Artigo

MACHINE LEARNING-BASED VIRTUAL SCREENING, MOLECULAR DOCKING, DRUG-LIKENESS, PHARMACOKINETICS AND TOXICITY ANALYSES TO IDENTIFY NEW NATURAL INHIBITORS OF THE GLYCOPROTEIN SPIKE (S1) OF SARS-CoV-2

Alexandre de F. Cobre[a,] **Beatriz Böger**[a,] **Mariana M. Fachi**[a,] **Carlos A. Ehrenfried**[a,] **Dile P. Stremel**[b,] **Eduardo B. De Melo**[c,] **Fernanda S. Tonin**[d,] **and Roberto Pontarolo**[a,]

[a]Departamento de Farmácia, Universidade Federal do Paraná, 80210-170 Curitiba – PR, Brasil

[b]Departamento de Engenharia e Tecnologia Florestal, Universidade Federal do Paraná, 80210-170 Curitiba – PR, Brasil

[c]Departamento de Farmácia, Universidade Estadual do Oeste do Paraná, 85819-110 Cascavel – PR, Brasil

[d]H&TRC- Health & Technology Research Center, ESTeSL, Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, 1990-096 Lisbon, Portugal

To identify natural bioactive compounds (NBCs) as potential inhibitors of spike (S1) by means of *in silico* assays. NBCs with previously proven biological *in vitro* activity were obtained from the ZINC database and analyzed through virtual screening and molecular docking to identify those with higher affinity to the spike protein. Eight machine learning models were used to validate the results: Principal Component Analysis (PCA), Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Partial Least Squares-Discriminant Analysis (PLS-DA), Gradient Boosted Tree Discriminant Analysis (XGBoostDA), Soft Independent Modelling of Class Analogies (SIMCA) and Logistic Regression Discriminate Analysis (LREG). Selected NBCs were submitted to drug-likeness prediction using Lipinski's and Veber's rule of five. A prediction of pharmacokinetic parameters and toxicity was also performed (ADMET). Antivirals currently used for COVID-19 (remdesivir and molnupiravir) were used as a comparator. A total of 170,906 compounds were analyzed. Of these, 34 showed greater affinity with the S1 (affinity energy < -7 kcal mol^{-1}). Most of these compounds belonged to the class of coumarins (benzopyrones), presenting a benzene ring fused to a lactone (group of heterosides). The PLS-DA model was able to reproduce the results of the virtual screening and molecular docking (accuracy of 97.0%). Of the 34 compounds, only NBC5 (feselol), NBC14, NBC15 and NBC27 had better results in ADMET predictions. These had similar binding affinity to S1 when compared to remdesivir and molnupirvir. Feselol and three other NBCs were the most promising candidates for treating COVID-19. *In vitro* and *in vivo* studies are needed to confirm these findings.

Keywords: *in silico*; COVID-19; spike glycoprotein; treatment.

INTRODUCTION

Several attempts have been made to manage the coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), including the recent vaccination programs rolled out worldwide.[1] However, there is still a need to identify effective treatments, considering that to this day there is not even a drug with proven efficacy data. Current treatments are limited to treating symptoms only, they are palliative treatments.[2-8]

The scarce knowledge of the pathogenesis and immunological peculiarity of SARS-CoV-2, especially regarding the interaction between viral antigens and human receptors and the triggering of cytokine storms, poses additional challenges for the development of successful treatments.[5,9] Although details of the cellular responses to this virus are unknown, a probable course of events can be postulated based on past studies with SARS-CoV-2.[5,9,10] Infections are initiated by the virus binding to the angiotensin-converting enzyme receptor-2 (ACE-2) cell-surface receptors, which is followed by fusion of the virus and cell membranes to release the virus RNA genome into the host cell through receptor-mediated endocytosis.[5,6] Both receptor binding and membrane fusion activities are mediated by the 'spike glycoprotein' of the virus.[10] As with other class-I membrane-fusion proteins (alpha-helical), the spike protein is post-translationally cleaved, in this case by furin, into the S1 and S2 components that

remain associated after cleavage. Each S1 component consists of two large domains, the N-terminal domain (NTD) and the receptor-binding domain (RBD).[11] The interaction between viral antigens and host immune cells finally results in the induction of pro-inflammatory responses that trigger vasodilation, increased vascular permeability and the accumulation of humoral factors, causing fever and interrupting gas exchange (*i.e.*, respiratory distress).[9]

Given the global emergency caused by COVID-19, there is great interest in drug repurposing (*i.e.*, drug repositioning or rediscovery) to accelerate the identification of drugs that can cure or prevent this disease.[1,8] One of the key drivers for the repositioning of drugs is the serendipitous discovery of pharmacological activity on new targets, which would then suggest a possible new indication of use. High-throughput screening of potential compounds available in databases is an emerging strategy that has already supported the discovering of new indications for marketed drugs (*e.g.*, lopinavir/ritonavir for HIV) and the development of additional therapeutic options against Ebola, hepatitis C and Zika virus infection.[10,12]

To accelerate the drug discovery process, several open source *in silico* platforms are available in the literature for the prediction of pharmacokinetic parameters (absorption, distribution, metabolism and elimination), toxicity and drug-likeness. These platforms were built using machine learning models. For example, the SwissADME *in silico* platform, built using the Support Vector Machine (SVM) machine learning algorithm, allows for drug-likeness predictions (for example, Lipinski's rule), and pharmacokinetic parameters, with an

*e-mail: beatrizboger@gmail.com

accuracy between 0.77 to 0.83.[13] The other *in silico* platform (also available online) is ADMETlab 2.0, which uses the algorithm of Artificial Neural Networks (deep learning) to predict pharmacokinetic and toxicity parameters, with an accuracy of > 85%.[14] The Pred-hERG platform uses the SVM algorithm to predict the probability of cardiac toxicity with an accuracy of 84%.[15] Finally, there is the PROTOX-II platform that combines three algorithms, SVM, Random Forest and Naive Bayes to predict different types of toxicity (organ-specific toxicity, acute toxicity and chronic toxicity), whose accuracy varies between 82-93%.[16] As noted, these platforms are based on mathematical models, and do not guarantee 100% certainty, so their interpretation must be done with caution.

Several computer-aided drug design/discovery (CADD) studies for COVID-19 are available in the literature, with remdesivir, hydroxychloroquine, and some anti-HIV/herpes, anti-inflammatories and immunomodulators being the most repositioned against COVID-19, and studies of their effectiveness are still ongoing.[17,18] Despite this great diversity, most of these studies were conducted using a small number of ligands (usually n < 100 ligands), which may exclude other bioactive compounds that may also be promising against COVID-19.[19-24] Thus, our aim was to assess the potential effects of over 170,000 natural bioactive compounds (obtained from the ZINC database using the following filters: substances/*in vitro*/biogenic) as inhibitors of the spike protein (S1) of COVID-19 by means of *in silico* assays (virtual screening based on the ligand structure, molecular docking, drug-likeness, pharmacokinetics and toxicity) combined with machine learning.

MATERIAL AND METHODS

Target protein selection

The structure of the spike glycoprotein (S1) (determined by electron microscopy) was obtained from the Research Collaboratory for Structural Bioinformatics (RCSB PDB) public database, where three-dimensional structures of macromolecules, including all SARS-CoV-2 proteins, are available. The PDB assists in the acquisition of 3D structures of therapeutic targets, which enables CADD studies to be conducted.[25] The S1 target was selected because of its SARS-CoV-2-RBD fragment that interacts with human ACE-2, which is responsible for the recognition and penetration of the virus in the lung cells.[26]

Several S1 glycoproteins obtained from the RCSB PDB database (PDB ID: 6M17, 7BZ5, 6VW1, 6LVN, 6VXX, 6YLA, 6LXT, SLGZ and 6M0J) were pre-evaluated, considering: (i) the result of the validation percentage ranking, (ii) the availability of information on the RBD with human ACE-2, (iii) the presence of a complete amino acid sequence and (iv) availability of the published article.

Protein target preparation

The 3D structure of the S1 glycoprotein (PDB ID: 6M17), in PDB format, was prepared in the AutoDock tools software, where (i) all water molecules were removed to avoid steric hindrance at the time of docking with the ligands, (ii) the polar hydrogens were added, and (iii) the Kolman charges were included.[27] After preparation, the molecule file was converted to PDBQT format, which is a suitable format for performing virtual screening and molecular docking. Discovery Studio Visualizer software was used to visualize the prepared protein structure.[26]

Collection and preparation of ligands

A total of 170,906 commercially available natural bioactive compounds (NBC) were obtained from the ZINC15 database. Of the 171,000 compounds obtained, we selected only those that were commercially available, which were a total of 72,885 compounds. These 72,885 commercially available compounds were the ones that were used for the *in silico* analysis. In the ZINC database, the NBC were obtained using the following filters: substances/for sale/*in vitro*/biogenic.[28] These are compounds that have shown biological activity against other diseases in *in vitro* studies. These compounds were obtained in 3D structure in a mol2 format file, which is recognized in most *in silico* study software.

All ligands were prepared in PyRx virtual screening tolls using two steps.[29] The first step was the minimization of the energy of each ligand using the following parameters: (i) force field: UFF; (ii) optimization algorithm: conjugate gradients; (iii) total number of steps: 200; (iv) number of steps for update: 1; stop if energy difference is less than 0.1, in order to obtain its most stable conformation, which could be used to interact with the target protein. The second step was the conversion of the format of the binder files, from mol2 to the PDBQT format, using the AutoDock Vina software, which is coupled with the PyRx virtual screening tolls software.

Virtual screening and molecular docking

Before performing the virtual screening (PyRx virtual screening tolls), the coordinates of the center X, Y, Z and the size (angstrom) of X, Y, Z of the grid box and the exhaustiveness of AutoDock Tools and AutoDock Vina were adjusted. The most promising bioactive compounds were those that had both lower protein and ligand binding energy values.[30]

The ligands that had the highest binding affinity for spike protein (S1) identified in the virtual screening were docked using AutoDock Tools 4.2.0 and AutoDock Vina software.[27,31,32] In this study, the protein and ligand binding energy results were compared with the root-mean-square deviation of atomic positions (RMSD) values, aiming to validate the results obtained in the virtual screening. RMSD measures the quality of different conformations (poses) of a ligand that bind to a single target protein.[33] In general, RMSD values less than 2 are considered acceptable.[27] In our study, docking modeling (AutoDock Tools and AutoDock Vina) was done using the random search function called genetic algorithm, and the scoring function, sum of parameterized energies. As the molecular docking results for each ligand show nine conformations (poses) that interacted with the molecular target (in this case the spike protein), the analysis of ligand-molecular target interactions was performed only using the conformation of ligand with the highest stability, that is, conformation with RMSD values < 2.[27]

Machine learning

In order to reproduce the results of virtual screening and molecular docking, the following machine learning (ML) algorithms were used to classify molecules with higher affinity with the spike (binary classification) and those with lower affinity: Principal Component Analysis (PCA), Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Partial Least Squares-Discriminant Analysis (PLS-DA), Gradient Boosted Tree Discriminant Analysis (XGBoostDA), Soft Independent Modelling of Class Analogies (SIMCA) and Logistic Regression Discriminate Analysis (LREG).[34,35]

Different physicochemical descriptors of the lipid solubility and water solubility of the compounds with high and low affinity with S1 were used as predictors for classification by the ML algorithms (Table 1S in the supplementary material). These descriptors were obtained from the SwissADME web server.[27]

The performance of the ML algorithms was evaluated using the following metrics: sensitivity (Equation 1), specificity (Equation 2), accuracy (Equation 3) and area under the ROC curve (receiver operating characteristic).[34] It is important to highlight that for the development of the ML algorithms, the physicochemical parameters, hydrosolubility and liposolubility were used. Decoy analysis was not performed.

$$\text{Sensitivity} = TP / (TP + FN) \qquad (1)$$
$$\text{Specificity} = TN / (TN + FP) \qquad (2)$$
$$\text{Accuracy} = (TP+TN) / (TP + TN + FP + FN) \qquad (3)$$

where: TP: true positive; TN: true negative; FP: false positive; FN: false negative.

Evaluation of the influence of the stereoisomeric, tautomeric and protonation state at physiological pH on protein-binding affinity

Considering that the drug's stereoisomeric and tautomeric states and its protonation state at physiological pH (pH = 7.4) are paramount for protein-binding affinity, an in-depth study was performed on the 34 ligands that had the highest affinity for the spike protein (S1) obtained from the virtual screening results.[36,37] The 3D structures of molecules in the protonated state at physiological pH, and of all stereoisomers and tautomers were automatically predicted using Marvin Sketch software. All converted chemical structures were also analyzed by molecular docking using AutoDock Tools 2.4.0, and the results were compared with those from the docking of the leader molecule.

Drug-likeness and pharmacokinetics

Drug-likeness is an important parameter that should be fulfilled by a synthetic molecule or natural product in order to be approved for use in clinical trials.[38] This criteria was evaluated in all natural products that showed a higher binding affinity with the S1 glycoprotein according to the rule of five of Lipinski and Veber.[39] In this study, the drug-likeness prediction was made using the SwissADME *in silico* platform which is based on the Support Vector Machine (SVM) algorithm with a predictive accuracy between 77-83%. The consensus between rules was used to consider a natural product as having drug-likeness characteristics.

For the natural compounds that proved to be drug-like, the pharmacokinetic parameters were predicted using the ADMETlab 2.0 (version 2021) *in silico* (online) platform. In the ADMETlab 2.0 platform, predictions are made by an Artificial Neural Networks (deep learning) model with accuracy greater than 85%. Thus, the following pharmacokinetic parameters were predicted: (i) Absorption: human intestinal absorption, transport by P-glycoprotein and bioavailability; (ii) Distribution: plasma protein binding, volume distribution (L kg^{-1}) and blood-brain barrier; (iii) Metabolism: biotransformation by cytochrome P450 enzymes (CYP1A2, CYP3A4, CYP2C9, CYP2C19 and CYP2D6); and (iv) Elimination: half life time ($T_{1/2}$) and clearance rate (mL/min/kg).[14]

Acute and chronic toxicity

For the natural compounds with better drug-likeness and pharmacokinetic predictions, the following types of toxicity were predicted on the Pred-hERG and ProTox-II online platforms. For Pred-hERG platform, these predictions are made using the Support Vector Machine (SVM) algorithm with an accuracy of 84%. For ProTox-II, these predictions are made using the following machine learning models: Suport Vector Machine, Random Forest and Naive Bayes with accuracy between 82-93%. For cardiac toxicity, the Pred-hERG platform was used, while ProTox-II was used for the remaining types of toxicity. Thus the following types of toxicity were predicted: (i) Organ-specific toxicity: hepatotoxicity and cardiac toxicity; (ii) Toxicity endpoint: carcinogenicity, immunotoxicity, mutagenicity and cytotoxicity; (iii) Stress response pathway: aryl hydrocarbon receptor (AhR), androgen receptor (AR), androgen receptor ligand binding domain (AR-LBD), peroxisome proliferator activated gamma receptor (PPAR-Gamma), estrogen receptor alpha (ER), androgen receptor ligand binding domain (AR-LBD) and aromatase; (iv) Nuclear receptor signaling pathways: ATPase family AAA domain containing protein 5 (ATAD5), Nuclear factor (erythroid-derived 2)-like 2/antioxidant responsive element (nrf2/ARE), phosphoprotein (tumor suppressor) p53, mitochondrial membrane potential (MMP) and heat shock factor response element (HSE).[15,16]

RESULTS

In this study, the spike protein was found to be complexed with ACE-2. PDB ID: 6M17 (resolution 2.9Å) was the only structure selected; given its completeness with the three above mentioned criteria (material and methods section).[26] This complex was formed by 22 peptide chains, alphabetically encoded from A-V. The "E" chain was selected for analysis because it corresponds to the RBD portion of the spike S1 of SARS-CoV-2 (Figure 1S, supplementary material).
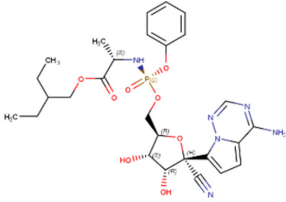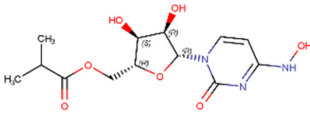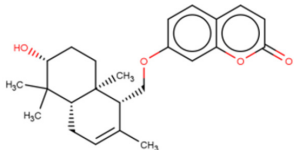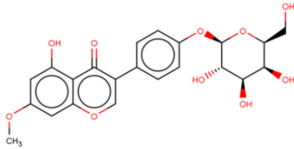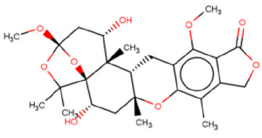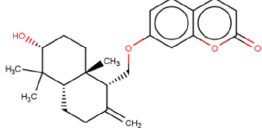
Virtual screening and molecular docking

The virtual screening and molecular docking were performed using the following parameters: (i) exhaustiveness of 8, (ii) coordinates of the center of the grid-box were optimized at x = 178.9, y = 109.6 and z = 260.1 and the grid-box size was optimized at x = 48.3 Å, y = 42.6 Å, z = 56.5Å. Table 2S (see in supplementary material) shows the results of the virtual screening and molecular docking of the 170,906 ligands (bioactive natural compounds) and S1 protein. A total of 34 compounds showed greater affinity with this glycoprotein (Table 2S in supplementary material). Of the 34 compounds, only 4 compounds, BNC5 (feselol), BN14, BN15 and BN27 showed better pharmacokinetic and toxicity results (Table 1). In terms of structural similarities, the four compounds belong to the class of coumarins (benzopyrones), as they have a benzene ring fused to a lactone (heterosides). Another structural similarity is that the four compounds present the benzopyrone ring that forms an ether bond with a polyhydroxylated cyclic alcohol group. Furthermore, these four compounds had similar binding affinity to spike (S1) with the antiviral drugs currently used in the treatment of COVID-19, namely, remdesivir and molnupinavir (Table 1). The four natural bioactive compounds (BNC5, BNC14, BN15 and BN27) and the standard antivirals (remdesivir and molnupinavir) both had affinity to block the following amino acid residues from the S1 region of the spike protein: R454, W436, N437, F464, E516, G482, F456, F374, S373, T430, R509 and V367 (Table 1), and this is illustrated in Figure 2S (supplementary material). The most important chemical interactions were hydrogen bonds (Table 1).

Machine learning

As only 34 chemical compounds had the highest affinity for the spike protein (see Table 2S), the machine learning models (including PCA) were built using these 34 compounds with the highest affinity for the spike protein. Still in the machine learning model, we also included 34 compounds that had lower affinity for the spike protein. The reason for choosing the 34 compounds with lower affinity for

Table 1. Comparison of the four most promising natural bioactive compounds against spike (S1) with the antiviral drugs currently used to treat COVID-19 (Remdesivir and Molnupiravir)

Name/ZINC ID	Molar weight (g mol^{-1})	Chemical structure	Affinity energy (kcal mol^{-1})	Amino acid	Chemical interations
Remdesivir*	602.6		-8.5	F456, R454, W436, N437, R509, V367, F464, E516	Hydrogen bounds, Pi-alkyl, Pi-pi T-shaped, Carbon-hydrogen bound
Molnupiravir*	329.31		-8.4	G482, F456, F374, S373, T430, R509, V367	Hydrogen bounds, Pi-alkyl, Pi-pi T-shaped, Carbon-hydrogen bound
BNC5 (ZINC000000704424)	382.5		-8.5	G482, F456, R454	Hydrogen bounds
BNC14 (ZINC000008662732)	446.4		-8.7	D364, V367, L368, W436, N437, N343, F374, S373	Hydrogen bounds, Pi-alkyl, Pi-pi T-shaped, Carbon-hydrogen bound
BNC15 (ZINC000013374469)	490.5		-8.4	T430, R355, F464, E516	Hydrogen bounds, Pi-pi T-shaped
BNC27 (ZINC000003022621)	382.5		-8.2	R509, V367	Hydrogen bounds, Pi-alkyl

*These compounds are antiviral drugs traditionally used for the treatment of COVID-19. In this study, they were used as a control group.

the spike protein was in order to balance the number of the two groups of compounds. The median values of the descriptors used for the implementation of ML models for classifying high and low compounds with S1 are summarized in Table 2.

The PCA model was able to differentiate between natural compounds (samples) with higher and lower affinity with the S1 protein of SARS-CoV-2 and to reproduce the results obtained by virtual screening and molecular docking (Figure 1(A)).

The leverage *versus* student residuals graph was built to detect outliers. In this graph, a sample is considered outliers, if it presents simultaneously high values of leverage and student residuals. Thus, sample 48, despite having a higher leverage value, cannot be considered outlier because it is within ± 2.5 standard deviations (Figure 1(B)).

For the training and testing of the classification models (ANN, DT, KNN, PLS-DA, LDA and SIMCA), two latent variables were selected, as they presented smaller calibration errors (RMSEC) and cross validation errors (RMSECV) (Figure 1(C)). Table 3 shows the performance evaluation of the seven machine learning models (ANN, DT, KNN, PLS-DA, LDA and SIMCA). The PLS-DA model had the best performance for the classification of compounds of greater and lower affinity with S1 (higher accuracy, sensitivity and specificity

values) (Table 3). The most important variables (physicochemical properties) in the prediction of high- and low-affinity compounds with the spike protein (S1) of SARS-CoV-2 by the PLS-DA model is illustrated in Figure 3S (see in supplementary material).

Drug-likeness

According to Lipinski's rule,[39] a molecule is considered to have drug-likeness if it meets at least three of the four established criteria; whereas, by Veber's rule, the drug should meet all three pre-defined criteria. In our study, consensus between the two rules was used to define a molecule as having drug-likeness. From the 34 natural products, 24 (70.5%) had drug-likeness characteristics. Thus, the remaining 10 compounds were not considered for further studies.

Pharmacokinetics

Table 4S shows the pharmacokinetic analysis (absorption, distribution, metabolism and excretion) of the 24 natural compounds that simultaneously showed greater affinity with spike S1 and presented drug-likeness characteristics. Of these, 21 (80.7%) had a significant probability of human intestinal absorption and bioavailability of 20-30%. Although all compounds presented distribution volume values

Table 2. Physicochemical, liposolubility and water solubility descriptors of natural products and their affinity with the S1 glycoprotein of SARS-CoV-2. These descriptors were predicted by the SwissADME online platform

Descriptors	NP with low affinity for spike - S1 (n = 34)			NP with high affinity for spike - S1 (n = 34)			p-value*
	Median	IQ1	IQ3	Median	IQ1	IQ3	
Physicochemical	…..	…..	…..	…..	…..	…..	…..
Molecular weight (g mol⁻¹)	106.17	83.62	140.60	444.51	375.46	470.62	0.000
Number of heavy atoms	6.50	5.00	9.00	31.50	28.00	34.25	0.000
Number of aromatic heavy atoms	0.00	0.00	0.00	12.00	3.75	16.00	0.000
Carbon fraction sp³ hybridization	0.85	0.67	1.00	0.44	0.24	0.745	0.000
Number of rotatable bonds	2.00	1.00	3.00	3.00	1.75	4.25	0.041
Number of hydrogen-bond acceptors	1.50	1.00	3.00	6.00	4.00	9.25	0.000
Number of hydrogen-bond donors	1.00	0.00	2.00	2.00	1.00	5.00	0.003
Molar refractivity	28.70	22.01	38.37	111.73	102.45	125.46	0.000
Topological polar surface area (Å²)	46.84	20.23	67.46	82.56	57.27	155.14	0.000
Liposolubility	…..	…..	…..	…..	…..	…..	…..
Log Po/w (iLOGP)	1.37	0.67	1.65	3.18	2.80	3.69	0.000
Log Po/w (XLOGP3)	0.06	-1.23	1.31	3.35	1.90	4.53	0.000
Log Po/w (WLOGP)	0.30	-0.65	1.38	3.51	1.71	4.98	0.000
Log Po/w (MLOGP)	-0.14	-1.13	1.09	2.65	-0.40	3.83	0.000
Log Po/w (SILICOS-IT)	0.35	-0.65	1.44	3.37	1.92	4.87	0.000
Consensus Log Po/w	0.09	-0.56	1.51	3.08	1.61	4.48	0.000
Water solubility	…..	…..	…..	…..	…..	…..	…..
Log S (ESOL)	-0.46	-1.27	0.55	-4.38	-5.28	-3.87	0.000
Solubility (mol L⁻¹)	0.35	0.05	3.58	0.00	0.00	0.00	0.000
Log S (Ali)	-0.41	-1.71	0.72	-4.78	-5.75	-4.23	0.000
Solubility (mol L⁻¹)	0.39	0.02	5.44	0.00	0.00	0.00	0.000
Log S (SILICOS-IT)	-0.24	-1.53	0.18	-4.55	-6.47	-3.70	0.000
Solubility (mol L⁻¹)	0.58	0.02	1.54	0.00	0.00	0.00	0.000

*Mann Whitney test. Values less than 0.05 indicate that there was a difference between low and high affinity compounds with spike S1. (NP): natural product.

within the recommended range (0.04-20 L kg⁻¹), compounds NBC3, NBC5, NBC12, NBC14, NBC15, NBC21, NBC27 and NBC33 were the only ones less likely to cross the human blood-brain barrier (*i.e.*, avoiding central nervous system toxicity). The fraction of the molecules of NBC3, NBC5, NBC12, NBC14, NBC15, NBC21, NBC27 and NBC33 that would be transported by plasma proteins during the distribution process was estimated to be within the recommended range (< 90%). Regarding the fraction of molecules not bound to plasma proteins, only six compounds (NBC3, NBC5, NBC12, NBC14, NBC15 and NBC33) were within the recommended range (Fu > 5%).[14]

According to the results of predictions made on the ADMETlab 2.0 *in silico* (online) platform shown in Table 4S, the compounds NBC3, NBC5, NBC12, NBC14, NBC15, NBC27 and NBC33 were more likely to act as substrates of the cytochrome P-450 enzymes (CYP1A2, CYP3A4, CYP2C9, CYP2C19 and CYP2D6), meaning that they can be biotransformed into soluble metabolites, which are easily eliminated by the body. Only the NBC5, NBC12, NBC14, NBC15, NBC27 and NBC34 compounds had clearance values (> 5 mL/min/kg) and an elimination half-life (T₁/₂ > 0.5 h) within the recommended ranges (*i.e.*, acceptable pharmacokinetic properties).

Acute and chronic toxicity

Given the pharmacokinetic results, compounds NBC5, NBC12, NBC14, NBC15, NBC27 and NBC33 were the only ones submitted for the toxicity predictions. NBC12 showed a high probability of causing two types of toxicity via nuclear receptor signalling pathways

(estrogen receptor alpha and peroxisome proliferator activated receptor gamma) and stress response pathways (mitochondrial membrane potential and phosphoprotein tumor suppressor p53. NBC33 was more likely to cause carcinogenicity, mutagenicity and toxicity by stress response pathways [Nuclear factor (erythroid-derived 2)-like 2/antioxidant responsive element (nrf2/ARE) and Heat shock factor response element (HSE)]. The compounds NBC5, NBC14, NBC15 and NBC27 were the only ones that did not show a probability of causing any type of the evaluated toxicities (see Table 5S in the supplementary material).

The predicted lethal dose (acute toxicity dose) of NBC5, NBC14, NBC15 and NBC27 was estimated to be 3200 mg kg⁻¹, 5000 mg kg⁻¹, 3000 mg kg⁻¹ and 3200 mg kg⁻¹, respectively (Table 1S in the supplementary material), showing that these compounds have low toxicity (*i.e.*, the LD50 is greater than 500 mg kg⁻¹), which means that they are promising candidates for evaluation in preclinical trials.

The poses (most stable conformations resulting from molecular docking results) of NBC5, NBC14, NBC15 and NBC27 bound to the spike S1 protein of SARS-CoV-2 are shown in Figure 2. The hydrogen bond-type bonds were the most important in the ligand-target protein interaction (Figure 3).

These four compounds (NBC5, NBC14, NBC15 and NBC27) are commercially available from AKos Consulting & Solutions (Germany) and Beijing FutureCeed Biotechnology Co., Ltd (China), which means they can be easily obtained for performing *in vitro* and *in vivo* assays. Company data are shown in Table 6S in supplementary material.
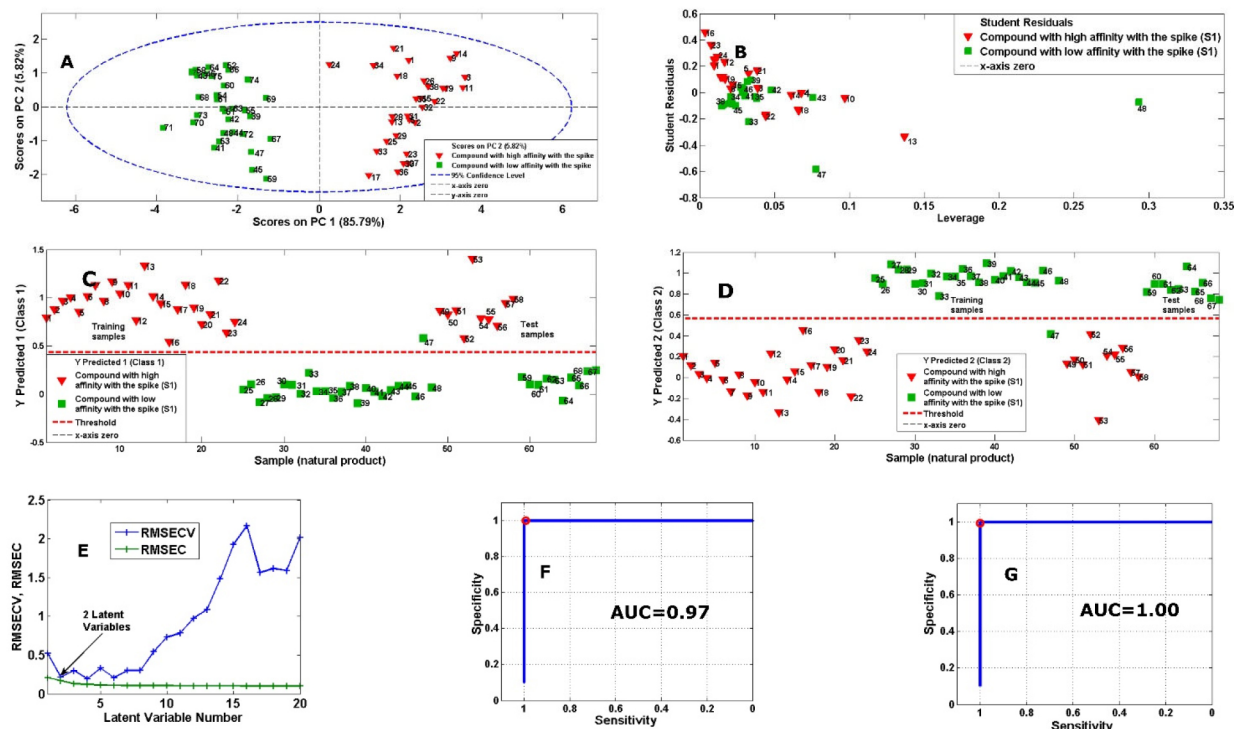
Figure 1. Machine learning models. In (A), the PCA model is shown in which it was possible to discriminate high affinity natural bioactive compounds (BNC) (represented by red coloured triangles) and low affinity (represented by green coloured squares) with the protein spike (S1) of SARS-CoV-2. In (B) is the graph of leverage versus student residues for sample detection (BNC) outliers. In this graph, a sample is considered outliers, if it presents simultaneously high values of leverage and student residuals. Thus, sample 48, despite having a higher leverage value, cannot be considered outlier because it is within ± 2.5 standard deviations. In (C) and (D), are the PLS-DA models for prediction of compounds of high (represented by red triangles) and low (represented by green squares) affinity with the spike protein (S1) of SARS-CoV-2, respectively. The dashed red line is the threshold of the models, and the BNC that are above the threshold line are the BNC of interest that are in classification. In (E), it is the graph of errors versus the number of latent variables. Two latent variables were selected for simultaneously showing lower values of calibration (RMSEC) and cross validation errors (RMSECV). In (F) and (G), are the areas under the ROC curve, of the accuracy of the PLS-DA models of classification of high-affinity (AUC = 0.97) and low-affinity (AUC = 1.00) compounds with the spike protein (S1) of SARS-CoV-2

Table 3. Evaluation of the performance of the seven machine learning models for classification of low and high affinity bioactive natural compounds from the spike (S1) protein of SARS-CoV-2. The PLS-DA model showed the highest performance (accuracy)

Model	TP	FN	TN	FP	Accuracy	Sensitivity	Specificity
ANN	23	11	25	9	0.71	0.68	0.74
SVM	29	5	22	12	0.75	0.85	0.65
PLS-DA	33	1	34	0	0.99	0.97	1.00
XGboost	24	10	27	7	0.75	0.71	0.79
SIMCA	19	15	21	13	0.59	0.56	0.62
KNN	28	6	20	14	0.71	0.82	0.59
LOGREG	25	9	22	12	0.69	0.74	0.65

(TP): true positive. (TN): true negative. (FN): false negative. (FP): false positive. (ANN): artificial neural networks. (KNN): k-nearest neighbors. (DT): decision tree. (PLS-DA): discriminant analysis by partial least squares. (LDA): linear discriminant analysis. (SIMCA): soft independent modelling of class analogies.

Evaluation of stereoisomeric, tautomeric and protonation states

NBC5 predominates as a neutral molecule (*i.e.*, the same structure previously used in docking analysis), with no stereoisomer nor dominant tautomer at physiological pH (pH = 7.4). On the other hand, 46.18% of NBC14 predominates as neutral microspecies, and the remaining 53.72% correspond to deprotonated microspecies (two dominant tautomers were found: one neutral tautomer with 46% dominance and a deprotonated specie with 54% dominance). A total of 32 stereoisomers were identified in compound NBC14.

For the compounds NBC15 and NBC27, both the protonation state at physiological pH and the dominant tautomer were in their neutral forms. A total of 128 and 16 stereoisomers were identified in NBC15 and NBC27, respectively. All stereoisomers were docked with the protein and showed similar affinity to spike S1 (Figure 2), meaning that these compounds can be used as racemic drugs.

DISCUSSION

In this study, more than 170,000 NBCs were evaluated against the spike (S1) of COVID-19, using various *in silico* and machine learning methods. According to the virtual screening and molecular docking analysis, out of more than 170,000 compounds, only 34 compounds
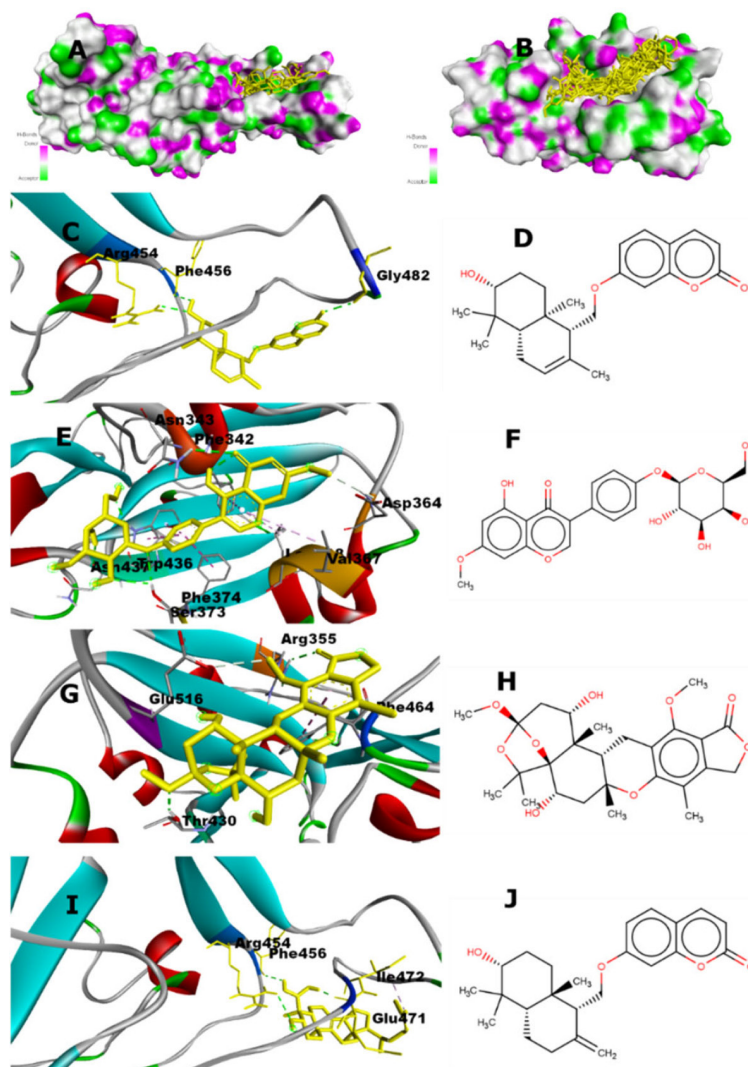
Figure 2. *Results of docking analysis of poses of compounds with higher affinity with the spike (S1) of SARS-CoV-2. Only the docking of ligands that showed drug-likeness, and which also presented better results in the ADMET analysis, are shown (NBC5, NBC14, NBC15 and NBC27). (A): all four ligands (NBC5, NBC14, NBC15 and NBC27) are shown docked in the same cavity of the spike surface (S1), according to hydrogen bonds. (B): all stereoisomers, all dominant tautomers and all microspecies in the protonation state at physiological pH (pH = 7.4) of the four ligands that have been shown to be promising against COVID-19 (NBC5, NBC14, NBC15 and NBC27) are bound in the same cavity of the spike, showing the large selectivity for the S1 site. The docking of the NBC5 ligand is illustrated in (C), and the chemical structure of the NBC5 ligand is illustrated in (D). The docking of the NBC14 ligand is illustrated in (E), and the chemical structure of the NBC14 ligand is illustrated in (F). The docking of the NBC15 ligand is illustrated in (G), and the chemical structure of the NBC15 ligand is shown in (H). The docking of the NBC27 ligand is illustrated in (I), and the chemical structure of the NBC27 ligand is shown in (J)*

(Table 1) were identified as having higher binding affinity with the spike protein, and these results were also reproduced by the machine learning models PCA and PLS-DA, the latter with an accuracy of 96%. However, according to drug-likeness and ADMET analysis, only four chemical compounds, namely NBC5, NBC14, NBC15 and NBC27 (See IUPAC names in the abbreviation list or in Table 1) were the most promising against the spike protein of COVID-19.

According to the literature, nearly 80% of the world's population depend on traditional medicines to manage a range of diseases. Past experience with the influenza outbreak, MERS-CoV and HIV infections has proven that natural products, including medicinal plants and their derivatives, are a valuable source for the synthesis of new antiviral drugs due to their availability and variety of substances with therapeutic potential. About 50% of all drugs approved and marketed worldwide between 1981 and 2014 were derived from natural products.[40,41] Substances such as flavonoids (*e.g*., neohesperidin, hesperidin, baicalin, kaempferol

3-*O*-rutinoside, rutin, neoandrographolide and 14-desoxy-11,12-dideshydroandrographolide), xanthones (*e.g*., substances from the *Swertia* genus plants) and alkaloids (*e.g*., ergotamine, nigellidine, noscapine and quinadoline B) have antiviral, antibacterial and anti-inflammatory activities.[42,43]

The availability of the virus RNA genome sequence (GenBank ID: MN908947.3) represents a valuable starting point for the identification of effective treatments against COVID-19 infections. Most importantly, SARS-CoV-2 features 82% similarity with SARS-CoV (GenBank ID: NC_004718.3), with a 90% resemblance in various essential enzymes.[43-46] The critical residues for receptor binding that were identified in the RBD of the SARS-CoV spike protein and the C-terminal SD-1 domain (CTD) of the SARS-CoV-2 spike protein, as well as in the interacting partner (hoster ACE-2), make them targets for the discovery and development of vaccines and drugs for the prevention and treatment of COVID-19 and other coronavirus infections.[47-50] This is important, because previous studies
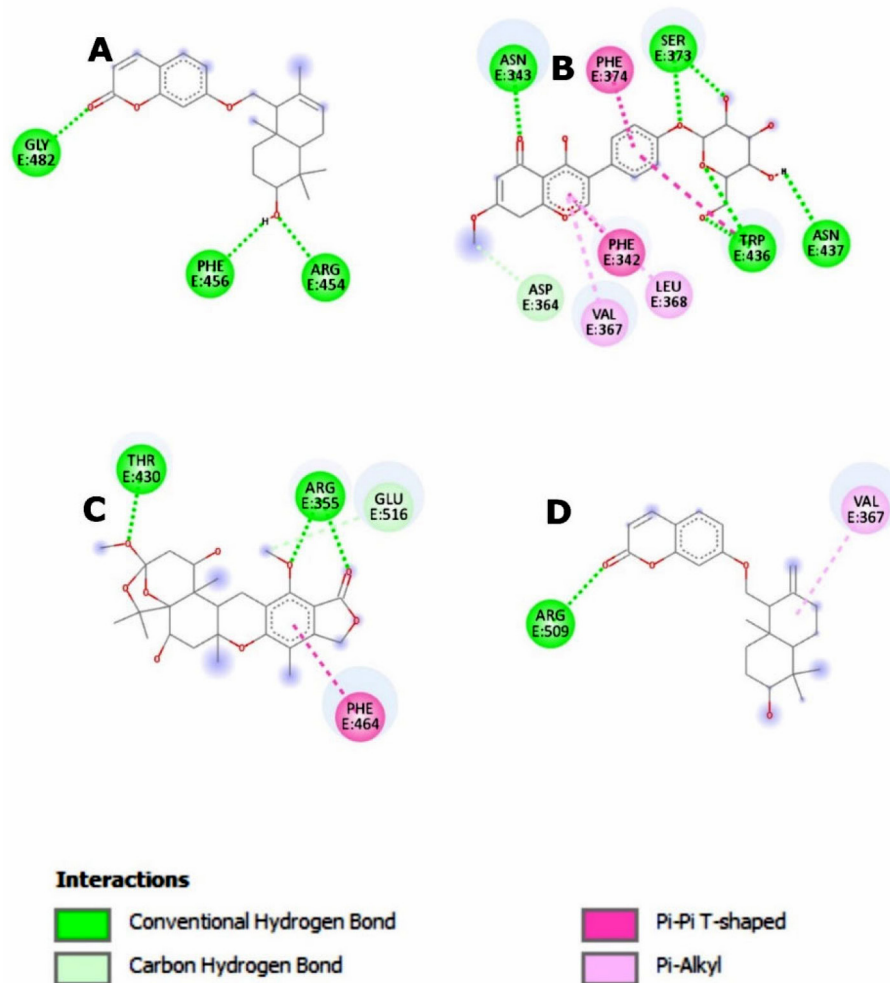
Figure 3. *2D structures of protein-ligand interactions. Only the structures of the ligands that were most promising against the spike protein of SARS-CoV-2 (NBC5, NBC14, NBC15 and NBC27) are shown. The interactions of the ligands NBC5, NBC14, NBC15 and NBC27 are shown in Figures A, B, C and D, respectively*

have reported that subjects with severe SARS-CoV-2 infection exhibit a larger antibody response against the spike and nucleocapsid protein and epitope, spreading to subdominant viral antigens (with a larger memory B cell response against the spike).[51]

Although we initially found 34 compounds with greater affinity with the spike (hydrogen bonds), only four (NBC5, NBC14, NBC15 and NBC27) were found to have drug-likeness characteristics, had promising pharmacokinetic profiles and low acute and chronic toxicities in *in silico* assays. Overall, the results suggest compounds NBC5, NBC14, NBC15 and NBC27 as potential drug candidates to be tested against COVID-19.[39,52]

Compound NBC5 is commonly known as feselol, a natural product found in plants of the genus *Ferula* (*e.g.*, *Ferula gummosa* Boiss. and *Ferula galbaniflua* Boiss.).[53] However, studies on the biological activities of this substance are limited in the literature. Only *in vitro* studies showing antimicrobial effects against *P. aeruginosa*, *S. epidermidis* and *S. aureus* and antiparasitic activity against *P. falciparum* are available.[54,55] *In vitro* studies show that the combination of feselol with antineoplasics have potentiated anticancer effects; this can be explained due to its ability to inhibit the P-glycoprotein, which is the main protein responsible for the resistance mechanism of many drugs (including anticancer drugs), favouring an increase of absorption rates and bioavailability of these drugs, and consequently obtaining the desired therapeutic activity.[56-58] No studies on the antiviral activities of feselol were found. Similarly, to our knowledge, no studies assessing the effects of the natural

compounds NBC14, NBC15 and NBC27 (see Table 1 for IUPAC names) exist, suggesting the need for further evaluations of the biological antiviral activities of these natural substances.

From a molecular point of view, the antiviral activity of the phytochemical compounds NBC5, NBC14, NBC15 and NBC27 (see Table 1 for IUPAC names) can be justified by the fact that they have carboxyl groups and hydroxyl groups whose oxygen and hydrogen atoms intermolecularly link by hydrogen bonds with the residues from the glycoprotein spike (S1) amino acids of SARS-CoV-2, namely, Gly482, F454, Arg454, N343, Ser373, W436, N437, Thr430, Arg355, F456, E471 and Arg454.[59,60]

The structural determination of ligands in their stereoisomeric, tautomeric states both at physiological pH (pH = 7.4) is very important in a docking study, as it would simulate the conditions of the human organism.[36,37] This analysis is feasible only in situations where the number of docked ligands is small, as this determination is performed manually by making one ligand at a time through specific software (*e.g.*, Chemdraw or Marvinsketch).[61,62] In situations where there are thousands (or even millions) of molecules to be docked, which is the case of our study (we used 171 thousands of ligands), it is not feasible to carry out the determination of ligands at physiological pH (pH = 7.4), due to the large volume of existing ligands in the database, there are even several similar studies in the literature.[63,64,65] In this situation of greater number of ligands (thousands or millions of ligands), the first task to be performed is virtual screening, which is a process that consists of investigating

which compounds have a greater binding affinity with the molecular target. After identifying the ligands with the greatest affinity with the molecular target (which are generally in small numbers), a docking analysis is then carried out using very rigorous criteria, which include the determination of the ligands structures at pH = 7.4, the influence of tautomery, stereoisomerism, drug-likeness, among others. In our study, we followed the same strategy; where initially we performed a virtual screening of 171 thousands of ligands, in which we selected 34 compounds. The 34 ligands were determined their structures at pH = 7.4, their stereoisomeric and tautomeric states and then performed a new consensual docking analysis (using two programs AutoDock Tools and AutoDock Vina) and machine learning to validate the results. We additionally demonstrated through stereoisomeric, tautomeric and protonation states (at physiological pH),[66,67] that NBC5, NBC14, NBC15 and NBC27 could be used as racemic drugs, meaning that no advanced technology is needed for the isolation of stereoisomers. Enantiomerically pure drugs (*e.g.*, naproxen, labetalol, warfarin) require high costs for development and production technologies (pure enantiomer with biological activity), which can be an important barrier in most countries.[68]

Renal toxicity is one of the very important toxicity that must be evaluated in compounds that have greater water solubility, such as the case of the coumarin derivatives identified in our study (for example, feselol).[69] However, the few online platforms (machine learning models) available in the literature that perform these predictions have low predictive accuracy. In addition, only a few recent studies are available that have developed machine learning models for predicting renal toxicity with greater accuracy, but the authors have not developed an application for the models to be used in practice, as can be seen in the recent study by Gong.[69,70] These were the reasons why it was not possible to make *in silico* predictions of the renal toxicity of the coumarin compounds identified in this study. This constituted one of the limitations of our study.

Considering that our study was based on *in silico* assays, several factors were not taken into account, such as the existence of multienzymatic systems and other biological factors that may interfere with the pharmacokinetics and pharmacodynamics of a drug candidate molecule. Considering these limitations, *in vitro* and *in vivo* studies are necessary in order to consolidate our findings.

CONCLUSION

Machine learning-combined *in silico* predictions of a database containing over 170,000 phytochemical compounds revealed four substances as the most promising inhibitors of the spike (S1) protein of SARS-CoV-2: NBC5 (feselol), NBC14, NBC15 and NBC27. In terms of structural similarities, the four compounds belong to the class of coumarins (benzopyrones), as they have a benzene ring fused to a lactone (heteroside group). Another structural similarity is that the four compounds present the benzopyrone ring that forms an ether bond with a polyhydroxylated cyclic alcohol group. Additionally, other similar characteristics among these compounds were: molecular weight less than 500 g mol^{-1}, LogP < 5.0, number of hydrogen donor and acceptor groups less than 10, number d and rotatable bonds less than 10, polar surface area (TPSA) less than 140. These compounds had binding affinity for spike (S1) similar to drugs currently used in the treatment of COVID-19 (remdesivir and molnupinavir).

Additionally, our machine learning model (PLS-DA) was able to classify compounds with low and high affinity with the spike protein (S1), with an accuracy of 99%, reproducing the results of molecular docking. This shows that the use of *in silico* methods combined with machine learning, provides more accurate and robust predictions in the design of new drug candidates.

In vitro and *in vivo* studies with these natural compounds are needed to confirm our findings and to support the development of drugs against COVID-19.

SUPPLEMENTARY MATERIAL

Figures 1S to 3S and Tables 1S to 6S are available at http://quimicanova.sbq.org.br in pdf format, with free access.

REFERENCES

1. Venkatesan, P.; *Lancet Respir. Med.* **2021**, *9*, e63. [Crossref]
2. Martinez, M. A.; *Front. Immunol.* **2021**, *12*, 635371. [Crossref]
3. Choudhury, A.; Mukherjee, S.; *J. Med. Virol.* **2020**, *92*, 2105. [Crossref]
4. Wang, C.; Horby, P. W.; Hayden, F. G.; Gao, G. F.; *Lancet* **2020**, *395*, 470. [Crossref]
5. Tizaoui, K.; Zidi, I.; Lee, K. H.; Ghayda, R. A.; Hong, S. H.; Li, H.; Smith, L.; Koyanagi, A.; Jacob, L.; Kronbichler, A.; Shin, J. I.; *Int. J. Biol. Sci.* **2020**, *16*, 2906. [Crossref]
6. Marroquín, G. C.; Saavedra, F.; Andrade, C. A.; Berrios, R. V.; Guilarte, L. R.; Opazo, M. C.; Riedel, C. A.; Kalergis, A. M.; *Front. Immunol.* **2020**, *11*, 1. [Crossref]
7. Eastman, R. T.; Roth, J. S.; Brimacombe, K. R.; Simeonov, A.; Shen, M.; Patnaik, S.; Hall, M. D.; *ACS Cent. Sci.* **2020**, *6*, 672. [Crossref]
8. Sultana, J.; Crisafulli, S.; Gabbay, F.; Lynn, E.; Shakir, S.; Trifirò, G.; *Front. Pharmacol.* **2020**, *11*, 588654. [Crossref]
9. Choudhury, A.; Mukherjee, S.; *J. Med. Virol.* **2020**, *92*, 2105. [Crossref]
10. Kumar, Y.; Singh, H.; Patel, C. N.; *J. Infect. Public Health* **2020**, *13*, 1210. [Crossref]
11. Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J.; *Nature* **2020**, *588*, 327. [Crossref]
12. Villas-Boas, G. R.; Rescia, V. C.; Paes, M. M.; Lavorato, S. N.; Magalhães-Filho, M. F.; Cunha, M. S.; Simões, R. C.; Lacerda, R. B.; Freitas-Júnior, R. S.; Ramos, B. H. S.; Mapeli, A. M.; Henriques, M. S. T.; Freitas, W. R.; Lopes, L. A. F.; Oliveira, L. G. R.; Silva, J. G.; Silva-Filho, S. E. S.; Silveira, A. P. S.; Leão, K. V.; Matos, M. M. S.; Fernandes, J. S.; Cuman, R. K. N.; Comar, F. M. S. S.; Comar, J. F.; Brasileiro, L. A.; Santos, J. N.; Oesterreich, S. A.; *Molecules* **2020**, *25*, 1. [Crossref]
13. Daina, A.; Michielin, O.; Zoete, V.; *Sci. Rep.* **2017**, *7*, 1. [Crossref]
14. Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Cao, D.; Yin, M.; Zeng, X.; Wu, C.; Lu, A.; Chen, X.; Hou, T.; Cao, D.; *Nucleic Acids Res.* **2021**, *49*, W5. [Crossref]
15. Braga, R. C.; Alves, V. M.; Silva, M. F. B.; Muratov, E.; Fourches, D.; Lião, L. M.; Tropsha, A.; Andrade, C. H.; *Mol. Inf.* **2015**, *34*, 698. [Crossref]
16. Banerjee, P.; Eckert, A. O.; Schrey, A. K.; Preissner, R.; *Nucleic Acids Res.* **2018**, *46*, W257. [Crossref]
17. Andrade, B. S.; Rangel, F. S.; Santos, N. O.; Freitas, A. S.; Soares, W. R. A.; Siqueira, S.; Barh, D.; Góes-Neto, A.; Birbrair, A.; Azevedo, V. A. C.; *Front. Pharmacol.* **2020**, *11*, 1. [Crossref]
18. Mohamed, K.; Yazdanpanah, N.; Saghazadeh, A.; Rezaei, N.; *Bioorg. Chem.* **2021**, *106*, 104490. [Crossref]
19. Vardhan, S.; Sahoo, S. K.; *Comput. Biol. Med.* **2020**, *124*, 103936. [Crossref]

20. Nag, A.; Paul, S.; Banerjee, R.; Kundu, R.; *Comput. Biol. Med.* **2021**, *137*, 104818. [Crossref]

21. Dey, D.; Borkotoky, S.; Banerjee, M.; *Comput. Biol. Med.* **2020**, *127*, 104063. [Crossref]

22. Kadioglu, O.; Saeed, M.; Greten, H. J.; Efferth, T.; *Comput. Biol. Med.* **2021**, *133*, 104359. [Crossref]

23. Mhatre, S.; Naik, S.; Patravale, V.; *Comput. Biol. Med.* **2021**, *129*, 104137. [Crossref]

24. Ibrahim, M. A. A.; Abdelrahman, A. H. M.; Hussien, T. A.; Badr, E. A. A.; Mohamed, T. A.; El-Seedi, H. R.; Pare, P. W.; Efferth, T.; Hegazy, M. E. F.; *Comput. Biol. Med.* **2020**, *126*, 104046. [Crossref]

25. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L.; *Nucleic Acids Res.* **2007**, *35*, D301. [Crossref]

26. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q.; *Science* **2020**, *367*, 1444. [Crossref]

27. Ravi, L.; Krishnan, K.; *Journal of the Medical Sciences* **2016**, *4*, 1. [Link] accessed in march 2023

28. Sterling, T.; Irwin, J. J.; *J. Chem. Inf. Model.* **2015**, *55*, 2324. [Crossref]

29. Dallakyan, S.; Olson, A. J. In *Chemical Biology*; Hempel, J. E.; Williams, C. H.; Hong C. C., eds.; Springer: New York, 2015, ch. 19. [Crossref]

30. Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T.; *J. Comput. -Aided. Mol. Des.* **2008**, *22*, 213. [Crossref]

31. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew R. K.; Goodsell D. S.; Olson A. J.; *J. Comput. Chem.* **2009**, *30*, 2785. [Crossref]

32. Allouche, A.; *J. Comput. Chem.* **2011**, *32*, 174. [Crossref]

33. Kufareva, I.; Abagyan, R.; *Methods Mol. Biol.* **2012**, *857*, 231. [Crossref]

34. Cobre, A. F.; Surek, A.; Stremel, D. P.; Fachi, M. M.; Borba, H. H. L.; Tonin, F. S.; Pontarolo, R.; *Comput. Biol. Med.* **2022**, *146*, 105659. [Crossref]

35. Dhall, D.; Kaur, R.; Juneja, M. In *Proceedings ICRIC 2019*; Singh, P. K.; Kar, A. K.; Singh, Y.; Kolekar, M. H.; Tanwar, S., eds.; Springer: Cham, 2020, ch. 2.

36. Brink, T. T.; Exner, T. E.; *J. Chem. Inf. Model.* **2009**, *49*, 1535. [Crossref]

37. Brink, T. T.; Exner, T. E.; *J. Comput. -Aided. Mol. Des.* **2010**, *24*, 935. [Crossref]

38. Tian, S.; Wang, J.; Li, Y.; Li, D.; Xu, L.; Hou, T.; *Adv. Drug Delivery Rev.* **2015**, *86*, 2. [Crossref]

39. Lipinski, C. A.; *Drug Discovery Today: Technol.* **2004**, *1*, 337. [Crossref]

40. Newman, D. J.; Cragg, G. M.; *J. Nat. Prod.* **2020**, *83*, 770. [Crossref]

41. Khan, H.; *J. Evidence-Based Complementary Altern. Med.* **2014**, *19*, 216. [Crossref]

42. Qamar, M. T.; Alqahtani, S. M.; Alamri, M. A.; Chen, L. L.; *J. Pharm. Anal.* **2020**, *10*, 313. [Crossref]

43. Antonio, A. S.; Wiedemann, L. S. M.; Veiga-Junior, V. F.; *RSC Adv.* **2020**, *10*, 23379. [Crossref]

44. Hilgenfeld, R.; Peiris, M.; *Antiviral Res.* **2013**, *100*, 286. [Crossref]

45. Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R.; *ChemBioChem* **2020**, *21*, 730. [Crossref]

46. Ghosh, A. K.; Brindisi, M.; Shahabi, D.; Chapman, M. E.; Mesecar, A. D.; *ChemMedChem* **2020**, *15*, 907. [Crossref]

47. Huang, Y.; Yang, C.; Xu, X. F.; Xu, W.; Liu, S. W.; *Acta Pharmacol. Sin.* **2020**, *41*, 1141. [Crossref]

48. Egieyeh, S.; Egieyeh, E.; Malan, S.; Christofells, A.; Fielding, B.; *PLoS One* **2021**, *16*, e0245258. [Crossref]

49. Zhou, Y.; Yang, Y.; Huang, J.; Jiang, S.; Du, L.; *Viruses* **2019**, *11*, 1. [Crossref]

50. Du, L.; Yang, Y.; Zhou, Y.; Lu, L.; Li, F.; Jiang, S.; *Expert Opin. Ther. Targets* **2017**, *21*, 131. [Crossref]

51. Guthmiller, J. J.; Stovicek, O.; Wang, J.; Changrob, S.; Li, L.; Halfmann, P.; Zheng, N. Y.; Utset, H.; Stamper, C. T.; Dugan, H. L.; Miller, W. D.; Huang, M.; Dai, Y. N.; Nelson, C. A.; Hall, P. D.; Jansen, M.; Shanmugarajah, K.; Donington, J. S.; Krammer, F.; Fremont, D. H.; Joachimiak, A.; Kawaoka, Y.; Tesic, V.; Madariaga, M. L.; Wilson, P. C.; *mBio* **2021**, *12*, e02940. [Crossref]

52. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith B. R.; Ward K. W.; Kopple K. D.; *J. Med. Chem.* **2002**, *45*, 2615. [Crossref]

53. Farhadi, F.; Iranshahi, M.; Mohtashami, L.; Asil, S. S.; Iranshahy, M.; *Phytochem. Anal.* **2021**, *32*, 811. [Crossref]

54. Amin, A.; Hanif, M.; Abbas, K.; Ramzan, M.; Rasheed, A.; Zaman, A.; Pieters, L.; *Microb. Pathog.* **2020**, *144*, 104184. [Crossref]

55. Amin, A.; Tuenter, E.; Cos, P.; Maes, L.; Exarchou, V.; Apers, S.; Pieters, L.; *Molecules* **2016**, *21*, 1287. [Crossref]

56. Mollazadeh, S.; Matin, M. M.; Bahrami, A. R.; Iranshahi, M.; Rassouli, M. B.; Rassouli, F. B.; Neshati, V.; *Zeitschrift für Naturforschung C* **2011**, *66*, 555. [Crossref]

57. Mollazadeh, S.; Matin, M. M.; Iranshahi, M.; Bahrami, A. R.; Neshati, V.; Behnam-Rassouli, F.; *J. Asian Nat. Prod. Res.* **2010**, *12*, 569. [Crossref]

58. Iranshahi, M.; Barthomeuf, C.; Bayet-Robert, M.; Chollet, P.; Davoodi, D.; Piacente, S.; Rezaee, R.; Sahebkar, A.; *J. Tradit. Complementary Med.* **2014**, *4*, 118. [Crossref]

59. Fujita, T.; Nishioka, T.; Nakajima, M.; *J. Med. Chem.* **1977**, *20*, 1071. [Crossref]

60. Dearden, J. C.; Cronin, M. T. D.; Zhao, Y. H.; Raevsky, O. A.; *Mol. Inf.* **2000**, *19*, 3. [Crossref]

61. Jena, N. R.; *Phys. Chem. Chem. Phys.* **2020**, *22*, 28115. [Crossref]

62. Umar, Y.; *Journal of Taibah University for Science* **2020**, *14*, 1613. [Crossref]

63. Fischer, A.; Sellner, M.; Neranjan, S.; Smieško, M.; Lill, M. A.; *Int. J. Mol. Sci.* **2020**, *21*, 3626. [Crossref]

64. Luttens, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D. D.; Akaberi, D.; Talibov, V. O.; Carlsson, J.; *J. Am. Chem. Soc.* **2022**, *144*, 2905. [Crossref]

65. Almes, F. J. M.; *Comput. Biol. Chem.* **2020**, *88*, 107351. [Crossref]

66. Kalliokoski, T.; Salo, H. S.; Lahtela-Kakkonen, M.; Poso, A.; *J. Chem. Inf. Model.* **2009**, *49*, 2742. [Crossref]

67. Camp, W. H.; *Chirality* **1989**, *1*, 2. [Crossref]

68. Pifferi, G.; Perucca, E.; *Eur. J. Drug Metab. Pharmacokinet.* **1995**, *20*, 15. [Crossref]

69. Al-Asmari, K. M.; Altayb, H. N.; Al-Attar, A. M.; Qahl, S. H.; Al-Thobaiti, S. A.; Abu-Zeid, I. M.; *Saudi J. Biol. Sci.* **2022**, *29*, 103307. [Crossref]

70. Gong, Y.; Teng, D.; Wang, Y.; Gu, Y.; Wu, Z.; Li, W.; Tang, Y.; Liu. G.; *J. Appl. Toxicol.* **2022**, *42*, 1639. [Crossref]