

# **Repurposing of drugs for combined treatment of COVID-19 cytokine storm using machine learning**

**Maanaskumar R. Gantla<sup>1</sup>, Igor F. Tsigelny<sup>2,3,4\*</sup>, Valentina L. Kouznetsova<sup>2,3</sup>**

<sup>1</sup>MAP program, UC San Diego, Calif., USA

<sup>2</sup>San Diego Supercomputer Center, UC San Diego, Calif., USA

<sup>3</sup>BiAna, La Jolla, Calif., USA

<sup>4</sup>Dept. of Neurosciences, UC San Diego, Calif., USA

\*Correspondence to: [itsigel@ucsd.edu](mailto:itsigel@ucsd.edu)

## **Abstract**

**Context:** SARS-CoV-2 induced cytokine storm is the major cause of COVID-19 related deaths. Patients have been treated with drugs that work by inhibiting a specific protein partly responsible for the cytokines production. This approach provided very limited success, since there are multiple proteins involved in the complex cell signaling disease mechanisms.

**Objective:** To elucidate using machine learning (ML) the set of drugs targeting a group of proteins involved in the mechanism of cytokine storm.

**Methods:** We selected for targeting five proteins: Angiotensin II receptor type 1 (AT1R), A disintegrin and metalloprotease 17 (ADAM17), Nuclear Factor-Kappa B (NF- $\kappa$ B), Janus kinase 1 (JAK1) and Signal Transducer and Activator of Transcription 3 (STAT3) that are involved in the SARS-CoV-2 induced cytokine storm pathway. We developed ML models for these five proteins, using known active inhibitors. After developing the model for each of these proteins, FDA-approved drugs were screened to find novel therapeutics for COVID-19.

**Results:** We identified twenty drugs that are active for four proteins and eight drugs active for five proteins. Mitomycin C is the most active drug across all five proteins with an average prediction score of 0.886. For further validation of these results, we used the PyRx software to conduct protein–ligand docking experiments and calculated the binding affinity. The docking results support findings by the ML model.

**Conclusions:** It is possible to elucidate the drugs, targeting simultaneously several proteins related to cytokine production to treat the cytokine storm in COVID-19 patients.

**Keywords:** COVID-19; SARS-CoV-2; docking; machine learning; multi-targeted drug discovery; screening of FDA-approved drugs.

**Acronyms and Abbreviations:** 1D, 2D, 3D, one-, two-, three-dimensional; ADAM17, A disintegrin and metalloprotease 17; ARDS, acute respiratory distress syndrome; AT1R, Angiotensin II receptor type 1; ARDS, acute respiratory distress syndrome; AUROC, area under receiver operator characteristic curve, COVID-19, coronavirus disease 2019; CRS, cytokine release syndrome; CXCL10, CXC-chemokine

ligand 10; FDA, Food and Drug Administration; G-CSF, granulocyte colony stimulating factor; IC<sub>50</sub>, half maximal inhibitory concentration; ICU, intensive care unit; IL, interleukin; JAK1, Janus kinase 1; MCP1, monocyte chemoattractant protein-1; MIP1 $\alpha$ , macrophage inflammatory protein 1; ML, machine learning; NF- $\kappa$ B, Nuclear Factor-Kappa B; PaDEL, Pharmaceutical data exploration laboratory; PDB, Protein Data Bank; ROC, receiver operator characteristic curve; SMILES, Simplified Molecular-Input Line-Entry System; STAT3, signal transducer and activator of transcription 3; TNF $\alpha$ , tumor necrosis factor  $\alpha$ ; WEKA, Waikato Environment for Knowledge Analysis.

## Introduction

The COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) resulted in millions of infected patients and deaths worldwide [1,2]. Patients frequently encountered complications with significant mortality, particularly by acute respiratory distress syndrome (ARDS) with a broad spectrum of issues such as multiple-organ failure, and blood clots [3,4]. There has been tremendous amount of research going on towards discovering therapeutics for the COVID-19, and few drugs have been approved by FDA such as remdesivir, Paxlovid and molnupiravir, and all of them mainly target viral proteins[5,6].

Mounting research data reveals that the severity of COVID-19 is mainly associated with an increased level of inflammatory mediators including cytokines and chemokines such as interleukin IL-2, IL-7, IL-8, IL-9, IL-10, IL-17, tumor necrosis factor  $\alpha$  (TNF $\alpha$ ), monocyte chemoattractant protein-1 (MCP1), macrophage inflammatory protein 1 alpha (MIP1 $\alpha$ ), granulocyte colony stimulating factor (G-CSF), CXC-chemokine ligand 10 (CXCL10), C-reactive protein, ferritin, and D-dimers in blood upon SARS-CoV-2 infection [7–13]. More specifically, patients in intensive care unit (ICU) showed higher levels of plasma inflammatory cytokines compared to non-ICU patients [14], and therefore fatal COVID-19 is characterized as a cytokine release syndrome (CRS) that is caused by a cytokine storm. Thus, targeting proteins responsible for cytokine storm serves as a possible mechanism of treatment for severe COVID-19 patients [15–17].

The SARS-CoV-2 induced cytokine storm pathway [18] shows that there are multiple proteins involved in the disease signaling mechanisms. Cytokines are cell signaling, small protein molecules that aid cell to cell communication in immune responses and stimulate the movement of cells towards sites of inflammation, infection, and trauma [19]. Cytokine Storm is essentially an unregulated immune response characterized by an excessive release of multiple pro-inflammatory Cytokines [20,21].

It has been identified that proteins such as Angiotensin II receptor type 1 (AT1R), A disintegrin and metalloprotease 17 (ADAM17), Nuclear Factor-Kappa B (NF- $\kappa$ B), Janus kinase 1 (JAK1) and Signal transducer and activator of transcription 3 (STAT3) are implicated in the production of proinflammatory

cytokines and are considered as a promising COVID-19 therapeutic targets [15]. Therefore, discovering a drug that can interfere with function of either all of the proteins or most of them synergistically would become an effective therapeutic. Based on literature search, as of now there are no such therapeutics exist. Discovery of novel effective drugs and therapies for COVID-19 is critical for tackling the disease. However, discovery and development of effective therapies can be costly and time-consuming. For this reason, it would be ideal to repurpose already existing FDA-approved drugs given the proven safety, if they can also interfere effectively with proteins responsible for cytokine storm.

In this pathway, we have investigated five proteins: AT1R, ADAM17, NF- $\kappa$ B1, JAK1, STAT3. AT1R signaling axis activates ADAM17, which results in the production of cytokines TNF $\alpha$  and IL-6. The IL-6 amplifier plays a critical role in chronic inflammatory diseases. Activation of NF- $\kappa$ B, JAK1 and STAT3 triggers the IL-6 amplifier which causes the cytokine storm and leads to the ARDS and multiple-organ failure. Targeting these five proteins would prevent cytokine storm to yield the best potent COVID-19 drug.

Conventional methods of drug discovery as shown in Fig. 1 are very expensive, complex processes that takes several years to bring drugs to the clinic. We used machine learning to expedite the drug discovery process by screening FDA drugs, so that the treatment for COVID-19 is available sooner.

Recently, machine learning (ML) has emerged as an important computational technique and has been applied to various tasks in drug discovery, such as molecular property prediction and drug–target interaction prediction. Given the great advantage of this computational tool in terms of the cost and time, in this project we have used ML classification model with a random forest algorithm in WEKA software [22] for repurposing of some FDA-approved drugs for use as COVID-19 therapeutics. These predictions can then be confirmed through structure-based virtual screening, specifically using docking simulators PyRX. The docking provides the binding energy for each conformer and helps validate the accuracy of prediction.

## **Methods**

All research was completed in silico. The programs, tools, and websites used: PubChem, ZINC database subsection covering FDA-approved drugs, Protein Data Bank (PDB), Pharmaceutical Data

Exploration Laboratory (PaDEL-Descriptor), Waikato Environment for Knowledge Analysis (WEKA), PyRx, Discovery Studio Visualizer. A flowchart of methods is presented in Fig. 1.

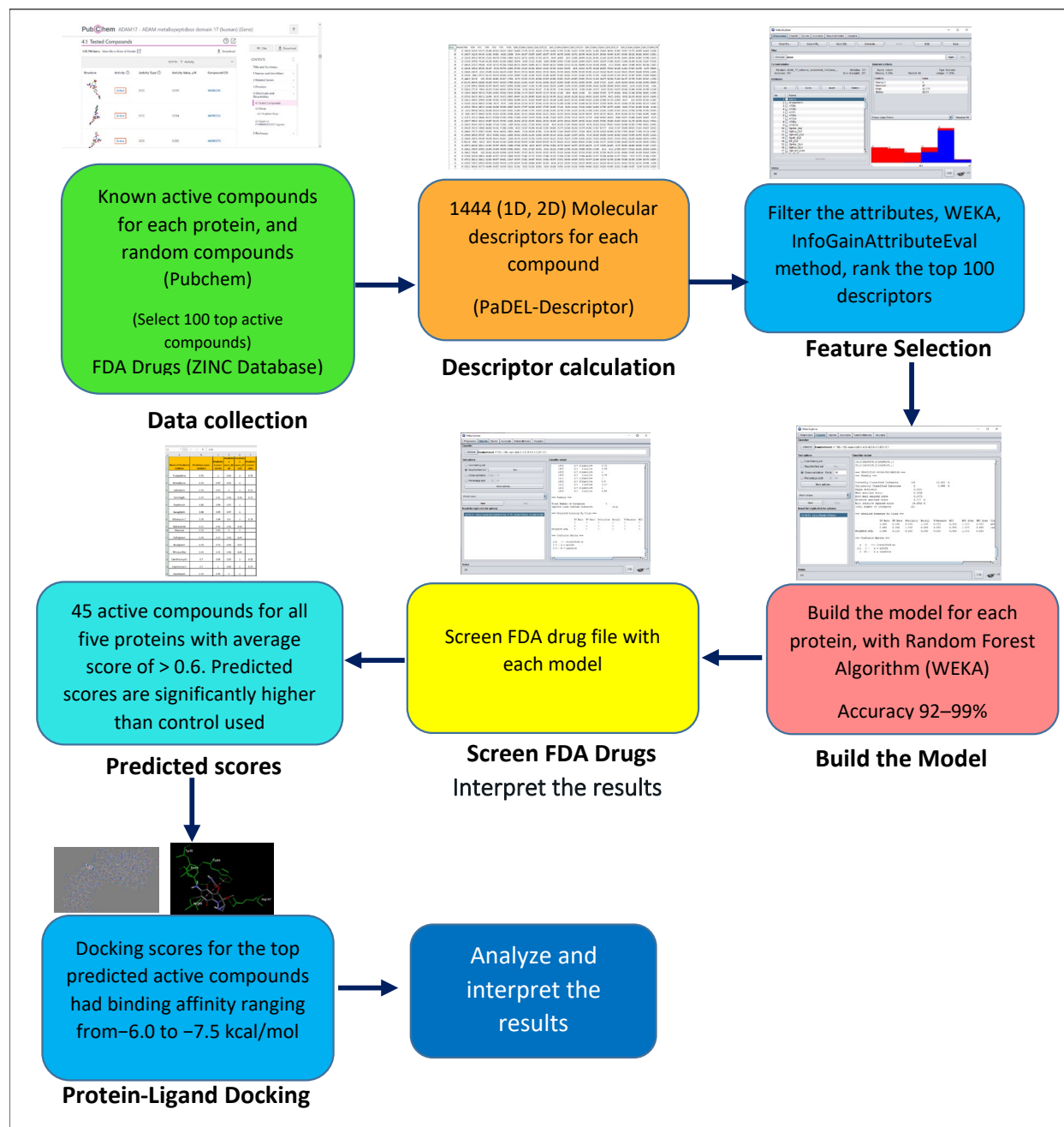


Figure 1. Flowchart of the methods.

## Data collection

Data for known active inhibitors and a control set of random compounds obtained from PubChem are listed in Table 1. Data for FDA-approved drugs obtained from the ZINC database. Activity values and SMILES (Simplified Molecular-Input Line-Entry System) [23] files for compounds tested with the proteins AT1R, ADAM17, NF- $\kappa$ B, JAK1, STAT3 were retrieved from PubChem. To limit the tested compounds to the strongest inhibitors, compounds with top 100 IC<sub>50</sub> values were chosen for training the model. One-thousand-six-hundred-fifteen FDA-approved drugs and their SMILES were retrieved from ZINC database.

**Table 1.** Known inhibitors obtained from PubChem.

Protein	Number of Known Inhibitors	IC <sub>50</sub> values range ( $\mu$ M)
AT1R	1192	0.00005–19.98
ADAM17	1813	0.000026–44.0
NF- $\kappa$ B	348	0.003–49.6
JAK1	4596	0.0000013–39.81
STAT3	588	0.0084–48.0

The chemical structures are obtained in SMILES format. These files are 1D ASCII strings that represent 3D molecular structure. An example of top ten inhibitors for AT1R are shown in Table 2.

**Table 2.** The top ten inhibitors for AT1R.

Compound	IC <sub>50</sub> (nM)
BDBM50049199	<b>0.05</b>
Saralasin	<b>0.06</b>
2Botbmip	<b>0.08</b>
CHEMBL42775	<b>0.01</b>
CHEMBL158809	<b>0.01</b>
CHEMBL298417	<b>0.01</b>
BDBM50283219	<b>0.01</b>
BDBM50283237	<b>0.01</b>
BDBM50283194	<b>0.01</b>
BDBM50283245	<b>0.01</b>

### ***Molecular descriptor calculation***

PaDEL-Descriptor software [24] is used to calculate the molecular descriptors for the compounds. These descriptors are the characteristics of the compound that are used for training of the ML model. For example, number of aromatic rings, number of pi bonds, molecular weight, atom count, etc. The software currently calculates 1875 descriptors (1444 1D and 2D descriptors and 431 3D descriptors) and 12 types of fingerprints. For our model building we have used 1444 1D and 2D descriptors.

### ***InfoGain filtering in WEKA to select top 100 descriptors***

To narrow down the calculated 1D and 2D descriptors from 1444 to only the most significant ones, we utilized attribute selection from WEKA [25], an open-source ML software. The descriptors were ranked by the Information Gain Attribute Evaluation (InfoGain) function, an unsupervised machine-learning algorithm, that measures how important each descriptor is in determining whether a given molecule is an inhibitor or not. InfoGain measures how each feature contributes to decreasing the overall computational entropy. Only the most significant descriptors were selected to be used by the ML model to reduce noise.

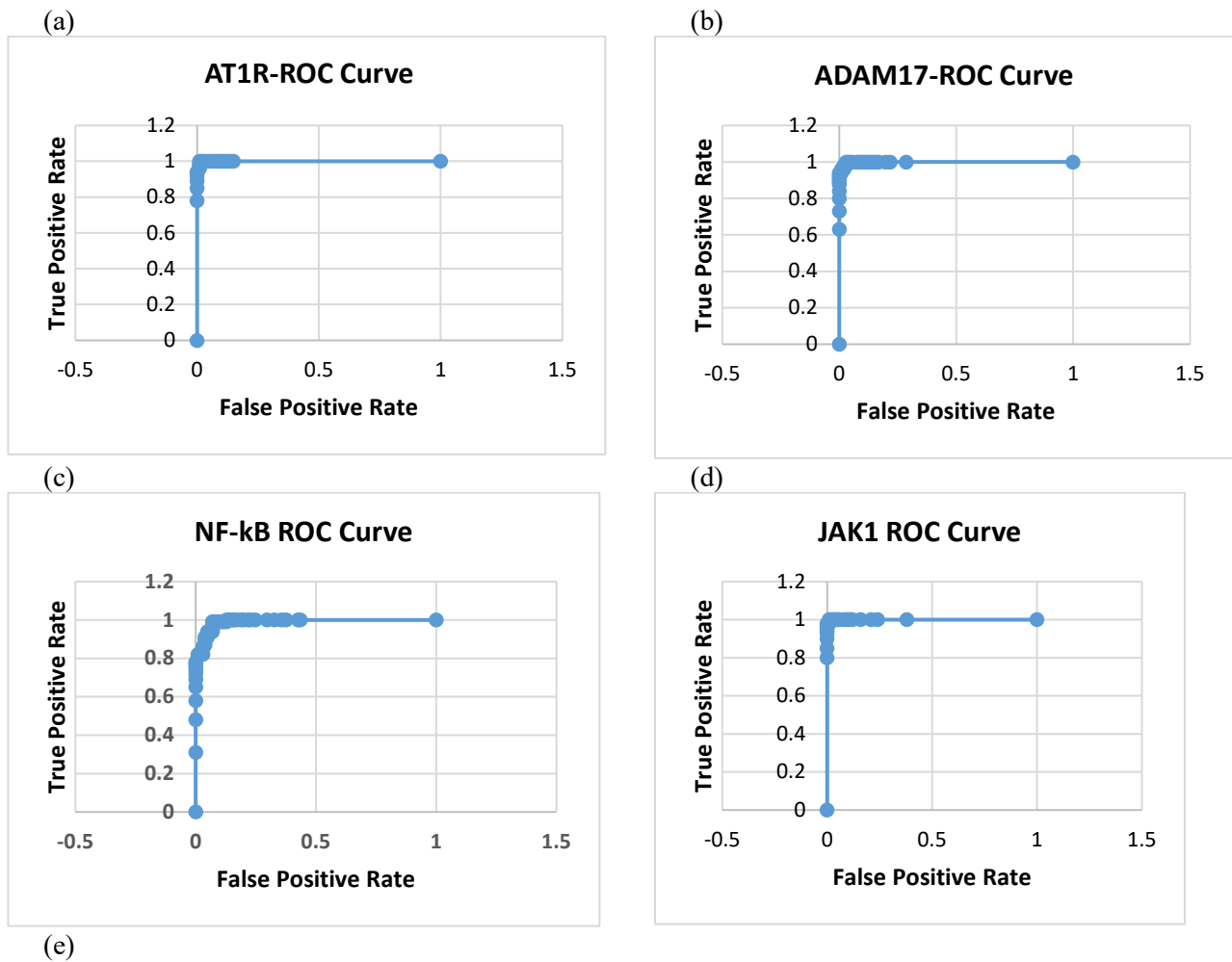
### ***Building a machine-learning model***

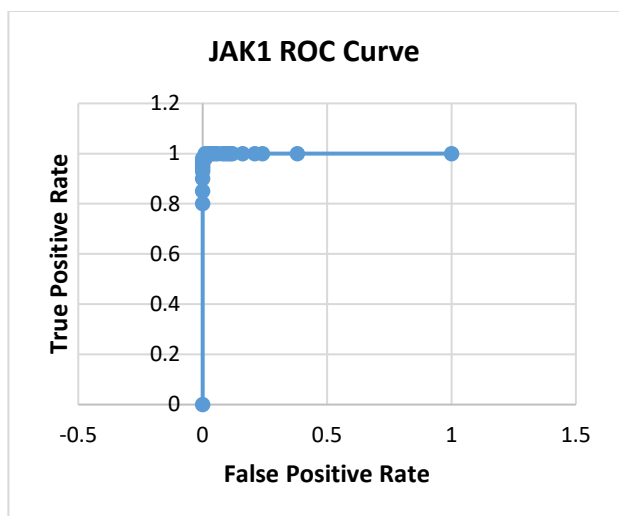
Machine-learning model for each protein was built using WEKA [22]. WEKA provides both standard and extensive ML functionality, integrated within classification, regression, clustering and other pattern recognition capabilities. Data for the model is prepared by taking top 100 descriptors of top 100 inhibitors for each of the proteins and 100 control set of random molecules.

First, we submitted the prepared inhibitor file containing selected and random compounds with their molecular descriptors into WEKA. Then used the Random Forest algorithm 10-folds cross validation to build the model. Also, we used Random Forest algorithm with an 80%/20% training–testing split to evaluate the performance of the model. Such the training–testing split ensures that there is no overfitting as 20% of the data. It was not used to build the model but used for testing. Then we analyzed the model accuracy and elucidated the ROC curves. Saved model was used in the next step to screen FDA approved drugs. The



Receiver Operating Characteristic (ROC) curves were calculated to measure the effectiveness of the model. ROC curve summarizes the prediction performance of a classification model at all classification thresholds. Figs. 2a–2e present the ROC graphs for machine-learning models of proteins AT1R, ADAM17, NF- $\kappa$ B, JAK1, and STAT3. Model accuracy is 91.5–99.0% range and Area Under the Receiver-Operating Characteristic curve (AUROC) is 0.97–1.00. That values confirm the accuracy of the models. Receiver-Operating Characteristic (ROC) curves for five proteins are shown in Fig. 2.





**Figure 2.** Accuracies and AUROC of the predictions of inhibitors for five proteins related to cytokine storm in COVID-19: (a) AT1R Accuracy 98.5% and AUROC 99% ;(b) ADAM17 Accuracy 98.5% and AUROC 99%; (c) NF- $\kappa$ B Accuracy 96%, AUROC 99%; (d) JAK1 Accuracy 98.5%, AUROC 100%; (e) STAT3 Accuracy 91.5%, AUROC 97.8%.

### *Screening of FDA-approved drugs using the model*

FDA approved drugs are downloaded from ZINC database [26]. Using PaDEL descriptor software, the molecular descriptors were calculated for all the downloaded 1665 FDA-approved drugs. Out of 1444 descriptors, the same 100 descriptors were selected as the training set of the corresponding protein inhibitors. These were then screened with the ML model built using WEKA. The output was analyzed for the prediction scores and the predicted drugs were ranked based on the ML predicted score. The predicted drugs were ranked for each protein and averaged the score among all five proteins listed in Table 3.

**Table 3.** Active compounds with the greater than 0.6 predictive score.

Name of Predicted Inhibitor	ZINCID	Average score for five proteins	Machine Learning (ML) rank
Mitomycin C	ZINC000030726187	0.886	1
Valrubicin	ZINC000028232755	0.882	2–3
Pomalidomide	ZINC000003940470	0.882	2–3
Fludarabine	ZINC000004216238	0.880	4
Clarithromycin	ZINC000085534098	0.872	5–6
Trabectedin	ZINC000150338708	0.872	5–6

Capreomycin	ZINC000150338698	0.868	7
Sonidegib	ZINC000068202099	0.852	8
Abacavir	ZINC000002015928	0.820	9
Raltegravir	ZINC000013831130	0.816	10
Saxagliptin	ZINC000013648755	0.810	11
Nimodipine	ZINC000019632713	0.790	12–13
Suvorexant	ZINC000049036447	0.790	12–13
Boceprevir	ZINC000014210457	0.783	14
Saquinavir	ZINC000026985532	0.780	15
Balsalazide	ZINC000003952881	0.760	16
Minocycline	ZINC000014879992	0.750	17
Eribulin	ZINC000169344691	0.742	18
Isradipine	ZINC000100001918	0.731	19
Cangrelor	ZINC000085537017	0.730	20
Tinidazole	ZINC000000113446	0.721	21
Oxymorphone	ZINC000003875483	0.720	22
Brivaracetam	ZINC000003979899	0.718	23
Naloxegol	ZINC000095564694	0.715	24
Roflumilast	ZINC592419	0.712	25–27
Methotrexate	ZINC1529323	0.712	25–27
Piperacillin	ZINC3913937	0.712	25–27
Darunavir	ZINC3955219	0.711	28
Apremilast	ZINC30691736	0.710	29
Tenidap	ZINC100001953	0.700	30–32
Dibrospidium	ZINC1842693	0.700	30–32
Ladostigil	ZINC27623417	0.700	30–32
Tibezonium	ZINC4217561	0.682	33
Taribavirin	ZINC3781686	0.673	34
Omoconazole	ZINC1532917	0.650	35–36
Irampanel	ZINC7363	0.650	35–36
Ciprostene	ZINC3800479	0.640	37
Nalfurafine	ZINC4393014	0.630	38–39
Ritonavir	ZINC3944422	0.630	38–39
Fingolimod	ZINC1542002	0.620	40
Lisinopril	ZINC3812863	0.620	41
Lesinurad	ZINC84757007	0.610	42–44
Ibrutinib	ZINC35328014	0.610	42–44
Ticagrelor	ZINC28957444	0.610	42–44
Ertapenem	ZINC3918453	0.600	45

### ***Docking of predicted FDA-approved drugs to selected proteins***

To confirm the activity and binding to the protein, docking of predicted FDA-approved drugs was performed using PyRx tool [27] with Discovery Studio software [28] to visualize the results. For docking the selected compounds, the crystal structure of the protein was downloaded from PDB [29,30] for each of the five proteins. PDB IDs for selected proteins are AT1R—4ZUD, ADAM17—2FV5, NF- $\kappa$ B—1SVC, JAK1—4EI4, and STAT3—6NUQ. A binding active site is defined for each protein based on the reported ligand interactions with protein.

To validate the specificity of the docked compounds, docking of random compounds was also conducted. A random number generator without repetition was used to obtain 100 random compound IDs and to select entries from the PubChem database that correspond to the random numbers obtained.

Each of the five proteins' 3D structure with a known ligand was downloaded from the PDB database. Each predicted FDA-approved drug's 3D structure was downloaded from PubChem. From each downloaded protein–ligand complex, the ligand was removed in Discovery Studio and remained protein was loaded into PyRx. The active sites of each protein were defined as a box that encompasses residues of the binding site. Then we ran the AutoDock Wizard for the top 12 predicted compounds, 12 best-activity known compounds, and 12 random compounds with each protein. For each compound nine conformers were generated and docked. In total there were generated 324 conformers, which were docked to each protein. The docked protein–ligand complexes were analyzed to elucidate the interactions of compounds with amino-acid residues. Binding Free Energy values are listed in Tables 10 and 11.

## **Results**

### ***Machine-learning prediction results***

The results of the ML models' predictions were evaluated using confusion matrices and their derivatives: the accuracy (ACC), precision (PREC), Matthews correlation coefficient (MCC), true-positive

rate (TPR) or recall (REC), false-positive rate (FPR), as well as the area under the receiver operating characteristic (ROC) curve (AUROC), and the area under the precision–recall curve (PRC area).

The weighted averages for each of these metrics are listed in Table 4. The ROC curve compares the sensitivity and specificity across a range of values. Thus, the vertical axis is the TPR, that is, the sensitivity or recall; and the horizontal axis is the FPR or (1–specificity). The FPR is the probability of falsely classifying a positive class. The model’s low FPR of 0.015 to 0.040 demonstrates a low probability of wrongly classifying an inactive compound to active compound. The TPR (sensitivity) is the probability of correctly classifying a positive class. The model’s high TPR of 0.915 to 0.985 indicates a high probability of correctly classifying an active compound. The large average AUROC value 0.978 to 1.0 indicates that the classification is accurate. Another way to evaluate the performance of the proposed method is the PRC area, which shows precision values for the corresponding sensitivity (recall, i.e., TPR) values. The model’s large PRC area value of 0.979 to 1.0 again shows the good performance of our method for all the five proteins.

**Table 4.** Performance of the developed ML models for the five proteins related to cytokine storm in COVID-19.

<b>Protein</b>	<b>ACC</b>	<b>TPR</b>	<b>FPR</b>	<b>PREC</b>	<b>MCC</b>	<b>AUROC</b>	<b>PRC Area</b>
AT1R	98.5%	0.985	0.015	0.985	0.970	0.999	0.999
ADAM17	98.5%	0.985	0.015	0.985	0.970	0.999	0.999
NF- $\kappa$ B	96%	0.960	0.040	0.961	0.921	0.993	0.993
JAK1	98.5%	0.985	0.015	0.985	0.970	0.999	0.999
STAT3	91.5%	0.915	0.085	0.915	0.830	0.978	0.979

*Note:* ACC, accuracy; TPR, true-positive rate; FPR, false-positive rate; PREC, precision; MCC, Matthews correlation coefficient; AUROC, area under the receiver-operating characteristic curve; PRC area, area under the precision–recall curve.

Total 45 compounds found to be active inhibiting all five proteins AT1R, ADAM17, NF- $\kappa$ B, STAT3, JAK1. Forty-five active compounds with the greater than 0.6 average predictive score is shown in Table 4.

Top eight FDA-approved drugs predicted active for all five proteins AT1R, ADAM17, NF- $\kappa$ B, STAT3, and JAK1 are shown in Table 5.

**Table 5.** Eight compounds active for five proteins with greater than 0.85 average predictive score.

Name of Predicted Inhibitor	Prediction score					Average score	ML Rank
	AT1R	ADAM17	NF- $\kappa$ B	JAK1	STAT3		
Mitomycin C	0.99	0.76	0.90	0.78	1.00	0.886	1
Valrubicin	0.97	0.73	1.00	0.71	1.00	0.882	2–3
Pomalidomide	1.00	0.76	0.92	0.73	1.00	0.882	2–3
Fludarabine	1.00	0.73	0.94	0.73	1.00	0.880	4
Clarithromycin	0.99	0.70	0.95	0.72	1.00	0.872	5–6
Trabectedin	0.99	0.75	0.91	0.71	1.00	0.872	5–6
Capreomycin	1.00	0.70	0.89	0.75	1.00	0.868	7
Sonidegib	0.91	0.72	0.94	0.71	0.98	0.852	8

Mitomycin C (Table 3) is top ranked FDA drug across all five proteins with average prediction score of 0.886. Top 20 FDA-approved drugs predicted active for four proteins (AT1R, ADAM17, NF- $\kappa$ B, STAT3) showed an average score of greater than 0.83 are shown in Table 6.

**Table 6.** Twenty compounds active for four proteins with greater than 0.83 average predictive score.

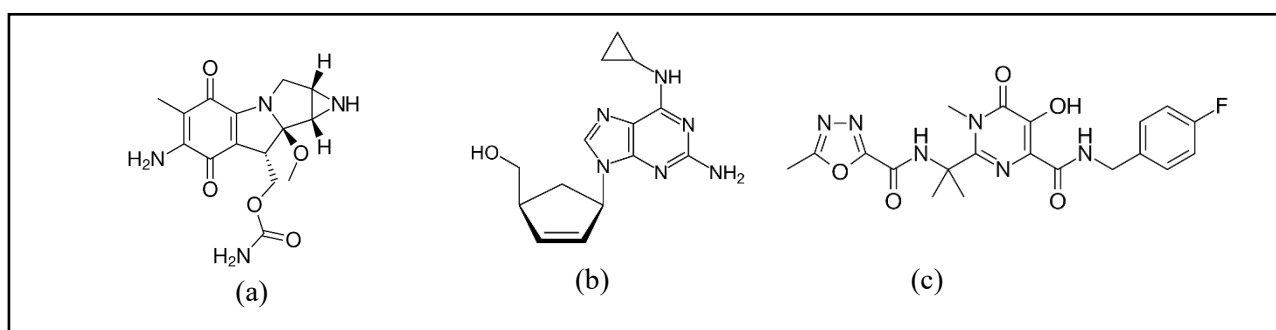
Name of Predicted Inhibitor	Prediction score				Average score	ML Rank
	AT1R	ADAM17	NF- $\kappa$ B	STAT3		
Abacavir	0.93	0.73	1.00	1.00	0.9600	1–2
Raltegravir	0.73	0.74	0.98	0.89	0.9600	1–2
Saxagliptin	0.99	0.88	0.97	1.00	0.9250	3–4
Valrubicin	0.97	0.73	1.00	1.00	0.9250	3–4
Pomalidomide	1.00	0.76	0.92	1.00	0.9200	5
Nimodipine	0.97	0.72	0.95	1.00	0.9175	6–7
Fludarabine	1.00	0.73	0.94	1.00	0.9175	6–7
Suvorexant	0.95	0.72	1.00	1.00	0.9150	8
Boceprevir*	0.72	0.72	0.99	0.97	0.9125	9–11
Mitomycin C	0.99	0.76	0.90	1.00	0.9125	9–11
Trabectedin	0.99	0.75	0.91	1.00	0.9125	9–11
Saquinavir	0.99	0.88	0.97	1.00	0.9100	12–13
Clarithromycin**	0.99	0.70	0.95	1.00	0.9100	12–13
Capreomycin	1.00	0.70	0.89	1.00	0.8975	14–16
Balsalazide	0.93	0.71	0.96	0.94	0.8875	14–16
Sonidegib	0.91	0.72	0.94	0.98	0.8875	14–16

Minocycline	0.73	0.74	0.98	0.89	0.8850	17
Eribulin	0.92	0.71	0.86	1	0.8500	18
Isradipine	0.98	0.72	0.94	1	0.8350	19–20
Cangrelor	0.96	0.70	0.95	1	0.8350	19–20

\*Boceprevir [31] showed in-vitro activity for COVID-19.

\*\*Clarithromycin [32,33] is in clinical trials to treat COVID-19.

Thirteen drugs (Table 6) showed average scores of greater than 0.9. Abacavir and raltegravir (Table 6) showed top average scores of 0.96. Chemical structures of top active compounds are shown in Fig. 3.



**Figure 3.** Chemical structures of the top predicted active for treatment of cytokine storm FDA-approved drugs: (a) mitomycin C; (b) abacavir; (c) raltegravir.

Current use of predicted active drugs is shown in Table 7.

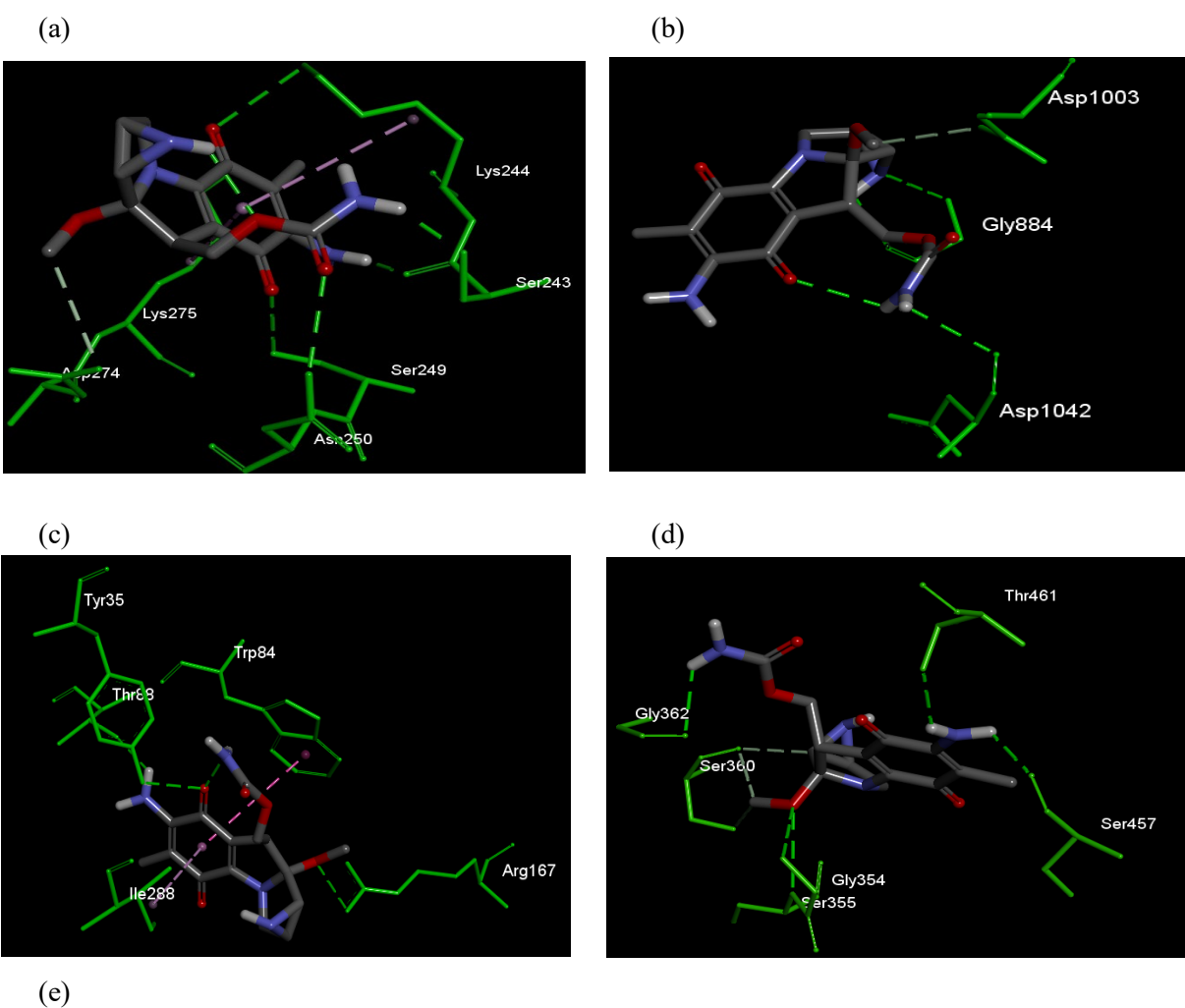
**Table 7.** Current use of predicted active drugs.

ZINCID	Name of Drug	Current Uses
ZINC000004216238	Fludarabine	Purine analogue and antineoplastic agent
ZINC000019632713	Nimodipine	Calcium channel blocker used for high blood pressure
ZINC000028232755	Valrubicin	Used to treat bladder cancer
ZINC000068202099	Sonidegib	A small molecule used to treat cancer
ZINC000026985532	Saquinavir	A peptide used to treat HIV
ZINC000013648755	Saxagliptin	A small molecule/peptide used for diabetes
ZINC000030726187	Mitomycin C	A small molecule used to treat cancer
ZINC000003952881	Balsalazide	Used to treat inflammatory bowel disease
ZINC000002015928	Abacavir	A nucleotide used to treat HIV
ZINC000013831130	Raltegravir	A small molecule used to treat HIV
ZINC000014210457	Boceprevir	A peptide used to treat HCV
ZINC000014879992	Minocycline	A tetracycline antibiotic used for bacterial infections
ZINC000085534098	Clarithromycin	A macrolide antibiotic used to treat bacterial infections

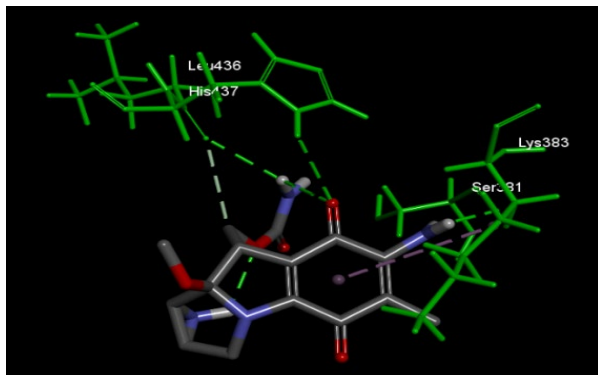
ZINC000150338698	Capreomycin	A macrolide antibiotic used to treat TB
ZINC000049036447	Suvorexant	A small molecule used to treat insomnia
ZINC000100001918	Isradipine	Used to treat high blood pressure
ZINC000003940470	Pomalidomide	A small molecule used to treat cancer
ZINC000150338708	Trabectedin	A small molecule used to treat cancer
ZINC000169344691	Eribulin	A small molecule used to treat cancer
ZINC000085537017	Cangrelor	A nucleotide used to treat cancer

## Docking results

Docking of the top three predicted FDA-approved drugs mitomycin C, abacavir and raltegravir with the protein binding site are shown in Fig. 4 and Fig. 5.

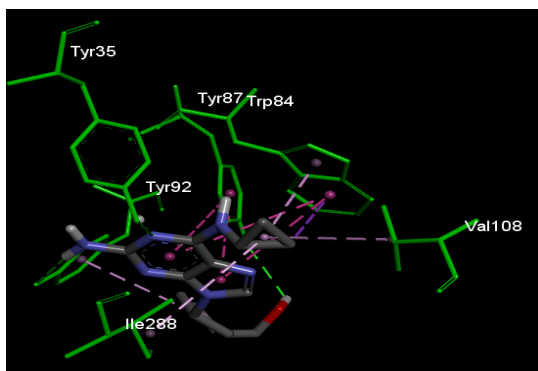




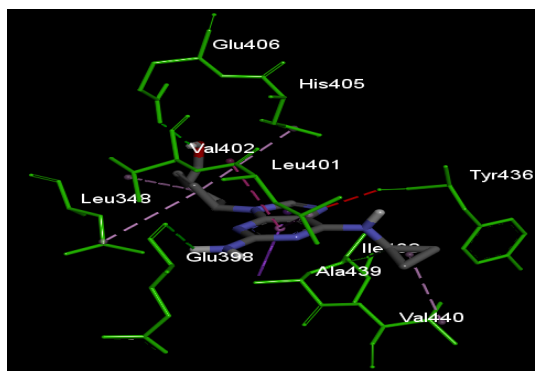


**Figure 4.** Docking of mitomycin C with proteins: (a) AT1R; (b) ADAM17; (c) NF- $\kappa$ B; (d) JAK1; (e) STAT3.

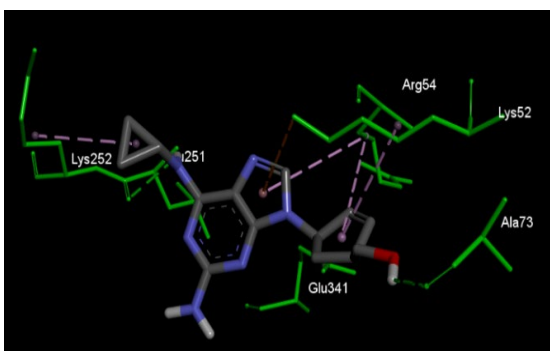
(a)



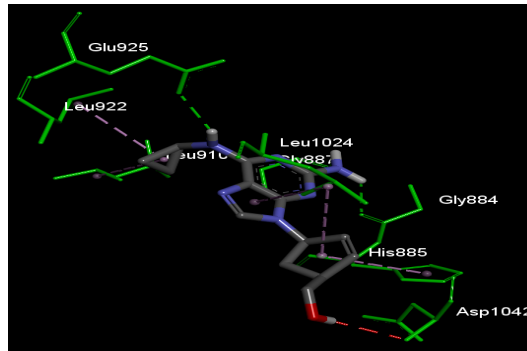
(b)



(c)



(d)

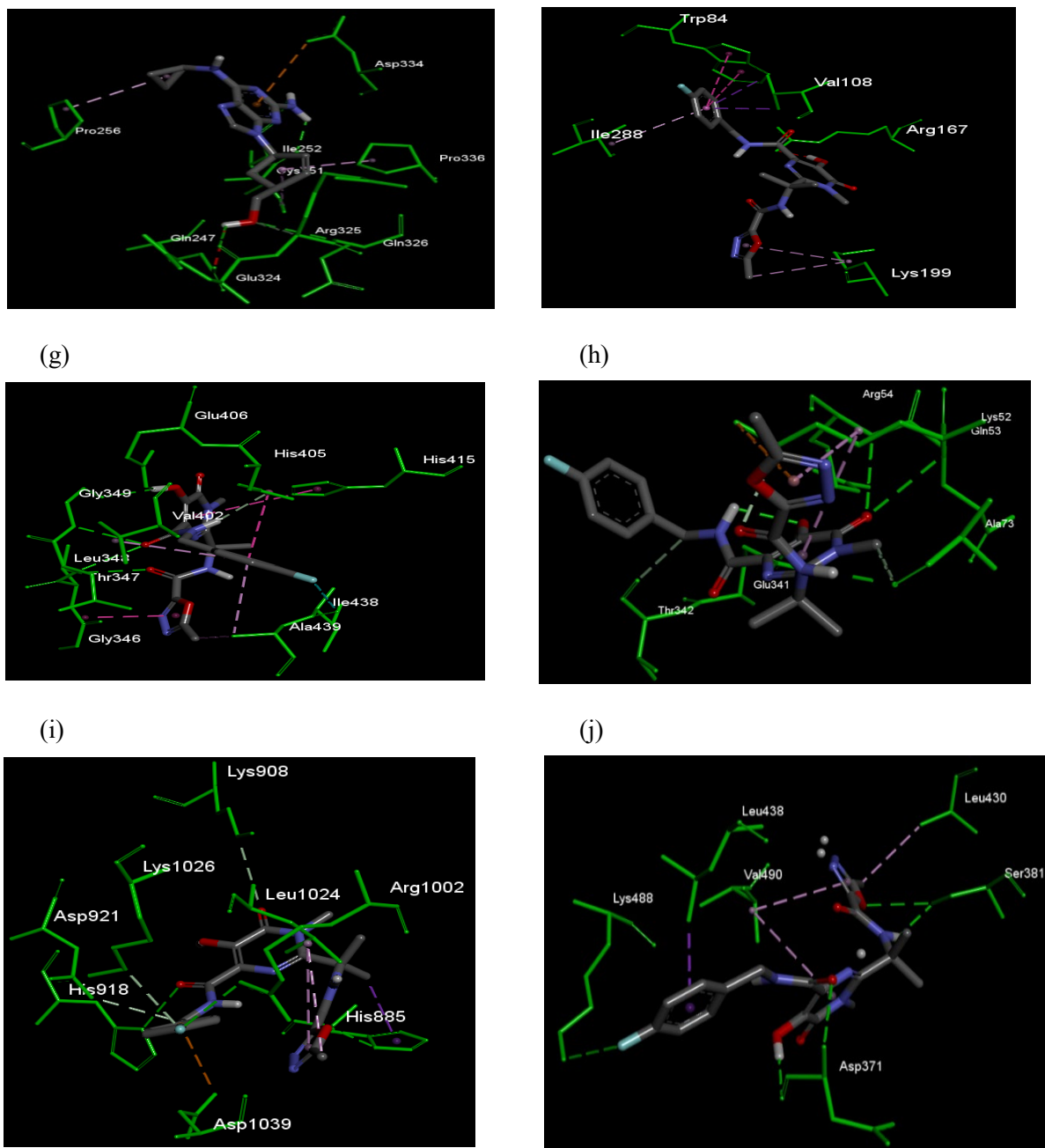


(e)



(f)





**Figure 5.** Docking of abacavir and raltegravir with proteins: (a) AT1R–abacavir; (b) ADAM17–abacavir; (c) NF-κB–abacavir; (d) JAK1–abacavir; (e) STAT3–abacavir; (f) AT1R–raltegravir; (g) ADAM17–raltegravir; (h) NF-κB–raltegravir; (i) JAK1–raltegravir; (j) STAT3–raltegravir.

Mitomycin C docked with all five proteins showed binding affinity ranging from  $-6.0$  to  $-7.5$  kcal/mol. (Table 8).

**Table 8.** Binding free energy for mitomycin C.

<b>Protein</b>	<b>Binding free energy (kcal/mol)</b>
AT1R	-7.5
ADAM17	-7.0
NF- $\kappa$ B	-6.0
JAK1	-7.2
STAT3	-6.5

Abacavir docked with four proteins showed binding free energy ranging from -6.1 to -8.6 kcal/mol (Table 9). Raltegravir docked with four proteins showed binding energy ranging from -7.4 to -9.6 kcal/mol (Table 9) which is considered as reasonable binding affinity.

**Table 9.** Binding free energy for abacavir and raltegravir.

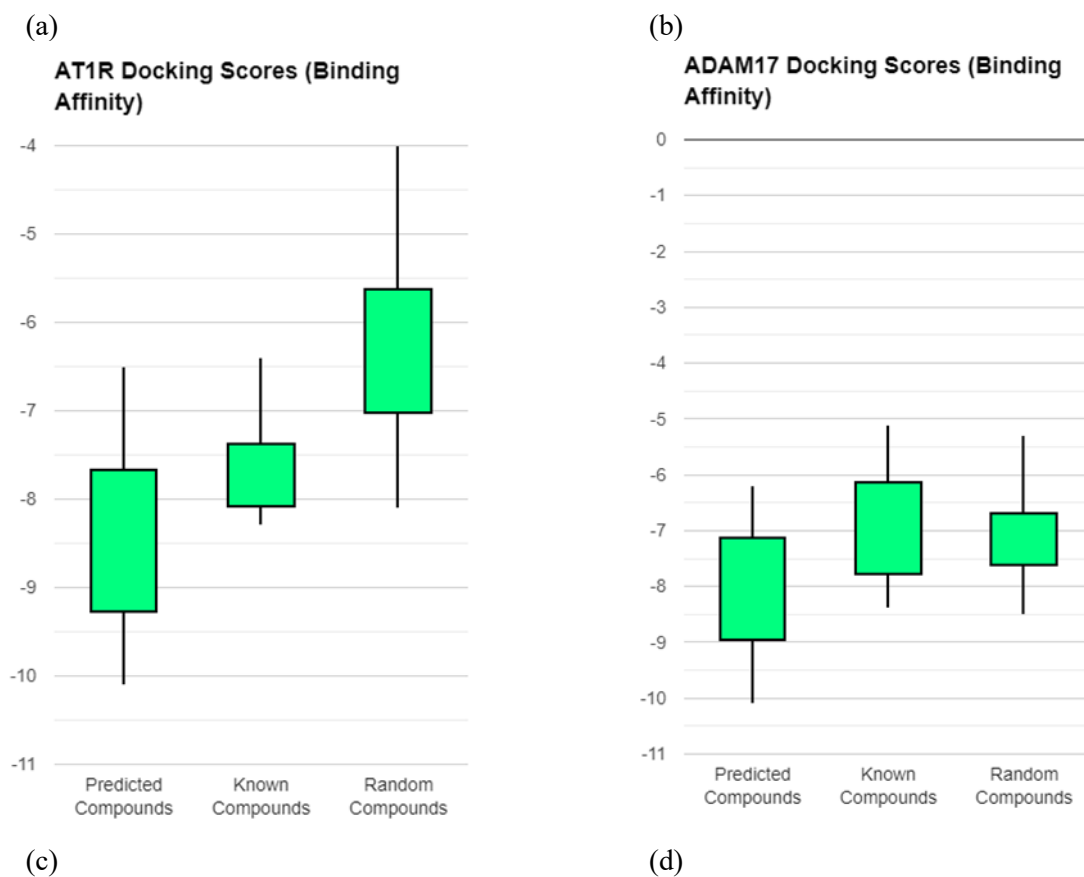
<b>Protein</b>	<b>Binding Free energy with abacavir (kcal/mol)</b>	<b>Binding Free energy with raltegravir (kcal/mol)</b>
AT1R	-7.8	-9.4
ADAM17	-8.6	-9.6
NF- $\kappa$ B	-6.1	-7.4
STAT3	-7.2	-8.3

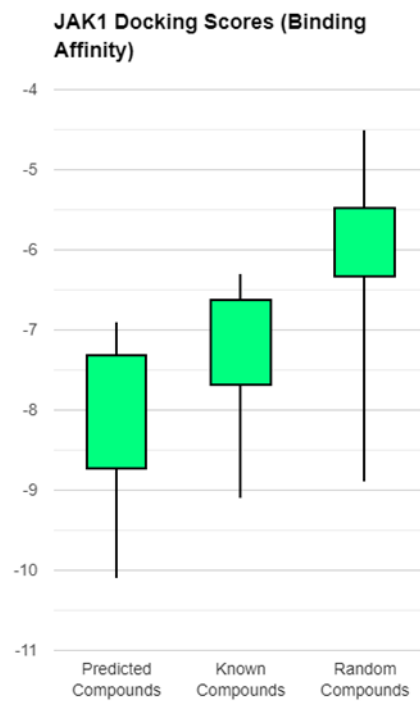
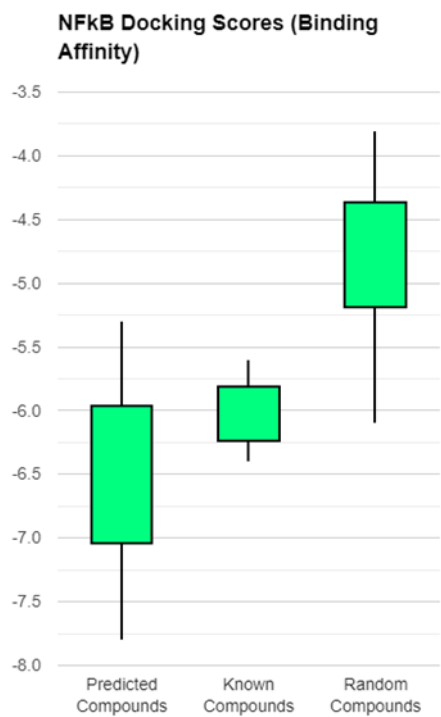
Binding affinities calculated using PyRx docking for the top 12 compounds. Table 10 lists the top predicted active compounds docked for all five proteins.

**Table 10.** Binding free energies of the top 12 ML-predicted FDA-approved drugs docked using PyRx software.

<b>Compound</b>	<b>Binding free energy to</b>				
	<b>AT1R</b>	<b>ADAM17</b>	<b>NF-<math>\kappa</math>B</b>	<b>JAK1</b>	<b>STAT3</b>
Mitomycin C	-9.2	-7.9	-6.1	-7.4	-6.7
Pomalidomide	-8.5	-7.1	-5.9	-8.3	-7.2
Fludarabine	-8.1	-9.1	-6.2	-8.1	-7.5
Sonidegib	-10.1	-10.1	-7.3	-9.1	-8.5
Abacavir	-9.6	-8.5	-7.8	-10.1	-7.9
Raltegravir	-7.8	-7.3	-6.1	-7.7	-7.2
Saxagliptin	-7.3	-8.6	-5.3	-6.9	-5.9
Nimodipine	-7.9	-6.2	-6.4	-8.4	-7.5
Suvorexant	-9.4	-8.9	-7.4	-9.7	-8.3
Boceprevir	-8.6	-9.6	-6.8	-7.4	-7.8
Balsalazide	-7.5	-7.1	-6.0	-7.2	-6.5
Minocycline	-6.5	-7.0	-5.6	-7.2	-6.1

The box plots (Fig 6) demonstrated that the predicted inhibitors had the better binding affinities than known and random compounds. They were constructed using binding affinities obtained from docking for predicted compounds, known inhibitors and control random compounds and are shown in Fig. 6. Docking free energies are listed in Table 11.





(c)



**Figure 6.** Free energies of docking interactions—docking scores—of predicted, known, and random compounds: (a) AT1R; (b) ADAM17; (c) NF- $\kappa$ B; (d) JAK1; and (e) STAT3.

**Table 11.** Summary of binding free energies (kcal/mol) for predicted FDA-approved drugs, known inhibitors and random compounds.

<b>Protein</b>	<b>min</b>	<b>q1</b>	<b>med</b>	<b>q3</b>	<b>max</b>	<b>mean</b>
<b>AT1R</b>						
Predicted	-10.10	-9.30	-8.30	-7.65	-6.50	-8.37
Known	-8.30	-8.10	-7.65	-7.35	-6.40	-7.63
Random	-8.10	-7.05	-6.35	-5.60	-4.00	-6.28
<b>ADAM 17</b>						
Predicted	-10.10	-9.00	-8.20	-7.10	-6.20	-8.10
Known	-8.40	-7.80	-7.20	-6.10	-5.10	-6.91
Random	-8.50	-7.65	-6.80	-6.65	-5.30	-7.08
<b>NF-κB</b>						
Predicted	-7.80	-7.05	-6.15	-5.95	-5.30	-6.40
Known	-6.40	-6.25	-6.00	-5.80	-5.60	-6.00
Random	-6.10	-5.20	-4.50	-4.35	-3.80	-4.70
<b>JAK1</b>						
Predicted	-10.10	-8.75	-7.90	-7.30	-6.90	-8.12
Known	-9.10	-7.70	-7.40	-6.60	-6.30	-7.37
Random	-8.90	-6.35	-5.80	-5.45	-4.50	-6.09
<b>STAT3</b>						
Predicted	-8.50	-7.85	-7.35	-6.60	-5.90	-7.25
Known	-7.00	-6.25	-5.90	-5.75	-5.30	-6.00
Random	-7.20	-6.05	-5.15	-4.85	-4.00	-5.45

The box plots (Fig 6) demonstrates that the predicted inhibitors had an average binding affinity of  $-6.40$  kcal/mol (NF- $\kappa$ B) to  $-8.37$  kcal/mol for AT1R (Fig. 6a), which was better than that of known inhibitors, which had an average of  $-6.0$  kcal/mol for STAT3(Fig. 6e) to  $-7.63$  kcal/mol for AT1R (Fig 6). The control group of random molecules had an average binding affinity of  $-4.78$  kcal/mol for NF- $\kappa$ B (Fig 6c) to  $-7.0$  kcal/mol for ADAM17 (Fig 6b). This confirms that the predicted inhibitors performed statistically better than the control group.

In this project, five crucial proteins—AT1R, ADAM17, NF- $\kappa$ B, JaK1, STAT3—playing the important roles in cytokine production pathway are targeted to predict the best potential drug for treatment of COVID-19 cytokine storm. Number of known inhibitors with reported  $IC_{50}$  values for each protein obtained from PubChem were 1192, 1813, 348, 4596, and 588 respectively. Machine-learning models developed

exhibited an accuracy ranging from 91.5–99.0%, with their AUROC values ranging from 0.98–1.0, which is considered as excellent predictive performance of the models.

The box plots (Fig. 6) show that the predicted active compounds have better binding energies than the already known inhibitors and control set of random compounds. One can see that the predicted inhibitors had an average binding affinity of  $-6.40$  kcal/mol (NF- $\kappa$ B) to  $-8.37$  kcal/mol (AT1R), which was better than that of known inhibitors, which had an average of  $-6.0$  kcal/mol (STAT3) to  $-7.63$  kcal/mol (AT1R). The control group of random molecules had an average binding affinity of  $-4.78$  kcal/mol (NF- $\kappa$ B) to  $-7.0$  kcal/mol (ADAM17). This confirms the predicted inhibitors performed statistically better than the control group.

From the docking results, binding free energy values ranging from  $-6.0$  to  $-9.6$  kcal/mole targeted for five and four proteins confirms that the predicted compounds bind at the active site of the proteins. The amino acids that showed interaction in the docking experiments for the top three drugs mitomycin C, abacavir and raltegravir are listed in Table 12.

**Table 12.** Summary of binding residues involved for each protein with top predicted drugs. Bold are residues that are involved in binding to more than one compound.

Protein binding residues	Mitomycin C	Abacavir	Raltegravir
AT1R	<b>Tyr35, Trp84</b> , Thr88, <b>Arg167, Ile288</b>	<b>Tyr35, Trp84</b> , Tyr87, Tyr92, <b>Ile288</b> , Val108	<b>Trp84</b> , Val108, <b>Arg167</b> , Lys199, <b>Ile288</b>
ADAM17	Gly354, Ser355, Ser360, Gly362, Thr461, Ser457	<b>Leu348</b> , Glu398, <b>Val402</b> , <b>Glu406</b> , <b>His405</b> , Leu401, Try436, <b>Ile438</b> , Ala439, Val440	Gly346, Thr347, <b>Leu348</b> , Gly349, <b>Val402</b> , <b>Glu406</b> , <b>His405</b> , His415, <b>Ile438</b> , Ala439
NF- $\kappa$ B	Ser243, Ser249, Asp250, Asp274, Lys275, Lys244	Arg54, <b>Lys52</b> , <b>Ala73</b> , Lys252, Leu251, <b>Glu341</b>	<b>Lys52</b> , Gln53, Arg54, <b>Ala73</b> , <b>Glu341</b> , Thr342
JAK1	Asp1003, <b>Gly884</b> , <b>Asp1042</b>	Gly887, Leu910, Leu922, Leu1024, <b>Gly884</b> , <b>His885</b> , <b>Asp1042</b>	Asp1039, His918, Asp921, Lys1026, Lys908, Leu1024, Arg1002, <b>His885</b>
STAT3	Lys383, <b>Ser381</b> , Leu436, His437	Gln247, Cys251, Ile252, Pro256, Glu324, Arg325, Asp334, Pro336	Asp371, <b>Ser381</b> , Leu430, Leu438, Lys488, Val490

From Table 12 one can see that AT1R's binding residues—Tyr35, Trp84, Val108, Arg167, and Ile288—are common for mitomycin C, abacavir and raltegravir; ADAM17's binding residues—Glu406, His405, Ile438, and Ala439—are common for abacavir and raltegravir ; NF- $\kappa$ B's binding residues—Lys52, Ala73, and Glu341—are for abacavir and raltegravir; JAK1's binding residues—Asp1042, Gly884, and His885—are common for abacavir and raltegravir; STAT3's binding residue—Ser381—is for abacavir and raltegravir binding to STAT3.

These results support the idea that the compounds are binding at the active site of protein and not at the non-bonding sites, thus proving that these compounds could act specifically on the protein makes them some of the most promising candidates to treat COVID-19.

## **Discussion**

The main goal of this project was to predict the drugs that can target as many as possible cytokine-related genes. We know that there is a number of such genes that are activated in response on SARS-CoV-2 viral proteins. Here we have a problem that is eternal for medicine. We cannot prescribe more drugs than a set that would be tolerated by the organism. Here we selected five genes that have to be targets of inhibiting agent to prevent the cytokine storm or just decrease the immune response to viral agents. Our model predicted drugs that can simultaneously target at least four of such cytokine-related genes. Based on model average prediction scores over 0.6, 45 FDA-approved drugs were predicted active, with 20 drugs having predicted scores greater than 0.8 for 4 proteins (AT1R, ADAM17, NF- $\kappa$ B, STAT3). Eight FDA-approved drugs had predicted scores over 0.85 for all five proteins involved in the mechanism of the cytokine storm (AT1R, ADAM17, NF- $\kappa$ B, JAK1, STAT3). Mitomycin C is the most active drug across all five proteins with an average prediction score of 0.886. Abacavir and raltegravir are the top active compounds for four proteins with average scores of 0.96.

We predicted several drugs that can target simultaneously several proteins in cytokine-storm-related pathway. These may be useful drugs to treat patients because these therapies can fight cytokine storm caused by the virus at multiple points of inhibiting, leading to synergistically effective treatments.



Mitomycin C ranked top with highest average scores for all five proteins suggesting that it possesses all required chemical functional groups with desired spatial arrangements, so that it can interact and bind well with all proteins. Abacavir and raltegravir with excellent predictive scores for four proteins (except JAK1) suggesting that changes in chemical structure of the drugs made the difference in biological activity.

Two of predicted drugs—boceprevir [31] and clarithromycin [32,33] are already tested for COVID-19 treatment.

## Conclusions

Our hypothesis that it is possible to develop a valid predicting machine-learning model to select the drugs that would target multiprotein pathways for treatment of the COVID-19 cytokine storm is confirmed and gave the model accuracy ranging from 91.5–99%, with AUROC ranging from 0.978 to 1.0, considered as excellent predictive performance of the models.

This project not only provides drug candidates that could treat COVID-19, but it also demonstrates the application of predictive models for multi-target drug discovery approach with machine learning.

Future steps for this project would be to confirm the inhibitory activity of the predicted drugs against the target proteins in animal models. The methods of this project could be extended to predictive models for discovering therapeutics for other disease areas, such as chronic inflammatory diseases.

## References

- [1] World Health Organization (WHO). Coronavirus Disease 2019 (COVID-19) Situation Report: Weekly epidemiological Update on COVID-19. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- [2] Johns Hopkins University & Medicine: Coronavirus Resource Center <https://coronavirus.jhu.edu>
- [3] Hsu CY, Lai CC, Yeh YP, Chuane CC, Chen HH. Progression from Pneumonia to ARDS as a Predictor for Fatal COVID-19. *Journal of Infection and Public Health* 2021;14:504-507
- [4] Aslan A, Aslan C, Zolbanin NM and Jafari R. Acute respiratory distress syndrome in COVID-19: possible mechanisms and therapeutic management. *Pneumonia* 2021;13:1-15

- [5] Robinsona PC, Liew DFL, Tannera HL, Graingerf JR, Dwekg RA, Reisleh RB, Steinmani L, Feldmannk M, Hol LP, Hussellf T, Mossm P, Richardsn D, and Zitzmannng N. COVID-19 therapeutics: Challenges and directions for the future. *PNAS* 2022;119:1-10
- [6] Rando HM, Wellhausen N, Ghosh S, Lee AJ, Dattoli AA, Hu F, Byrd JB, Rafizadeh DN, Lordan R, Qi Y, Sun Y, Brueffer C, Field JM, Guebila MB, Jadavji NM, Skelly AN, Ramsundar B, Wang J, Goel RR, Park YS, Boca SM, Gitter A, Greene CS. Identification and Development of Therapeutics for COVID-19. *mSystems* 2021;6:1-52
- [7] Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: Immunity, inflammation and intervention. *Nat Rev Immunol.* 2020; 20(6):363–374. <https://doi.org/10.1038/s41577-020-0311-8>
- [8] Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest.* 2020;130(5):2620–2629. <https://doi.org/10.1172/JCI137244>
- [9] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet.* 2020;395(10223):507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
- [10] Kim JY, Ko JH, Kim Y, Kim YJ, Kim JM, Chung YS, et al. Viral load kinetics of SARS-CoV-2 infection in first two patients in Korea. *J Korean Med Sci.* 2020;35(7):e86. <https://doi.org/10.3346/jkms.2020.35.e86>
- [11] Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci.* 2020;63(3):364–374. <https://doi.org/10.1007/s11427-020-1643-8>
- [12] Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect Dis.* 2020;20(4): 411–412. [https://doi.org/10.1016/S1473-3099\(20\)30113-4](https://doi.org/10.1016/S1473-3099(20)30113-4)
- [13] Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and human-to-human transmission of a novel coronavirus in Vietnam. *N Engl J Med.* 2020; 382(9):872–874. <https://doi.org/10.1056/NEJMc2001272>
- [14] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395(10223):497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- [15] Hirano T, Murakami M. COVID-19: A new virus, but a familiar receptor and cytokine release syndrome. *Immunity.* 2020; 52(5):731–733. <https://doi.org/10.1016/j.immuni.2020.04.003>
- [16] Mahmudpour M, Roozbeh J, Keshavarz M, Farrokhi S, Nabipour I. COVID-19 cytokine storm: The anger of inflammation. *Cytokine.* 2020;133:155151. <https://doi.org/10.1016/j.cyto.2020.155151>
- [17] McGonagle D, Sharif K, O'Regan A, Bridgewood C. The role of cytokines including interleukin-6 in COVID-19 induced pneumonia and macrophage activation syndrome-like disease. *Autoimmun Rev.* 2020;19(6):102537. <https://doi.org/10.1016/j.autrev.2020.102537>
- [18] Hojyo S, Uchida M, Tanaka K, Hasebe R, Tanaka Y, Murakami M, Hirano T. How COVID-19 induces cytokine storm with high mortality. *Inflamm Regen* 2020;40:37 <https://doi.org/10.1186/s41232-020-00146-3>
- [19] Zhang JM, An J. Cytokines, inflammation and pain. *Int Anesthesiol Clin.* 2007;45(2):27–37 <https://doi.org/10.1097/AIA.0b013e318034194e>
- [20] Ragab D, Eldin HS, Taeimah M, Khattab R, Salem R. The COVID-19 cytokine storm; What we know so far. *Front Immunol.* 2020;11:1446. <https://doi.org/10.3389/fimmu.2020.01446>
- [21] Tisoncik JR, Korth MJ, Simmons CP, Farrar J, Martin TR, Katze MG. Into the eye of the cytokine storm. *Microbiol Molecular Biol Rev.* 2012;76:16–32. <https://doi.org/10.1128/MMBR.05015-11>
- [22] Frank E, Hall MA, Whitten IH. The WEKA Workbench. Online Appendix for *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Fourth Edition, 2016; <https://www.cs.waikato.ac.nz/ml/weka/book.html>

- [23] NCI/CADD Group. Online SMILES Translator and Structure File Generator. National Institutes of Health, U.S. Department of Health and Human Services, National Cancer Institute. 21 Apr. 2020; <https://cactus.nci.nih.gov/translate> (Last accessed 25 May 2022).
- [24] Yap CW. PaDEL-descriptor: An open-source software to calculate molecular descriptors and fingerprints. *J Comput Chem*. 2010;32(7):1466–1474. <https://doi.org/10.1002/jcc.21707>
- [25] WEKA 3.8.5 University of Waikato; software download; <https://www.cs.waikato.ac.nz/~ml/weka/>
- [26] (a) Irwin JJ, Soichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45(1):177–182. <https://doi.org/10.1021/ci049714+>. (b) Sterling T, Irwin JJ. ZINC 15--Ligand discovery for everyone. *J Chem Inf Model*. 2015; 55(11):2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>. (c) <https://zinc.docking.org/> (Last accessed 25 May 2022).
- [27] (a) Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol*. 2015; 1263:243–250. [https://doi.org/10.1007/978-1-4939-2269-7\\_19](https://doi.org/10.1007/978-1-4939-2269-7_19); (b) Python Prescription: Virtual Screening Tool. <https://pyrx.sourceforge.io/> (Last accessed 23 May 2022).
- [28] Dassault Systemes, Free Download: BIOVIA Discovery Studio Visualizer. <https://discover.3ds.com/discovery-studio-visualizer-download> (Last accessed 27 May 2022).
- [29] (a) Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28(1):235–242. <https://doi.org/10.1093/nar/28.1.235>; (b) RCSB PDB: Protein Data Bank. <http://www.rcsb.org/> (Last accessed 23 May 2022).
- [30] Mukund V, Behera SK, Alam A, Nagaraju GP. Molecular docking analysis of nuclear factor- $\kappa$ B and genistein interaction in the context of breast cancer. *Bioinformation*. 2019;15(1):11–17. <https://doi.org/10.6026/97320630015011>
- [31] Fu L, Ye F, Feng Y, Yu F, Wang Q, Wu Y, et al. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat Commun*. 2020;11(1):4417. <https://doi.org/10.1038/s41467-020-18233-x>
- [32] Tsiakos K, Tsakiris A, Tsibris G, Voutsinas PM, Panagopoulos P, Kosmidou M, et al. Early start of oral clarithromycin is associated with better outcome in COVID-19 of moderate severity: The ACHIEVE Open-Label Single-Arm Trial. *Infect Dis Ther*. 2021;10(4):2333–2351. <https://doi.org/10.1007/s40121-021-00505-8>
- [33] Yamamoto K, Hosogaya N, Sakamoto N, Yoshida H, Ishii H, Yatera K, Izumikawa K, Yanagihara K, Mukae H. Efficacy of clarithromycin in patients with mild COVID 19 pneumonia not receiving oxygen administration: Protocol for an exploratory, multicentre, open-label, randomised controlled trial (CAME COVID 19 study). *BMJ Open*. 2021;11(9):e053325. <https://doi.org/10.1136/bmjopen-2021-053325>