bonding with reported substrate-binding residue GLY462, as well as with other residues like SER436, GLY439, and GLY464 of TMRSS2.

On the other hand, the re-docked ligand NOW also retained hydrogen bonding interactions with the catalytic dyad CYS25 and HIS163, as described for the experimental structure. Other interactions with substrate-binding and catalytic residues forming a deep hydrophobic pocket in the binding site were also retained. These features include hydrogen bonding and attractive charge with ASP162, van der Waals interactions with MET161, pi-alkyl bonding with ALA135, ALA214, and LEU69, pi-sulfur bonding with MET70, and hydrogen bonding with GLY68.

After optimizing the grid box parameters within the docking site and validating the protein preparation process and AutoDock Vina protocol, the 115 common hits from pharmacophore-based virtual screening hits were individually docked into the active sites of TMRSS2 and CTSL. Known inhibitors of each receptor, camostat on TMRSS2 and CLIK-148 on CTSL, were also



docked as controls in addition to the co-crystallized ligands. Ligands with more negative binding energies versus all controls were identified as hits since higher magnitudes of binding energy denote more stable receptor-ligand complexes while the negative sign is indicative of the exergonic process of complex formation [24]. The controls for TMPRSS2, namely the re-docked ligand GBS and known inhibitor camostat, yielded binding energy values of -7.6 and -7.5 kcal/mol, respectively. Meanwhile, the controls for CTSL, specifically the re-docked ligand NOW and known selective inhibitor CLIK-148, yielded binding energy values of -7.0 and -6.7 kcal/mol, respectively.

Docking of 54 drugs and 61 natural products yielded 5 drugs and 12 natural products that have equal or higher (more negative) binding energies than the co-crystallized ligands and known inhibitors of TMPRSS2 and CTSL (**Table 1**). In summary, 17 compounds from the 115 overlapping hits from pharmacophore-based virtual screening were able to pass both screens and can potentially serve as dual inhibitors of TMPRSS2 and CTSL, two crucial targets in preventing SARS-CoV-2 cell entry. The binding energies of the putative dual inhibitors are in the range of -7.6 to -9.1 kcal/mol for TMPRSS2 and -7.0 to -8.7 kcal/mol for CTSL.

**Table 1.** Binding energies of shortlisted drugs and natural products after docking into TMPRSS2 and CTSL active sites.

Database ID	Name	Binding Energy (kcal/mol)	
		TMPRSS2	CTSL
Drugs			
CHEMBL3360203	Pilaralisib	-8.8	-7.9
CHEMBL9509	Silibinin	-8.5	-8.1
CHEMBL4297315	PCI-27483	-8.1	-7.3
CHEMBL1697687	Cloguanamil	-7.9	-7.2
CHEMBL474260	Carubicin	-7.9	-7.0
Natural Products			
NPC65003	Rhoifolin	-9.1	-8.7
NPC471536	(Cyclo[(6-Bromo-8-(6-Bromobenzioxazol-3(1h)-One)-8-Hydroxy)Tryptophan])Arginine	-8.4	-7.7
NPC304187	Barettin	-8.4	-7.6
NPC470135	Phelligrin A	-7.8	-7.8
NPC476538	Tropeoside A1	-8.5	-7.3
NPC135277	Hypolaetin 8-O-Beta-D-Glucuronide	-8.0	-7.4
NPC190637	6,8-Diprenylkaempferol	-7.7	-7.6
NPC262038	Neophellamuretin	-7.6	-7.6
NPC268204	Uncinanone A	-7.7	-7.4
NPC20907	Gericudranin E	-7.7	-7.4
NPC470136	Epi-Phelligrin A	-7.6	-7.4
NPC323123	1-(3,4-Dihydroxy-Benzyl)-1,2,3,4-Tetrahydro-Isoquinoline-6,7-Diol	-7.6	-7.3
Controls	Ligand <i>GBS</i>	-7.6	
	Camostat	-7.5	
	Ligand <i>NOW</i>		-7.0
	CLIK-148		-6.7

In silico analysis of the pharmacokinetics, medicinal chemistry, and toxicity of top hits is an important procedure in identifying leads that may be optimized and/or undergo further processing along the drug discovery pipeline. The overlapping hits from the docking procedure were submitted to the SwissADME server to determine if they have potential drug properties, and to OSIRIS Property Explorer to check for mutagenic, tumorigenic, irritant, and reproductive toxicity risks (**Table 2**). The drugs were also included in this screening, albeit some were studied already in clinical trials to validate their drug-likeness and safety.

**Table 2.** ADME, drug-likeness, synthetic accessibility, and toxicity profiles of docking hits.

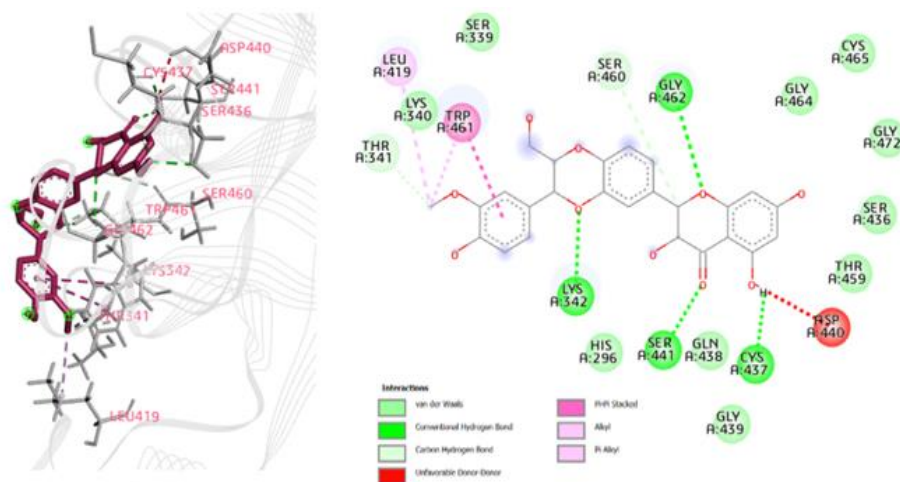
Compound	ADME Properties, Drug-likeness, and Synthetic Accessibility							Toxicity Risk			
	Water solubility	GI absorption	BBB permeant	P-gp substrate	CYP2D6 inhibitor	Drug-likeness (Lipinski)	SAscore <sup>a</sup>	Mutagenic	Tumorigenic	Irritant	Reproductive
Pilaralisib	Moderate	Low	No	No	No	Yes; 1 violation	4.00	Medium Risk	No	High Risk	No

Silibinin	Moderate	Low	No	No	No	Yes; 0 violation	4.92	No	No	No	No
PCI-27483	Soluble	Low	No	No	No	No; 3 violations	4.27	No	No	No	No
Carubicin	Moderate	Low	No	Yes	No	No; 3 violations	5.67	No	No	High Risk	High Risk
Cloguanamil	Soluble	High	No	No	No	Yes; 0 violation	2.17	No	No	No	Medium Risk
NPC65003	Soluble	Low	No	Yes	No	No; 3 violations	6.33	No	No	No	No
NPC471536	Moderate	Low	No	No	No	No; 3 violations	5.29	High Risk	High Risk	No	Medium Risk
NPC304187	Soluble	Low	No	Yes	No	Yes; 1 violation	4.19	No	No	No	No
NPC470135	Moderate	High	No	No	Yes	Yes; 0 violation	3.65	No	Medium Risk	No	Medium Risk
NPC476538	Moderate	Low	No	Yes	No	No; 3 violations	8.98	No	No	No	No
NPC135277	Soluble	Low	No	Yes	No	No; 2 violations	5.16	High Risk	No	No	No
NPC190637	Poor	Low	No	No	No	Yes; 0 violation	4.21	No	No	No	No
NPC262038	Moderate	High	No	Yes	Yes	Yes; 0 violation	4.04	No	No	No	No
NPC268204	Moderate	High	No	No	Yes	Yes; 0 violation	3.45	No	No	No	No
NPC20907	Moderate	High	No	No	Yes	Yes; 0 violation	4.03	No	No	No	No
NPC470136	Moderate	High	No	No	Yes	Yes; 0 violation	3.65	No	Medium Risk	No	Medium Risk
NPC323123	Soluble	High	No	Yes	Yes	Yes; 0 violation	2.66	No	No	No	No

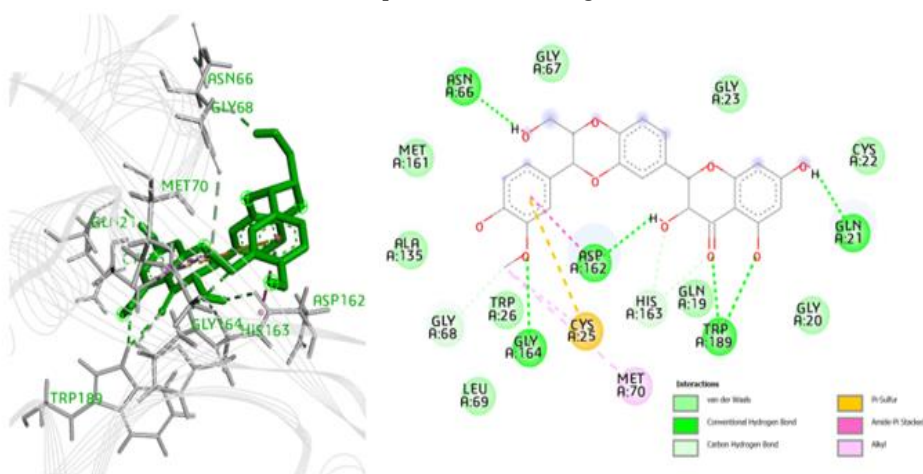
\*Synthetic Accessibility score of 6.0 was set as a threshold between easy- and hard-to-synthesize compounds [25].

Among the five selected ChEMBL drugs, almost all were shown to have favorable pharmacokinetics despite low gastrointestinal absorption but only three passed the Lipinski filter for drug-likeness. Pilaralisib, Cloguanamil, and Carubicin have been predictably found to exert medium to high toxicity risks of mutagenic, irritant, and/or reproductive effects, while PCI-27483 and Carubicin violate the Lipinski rule indicating less likelihood of pharmaceutical processing. Only Silibinin shows the most promising drug properties with its oral availability albeit having low gastrointestinal absorption, and it is neither penetrative of the blood-brain barrier nor a binder to P-gp and CYP2D6, complies with the Lipinski rule of drug-likeness, has tolerable SAscore, and without any predicted risk of toxicity. This suggests that Silibinin may be repurposed as a COVID-19 drug that blocks SARS-CoV-2 entry into cells via dual inhibition of TMPRSS2 and CTSL.

Based on the two- and three-dimensional interaction diagrams of Silibinin (**Figure 4**), the drug formed the important hydrogen bond with SER441 and van der Waals interaction with HIS296, which are both included in the catalytic triad of TMPRSS2. Its ether and hydroxyl groups also interacted with LYS342, CYS437, and GLY462 via conventional hydrogen bonding. However, it showed unfavorable donor-donor bonds with ASP440 which indicates repulsive forces that may reduce the stability of the complex. On the other hand, in the CTSL binding site, Silibinin formed important interactions with the catalytic dyad, namely pi-sulfur bonding with catalytic residue CYS25 and weak C-H bonding with HIS163. It also showed conventional hydrogen bonding with ASN66, GLN21, TRP189, and GLY164, as well as hydrophobic interaction with MET70 of the enzyme binding pocket.



a)



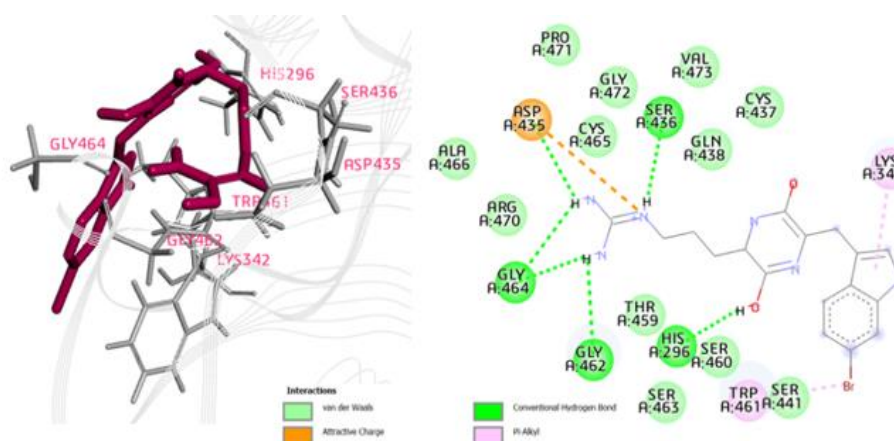
b)

**Figure 4.** 3D and 2D interaction diagrams of Silibinin with a) TMPRSS2 and b) CTSL.

Silibinin is the major active constituent of silymarin, a standardized extract of the milk thistle (*Silybum marianum*) seeds. This drug is available in European countries as a treatment for hepatotoxicity, chronic hepatitis, and cirrhosis. It is also investigated as an anticancer treatment due to its inhibition of STAT3 signaling and chemoprotective effects from environmental toxins and UV protection, as well as anti-inflammatory and immunomodulatory effects [26]. Interestingly, Silibinin was proposed to be a treatment for COVID-19 patients due to its dual activity in limiting host cytokine storm via STAT3 inhibition, and in inhibiting RdRp necessary for viral replication. Silibinin is currently undergoing clinical trials for oncohematological patients infected with the coronavirus in Spain [27]. Previous docking studies have also identified Silibinin to be able to form stable complexes with SARS-CoV-2 spike protein RBD and M<sup>pro</sup> [28]. This suggests that Silibinin may inhibit SARS-CoV-2 pathogenesis by acting against multiple mechanisms like viral attachment, cell entry, viral replication, post-translational modification, and immunomodulation.

This study also screened a natural products database for novel prophylactic or early-stage treatments for COVID-19. However, among the 12 selected natural products, all structures had either undesirable ADME properties or poor drug-likeness, and four were predictably found to exert medium to high toxicity risks of mutagenic, irritant, and/or reproductive effects. Nonetheless, NPC304187 or Baretin shows the most acceptable pharmacokinetic profile and lead-likeness, although it's a substrate of P-gp, it has no risks for any toxicity effect and is thus the best candidate among selected natural products for further processing to develop dual TMPRSS2 and CTSL inhibitors.

According to its interaction profiles (**Figure 5**), Baretin possesses a guanidine moiety which formed several hydrogen bonds and attractive charge with residues of TMPRSS2 binding site, namely ASP435, SER436, GLY464, and GLY462, and van der Waals forces with several hydrophobic side chains. The catalytic residues HIS296 interacted with the pyrazinol moiety via hydrogen bonding, and SER441 with the bromoindole moiety via van der Waals forces. Meanwhile, in the CTSL binding site, the pyrazinol moiety of Baretin displayed conventional hydrogen bonding with catalytic residue CYS25 and with MET161, as well as van der Waals interaction with the other catalytic residue HIS163. The guanidine moiety formed hydrogen bonds with ASP114, and another with ASP71 along with an attractive charge. Aside from the bromoindole moiety which displayed amide- $\pi$  stacking with GLY67 and hydrogen bonding with ASN66, several van der Waals forces were also evident with the hydrophobic side chains of the enzyme binding pocket.



a)



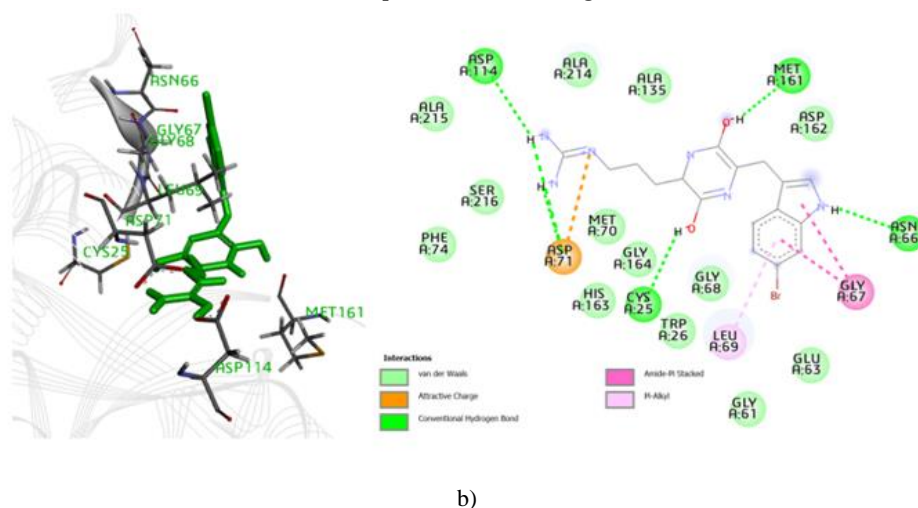


Figure 5. 3D and 2D interaction diagrams of Baretin with a) TMPRSS2 and b) CTSL.

Baretin is the major compound in the deep-sea sponge *Geodia barretti* found in the North Atlantic Ocean. The metabolite was found to possess antioxidant and anti-inflammatory properties via inhibition of protein kinases RIPK2 and CAMK1 $\alpha$  [29] but had not yet been classified as an antiviral agent before this work. This compound may be further elaborated *in silico* and screened using *in vitro* and *in vivo* methods to validate its antiviral activity against SARS-CoV-2.

## Conclusion

This study offers the first comprehensive *in silico* approach to identify potential compounds that simultaneously target TMPRSS2 and CTSL, each having independent roles in the entry mechanism of the SARS-CoV-2. An integrative structure-based pharmacophore modeling, virtual screening, molecular docking, and ADMET prediction revealed Silibinin (Phase 3 drug) and Baretin (natural product) as potential dual inhibitors of TMPRSS2 and CTSL. The outcomes of this work may stimulate further experimental and clinical studies on the development of COVID-19 drugs based on the molecular architecture of the top hits.

**Acknowledgments:** The authors are grateful to Inte: Ligand for the LigandScout software and to The Scripps Research Institute for the AutoDock Vina. ILSV also thanks the Department of Science and Technology – Science Education Institute for the scholarship.

**Conflict of interest:** None

**Financial support:** None

**Ethics statement:** None

## References

1. [WHO] World Health Organization. WHO Coronavirus (COVID-19) Dashboard. 2022. Available from: <https://covid19.who.int>.
2. Bhimraj A, Morgan RL, Shumaker AH, Laverigne V, Baden L, Cheng VC, et al. Infectious Diseases Society of America Guidelines on the Treatment and Management of Patients with COVID-19. Clin Infect Dis. 2020:ciaa478. doi:10.1093/cid/ciaa478
3. Bryant A, Lawrie TA, Dowswell T, Fordham EJ, Mitchell S, Hill SR, et al. Ivermectin for Prevention and Treatment of COVID-19 Infection: A Systematic Review, Meta-analysis, and Trial Sequential Analysis to Inform Clinical Guidelines. Am J Ther. 2021;28(4):e434-60. doi:10.1097/MJT.0000000000001402
4. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. 2020;41(9):1141-9. doi:10.1038/s41401-020-0485-4
5. Davidson A, Wysocki J, Batlle D. Interaction of SARS-CoV-2 and other coronaviruses with ACE (angiotensin-converting enzyme)-2 as their main receptor. Hypertension. 2020;76(5):1339-49. doi:10.1161/HYPERTENSIONAHA.120.15256
6. Xiang R, Yu Z, Wang Y, Wang L, Huo S, Li Y, et al. Recent advances in developing small-molecule inhibitors against SARS-CoV-2. Acta Pharm Sin B. 2022;12(4):1591-623. doi:10.1016/j.apsb.2021.06.016

7. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020;181(2):271-80. doi:10.1016/j.cell.2020.02.052
8. Zhao MM, Yang WL, Yang FY, Zhang L, Huang WJ, Hou W, et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduct Target Ther*. 2021;6(1):134. doi:10.1038/s41392-021-00558-8
9. Shukla D. New antiviral drug combo may fight off coronavirus infection. *Medical News Today*. 2021. Available from: <https://www.medicalnewstoday.com/articles/new-antiviral-drug-combo-may-fight-off-coronavirus-infection>
10. Macalino SJ, Billones JB, Organo VG, Carrillo MC. In silico strategies in tuberculosis drug discovery. *Molecules*. 2020;25(3):665. doi:10.3390/molecules25030665
11. Stahura F, Bajorath J. Virtual screening methods that complement HTS. *Comb Chem High Throughput Screen*. 2004;7(4):259-69. doi:10.2174/1386207043328706
12. Sampaco III ARB, Billones JB. Virtual screening of natural products, molecular docking and dynamics simulations on M. tuberculosis S-adenosyl-L-homocysteine. *Orient J Chem*. 2015;31(4):1859-65. doi:10.13005/ojc/310402
13. Billones JB, Carrillo MCO, Organo VG, Macalino SJY, Emnacen IA, Sy JBA. Virtual Screening against Mycobacterium tuberculosis Lipoate Protein Ligase B (MtbLipB) and In Silico ADMET Evaluation of Top Hits. *Orient J Chem*. 2013;29(4):1457-68. doi:10.13005/ojc/290423
14. Uy VCC, Billones JB. Towards antituberculosis drugs: virtual screening for potential inhibitors of pantothenate synthetase of Mycobacterium tuberculosis. *Philipp Sci Lett*. 2012;5(2):122-30. Available from: <https://philsciletters.net/2012-122/>
15. Yang CTM, Billones JB. Towards antituberculosis drugs: molecular docking of curcumin and its analogues to pantothenate synthetase. *Philipp J Sci*. 2012;141(2):187-96. Available from: <https://philjournalsci.dost.gov.ph/home-1/36-vol-141-no-2-december-2012/473-towards-antituberculosis-drugs-molecular-docking-of-curcumin-and-its-analogues-to-pantothenate-synthetase>
16. Billones JB. Reverse docking study unravels the potential Mycobacterium tuberculosis enzyme targets of agelastine F. *Orient J Chem*. 2016;32(2):851-8. doi:10.13005/ojc/320210
17. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*. 2005;45(1):160-9.
18. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform*. 2011;3:33. doi:10.1186/1758-2946-3-33
19. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455-61. doi:10.1002/jcc.21334
20. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem*. 2012;55(14):6582-94. doi:10.1021/jm300687e
21. Chan C. What is a ROC Curve and How to Interpret It. *Displayr Blog*. 2018. Available from: <https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/>
22. Bouback TA, Pokhrel S, Albeshri A, Aljohani AM, Samad A, Alam R, et al. Pharmacophore-based virtual screening, quantum mechanics calculations, and molecular dynamics simulation approaches identified potential natural antiviral drug candidates against MERS-CoV S1-NTD. *Molecules*. 2021;26(16):4961. doi:10.3390/molecules26164961
23. Hevener KE, Zhao W, Ball DM, Babaoglu K, Qi J, White SW, et al. Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *J Chem Inf Model*. 2009;49(2):444-60. doi:10.1021/ci800293n
24. Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, et al. Insights into protein-ligand interactions: mechanisms, models, and methods. *Int J Mol Sci*. 2016;17(2):144. doi:10.3390/ijms17020144
25. Vošilák M, Kolár M, Čmelo I, Svozil D. SYBA: Bayesian estimation of synthetic accessibility of organic compounds. *J Cheminform*. 2020;12(1):35. doi:10.1186/s13321-020-00439-2
26. Bosch-Barrera J, Sais E, Cañete N, Marruecos J, Cuyàs E, Izquierdo A, et al. Response of brain metastasis from lung cancer patients to an oral nutraceutical product containing silibinin. *Oncotarget*. 2016;7(22):32006-14. doi:10.18632/oncotarget.7900
27. Bosch-Barrera J, Martín-Castillo B, Buxó M, Brunet J, Encinar JA, Menéndez JA. Silibinin and SARS-CoV-2: dual targeting of host cytokine storm and virus replication machinery for clinical management of COVID-19 patients. *J Clin Med*. 2020;9(6):1770. doi:10.3390/jcm9061770
28. Speciale A, Muscarà C, Molonia MS, Cimino F, Saija A, Giofrè SV. Silibinin as potential tool against SARS-Cov-2: In silico spike receptor-binding domain and main protease molecular docking analysis, and in vitro endothelial protective effects. *Phytother Res*. 2021;35(8):4616-25. doi:10.1002/ptr.7107
29. Lind KF, Østerud B, Hansen E, Jørgensen TØ, Andersen JH. The immunomodulatory effects of baretin and involvement of the kinases CAMK1 $\alpha$  and RIPK2. *Immunopharmacol Immunotoxicol*. 2015;37(5):458-64. doi:10.3109/08923973.2015.1082584