









In silico analysis of the interactions of certain flavonoids with the receptor-binding domain of 2019 novel coronavirus and cellular proteases and their pharmacokinetic properties

Erman Salih Istifli^a , Paulo A. Netz^b , Arzuhan Sihoglu Tepe^c , Mehmet Tahir Husunet^a , Cengiz Sarikurkcu^d  and Bektas Tepe^e 

^aDepartment of Biology, Faculty of Science and Literature, Cukurova University, Adana, Turkey; ^bTheoretical Chemistry Group, Institute of Chemistry, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; ^cDepartment of Biology, Faculty of Science and Literature, Gaziantep University, Gaziantep, Turkey; ^dDepartment of Analytical Chemistry, Faculty of Pharmacy, Afyonkarahisar Health Sciences University, Afyonkarahisar, Turkey; ^eDepartment of Molecular Biology and Genetics, Faculty of Science and Literature, Kilis 7 Aralik University, Kilis, Turkey

ABSTRACT

Coronavirus Disease 2019 (COVID-19) has infected more than thirty five million people worldwide and caused nearly 1 million deaths as of October 2020. The microorganism causing COVID-19 was named as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2 or 2019-nCoV). The aim of this study was to investigate the interactions of twenty-three phytochemicals belonging to different flavonoid subgroups with the receptor binding domain (RBD) of the spike glycoprotein of 2019-nCoV, and cellular proteases [transmembrane serine protease 2 (TMPRSS2), cathepsin B and L (CatB/L)]. The compounds interacted more strongly with CatB and CatL than with the other proteins. Van der Waals and hydrogen bonds played an important role in the receptor-ligand interactions. As a result of RBCI (relative binding capacity index) analysis conducted to rank flavonoids in terms of their interactions with the target proteins, (-)-epicatechin gallate interacted strongly with all the proteins studied. The results obtained from molecular dynamics and molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) methods also supported this data. According to Lipinski's rule of five, (-)-epicatechin gallate showed drug-likeness properties. Although this molecule is not capable of crossing the blood-brain barrier (BBB), it was concluded that (-)-epicatechin gallate can be evaluated as a candidate molecule in drug development studies against 2019-nCoV since it was not the substrate of P-gp (P-glycoprotein), did not inhibit any of the cytochrome Ps, and did not show AMES toxicity or hepatotoxicity on eukaryotic cells.

ARTICLE HISTORY

Received 6 August 2020
Accepted 17 October 2020

KEYWORDS

COVID-19; flavonoid;
molecular docking;
molecular dynamics;
MM/PBSA

1. Introduction

Coronavirus Disease 2019 (COVID-19), which first appeared in the Wuhan region of China and then spread all over the world rapidly, has infected more than thirty five million people worldwide and caused nearly 1 million deaths as of October 2020 (Worldometers.info, 2020; F. Wu et al., 2020b; Zhu et al., 2020). The microorganism causing COVID-19 was named as new type Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2 or 2019-nCoV) (F. Wu et al., 2020b; Zhou et al., 2020; Zhu et al., 2020). Some outbreaks caused by SARS-CoV and MERS-CoV in various parts of the world over the past decade were recorded. 2019-nCoV was the third important member of the coronavirus family that caused severe respiratory disease and human deaths (Zaki et al., 2012; Zhong et al., 2003).

The SARS-CoV and 2019-nCoV classified in the sarbecovirus subgenus of Coronaviridae were called human SARS-

related coronaviruses (SARSr-CoV). Genome analysis has shown that the genetic material of this virus is a single-stranded RNA molecule consisting of approximately 30,000 nucleotides. As a result of genomic analysis, the genes encoding the four basic proteins of 2019-nCoV were clarified as follows: nucleocapsid protein, envelope protein, membrane protein, and spike glycoprotein (A. Wu et al., 2020a). Spike glycoprotein is actually a type I glycoprotein that extends out of the surface of the virus and is the first component to come into contact with the host cell. Since the spike glycoprotein is the main component the virus uses to bind to receptors on the host cell surface, it is of great interest in the development of the treatment strategies of COVID-19 (Luan et al., 2020).

Angiotensin converting enzyme 2, also known as ACE2, binds to the receptor binding domain (RBD) of SARS-CoV. Thus, this enzyme acts as the receptor that the virus uses to enter the cell (F. Li et al., 2005; W. Li et al., 2003). ACE2

shows a wide distribution in many tissues in the body, especially intestine, kidney, testis, liver, lungs, and heart. ACE2 is known to play an important role in regulating blood pressure. In addition, depending on the control of blood pressure, it also has important physiological tasks in the regulation of kidney and heart functions (Anguiano et al., 2017). Since the RBD of 2019-nCoV interacted with human ACE2, many studies have shown that human ACE2 plays a critical role in the entry of 2019-nCoV into cells (Letko et al., 2020; Zhou et al., 2020).

The receptor-ligand relationship is vital in host specificity and the life cycle of the virus. Some researchers claim that 2019-nCoV was transferred from bats to humans (Zhou et al., 2020). However, what we know about the intermediate host, which was thought to play a role in this transfer process, is still limited. Some studies suggested that pangolin may have played an important role in the evolution of 2019-nCoV (Lam et al., 2020; Wong et al., 2020). As can be understood from these speculative findings, the questions of which animal was involved in the evolution of 2019-nCoV and which other animal species were infected by the virus are still unanswered. The interaction between spike glycoprotein and ACE2 has been clarified as a result of studies on what amino acids bound to ACE2 in the receptor-binding domain of 2019-nCoV (Andersen et al., 2020). According to these studies, Leu455, Phe486, Glu493, Ser494, Asp501, and Tyr505 were the amino acids that interacted with mammalian ACE2. Considering the interaction between the 2019-nCoV spike glycoprotein and ACE2, it was understood that both SARS-CoV and 2019-nCoV have the capacity to infect many mammalian species such as Chinese hamsters, pangolin, dogs, cats, etc. (Luan et al., 2020).

Besides ACE2, the transmembrane protease, serine 2 (TMPRSS2) and cathepsins (CatB and L) were also thought to mediate the entry of both SARS-CoV and SARS-CoV-2 into the host cell. TMPRSS2 is an androgen-responsive serine protease that cleaves the spike glycoprotein from the S1/S2 region, facilitating viral entry and activation (Hoffmann et al., 2020). CatL is an important lysosomal endopeptidase. It plays a critical role in initiating protein breakdown. There is evidence that this cysteine protease mediated the entry of SARS-CoV and related viruses into the host cell (Hoffmann et al., 2020; Huang et al., 2006; Sudhan & Siemann, 2015).

Plants produce a wide variety of compounds to perform their normal physiological functions. These include important phytochemicals such as tannins, lignans, phytoalexins, polyphenols, and flavonoids. Polyphenols and flavonoids make up the largest group of secondary metabolites and are abundant in many sources such as spices, nuts, seeds, fruits, vegetables, stems, red wine, and tea that people consume frequently (Middleton et al., 2000). Plants synthesize these compounds in response to abiotic and biotic stress conditions such as UV rays, pathogens and insects (Carletti et al., 2014; Mandal et al., 2010). Studies have revealed that flavonoids have a wide range of biological/pharmacological properties such as antifungal, antibacterial, anticancer, anti-inflammatory, and antioxidant. In recent *in vivo* and *in vitro*

studies, some flavonoids have been also found to exhibit considerable antiviral activity (Zakaryan et al., 2017).

The aim of this study was to determine the interaction of certain flavonoids (Figure 1) with the spike glycoprotein of 2019-nCoV, and host proteases (TMPRSS2, CatB and CatL). As a result of the receptor-ligand interactions, binding affinities and inhibition constant (K_i) values were determined. In order to rank flavonoids in terms of their interactions with all target proteins, RBCI (relative binding capacity index) analysis were conducted and 'hit' flavonoids were determined. Results of the docking analysis of the best molecules were also confirmed by subjecting them to the molecular dynamics simulations for 150 ns with all receptor proteins with subsequent estimation of the binding free energy using molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) methods. Drug-likeness properties according to the Lipinski's rule of five, ADMET (absorption, distribution, metabolism, excretion, and toxicity) profiles and intracellular target predictions of the flavonoids were also given.

2. Material and methods

2.1. Structural optimization of ligands

The protein data bank (pdb) files of all the ligands (quercetin, kaempferol, isorhamnetin, rutin, myricetin, resveratrol, naringenin, hesperidin, naringin, eriodictyol, genistein, daidzein, glycitein, apigenin, luteolin, tangeretin, epicatechin gallate, catechin, delphinidin, malvidin, peonidin, pelargonidin, petunidin) have been downloaded from PubChem (<https://pubchem.ncbi.nlm.nih.gov>) using Vega ZZ 3.2.0.9 download module. In the Vega ZZ platform, the atom types and electrical charges of the ligands were optimized with MMFF94 force field and Gasteiger-Marsili parameters (Pedretti et al., 2004).

2.2. Energy minimization of 2019-nCoV ACE2-RBD, TMPRSS2, CatB/L using nanoscale molecular dynamics (NAMD)

The structure of the spike glycoprotein was gained by removing the ACE2 subunit from the angiotensin-converting enzyme 2 - 2019nCoV RBD complex in the Vega ZZ environment. This model was downloaded from the following URL: <https://swissmodel.expasy.org/interactive/HLkhkP/models/03> (PDBID: model_03.pdb) (Camacho et al., 2009; Remmert et al., 2012). Since the structure of the spike glycoprotein in model_03 shows a sequence identity of 100% to the 2019-nCoV ACE2 binding domain, this model was chosen as an appropriate 3D structure in the molecular docking analyses. During the protein preparation step, the atom types and electrical charges of the spike glycoprotein were fixed using CHARMM22_PROT force field and Gasteiger-Marsili charges. Next, for the energy minimization of the spike glycoprotein with NAMD, each parameter was loaded from a template file. The number of time steps (number of minimization steps) were set to 10,000 and CHARMM22_PROT was set as the force field. When the energy minimization was completed,

the 3D structure corresponding to the last minimization step was saved as the lowest energy conformation. Also, to keep the spike glycoprotein structurally closer to the original crystallographic data, atom constraints were also applied to the protein backbone. In the energy minimization of TMPRSS2, CatB, and CatL, the same steps described above for the minimization of the 2019-nCoV spike glycoprotein were applied.

2.3. Homology modeling of TMPRSS2

The crystallographic data of TMPRSS2 enzyme structure has not been resolved until today, therefore we generated a homology model of this enzyme to utilize in further molecular docking and molecular dynamics analyses. The amino acid sequence of TMPRSS2 was downloaded from UniProtKB (<https://www.uniprot.org/uniprot/O15393>). Template search for TMPRSS2 catalytic domain was performed against the SWISS-MODEL template library with BLAST and HHblits. BLAST was used to search the TMPRSS2 catalytic domain target sequence against the primary amino acid sequence in the SMTL. As a result of the BLAST search, a total of 788 templates were found. An initial HHblits profile has been built using the procedure as described in (Remmert et al., 2012). This procedure was followed by 1 iteration of HHblits against NR20. The obtained profile was then searched against all profiles of the SMTL and, finally, a total of 1167 templates were found.

ProMod3 was used to carry out model building for TMPRSS2 catalytic domain based on the target-template alignment. The rest of the procedure was carried out as previously described (Guex et al., 2009).

The model quality (global and per-residue) of TMPRSS2 obtained was evaluated with the QMEAN scoring function (Studer et al., 2020). A near-zero QMEAN score is a good value in terms of the quality of the fit between model structure and the experimental structure. According to the QMEAN score, however, scores of 4.0 and below indicate that the model is of poor quality. Therefore, among the top 5 TMPRSS2 models we obtained as a result of homology modeling, we determined the 5ce1.1.A (model 06) model as the target structure in the molecular docking analysis.

In addition, whether our model has an energetically favorable conformation, we generated a Ramachandran plot (Figure S2) using the PROCHECK web server (Laskowski et al., 1993). Also, ERRAT web-based tool (Figure S1) was also deployed to calculate the overall quality factor (OQF) for non-bonded atomic interactions (Colovos & Yeates, 1993).

2.4. Molecular docking analyses

Firstly, we thought that it could be helpful to explain more clearly the expressions of RBM (receptor binding motif) and RBD (receptor binding domain) in the literature: Spike's RBD is the smallest part of this protein which is supposed to be able to capture the interaction with the receptor, hACE2. The RBD itself is on its hand composed by a core (residues 333–438 and residues 507–527) and a Receptor Binding Motif (RBM, residues 438–506). This RBM is the part of the RBD which is directly

interacting with the hACE2 (de Andrade et al., 2020; Lan et al., 2020; Omotuyi et al., 2020; Y. Wang et al., 2020; Woo et al., 2020). In this sense, throughout this study, we envisioned that RBM is a kind of "active site" of the RBD.

Molecular docking analyses between the target structures and the ligands were performed using AutoDock 4.2.6 and the corresponding docking scores (binding free energies) of the ligands with 2019-nCoV RBM, TMPRSS2, CatB (PDB ID: 1GMY), and CatL (PDB ID: 2YJ9) were calculated. AutoDockTools-1.5.6 was used to prepare the target and ligand molecules and also the parameters prior to initiating the docking analysis using AutoDock 4.2.6 (Sanner, 1999). In this study, the grid box coordinates used in molecular docking analyses were adjusted to ensure that all the tested phytochemicals interact with amino acids in the active sites of the enzymes in question (Andersen et al., 2020; Greenspan et al., 2001; Hardegger et al., 2011; Wilson et al., 2005).

Prior to molecular docking analyses, polar hydrogen atoms in the receptor and the ligand molecules were retained, while nonpolar hydrogens were merged and then, the Gasteiger charges of the ligands were calculated with AutoDockTools (Morris et al., 2009; Nasab et al., 2017; Sanner, 1999). Also, the Kollmann charges were added for the receptor. During the docking experiments, all the rotatable bonds of the ligands were allowed to rotate and then the optimized protein (rigid) and ligand (flexible) structures were saved in PDBQT format. Grid box coordinates were adjusted as: a) $80 \times 90 \times 40$ Å points for the spike glycoprotein; b) $60 \times 110 \times 86$ Å points for TMPRSS2; c) $86 \times 84 \times 44$ Å points for CatB; and d) $54 \times 52 \times 60$ Å points for CatL. These grid box sizes were previously determined to cover the active amino acid residues of the enzymes in question.

In all docking analyses, 50 genetic algorithm (GA) runs using an initial population of 150 individuals, maximum number of 2,500,000 energy evaluations, and a maximum number of 27,000 generations were selected. The values of 0.02 and 0.8 were chosen as the default parameters for mutation and crossover rates, respectively. After 50 independent docking runs, all the possible binding modes (conformations) of the ligands were clustered by the program and were ranked based on the most negative binding free energy (kcal/mol) of the ligand conformation. The best docking poses obtained using the AutoDock 4.2.6 between the ligand and receptor structures were analyzed with the BIOVIA Discovery Studio Visualizer 2016.

Furthermore, according to the results obtained from molecular docking experiments, heatmaps of ligands (Supplementary Figures S3, S4, S5 and S6) that interact with specific amino acid residues of each receptor were created. Heatmaps are built on the logic of how many times each ligand interacts with each residue (regardless of bond type). Thus, heatmaps provide an overview of the frequency of interactions of 23 different ligands with certain amino acids of these receptors.

2.5. Success criteria set in docking analysis

In the current study, the lowest of all clusters in terms of the binding free energy was considered as the energetically

Table 1. PubChem CID, molecular weight and molecular formula of the compounds.

No	Compound	PubChem CID	Molecular weight (g/mol)	Molecular formula
<i>Flavonols</i>				
1	Quercetin	5284452	302.23	C ₁₅ H ₁₀ O ₇
2	Kaempferol	5280863	286.24	C ₁₅ H ₁₀ O ₆
3	Isorhamnetin	44259381	478.40	C ₂₂ H ₂₂ O ₁₂
4	Rutin	5280805	136.23	C ₂₇ H ₃₀ O ₁₆
5	Myricetin	5281672	318.23	C ₁₅ H ₁₀ O ₈
6	Resveratrol	445154	228.24	C ₁₄ H ₁₂ O ₃
<i>Flavanones</i>				
7	Naringenin	932	272.25	C ₁₅ H ₁₂ O ₅
8	Hesperidin	10621	610.60	C ₂₈ H ₃₄ O ₁₅
9	Naringin	442428	580.50	C ₂₇ H ₃₂ O ₁₄
10	Eriodictyol	102370911	450.40	C ₂₁ H ₂₂ O ₁₁
<i>Isoflavones</i>				
11	Genistein	5280961	270.24	C ₁₅ H ₁₀ O ₅
12	Daidzein	5281708	254.24	C ₁₅ H ₁₀ O ₄
13	Glycitein	5317750	284.26	C ₁₆ H ₁₂ O ₁₅
14	Apigenin	5280443	270.24	C ₁₅ H ₁₀ O ₅
15	Luteolin	5280445	286.24	C ₁₅ H ₁₀ O ₆
16	Tangeretin	68077	372.40	C ₂₀ H ₂₀ O ₇
17	(-)-Epicatechin gallate	107905	442.40	C ₂₂ H ₁₈ O ₁₀
18	(+)-Catechin	9064	290.27	C ₁₅ H ₁₄ O ₆
<i>Anthocyanidins</i>				
19	Delphinidin	68245	338.69	C ₁₅ H ₁₁ ClO ₇
20	Malvidin	159287	331.30	C ₁₇ H ₁₅ O ₇ ⁺
21	Peonidin	441773	301.27	C ₁₆ H ₁₃ O ₆ ⁺
22	Pelargonidin	440832	271.24	C ₁₅ H ₁₁ O ₅ ⁺
23	Petunidin	73386	352.72	C ₁₆ H ₁₃ ClO ₇

most stable configuration and taken as reference (Morris & Lim-Wilby, 2008). The calculated inhibition constant (K_i) obtained with AutoDock 4.2.6 for each docked phytochemical were also represented.

2.6. Calculation of relative binding capacity index (RBCI) values

In this study, we have developed a new analysis method called RBCI and applied it to statistically rank the activity potentials of phytochemicals by using binding free energy and inhibition constant values obtained from the docking analyses (Istifli et al., 2020). Using the RBCI parameter, it is possible to compare statistically related data with different scientific meanings. If the ranking of the interactions of the ligands with proteins is done only in light of one parameter (e.g. binding free energy or inhibition constant only), the molecules can only be ordered in terms of their potential in this parameter. However, ranking based on only one of these parameters cannot represent the full activity potential of these molecules. The most common method used to calculate the interaction between the receptor and the ligand in multiple measurements is the 'central tendency', in which the components are ranked based on the average value of each component. However, since the units and scales of the data obtained from each parameter are different, it is not possible to obtain a universal value for all components.

If the values in each data set (binding energy and inhibition constant) are converted to standard scores, it is possible to compare them with each other. In order to calculate the arithmetic mean values, first of all, binding energy and inhibition constant data of each phytochemical were used regardless of their units and raw values were obtained. These raw values calculated for each component were subtracted from

the arithmetic mean and divided by standard deviation, and standard scores were obtained (see equation given below) (Sharma, 1996). RBCI values of each phytochemical were calculated by averaging these standard scores obtained separately for each protein target. where 'x' is the raw data, 'μ' is the mean, and 'σ' is the standard deviation.

Although RBCI is a relative index and does not represent the specific binding capacities of the components, it makes it possible to rank components reasonably based on their binding energy and inhibition constant values. Therefore, it can be used as an integrated approach to evaluate the molecular interaction of the components, considering all parameters.

2.7. Molecular dynamics of top-ranked receptor-ligand complexes

The configurations with the highest affinity (most negative binding free energy) obtained from the docking of receptor-ligand complexes were stored and prepared for the molecular dynamics simulations as follows: The receptor coordinates files were converted from pdbqt format to pdb format using Babel (O'Boyle et al., 2011) and thereafter to use in molecular dynamics simulations using standard GROMACS (Abraham et al., 2015) tools, choosing the AMBER03 force field (Duan et al., 2003). The ligand coordinates files were read with AutoDockTools (Morris et al., 2009) to add the hydrogen atoms, and the obtained pdb files were then submitted to the AcPype online server (Sousa da Silva & Vranken, 2012) to generate GROMACS topology files, employing the parameters of the General Amber Force Field, GAFF (J. Wang et al., 2004) and AM1-BCC partial charges (Jakalian et al., 2002). The topologies and coordinates of receptors and ligands were then combined to construct the topologies and coordinates of the complexes.

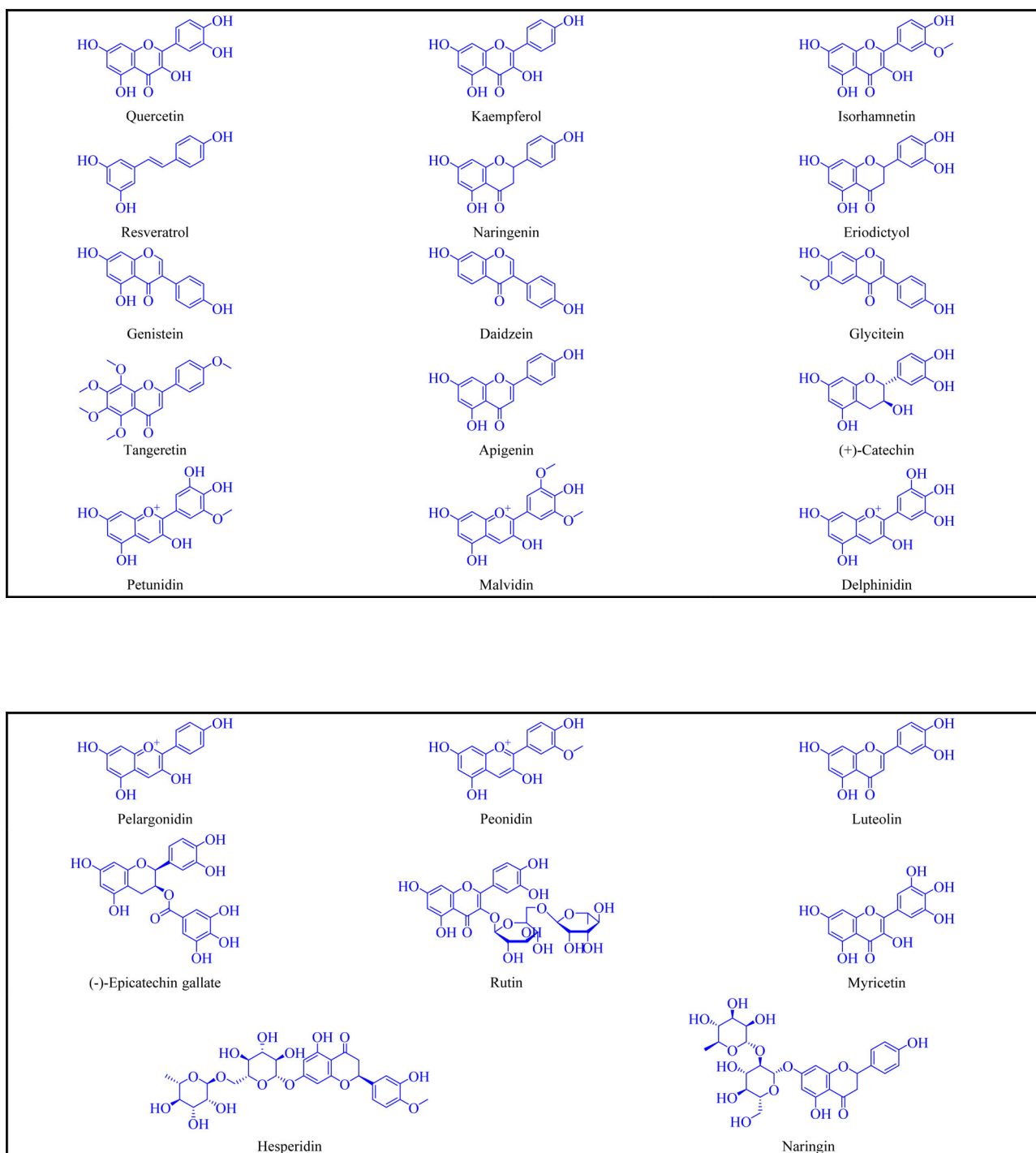


Figure 1. Chemical structures of the flavonoids.

The complexes were then solvated using TIP3P water molecules (Jorgensen et al., 1983) as well as sodium and chloride ions corresponding to physiological concentration in cubic boxes under periodic boundary conditions. The van der Waals interactions were calculated until a cutoff of 1.2 nm and the electrostatic interactions were calculated using the PME method (Darden et al., 1993; Essmann et al., 1995). The systems were initially energetically minimized using the steepest descent algorithm, submitted to a 500 ps simulation employing position restraints for both receptor and ligand and to a sequence of three unrestricted molecular dynamics simulations with 5 ns each in temperatures of 200 K, 240 K and 280 K

for the thermalization (heating) of the system. After these steps, 150 ns long production simulation runs were carried out. The simulation sampled the NPT ensemble, employing a Nosé-Hoover thermostat (Hoover, 1985; Nosé, 1984) and a Parrinello-Rahman barostat (Parrinello & Rahman, 1981). The trajectories were analyzed to quantify the structural and thermodynamic stability of the complexes and to identify the pattern of intermolecular interactions.

Detailed qualitative visual analyses of the time evolution of the interactions in the trajectories were also carried out. The strength of the interactions between ligands and receptors was estimated with Molecular mechanics Poisson-Boltzmann Surface

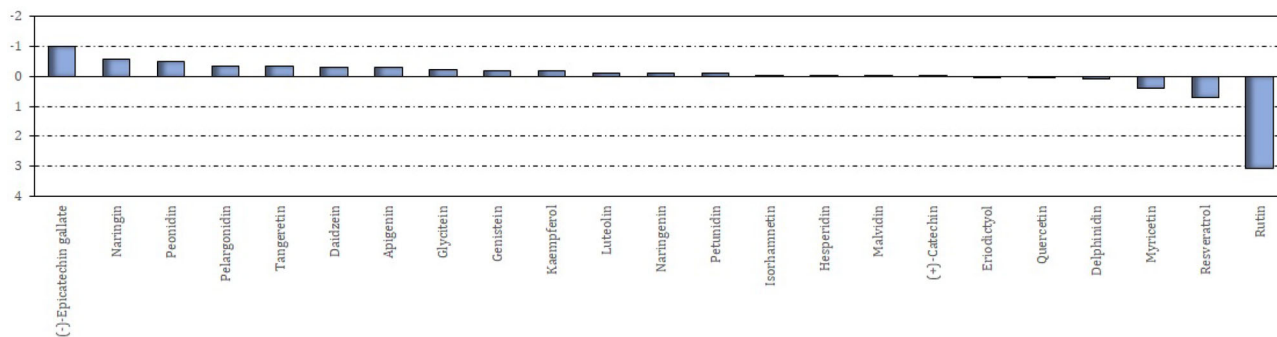


Figure 2. RBCE values of the flavonoids.

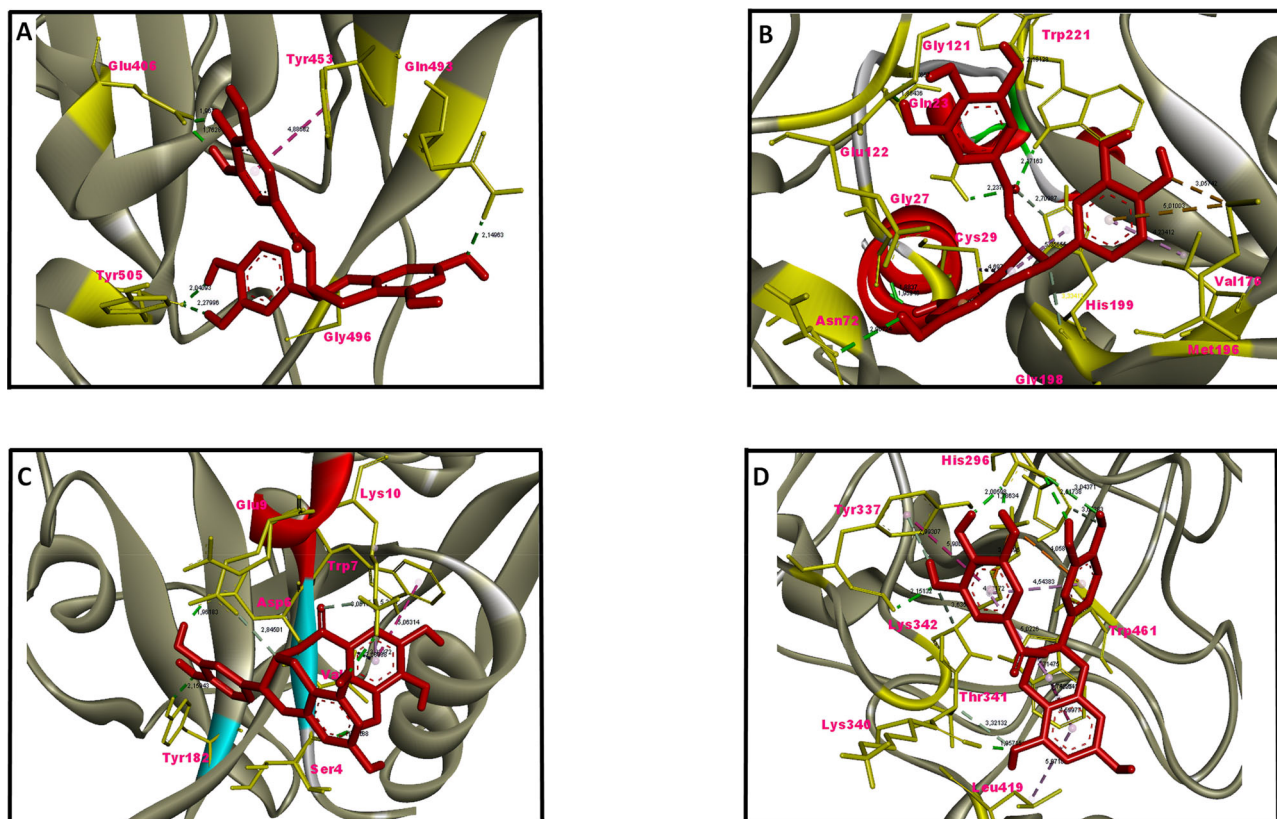


Figure 3. Top ranked conformations of (-)-epicatechin gallate. (A- RBM of the spike glycoprotein of 2019-nCoV, B- TMRSS2, C- CatB, D- CatL).

Area (MM/PBSA) binding free energy calculations (Baker et al., 2001; Homeyer & Gohlke, 2012). Sets of 150 configurations for each system were obtained from 1 ns spaced snapshots obtained from the molecular dynamics trajectories. The calculations were carried out with g_mmpbsa (Kumari et al., 2014) using a grid-space of 0.5 Å, salt concentration of 0.150 M, solute dielectric constant of 2 and using the solvent accessible surface area (SASA) as estimate of the nonpolar solvation energy.

2.8. Drug-likeness, ADMET profile and target prediction

The drug-likeness, ADMET and target profiles of potential hit compounds are very important in terms of reducing side effects in the pharmaceutical industry. In the current study, web-based SwissADME, SwissTargetPrediction and pkCSM

tools were used to determine such effects of flavonoids analyzed (Daina et al., 2017, 2019; Pires et al., 2015).

3. Results and discussion

In this study, the molecular interactions of twenty-three phytochemicals belonging to four different flavonoid subgroups (flavonols, flavanones, isoflavones, and anthocyanidins) (Table 1, Figure 1) with the RBD of the spike glycoprotein of 2019-nCoV, TMRSS2, CatB and CatL were analyzed.

3.1. Molecular docking analysis

In receptor-ligand interactions, binding energy can vary depending on the three-dimensional structure of the

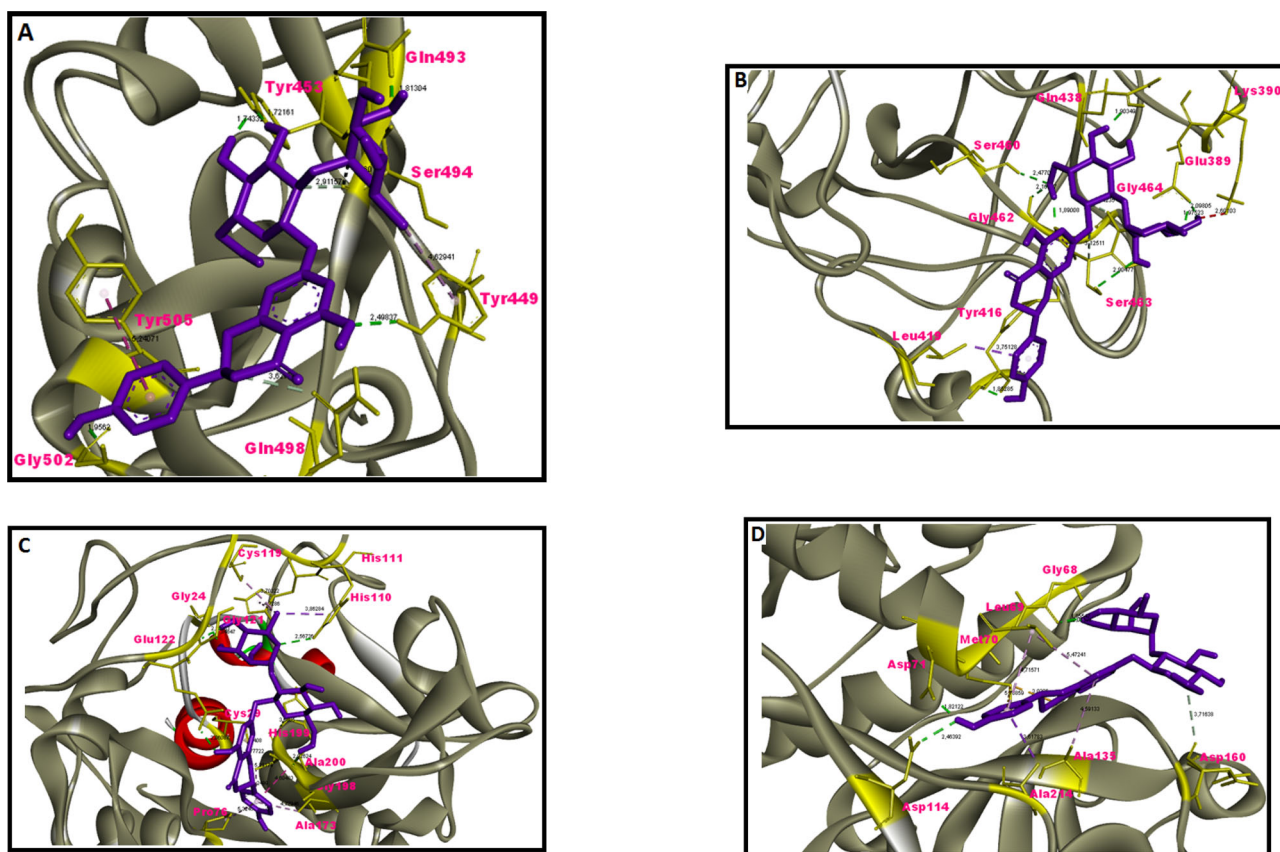


Figure 4. Top ranked conformations of naringin. (A- RBM of the spike glycoprotein of 2019-nCoV, B- TMRSS2, C- CatB, D- CatL)

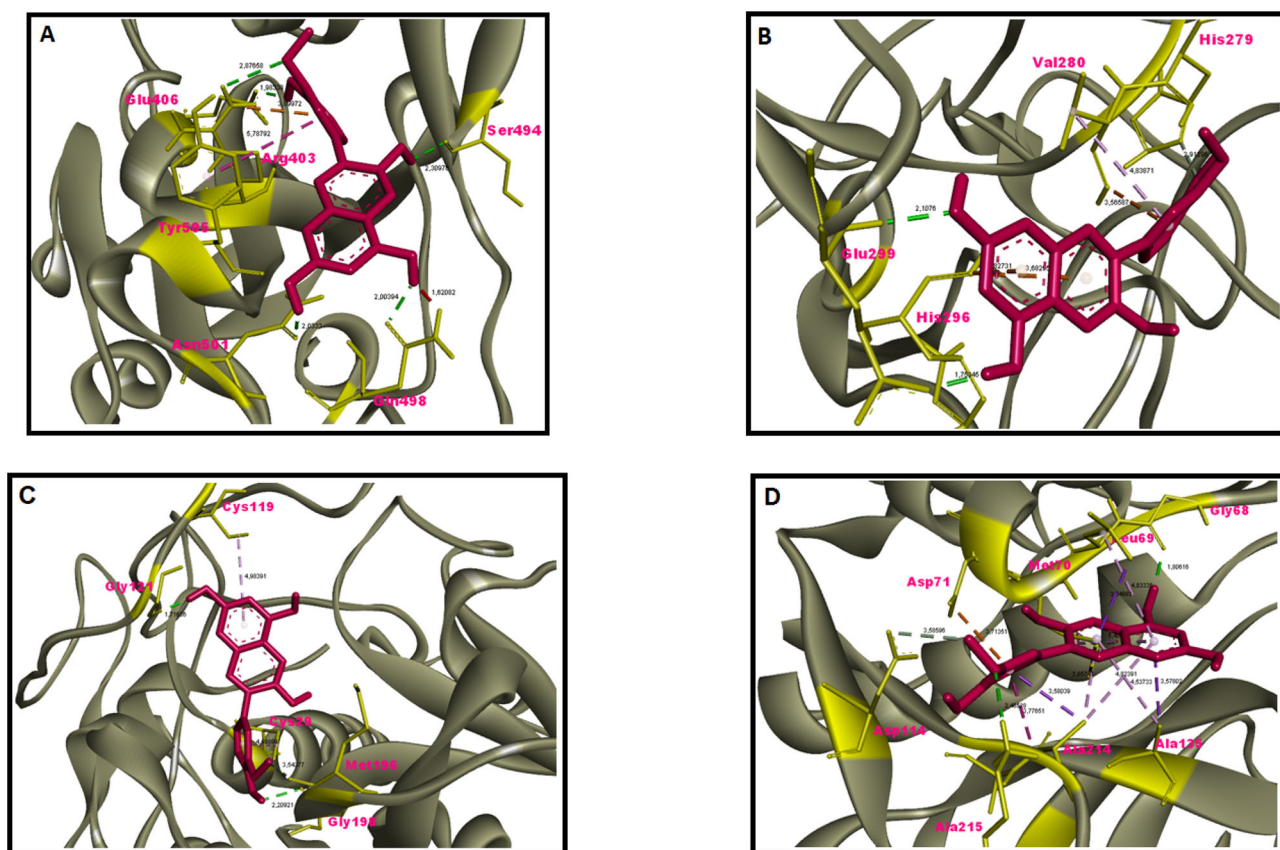


Figure 5. Top ranked conformations of peonidin. (A- RBM of the spike glycoprotein of 2019-nCoV, B- TMRSS2, C- CatB, D- CatL).

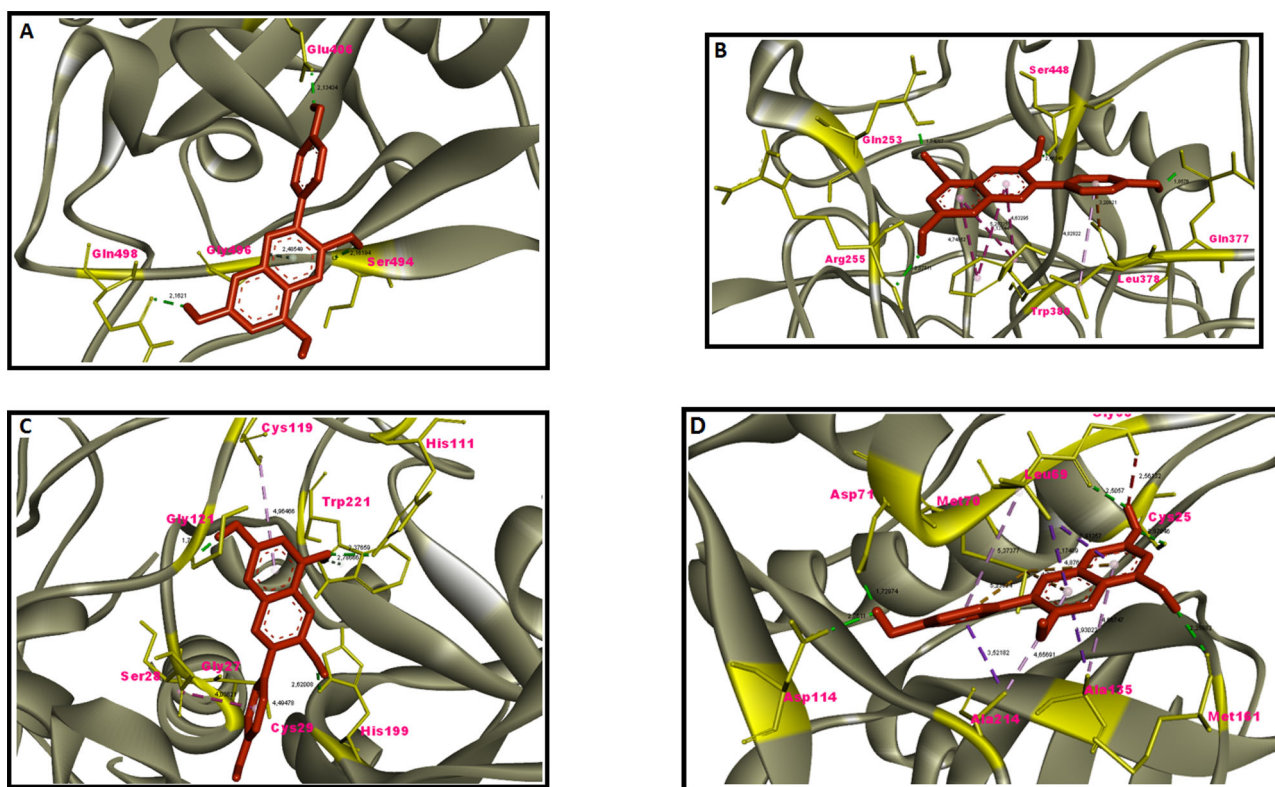


Figure 6. Top ranked conformations of pelargonidin. (A- RBM of the spike glycoprotein of 2019-nCoV, B- Tmprss2, C- CatB, D- CatL).

receptor, the type of amino acids present in the active center and their locations. Therefore, different binding energy profiles were detected in each of the receptor-ligand pairs analyzed in the current study (Tables S1, S2, S3, and S4). According to these results, (-)-Epicatechin gallate, naringin, peonidin, pelargonidin showed high affinity to target proteins. On the other hand, rutin was one of the compounds with the lowest affinity for almost all protein targets. Ligands showed higher affinity to CatB and CatL than other proteins.

In order to determine the effective concentrations of ligands on protein targets, inhibition constants were calculated by AutoDock 4.2.6. The predicted inhibition constants of the flavonoids on the spike glycoprotein, Tmprss2, CatB, and CatL were found to be in the ranges of 0.02-0.58, 0.04-339.73, 0.001-11.82 and 0.001-11.04 mM, respectively. (-)-Epicatechin gallate was found to be highly effective against all receptors.

Van der Waals interactions played an important role in receptor-ligand interplay. Classical and non-classical H bonds, hydrophobic and electrostatic interactions were also frequently found.

In the interaction between flavonoids and the RBD of the spike glycoprotein, the interaction with the amino acid residues of the receptor binding motif (RBM) inside the RBD such as Phe486 and Leu455 remained weak, while the ligands interacted with other residues of the RBM such as Tyr505, Asn501, Ser494, and Gln493 more effectively. Flavonoids also interacted intensely with Arg403, which is not one of the amino acid residues of the RBM of the spike glycoprotein (Figure S3).

Flavonoids interacted with His296 and Ser441, active amino acid residues of Tmprss2, but did not establish any interactions with Asp 345. Other amino acids that interacted with flavonoids were Val280, Lys300, Tyr337, Lys340, Thr341, and Lys342 (Figure S4).

Cys29, Gly27, His111, His199, and Trp221, active amino acid residues, played an important role in the interaction of flavonoids with CatB (Figure S5).

Cys25, Trp26, Gly68, Leu69, Met70, Ala135, Met161, and Asp162, active amino acid residues, played an important role in flavonoid/CatL interactions. Asp114, His163, Gly164, Ala214, Ala215, and Ser216, which are not active amino acid residues of CatL, also contributed significantly to these interactions (Figure S6).

As a result of the RBCl analysis, (-)-epicatechin gallate was the most effective molecule against all protein targets (Figure 2). It was followed by naringin, peonidine and pelargonidine, respectively. The top ranked conformations of these flavonoids ('hit' compounds) were given as Figures 3–6, respectively.

3.2. Drug-likeness properties and ADMET profiles of flavonoids

The drug-likeness properties of flavonoids were given in Table 2. Although rutin, hesperidin and naringin have three violations ($MW > 500$, N or $O > 10$, NH or $OH > 5$) and myricetin, (-)-epicatechin gallate and delphinidin have one violation (NH or $OH > 5$), all other flavonoids were found to have drug-likeness properties without any violation.

Table 2. Drug-likeness properties of docked flavonoids.

No	Compound	Number of rotatable bonds	TPSA ¹	Consensus Log P	Log S (ESOL ²)	Drug-likeness (Lipinski's rule of five)
1	Quercetin	1	131.36 Å ²	1.23	-3.16	Yes (0 Violation)
2	Kaempferol	1	111.13 Å ²	1.58	-3.31	Yes (0 violation)
3	Isorhamnetin	2	120.36 Å ²	1.65	-3.36	Yes (0 violation)
4	Rutin	6	269.43 Å ²	-1.12	-3.30	No (3 violations; MW > 500, N or O > 10, NH/OH > 5)
5	Myricetin	6	151.59 Å ²	0.79	-3.01	Yes (1 Violation; NH or OH > 5)
6	Resveratrol	2	60.69 Å ²	2.48	-3.62	Yes (0 Violation)
7	Naringenin	1	86.99 Å ²	1.84	-3.49	Yes (0 Violation)
8	Hesperidin	7	234.29 Å ²	-0.72	-3.28	No (3 violations; MW > 500, N or O > 10, NH or OH > 5)
9	Naringin	6	225.06 Å ²	-0.79	-2.98	No (3 violations; MW > 500, N or O > 10, NH or OH > 5)
10	Eriodictyol	6	107.22 Å ²	1.45	-3.72	Yes (0 Violation)
11	Genistein	1	90.90 Å ²	2.04	-3.72	Yes (0 Violation)
12	Daidzein	1	70.67 Å ²	2.24	-3.53	Yes (0 Violation)
13	Glycitein	2	79.90 Å ²	2.30	-3.57	Yes (0 Violation)
14	Apigenin	1	90.90 Å ²	2.11	-3.71	Yes (0 Violation)
15	Luteolin	1	111.13 Å ²	1.73	-3.71	Yes (0 Violation)
16	Tangeretin	6	76.36 Å ²	3.02	-4.11	Yes (0 Violation)
17	(-)-Epicatechingallate	4	177.14 Å ²	1.23	-3.70	Yes (1 violation; NH or OH > 5)
18	(+)-Catechin	1	110.38 Å ²	0.85	-2.22	Yes (0 Violation)
19	Delphinidin	1	134.52 Å ²	-0.79	-3.89	Yes (1 violation; NH or OH > 5)
20	Malvidin	3	112.52 Å ²	0.71	-2.86	Yes (0 Violation)
21	Peonidin	2	103.29 Å ²	0.76	-2.81	Yes (0 Violation)
22	Pelargonidin	1	94.06 Å ²	0.73	-2.76	Yes (0 Violation)
23	Petunidin	2	123.52 Å ²	-0.55	-3.36	Yes (0 Violation)

¹TPSA: Topological polar surface area.²ESOL: Estimated aqueous solubility.Data source: <https://www.swissadme.ch>.**Table 3.** ADMET profiles of flavonoids.

No	Compound	BBB ¹ permeation*	P-gp substrate ^{2*}	CYP ³ inhibition [†]	AMES Toxicity [†]	Hepatotoxicity [†]	LD ₅₀ in rat (mol/kg) [†]
1	Quercetin	Yes	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.471
2	Kaempferol	No	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.449
3	Isorhamnetin	No	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.407
4	Rutin	No	Yes	No	No	No	2.491
5	Myricetin	No	No	Yes (CYP1A2, CYP3A4)	No	No	2.497
6	Resveratrol	Yes	No	Yes (CYP1A2, CYP2C9, CYP3A4)	Yes	No	2.529
7	Naringenin	No	Yes	Yes (CYP1A2, CYP3A4)	No	No	1.791
8	Hesperidin	No	Yes	No	No	No	2.506
9	Naringin	No	Yes	No	No	No	2.495
10	Eriodictyol	No	Yes	Yes (CYP3A4)	No	No	2.030
11	Genistein	No	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.268
12	Daidzein	Yes	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.164
13	Glycitein	No	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.170
14	Apigenin	No	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.450
15	Luteolin	No	No	Yes (CYP1A2, CYP2D6, CYP3A4)	No	No	2.455
16	Tangeretin	Yes	No	Yes (CYP2C9, CYP3A4)	No	No	2.368
17	(-)-Epicatechin gallate	No	No	No	No	No	2.558
18	(+)-Catechin	No	Yes	No	No	No	2.428
19	Delphinidin	No	Yes	No	No	No	2.548
20	Malvidin	No	Yes	Yes (CYP1A2)	No	No	2.346
21	Peonidin	No	Yes	Yes (CYP1A2)	No	No	2.408
22	Pelargonidin	No	Yes	Yes (CYP1A2)	No	No	2.432
23	Petunidin	No	Yes	No	No	No	2.453

¹BBB: Blood Brain Barrier.²P-gp: P-glycoprotein substrate.³CYP: Cytochrome P.*<https://www.swissadme.ch>.†<http://biosig.unimelb.edu.au/pkcs/m/prediction>.

ADMET profiles of flavonoids were also given in Table 3. According to the data given in the table, it was understood that only quercetin, resveratrol, daidzein, and tangeretin could cross the blood-brain barrier. Rutin, hesperidin, naringin, (-)-epicatechin gallate, (+)-catechin, delphinidin, and petunidine did not have inhibitory effect on any of cytochrome P. None of the flavonoids showed hepatotoxicity. Apart from resveratrol, all other flavonoids

were also found to show no AMES toxicity. LD₅₀ values of the compounds in rats were in the range of 1.791-2.558 mol/kg.

3.3. Target predictions of 'hit' flavonoids

The possible intracellular targets of (-)-epicatechin gallate, naringin, peonidine, and pelargonidine, were given in

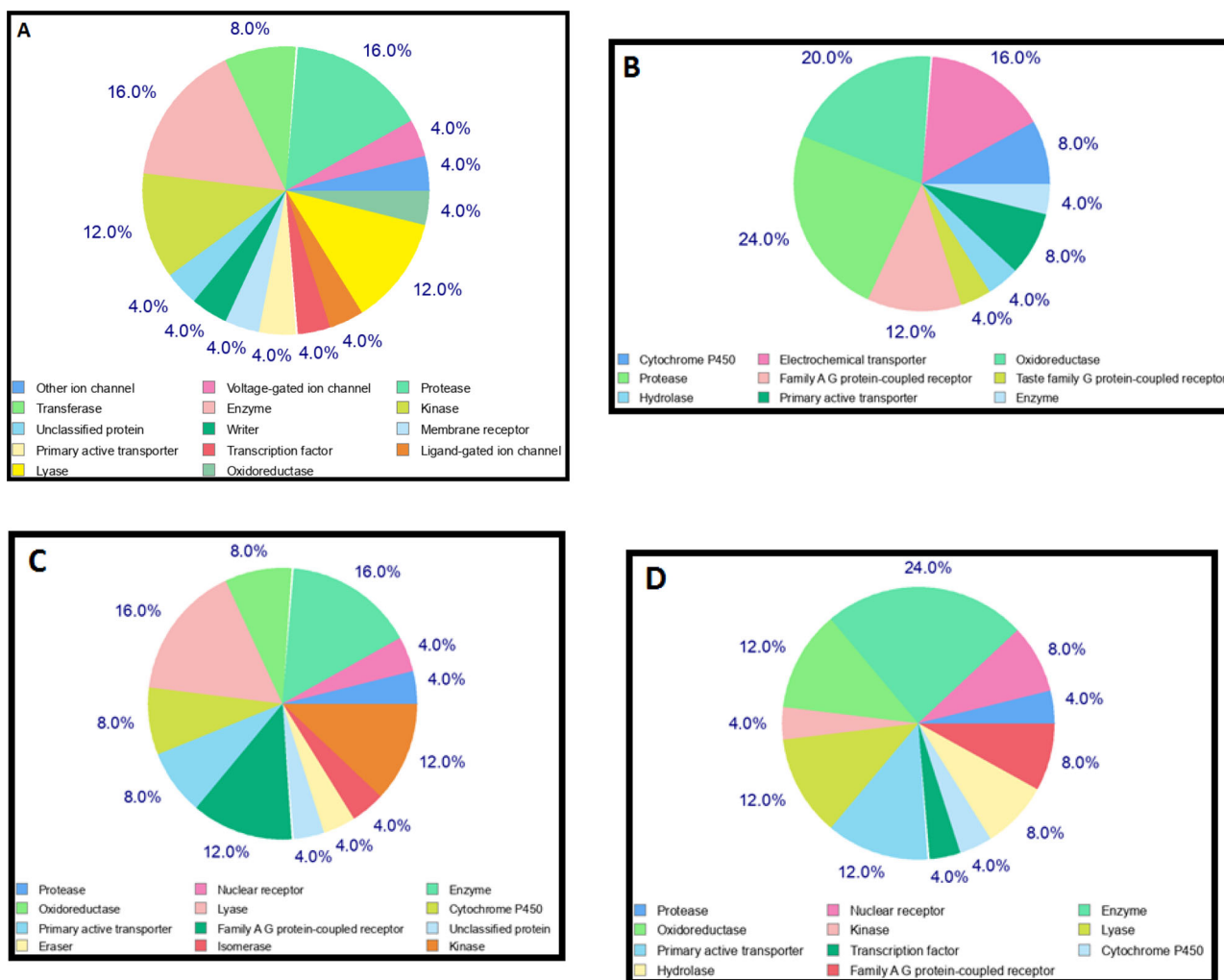


Figure 7. Target predictions of A- (-)-epicatechin gallate, B- Naringin, C- Peonidin, and D- Pelargonidin.

Figure 7(A–D), respectively. (-)-Epicatechin gallate appeared to act on proteases and kinases. While naringin was effective on proteases, oxidoreductases and electrochemical transporters, peonidine had an effect on lyase. Pelargonidine also appeared to be effective on many intracellular enzymes, especially oxidoreductases.

3.4. Results of the molecular dynamics analysis of 'hit' flavonoids

The complex (-)-epicatechin gallate (ECG)-spike protein remained stable and the ECG molecule remained in the proximity of the receptor during all the simulation, with small rearrangements (Figure 8(A)). The ECG molecule started interacting with the RBM of the spike protein (mainly with residues Gln493, Ser494, Gly496, and Tyr505), but also with Glu406, Lys417 and Tyr453. During the simulation, the ECG molecule remained interacting in the RBM, which suffered only small rearrangements. At the end of the simulation the ECG molecule interacted with essentially the same residues as in the beginning, losing only the interactions with Lys417 and Gly496, but interacting with Ile418. The number of

hydrogen bonds remained roughly conserved through the simulation (Figure 8(A)).

The complex ECG-TMPRSS2 was stable during the simulation, with the ligand keeping essentially the same place interacting with the receptor (Figure 8(B)). The ECG molecule started interacting mainly with the residues His296, Tyr337, Asp338, Lys340, Thr341, Leu419 and Trp461 of the protein TMPRSS2. The molecule changed only slightly the position, although some structural rearrangements took place. At the end of the simulation the molecule interacted with the same residues as in the beginning (but losing the interaction with Asp338). There was some fluctuation on the number of hydrogen bonds along the simulation, nevertheless the interactions remained strong (Figure 8(B)).

The ECG molecule interacted strongly with the CatB, maintaining the distance and the majority of the interactions along the simulation (Figure 8(C)), as well as the high number of H-bonds. The ECG molecule remained almost in the same place: it started interacting with the residues Ser25, Cys26, Gly27, Asn72, Gly73, His110, Gly121, Glu122, Ser178, Met196, Gly197 and Gly198 of the receptor. At the end of the simulation, the interactions with Ser25 and Cys26 were lost, but new interactions with Trp30, His199 and Ala200

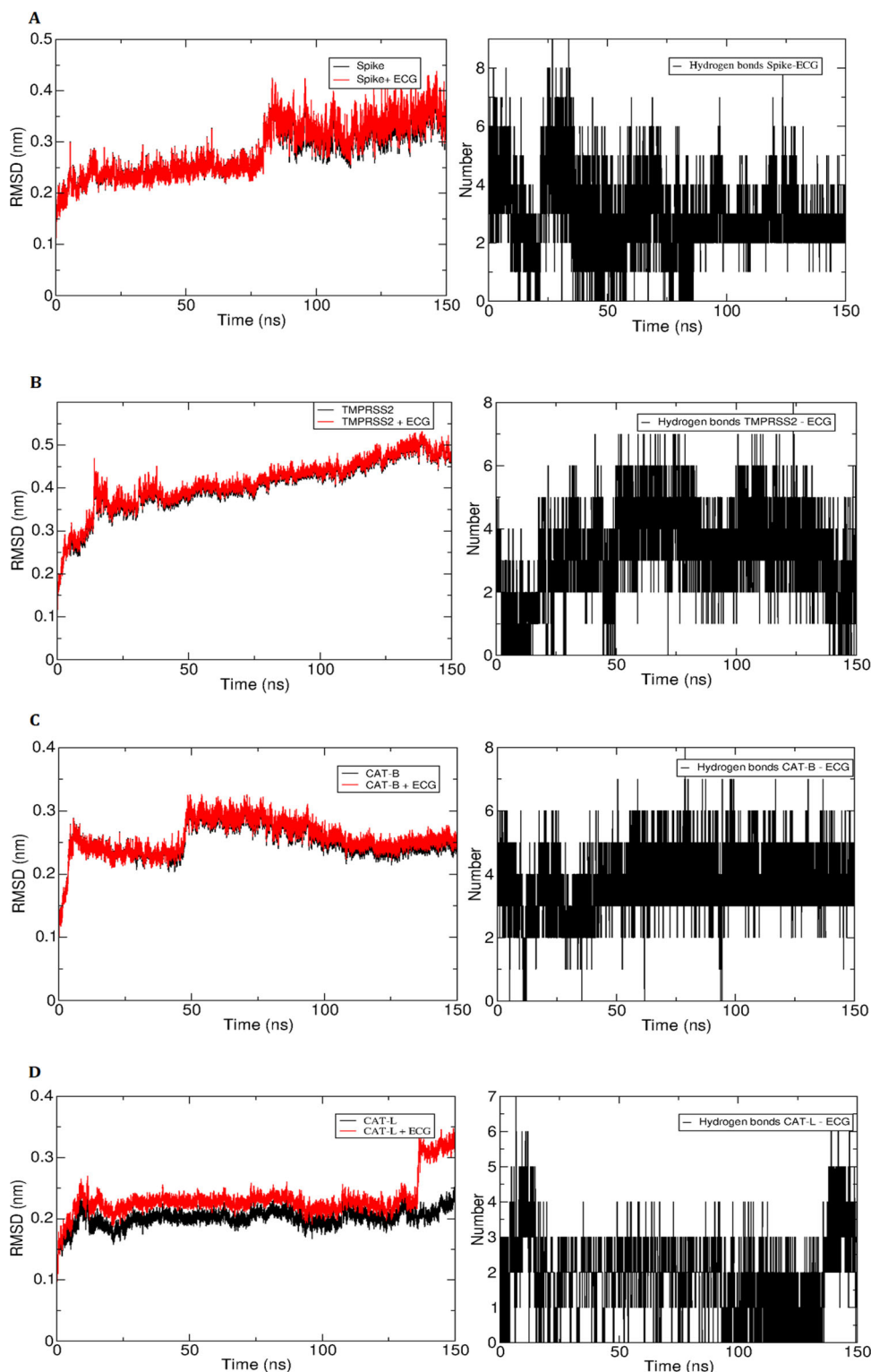


Figure 8. Time evolution of the structure and interactions in the complexes of (-)-epicatechin gallate (ECG) with A- RBD of the spike glycoprotein of 2019-nCoV, B- TMPRSS2, C- CatB, and D- CatL in molecular dynamics simulations. In the left the root-mean-square deviation (RMSD) of atomic positions of the receptor (black) and of the complex ligand + receptor (red) are shown, whereas in the right the number of hydrogen bonds between ligand and receptor is shown.

appeared, with the ECG molecule migrating slightly from the surface to a position more buried inside the protein. However, there were almost no structural changes on the protein (Figure 8(C)).

ECG and CatL, the interactions were moderately strong and some structural changes occurred (Figure 8(D)). There was some loss of hydrogen bonds along the

simulation, with a recovery at the end. In the beginning of the simulation the ECG molecule was interacting mainly with the residues Ser4, Val5, Asp6, Trp7, Arg8, Glu9 and Lys10. The ECG molecule migrated at about 135 ns, losing contacts with the residues 4 until 9 and starting to interact with Gly11, Tyr12, Val13, Thr14 and Pro15 (Figure 8(D)).

Table 4. Binding free energies and their components, calculated using MM/PBSA, for the interaction between ECG and the receptors A- Spike, B- Tmprss2, C- CatB and D- CatL.

Ligand	Protein	Van der Waals Energy	Electrostatic Energy	Polar Solvation Energy	SASA Energy	Binding Energy
ECG	Spike	-116.31 ± 2.30 kJ/mol	-81.18 ± 2.82 kJ/mol	160.83 ± 2.53 kJ/mol	14.71 ± 0.15 kJ/mol	-51.43 ± 1.38 kJ/mol
ECG	Tmprss2	-121.51 ± 1.72 kJ/mol	-93.07 ± 2.89 kJ/mol	173.74 ± 3.06 kJ/mol	-14.87 ± 0.13 kJ/mol	-55.67 ± 1.60 kJ/mol
ECG	CatB	-173.72 ± 2.00 kJ/mol	-119.03 ± 1.71 kJ/mol	202.52 ± 2.32 kJ/mol	-19.02 ± 0.13 kJ/mol	-109.15 ± 1.48 kJ/mol
ECG	CatL	-88.08 ± 1.50 kJ/mol	-75.10 ± 3.21 kJ/mol	155.12 ± 4.02 kJ/mol	-12.13 ± 0.12 kJ/mol	-20.11 ± 1.99 kJ/mol

The binding free energies calculated using the MM/PBSA method (Table 4) agree with the findings of the docking and molecular dynamics: there are favorable interactions between ECG and all receptors, as can be seen in the negative binding free energies. The strongest interaction was found for ECG with CatB and the weakest for ECG with CatL, in agreement with the time evolution of the number of hydrogen bonds and with the qualitative analyses of the simulations.

Results of the molecular dynamics and binding free energy analysis of other 'hit' flavonoids (naringin, peonidin and pelargonidin) were also given in Figures S7, S8 and S9 and Table S6, respectively in the Supplementary file.

It is impossible to discuss the literature data on all flavonoids studied in the current study. Instead, it would be more logical to discuss the literature data of 'hit' flavonoids on target proteins. Literature data obtained from the docking analysis of the phytochemicals analyzed in the current study was also presented in Table S5 in detail (Supplementary file).

Some recent studies have shown that catechins can be used as alternative agents in COVID-19 infection. According to Maiti and Banerjee (2020), catechins, such as epigallocatechin gallate or theaflavin gallate, interact more strongly with the amino acids of the spike protein than hydroxychloroquine. There are also some data in the literature that some catechins, such as theaflavin digallate, epigallocatechin gallate and galocatechin gallate, inhibit the main protease of 2019-nCoV together with the spike glycoprotein (Manish, 2020; Peterson, 2020; Sayed et al., 2020; Tallei et al., 2020). In addition to 2019-nCoV, molecular docking studies performed on (-)-epicatechin 3-O-(3'-O-methyl) gallate, a green tea polyphenol, showed that the molecule in question can inhibit Tmprss2, a cellular protease (Rahman et al., 2020). Catechins such as (-)-catechin-3-gallate, (+)-epicatechin-3-gallate and (-)-epigallocatechin-gallate also have an inhibitory effect on cathepsins, a group of lysosomal enzymes (Majumdar et al., 2011; Zhang et al., 2012). In a study carried out by Devika and Prince (2008), a significant decrease in cathepsin B and D activities were detected in Wistar rats given epigallocatechin gallate at a dose of 30 mg/kg for 21 days compared to the control group. However, there are no reports in the literature directly about the effect of epicatechin gallate itself on CatB and especially CatL. Therefore, data presented in the current study on the interaction of (-)-epicatechin gallate with CatB and CatL could be assumed as the first record for the literature.

In the literature, there were no reports of the interaction of peonidine and pelargonidine with target proteins. On the other hand, only one study was found regarding the interaction of naringin with the RBD of the spike protein of 2019-

nCoV (Bhowmik et al., 2020). According to the results of this study conducted using PyRx and iGEMDOCK, it was determined that the binding affinity of naringin to spike protein was -8.3 kcal/mol.

4. Conclusion

In this computational study, the interactions of 23 different phytochemicals belonging to different flavonoid subgroups with the RBD of the spike glycoprotein of 2019-nCoV, Tmprss2, CatB and CatL were analyzed. Anthocyanidins, isoflavones and flavanones were found to have more powerful interactions with the target proteins. Considering the interactions of phytochemicals with all target proteins, it was concluded that (-)-epicatechin gallate was more effective than other compounds. This molecule showed drug-likeness properties according to Lipinski's rule of five. Although (-)-epicatechin gallate cannot cross the blood-brain barrier, it was concluded that it can be considered as a candidate molecule in the drug development processes against 2019-nCoV since it did not have any toxic effect on eukaryotic cells.

Acknowledgements

P. A. N. acknowledges financial support from CNPq and CAPES (Financial code 001), as well as the Centro Nacional de Supercomputação (CESUP), Universidade Federal do Rio Grande do Sul (UFRGS) for providing HPC resources.

Disclosure statement

The authors have no conflict of interest to declare.

Author contributions

E.S.İ. performed homology modeling of Tmprss2 and molecular docking analysis, contributed to the writing and revision of the manuscript; P.A.N. conducted the molecular dynamics simulations and MM/PBSA analysis, interpreted the results, and contributed to the writing and revision of the manuscript; A.S.T. contributed to the clarification of the structure-effect relationship of phytochemicals and the preparation of the manuscript; M.T.H. analyzed the results of experiments, proofread and reviewed the manuscript; C.S. carried out two-dimensional drawing of phytochemicals, contributed to the development of the RBCI parameter; B.T. contributed to the clarification of the structure-effect relationship of phytochemicals and drafted the manuscript.

ORCID

Erman Salih Istifli  <http://orcid.org/0000-0003-2189-0703>
Paulo A. Netz  <http://orcid.org/0000-0003-4242-0591>

Arzuhan Sihoglu Tepe  <http://orcid.org/0000-0001-8290-9880>
 Mehmet Tahir Husunet  <http://orcid.org/0000-0003-1424-5132>
 Cengiz Sarikurkcu  <http://orcid.org/0000-0001-5094-2520>
 Bektaş Tepe  <http://orcid.org/0000-0001-8982-5188>

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Anguiano, L., Riera, M., Pascual, J., & Soler, M. (2017). Circulating ACE2 in cardiovascular and kidney diseases. *Current Medicinal Chemistry*, 24(30), 3231–3241. <https://doi.org/10.2174/0929867324666170414162841>
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), 10037–10041. <https://doi.org/10.1073/pnas.181342398>
- Bhowmik, D., Nandi, R., & Kumar, D. (2020). Evaluation of flavonoids as 2019-nCoV cell entry inhibitor through molecular docking and pharmacological analysis. *ChemRxiv*. Preprint. <https://doi.org/10.26434/chemrxiv.12071508.v1>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Carletti, G., Nervo, G., & Cattivelli, L. (2014). Flavonoids and melanins: A common strategy across two kingdoms. *International Journal of Biological Sciences*, 10(10), 1159–1170. <https://doi.org/10.7150/ijbs.9672>
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society*, 2(9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>
- Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, 7, 42717. <https://doi.org/10.1038/srep42717>
- Daina, A., Michielin, O., & Zoete, V. (2019). Swiss target prediction: Updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Research*, 47(W1), W357–W364. <https://doi.org/10.1093/nar/gkz382>
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98(12), 10089–10092. <https://doi.org/10.1063/1.464397>
- de Andrade, J., Gonçalves, P. F. B., & Netz, P. A. (2020). How does the novel coronavirus interact with the human ACE2 enzyme? A thermodynamic answer. *ChemRxiv*. Preprint. <https://doi.org/10.26434/chemrxiv.12361238.v2>
- Devika, P., & Prince, P. S. M. (2008). Preventive effect of (–) epigallocatechin-gallate (EGCG) on lysosomal enzymes in heart and subcellular fractions in isoproterenol-induced myocardial infarcted Wistar rats. *Chemico-Biological Interactions*, 172(3), 245–252.
- Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., & Kollman, P. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16), 1999–2012. <https://doi.org/10.1002/jcc.10349>
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *Journal of Chemical Physics*, 103(19), 8577–8593.
- Greenspan, P. D., Clark, K. L., Tommasi, R. A., Cowen, S. D., McQuire, L. W., Farley, D. L., van Duzer, J. H., Goldberg, R. L., Zhou, H., Du, Z., Fitt, J. J., Coppa, D. E., Fang, Z., Macchia, W., Zhu, L., Capparelli, M. P., Goldstein, R., Wigg, A. M., Doughty, J. R., Bohacek, R. S., & Knap, A. K. (2001). Identification of dipeptidyl nitriles as potent and selective inhibitors of cathepsin B through structure-based drug design. *Journal of Medicinal Chemistry*, 44(26), 4524–4534. <https://doi.org/10.1021/jm010206q>
- Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30(S1), S162–S173.
- Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., Gsell, B., Thoma, R., Diez, J., Benz, J., Plancher, J.-M., Hartmann, G., Isshiki, Y., Morikami, K., Shimma, N., Haap, W., Banner, D. W., & Diederich, F. (2011). Halogen bonding at the active sites of human cathepsin L and MEK1 kinase: Efficient interactions in different environments. *ChemMedChem*, 6(11), 2048–2054. <https://doi.org/10.1002/cmdc.201100353>
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2), 271–280. <https://doi.org/10.1016/j.cell.2020.02.052>
- Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics poisson-boltzmann surface area method. *Molecular Informatics*, 31(2), 114–122. <https://doi.org/10.1002/minf.201100135>
- Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical Review. A, General Physics*, 31(3), 1695–1697. <https://doi.org/10.1103/PhysRevA.31.1695>
- Huang, I.-C., Bosch, B. J., Li, F., Li, W., Lee, K. H., Ghiran, S., Vasileva, N., Dermody, T. S., Harrison, S. C., Dormitzer, P. R., Farzan, M., Rottier, P. J. M., & Choe, H. (2006). SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. *The Journal of Biological Chemistry*, 281(6), 3198–3203. <https://doi.org/10.1074/jbc.M508381200>
- Istifli, E. S., Sihoglu Tepe, A., Sarikurkcu, C., & Tepe, B. (2020). Interaction of certain monoterpene hydrocarbons with the receptor binding domain of 2019 novel coronavirus (2019-nCoV), transmembrane serine protease 2 (TMPRSS2), cathepsin B, and cathepsin L (CatB/L) and their pharmacokinetic properties. *Turkish Journal of Biology = Turk Biyoloji Dergisi*, 44(3), 242–264. <https://doi.org/10.3906/biy-2005-46>
- Jakalian, A., Jack, D. B., & Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16), 1623–1641. <https://doi.org/10.1002/jcc.10128>
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2), 926–935. <https://doi.org/10.1063/1.445869>
- Kumari, R., Kumar, R., & Lynn, A., Open Source Drug Discovery Consortium. (2014). g_mmpbsa-a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, 54(7), 1951–1962. <https://doi.org/10.1021/ci500020m>
- Lam, T. T.-Y., Shum, M. H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., Wei, W., Cheung, W. Y.-M., Li, W.-J., Li, L.-F., Leung, G. M., Holmes, E. C., Hu, Y.-L., & Guan, Y. (2020). Identification of 2019-nCoV related coronaviruses in Malayan pangolins in Southern China. *BioRxiv*. <https://doi.org/10.1101/2020.1102.1113.945485>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215–220. <https://doi.org/10.1038/s41586-020-2180-5>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291.
- Letko, M., Marzi, A., & Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, 5(4), 562–569. <https://doi.org/10.1038/s41564-020-0688-y>

- Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science (New York, N.Y.)*, 309(5742), 1864–1868. <https://doi.org/10.1126/science.1116480>
- Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965), 450–454. <https://doi.org/10.1038/nature02145>
- Luan, J., Lu, Y., Jin, X., & Zhang, L. (2020). Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochemical and Biophysical Research Communications*, 526(1), 165–169. doi: 10.1016/j.bbrc.2020.1003.1047
- Maiti, S., & Banerjee, A. (2020). Epigallocatechin-gallate and theaflavin-gallate interaction in SARS CoV-2 spike-protein central-channel with reference to the hydroxychloroquine interaction: bioinformatics and molecular docking study. *Preprints 2020*, 2020040247. <https://doi.org/10.20944/preprints202004.0247.v1>
- Majumdar, I. D., Devanabanda, A., Fox, B., Schwartzman, J., Cong, H., Porco, J. A. Jr, & Weber, H. C. (2011). Synthetic cyclohexenyl chalcone natural products possess cytotoxic activities against prostate cancer cells and inhibit cysteine cathepsins in vitro. *Biochemical and Biophysical Research Communications*, 416(3–4), 397–402. <https://doi.org/10.1016/j.bbrc.2011.11.052>
- Mandal, S. M., Chakraborty, D., & Dey, S. (2010). Phenolic acids act as signaling molecules in plant-microbe symbioses. *Plant Signaling & Behavior*, 5(4), 359–368. <https://doi.org/10.4161/psb.5.4.10871>
- Manish, M. (2020). Studies on computational molecular interaction between SARS-CoV-2 main protease and natural products. *ChemRxiv*. Preprint. <https://doi.org/10.26434/chemrxiv.12024789.v2>
- Middleton, E., Kandaswami, C., & Theoharides, T. C. (2000). The effects of plant flavonoids on mammalian cells: Implications for inflammation, heart disease, and cancer. *Pharmacological Reviews*, 52(4), 673–751.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791. <https://doi.org/10.1002/jcc.21256>
- Morris, G. M., & Lim-Wilby, M. (2008). *Molecular docking molecular modeling of proteins*. (pp. 365–382): Springer.
- Nasab, R. R., Hassanzadeh, F., Khodarahmi, G. A., Rostami, M., Mirzaei, M., Jahanian-Najafabadi, A., & Mansourian, M. (2017). Docking study, synthesis and antimicrobial evaluation of some novel 4-anilinoquinazoline derivatives. *Research in Pharmaceutical Sciences*, 12(5), 425–433. <https://doi.org/10.4103/1735-5362.213988>
- Nosé, S. (1984). A unified formulation of the constant temperature molecular-dynamics methods. *Journal of Chemical Physics*, 81(1), 511–519. doi: 10.1063/1.447334
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3, 33 <https://doi.org/10.1186/1758-2946-3-33>
- Omotuyi, I. O., Nash, O., Ajiboye, O. B., Iwegbulam, C. G., Oyinloye, E. B., Oyediji, O. A., Kashim, Z. A., & Okaiyeto, K. (2020). Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein-enhanced fitness in SARS-COV-2. *Journal of Computational Chemistry*, 41(24), 2158–2161. doi: 10.1002/jcc.26383
- Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12), 7182–7190.
- Pedretti, A., Villa, L., & Vistoli, G. (2004). VEGA—an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *Journal of Computer-Aided Molecular Design*, 18(3), 167–173.
- Peterson, L. (2020). COVID-19 and flavonoids: In silico molecular dynamics docking to the active catalytic site of SARS-CoV and SARS-CoV-2 main protease. Available at SSRN 3599426.
- Pires, D. E., Blundell, T. L., & Ascher, D. B. (2015). pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based Signatures. *Journal of Medicinal Chemistry*, 58(9), 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>
- Rahman, N., Basharat, Z., Yousuf, M., Castaldo, G., Rastrelli, L., & Khan, H. (2020). Virtual screening of natural products against Type II transmembrane serine protease (TMPRSS2), the priming agent of coronavirus 2 (SARS-CoV-2). *Molecules*, 25(10), 2271.
- Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, 9(2), 173–175. <https://doi.org/10.1038/nmeth.1818>
- Sanner, M. F. (1999). Python: A programming language for software integration and development. *Journal of Molecular Graphics & Modelling*, 17(1), 57–61.
- Sayed, A. M., Khattab, A. R., AboulMagd, A. M., Hassan, H. M., Rateb, M. E., Zaid, H., & Abdelmohsen, U. R. (2020). Nature as a treasure trove of potential anti-SARS-CoV drug leads: A structural/mechanistic rationale. *RSC Advances*, 10(34), 19790–19802.
- Sharma, S. (1996). *Two group discriminant analysis. Applied multivariate techniques* (pp. 237–286). John Wiley & Sons, Inc.
- Sousa da Silva, A. W., & Vranken, W. F. (2012). ACPYPE - AnteChamber python parser interface. *BMC Research Notes*, 5, 367. <https://doi.org/10.1186/1756-0500-5-367>
- Studer, G., Rempfer, C., Waterhouse, A. M., Gumieny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics (Oxford, England)*, 36(6), 1765–1771. <https://doi.org/10.1093/bioinformatics/btz828>
- Sudhan, D. R., & Siemann, D. W. (2015). Cathepsin L targeting in cancer treatment. *Pharmacology & Therapeutics*, 155, 105–116. <https://doi.org/10.1016/j.pharmthera.2015.08.007>
- Tallei, T. E., Tumilaar, S. G., Niode, N. J., Fatimawali, F., Kepel, B. J., Idroes, R., & Effendi, Y. (2020). Potential of plant bioactive compounds as SARS-CoV-2 main protease (Mpro) and Spike (S) glycoprotein inhibitors: A molecular docking study. *Preprints 2020*, 2020040102. <https://doi.org/10.20944/preprints202004.0102.v1>
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157–1174. <https://doi.org/10.1002/jcc.20035>
- Wang, Y., Liu, M., & Gao, J. (2020). Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 117(25), 13967–13974. <https://doi.org/10.1073/pnas.2008209117>
- Wilson, S., Greer, B., Hooper, J., Zijlstra, A., Walker, B., Quigley, J., & Hawthorne, S. (2005). The membrane-anchored serine protease, TMPRSS2, activates PAR-2 in prostate cancer cells. *The Biochemical Journal*, 388(Pt 3), 967–972. <https://doi.org/10.1042/BJ20041066>
- Wong, M. C., Cregeen, S. J. J., Ajami, N. J., & Petrosino, J. F. (2020). Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *BioRxiv*. <https://doi.org/10.1101/2020.1102.1107.939207>
- Woo, H., Park, S. J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., Kern, N. R., Lee, J., Yeom, M. S., Croll, T. I., Seok, C., & Im, W. (2020). Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *Journal of Physical Chemistry B*, 124(33), 7128–7137. <https://doi.org/10.1021/acs.jpcc.0c04553>
- Worldometers.info. (2020). COVID-19 Coronavirus Pandemic. <https://www.worldometers.info/coronavirus/>
- Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., & Jiang, T. (2020a). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host & Microbe*, 27(3), 325–328. <https://doi.org/10.1016/j.chom.2020.02.001>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020b). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265–269.
- Zakaryan, H., Arabyan, E., Oo, A., & Zandi, K. (2017). Flavonoids: Promising natural compounds against viral infections. *Archives of*

- Virology*, 162(9), 2539–2551. <https://doi.org/10.1007/s00705-017-3417-y>
- Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine*, 367(19), 1814–1820. <https://doi.org/10.1056/NEJMoa1211721>
- Zhang, Y., Yang, N.-D., Zhou, F., Shen, T., Duan, T., Zhou, J., Shi, Y., Zhu, X.-Q., & Shen, H.-M. (2012). (–) (-)-Epigallocatechin-3-gallate induces non-apoptotic cell death in human cancer cells via ROS-mediated lysosomal membrane permeabilization. *PLoS One*, 7(10), e46749. <https://doi.org/10.1371/journal.pone.0046749>
- Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, L. L. M., Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, J. S. M., & Guan, Y. (2003). Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *The Lancet*, 362(9393), 1353–1358.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273.
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W., China Novel Coronavirus Investigating and Research Team. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*, 382(8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>