

Molecular interactions of some phenolics with 2019-nCoV and related pathway elements

Erman Salih Istifli¹, Arzuhan Sihoglu-Tepe², Cengiz Sarikurkcu³, Bektas Tepe^{4,*}

¹Cukurova University, Faculty of Science and Literature, Department of Biology, TR-01330, Adana-TURKEY

²Kilis 7 Aralik University, Vocational High School of Health Services, Department of Pharmacy Services, Kilis, Turkey

³Afyonkarahisar Health Sciences University, Faculty of Pharmacy, Department of Analytical Chemistry, TR-03100, Afyonkarahisar-TURKEY

⁴Kilis 7 Aralik University, Faculty of Science and Literature, Department of Molecular Biology and Genetics, TR-79000, Kilis-TURKEY

Abstract: As of June 2021, the novel coronavirus disease (SARS-CoV-2) resulted in 180 million cases worldwide and resulted in the death of approximately 4 million people. However, an effective pharmaceutical with low side effects that can be used in the treatment of SARS-CoV-2 infection has not been developed yet. The aim of this computational study was to analyze the interactions of twenty-two hydroxycinnamic acid and hydroxybenzoic acid derivatives with the SARS-CoV-2 receptor binding domain (RBD) and host organism's proteases, transmembrane serine protease 2 (TMPRSS2), and cathepsin B and L (CatB/L). According to the RBCI analysis, the ligands with the highest affinity against 4 enzymes in the molecular docking study were determined as 1-caffeoyl- β -D-glucose, rosmarinic acid, 3-*p*-coumaroylquinic acid and chlorogenic acid. It has also been observed that these compounds interacted more strongly with spike RBD, CatB and CatL enzymes. Although the top-ranked ligand, 1-caffeoyl- β -D-glucose, violated the drug-likeness criteria at 1 point (NH or OH>5) and ADMET in terms of AMES toxicity, the second top-ranked ligand rosmarinic acid neither violated drug-likeness nor exhibited incompatibility in terms of ADMET. In conclusion, with its anti-inflammatory properties, rosmarinic acid can be considered and further investigated as a plant-based pharmaceutical that can offer a treatment option in SARS-CoV-2 infection. However, our findings should be supported by additional *in vitro* and *in vivo* studies.

ARTICLE HISTORY

Received: June 28, 2021

Accepted: July 26, 2021

KEYWORDS

SARS-CoV-2,
Molecular docking,
Rosmarinic acid,
Spike glycoprotein,
TMPRSS2,
Cathepsins.

1. INTRODUCTION

Corona Virus Disease 2019 (COVID-19), which spread all over the world within months after emerging in China, infected more than 180 million people as of June 2021 and caused the death of approximately four million people (Wu *et al.*, 2020; Zhu *et al.*, 2020; Worldometers.info, 2021). The World Health Organization (WHO) defined the microorganism causing this disease as the new type of Severe Acute Respiratory Syndrome Corona Virus 2 (2019-nCoV) (Wu *et al.*, 2020; Zhou *et al.*, 2020; Zhu *et al.*, 2020).

*CONTACT: Bektas TEPE ✉ bektastepe@yahoo.com 📍 Kilis 7 Aralik University, Faculty of Science and Literature, Department of Molecular Biology and Genetics, TR-79000, Kilis-TURKEY

It has been determined that the genome of 2019-nCoV consists of a single-stranded RNA molecule with a length of approximately 30,000 nucleotides. As a result of genome analysis, genes encoding the four main components of the virus (nucleocapsid protein, envelope protein, membrane protein and spike glycoprotein) were identified (Wu *et al.*, 2020). Among these proteins, especially the spike glycoprotein is the basic structural molecule that enables the virus to bind to receptors on the host cell surface. Therefore, it is one of the most targeted molecules in treatment development strategies against COVID-19 (Luan *et al.*, 2020).

Angiotensin converting enzyme 2 (ACE2) is the most critical host component that plays a role in the binding of 2019-nCoV to the host cell. The virus binds to ACE2 via spike glycoprotein, thus ACE2 functions as the key for virus entry into the cell (Li *et al.*, 2003; Li *et al.*, 2005). The majority of researchers trying to develop an effective drug against COVID-19 have turned to targeting the 2019-nCoV-ACE2 interaction (Letko *et al.*, 2020; Zhou *et al.*, 2020). Focusing on elucidating the details of the binding interaction between the spike glycoprotein and ACE2, researchers found contact between some amino acids of the viral component in question (Leu455, Phe486, Glu493, Ser494, Asp501 and Tyr505) and mammalian ACE2 (Zakaryan *et al.*, 2017; Andersen *et al.*, 2020).

Of course, ACE2 is not the only molecule mediating the entry of 2019-nCoV into the host cell. In addition to this molecule, transmembrane protease, serine 2 (TMPRSS2) and cathepsins (CatB and L) were also found to assist in the entry of 2019-nCoV into the cell. TMPRSS2 is a serine protease and cleaves the spike glycoprotein from the S1/S2 site, thus facilitating virus entry and activation (Hoffmann *et al.*, 2020). Cathepsins (CatB and CatL) are lysosomal endopeptidases that function in the host cell (Huang *et al.*, 2006; Sudhan & Siemann, 2015; Hoffmann *et al.*, 2020).

In this study, it was aimed to determine the interaction of some hydroxybenzoic and hydroxycinnamic acids with 2019-nCoV spike glycoprotein and host proteases (TMPRSS2, CatB and CatL).

2. MATERIAL and METHODS

2.1. Structural optimization of ligands

The Protein Data Bank (PDB) files of benzoic acid, 4-hydroxybenzoic acid, *p*-coumaric acid, caffeic acid, salicylic acid, protocatechuic acid, ferulic acid, cinnamic acid, vanillic acid, isovanillic acid, chlorogenic acid, gallic acid, syringic acid, rosmarinic acid, 1-*o*-feruloyl- β -D-glucose, 3-feruloylquinic acid, N-caffeoyl-L-aspartic acid, 1-caffeoyl- β -D-glucose, 3-*p*-coumaroylquinic acid, 6-*o-p*-coumaroyl-D-glucose, *p*-coumaroyltartaric acid, and chicoric acid (Figure 1) were downloaded from PubChem (<https://pubchem.ncbi.nlm.nih.gov>). Detailed information on the structural optimization of ligands was given in the [Supplementary file](#).

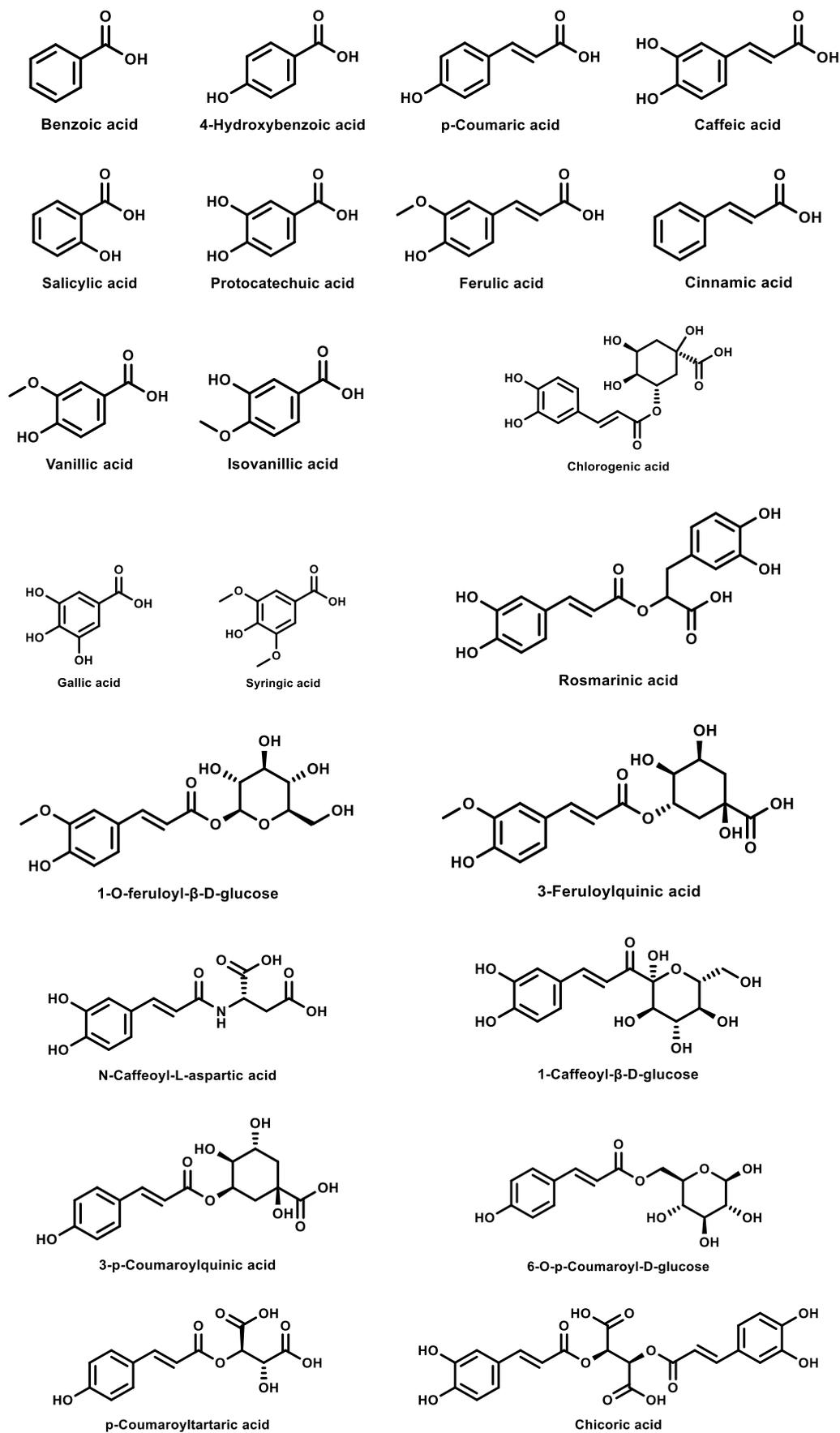
2.2. Protein retrieval and energy minimization of the Spike, TMPRSS2, CatB and CatL proteins using Nanoscale Molecular Dynamics (NAMD)

Detailed information on protein retrieval and energy minimization of the spike and other target proteins are given in the [Supplementary file](#).

2.3. Molecular docking analyses of phenolic acids

Detailed information on molecular docking analyses of phenolic acids is given in the [Supplementary file](#).

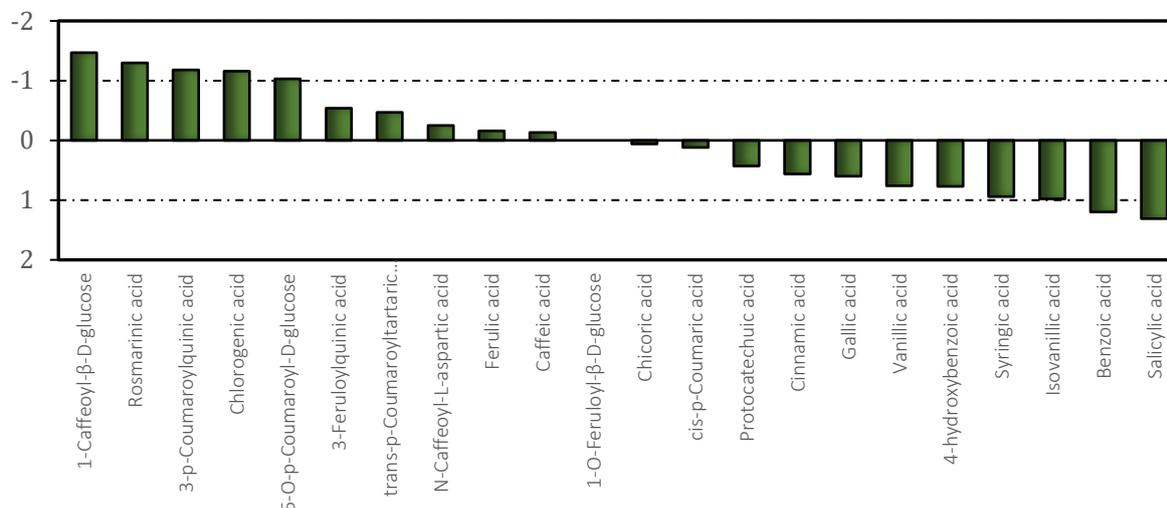
Figure 1. Chemical structures of the phenolic acids.



2.4. Calculation of relative binding capacity index (RBCI)

Detailed information on calculation of RBCI is given in the supplementary file ([Figure 2](#) and [Table S1](#)).

Figure 2. RBCI of the phenolic acids.



2.5. Drug-likeness, ADMET profile and target prediction

Detailed information on drug-likeness, ADMET profile and target prediction is given in the [Supplementary file](#).

2.6. Network pharmacology analysis

Detailed information on network pharmacology analysis is given in the [Supplementary file](#).

3. RESULTS / FINDINGS

3.1. Molecular docking results of phenolic acids

In this study, the molecular level interactions of 22 phenolic acids based on molecular docking analyses with target proteins spike glycoproteins, TMPRSS2, CatB and CatL were calculated; however, based on the RBCI analysis, the non-bonded interactions of only 4 compounds (1-caffeoyl-β-D-glucose, rosmarinic acid, 3-*p*-coumaroylquinic acid and chlorogenic acid) determined as 'hit' as a result of RBCI analysis are given ([Tables S2](#), [S3](#), [S4](#) and [S5](#)). On the other hand, the binding free energies (ΔG , kcal/mol) and calculated inhibition constants (K_i , μM) of the 22 phytochemicals are given in [Table 1](#).

The interactions of 1-caffeoyl-β-D-glucose with the receptor binding motif (RBM) of spike glycoprotein are shown in ([Figure 3A](#)). 1-caffeoyl-β-D-glucose formed a total of nine H-bonds with residues Arg403, Gln409, Lys417, Tyr453, Gly496, Tyr505 and Ser494. 1-caffeoyl-β-D-glucose also formed a pi-alkyl interaction with Lys417 of the RBM ([Figure 3A](#)). The interactions of 1-caffeoyl-β-D-glucose with the active site of TMPRSS2 are given in [Figure 3B](#). Analysis of the docking results of this compound revealed that the 1-caffeoyl-β-D-glucose formed mainly 6 H-bonds with residues Val280, His2961, Ser394, Gly439, Ser441 and an electrostatic interaction with Thr393 ([Figure 3B](#)). The molecular scale interactions of 1-caffeoyl-β-D-glucose with the active site of CatB mainly consisted of H-bonds with residues Gln231, Gly241, Gly271, Cys291, Gly72 and His197. It was also detected that 1-caffeoyl-β-D-glucose firmly fit the active site of this enzyme ([Figure 3C](#)). 1-caffeoyl-β-D-glucose formed a total of 7 H-bonds, a pi-sulfur interaction and two pi-alkyl interactions with residues Cys25, Met70, Asp71, Met161, Asp114, Ala135, Asp162 and Ala215 of CatL ([Figure 3D](#)).

Table 1. Free energy of binding and calculated inhibition constant values of the compounds.

No	Compound	Free energy of binding (kcal/mol)				Calculated inhibition constant (μM)			
		Spike RBD	TMPRSS2	CatB	CatL	Spike RBD	TMPRSS2	CatB	CatL
1	1-Caffeoyl- β -D-glucose	-5.97	-4.71	-6.42	-7.13	42.22	354.40	19.54	5.92
2	Rosmarinic acid	-6.13	-5.10	-5.44	-6.21	31.88	181.71	103.50	28.01
3	3- <i>p</i> -Coumaroylquinic acid	-6.38	-4.57	-5.96	-6.31	21.14	449.38	43	23.78
4	Chlorogenic acid	-5.78	-4.92	-6.19	-5.90	57.51	246.66	29.23	47.02
5	6- <i>o-p</i> -Coumaroyl-D-glucose	-4.87	-4.79	-6.30	-6.75	268.59	310.29	24.16	11.25
6	3-Feruloylquinic acid	-5.61	-4.46	-5.10	-5.90	76.70	534.14	184.15	47.29
7	<i>p</i> -Coumaroyltartaric acid	-5.47	-5.16	-4.24	-4.82	97.13	163.79	780.79	290.71
8	N-Caffeoyl-L-aspartic acid	-4.64	-5.16	-4.57	-4.75	395.03	165.48	449.57	328.05
9	Ferulic acid	-4.66	-4.61	-4.67	-5.66	385.89	418.48	378.67	70.64
10	3,4-Dihydroxycinnamic acid	-4.45	-4.69	-4.44	-5.78	544.16	362.04	556.84	57.82
11	1- <i>o</i> -Feruloyl- β -D-glucose	-4.83	-4.02	-5.55	-5.52	288.37	1140	85.52	89.98
12	Chicoric acid	-5.63	-4.32	-4.31	-4.66	74.48	675.85	690.09	386.65
13	<i>p</i> -Coumaric acid	-4.42	-4.57	-4.02	-5.56	577.62	447.42	1130	83.34
14	Protocatechuic acid	-4.30	-4.75	-3.60	-4.49	706.93	329.83	2310	507.99
15	Cinnamic acid	-4.42	-4.14	-3.80	-5.17	573.77	925.57	1650	162.17
16	Gallic acid	-4.29	-4.61	-3.45	-4.34	722.07	418.15	2970	657.23
17	Vanillic acid	-4.28	-4.33	-3.61	-4.30	727.49	669.40	2270	699.55
18	4-hydroxybenzoic acid	-4.19	-4.42	-3.45	-4.27	846.04	572.24	2970	738.17
19	Syringic acid	-4.21	-4.07	-3.83	-4.14	817.59	1040	1560	918.80
20	Isovanillic acid	-4.20	-4.14	-3.41	-4.20	839.55	915.66	3190	828.60
21	Benzoic acid	-4.15	-3.91	-3.42	-4.00	906.77	1370	3130	1170
22	Salicylic acid	-3.71	-3.96	-3.29	-4.12	1910	1240	3910	961.79

The favorable interactions of rosmarinic acid with the active site (RBM) of spike protein mainly consisted of 6 H-bonds and one pi-alkyl interaction with the residues Arg403, Tyr449, Ser494, Gly496, Gln498 and Tyr505 (Figure 4A). The rosmarinic acid-TMPRSS2 interactions included a total of 4 conventional, 2 carbon-hydrogen and one pi-donor H-bonds with residues His279, Val280, Thr393, Ser460, Gly462 and Ser441. Rosmarinic acid also formed hydrophobic contacts with His296 and Thr293 of TMPRSS2 (Figure 4B). Rosmarinic acid interacted with Gln23, Gly24, Cys29, Asn70, His108, Gly119, Glu120, Gly196 and His197 residues of CatB via a total of 9 H-bonds (Figure 4C). The binding interactions of rosmarinic acid in complex with CatL were mainly consisted of a total of 4 H-bonds with the residues Gly68, Met70, Lys117 and Ser213, an electrostatic interaction with Asp71, two pi-sigma interactions with Ala135 and 214, one pi-sulfur interaction with Met70, and two hydrophobic amide-pi stacked and pi-alkyl interactions with Ala214, 215 and Leu69, respectively (Figure 4D).

Figure 3. Top ranked conformations of 1-caffeoyl- β -D-glucose (A- RBM of the spike glycoprotein of SARS-CoV-2, B- Tmprss2, C- CatB, D- CatL).

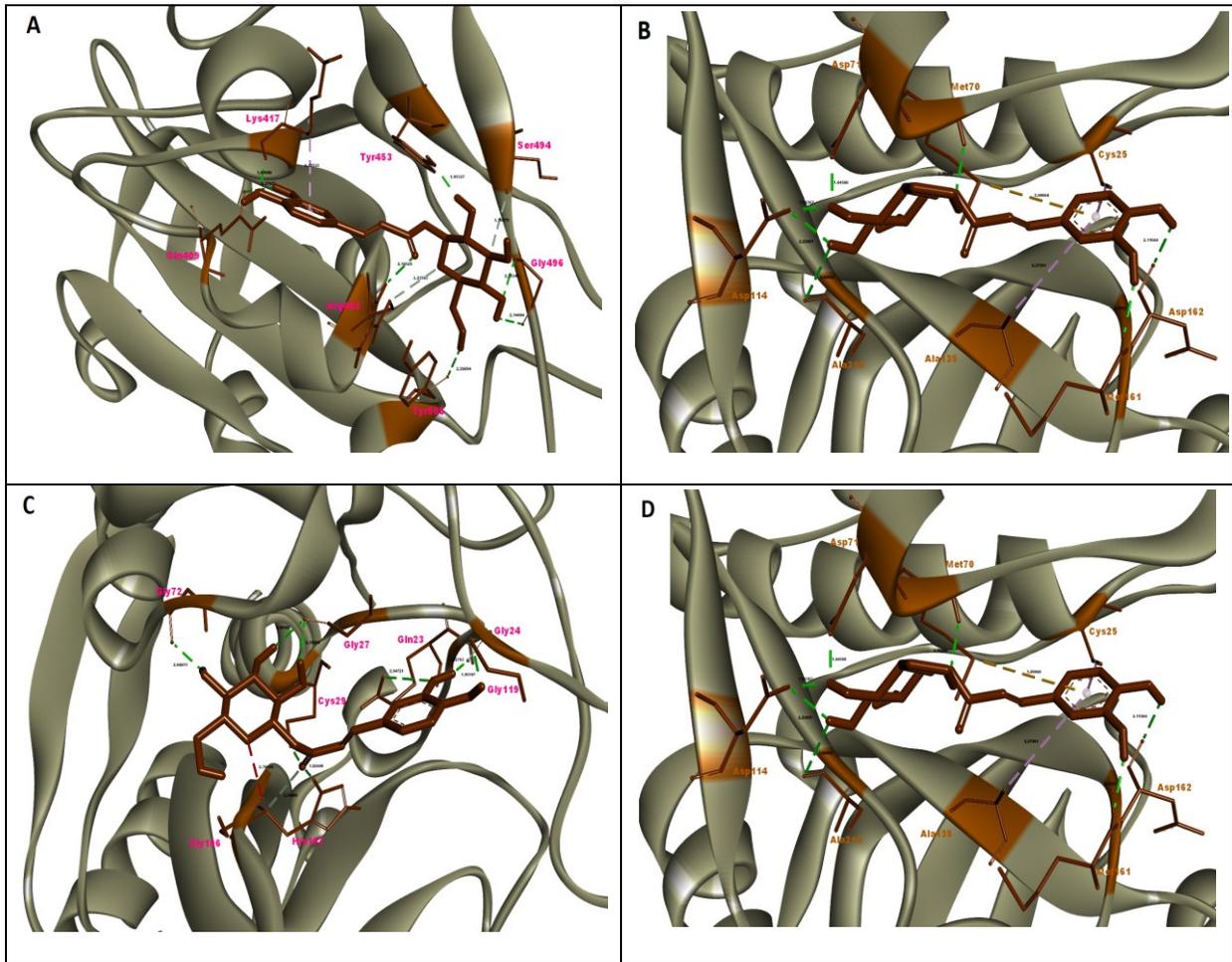
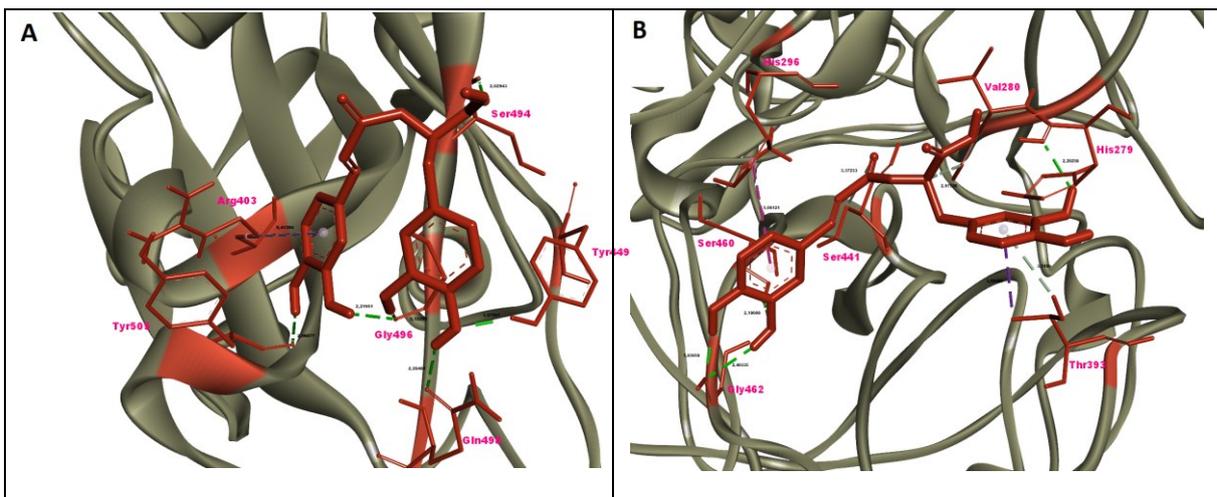
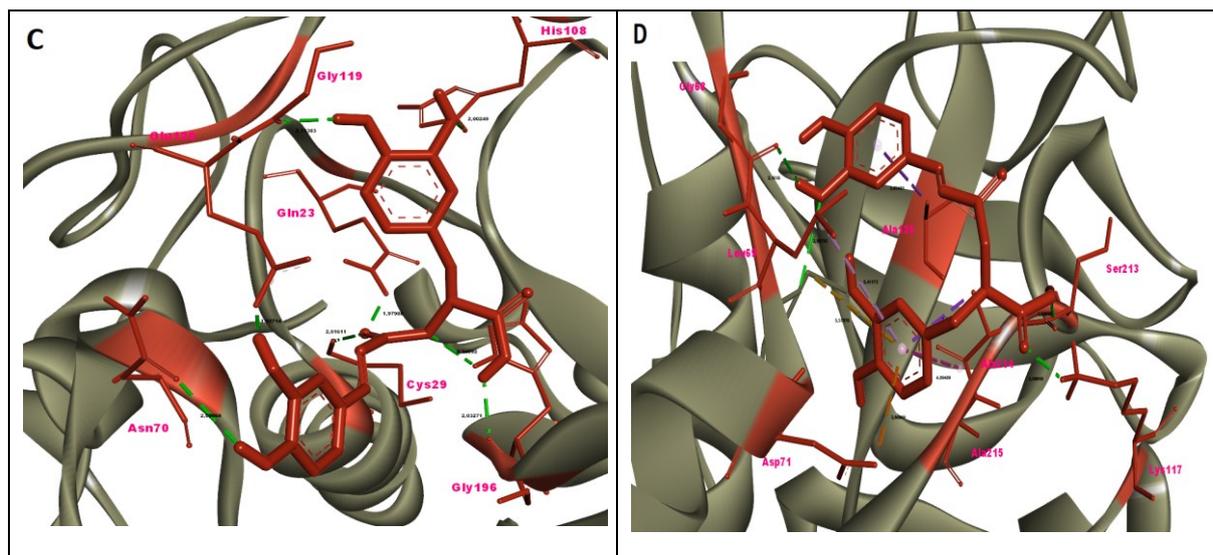


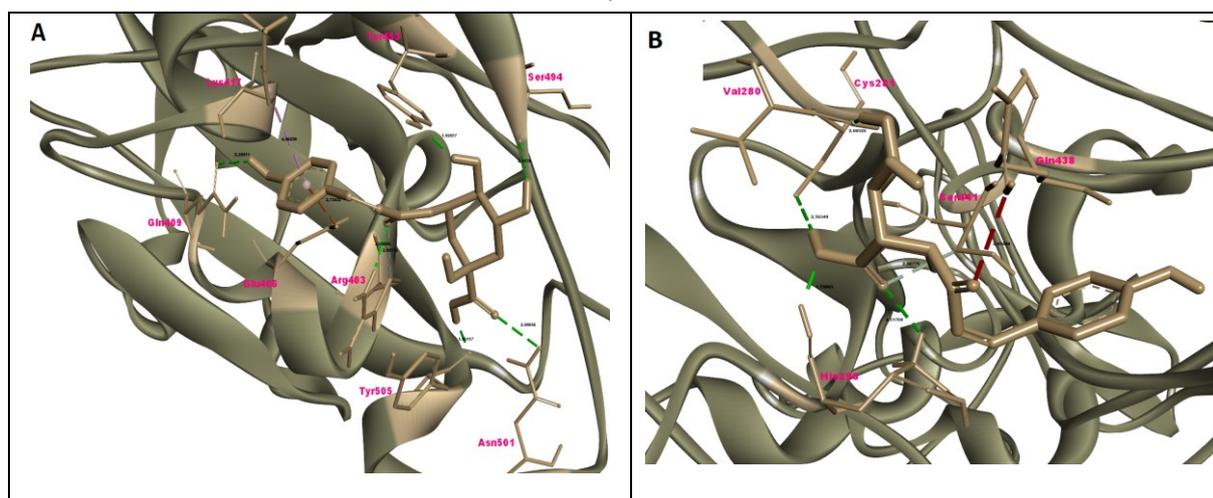
Figure 4. Top ranked conformations of rosmarinic acid (A- RBM of the spike glycoprotein of SARS-CoV-2, B- Tmprss2, C- CatB, D- CatL).

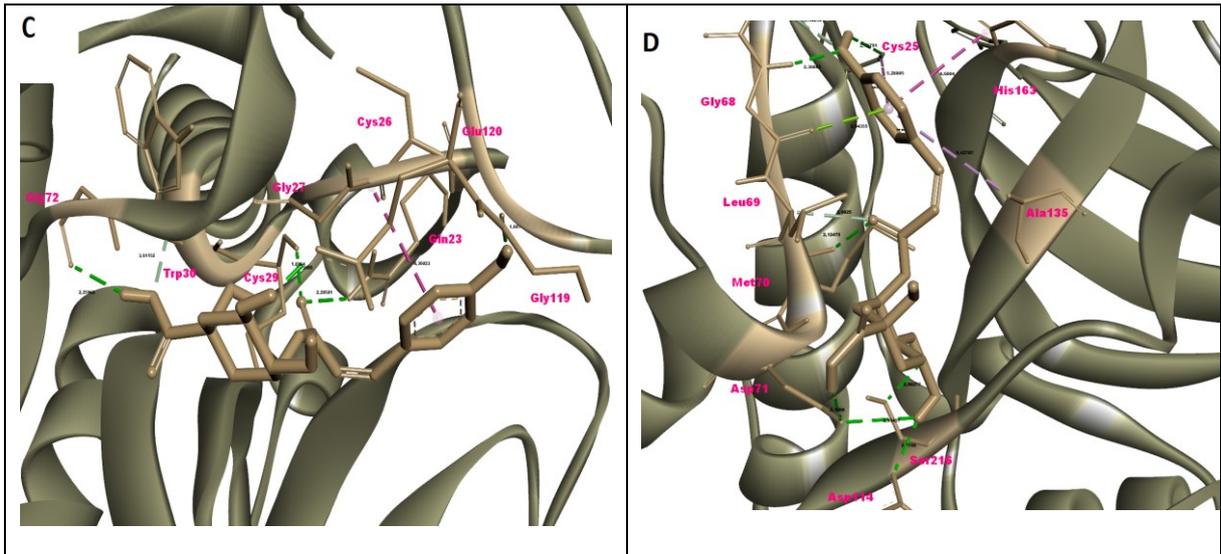




3-*p*-coumaroylquinic acid formed a total of 7 H-bonds with residues Arg403, Gln409, Tyr453, Ser494, Asn501 and Tyr505, one electrostatic interaction with Glu406 and one pi-alkyl interaction with Lys417 of RBM of the spike glycoprotein (Figure 5A). Analysis of the interactions of 3-*p*-coumaroylquinic acid with TMPRSS2 revealed that the compound in question mainly formed 5 H-bonds with residues Val280, Cys281, His296 and Ser441 (Figure 5B). Figure 5C depicts the binding mode of 3-*p*-coumaroylquinic acid with CatB enzyme. 3-*p*-coumaroylquinic acid interacts with Gln23, Cys29, Trp30, Gly72, Gly119 and Glu120 residues via a total of 7 H-bonds and forms hydrophobic contact with Cys26 and Gly27 (Figure 5C). 3-*p*-coumaroylquinic acid interacts with Cys25, Gly67, Gly68, Leu69, Met70, Asp71, Asp114 and Ser216 residues of CatL via a total of 9 H-bonds. It also shows hydrophobic contacts with Cys25, Ala135, Asp162 and His163 residues of the enzyme (Figure 5D).

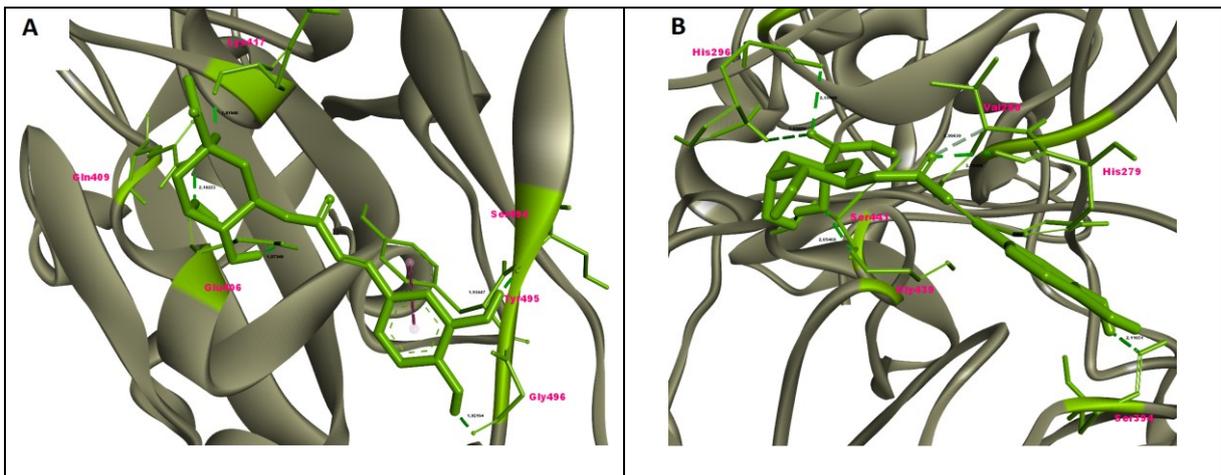
Figure 5. Top ranked conformations of 3-*p*-coumaroylquinic-acid (A- RBM of the spike glycoprotein of SARS-CoV-2, B- TMPRSS2, C- CatB, D- CatL).

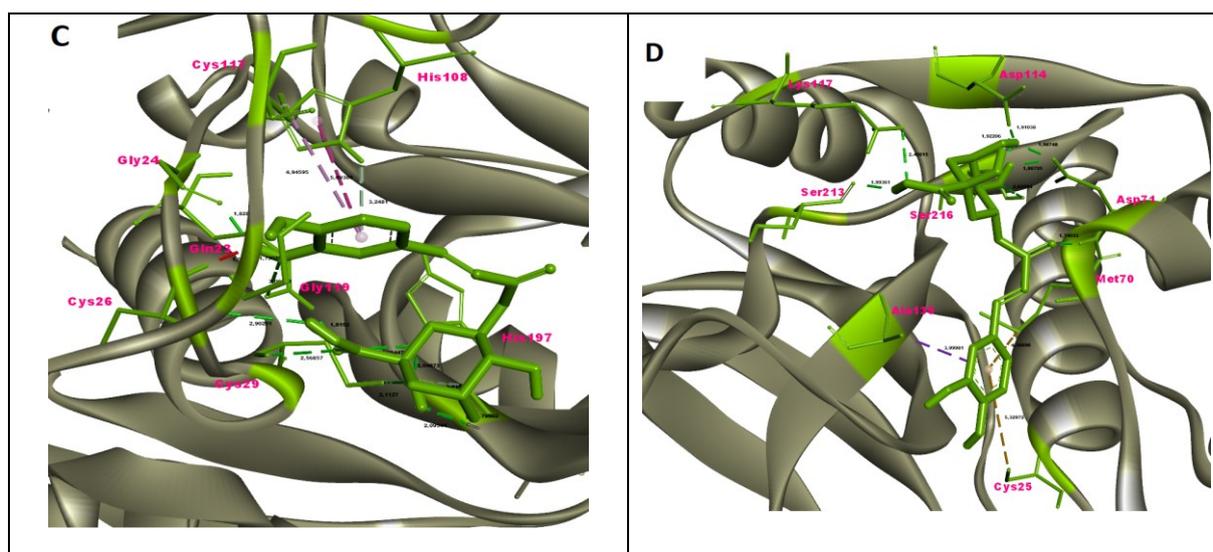




The interactions of chlorogenic acid with the active site of spike glycoprotein mainly consists of 5 H-bonds with residues Glu406, Gln409, Lys417, Ser494 and Gly496 as well as a hydrophobic pi-pi interaction with Tyr495 (Figure 6A). Chlorogenic acid formed a total of 7 H-bonds with the active site residues His279, Val280, His296, Ser394, Gly439 and Ser441 of TMPRSS2 (Figure 6B). Chlorogenic acid-CatB interactions included a total of 13 H-bonds with residues Gln23, Gly24, Cys26, Cys29, His108, Gly119, Gly196 and His197. Chlorogenic acid also formed two hydrophobic interactions with His108 and Cys117 (Figure 6C). The binding interactions of chlorogenic acid in complex with CatL showed a total of 8 H-bonds with residues Asp71, Asp114, Lys117, Ser213 and Ser216 (Figure 6D). Chlorogenic acid formed a hydrophobic pi-sigma interaction with Ala135 and two pi-sulfur interactions with Cys25 and Met70.

Figure 6. Top ranked conformations of chlorogenic acid (A- RBM of the spike glycoprotein of SARS-CoV-2, B- TMPRSS2, C- CatB, D- CatL).





As a result, hydrogen bonds constitute the greatest contribution to the interactions of the top-ranked 4 ligands with their receptors in molecular docking analyzes.

3.2. RBCI values of phenolic acids

In this study, molecular scale interactions between 22 different phenolic acids and 4 different protein targets were investigated using molecular docking analysis. As a result of molecular docking calculations, the binding free energy values (kcal/mol) of each phytochemical against 4 proteins in question were used to calculate the relative binding capacity index (RBCI). The details of this method have been explained in detail in the 'Materials and methods' section. Using the RBCI value, the efficiency of phytochemicals on different proteins can be calculated using different data sets at the same time (Figure 2 and Table S1).

As a result of the RBCI analysis we applied, it was determined that 1-caffeoyl- β -D-glucose, rosmarinic acid, 3-*p*-coumaroylquinic acid and chlorogenic acid were the 'hit' compounds among 22 phytochemicals. Top ranked poses of 1-caffeoyl- β -D-glucose, rosmarinic acid, 3-*p*-coumaroylquinic acid and chlorogenic acid on the RBD of spike glycoprotein, TMPRSS2, CatB and CatL are presented in Figures 3, 4, 5, and 6, respectively. RBCI analysis also revealed that isovanillic acid, benzoic acid and salicylic acid were the ligands with the weakest affinity for 4 different protein targets.

3.3. Pharmacokinetic properties of phenolic acids

Drug-likeness, ADMET and target profiles of analyzed phenolic acids against spike glycoprotein, TMPRSS2, CatB and CatL are given in Tables 2 and 3, respectively. Except for chicoric acid, all phenolic acids were determined to obey the Lipinski's rule-of-five. Chicoric acid violates this rule because it has N or O > 10, or NH or OH > 5. It was determined that no ligand other than ferulic acid, *p*-coumaric acid, cinnamic acid, 4-hydroxybenzoic acid, benzoic acid and salicylic acid could pass blood-brain barrier (BBB). It has also been determined that no compound other than chicoric acid is a substrate of P-glycoprotein (P-gp). Protocatechuic acid and gallic acid were observed to inhibit CYP3A4, while no other compounds showed an inhibitory effect on cytochrome enzymes. While 1-caffeoyl- β -D-glucose, 6-*o*-*p*-coumaroyl-D-glucose and 1-*o*-feruloyl- β -D-glucose show mutagenicity in the AMES bacterial test system, no other phenolic acid studied showed AMES toxicity. On the other hand, only *p*-coumaroyltartaric acid and N-caffeoyl-L-aspartic acid showed hepatotoxic effect data among 22 ligands studied. The LD₅₀ doses of the molecules in the rat ranged from 1.737 mol/kg to 2.811 mol/kg. While the most toxic molecule in the rat was 3-*p*-coumaroylquinic acid (1.737 mol/kg), the phytochemical with the least toxicity was rosmarinic acid (2.811 mol/kg).

Table 2. Drug-likeness properties of docked phenolic acids.

No	Compound	Number of rotatable bonds	TPSA ¹	Consensus Log P	Log S (ESOL ²)	Drug likeness (Lipinski's rule of five)
1	1-Caffeoyl-β-D-glucose	5	156.91	-0.98	-1.41	Yes; 1 violation: NH or OH>5
2	Rosmarinic acid	7	144.52	1.52	-3.44	Yes; 0 violation
3	3-p-Coumaroylquinic acid	5	144.52	-0.03	-1.75	Yes; 0 violation
4	Chlorogenic acid	5	164.75	-0.38	-1.62	Yes; 1 violation: NH or OH>5
5	6-O-p-Coumaroyl-D-glucose	5	136.68	-0.52	-1.20	Yes; 0 violation
6	3-Feruloylquinic acid	6	153.75	0.01	-1.84	Yes; 0 violation
7	p-Coumaroyltartaric acid	7	147.02	-0.42	-1.68	Yes; 0 violation
8	N-Caffeoyl-L-aspartic acid	7	144.16	0.05	-1.45	Yes; 0 violation
9	Ferulic acid	3	66.76	1.36	-2.11	Yes; 0 violation
10	3,4-Dihydroxycinnamic acid	2	77.76	0.93	-1.89	Yes; 0 violation
11	1-O-Feruloyl-β-D-glucose	6	145.91	-0.37	-1.64	Yes; 0 violation
12	Chicoric acid	11	208.12	1.01	-3.58	No; 2 violations: N or O>10, NH or OH>5
13	p-Coumaric acid	2	57.53	1.26	-2.02	Yes; 0 violation
14	Protocatechuic acid	1	77.76	0.65	-1.86	Yes; 0 violation
15	Cinnamic acid	2	37.30	1.79	-2.37	Yes; 0 violation
16	Gallic acid	1	97.99	0.21	-1.64	Yes; 0 violation
17	Vanillic acid	2	66.76	1.08	-2.02	Yes; 0 violation
18	4-hydroxybenzoic acid	1	57.53	1.05	-2.07	Yes; 0 violation
19	Syringic acid	3	75.99	1.02	-1.84	Yes; 0 violation
20	Isovanillic acid	2	66.76	1.07	-2.02	Yes; 0 violation
21	Benzoic acid	1	37.30	1.44	-2.20	Yes; 0 violation
22	Salicylic acid	1	57.53	1.24	-2.50	Yes; 0 violation

¹ TPSA: Topological polar surface area (Å²)

² ESOL: Estimated aqueous solubility [(Insoluble < -10 < Poorly < -6 < Moderately < -4 < Soluble < -2 Very < 0 < Highly), according to Delaney, J.S. (2004)].

Data source: <http://www.swissadme.ch/index.php#>

3.4. Target predictions and network pharmacology of hit phenolic acids

In this study, the screening of hit phytochemicals against possible intracellular targets was performed using Swiss Target Prediction (<http://www.swisstargetprediction.ch/>) and STITCH 5.0 (<http://stitch.embl.de/>) public databases.

In **Figure 7A**, the intracellular targets of 1-caffeoyl-β-D-glucose are given. According to **Figure 7A**, various enzymes (28%), lyases (18%) and proteases (16%) appear to make up the majority of the intracellular targets of 1-caffeoyl-β-D-glucose. However, the interaction probability of 1-caffeoyl-β-D-glucose for these enzyme groups was found to be no more than 10% ($p = 0.1$). Therefore, the interaction of this phytochemical with possible intracellular targets is not statistically significant.

Table 3. ADMET profiles of phenolic acids.

No	Compound	BBB permeation ^{1,*}	P-gp substrate ^{2,*}	CYP inhibition ^{3,*}	AMES Toxicity ⁴	Hepatotoxicity ⁴	LD ₅₀ in rat (mol/kg) ⁴
1	1-Caffeoyl-β-D-glucose	No	No	No	Yes	No	1.938
2	Rosmarinic acid	No	No	No	No	No	2.811
3	3-p-Coumaroylquinic acid	No	No	No	No	No	1.737
4	Chlorogenic acid	No	No	No	No	No	1.973
5	6-O-p-Coumaroyl-D-glucose	No	No	No	Yes	No	1.933
6	3-Feruloylquinic acid	No	No	No	No	No	2.025
7	p-Coumaroyltartaric acid	No	No	No	No	Yes	2.295
8	N-Caffeoyl-L-aspartic acid	No	No	No	No	Yes	2.084
9	Ferulic acid	Yes	No	No	No	No	2.282
10	3,4-Dihydroxycinnamic acid	No	No	No	No	No	2.383
11	1-O-Feruloyl-β-D-glucose	No	No	No	Yes	No	2.157
12	Chicoric acid	No	Yes	No	No	No	2.445
13	p-Coumaric acid	Yes	No	No	No	No	2.155
14	Protocatechuic acid	No	No	CYP3A4	No	No	2.423
15	Cinnamic acid	Yes	No	No	No	No	2.094
16	Gallic acid	No	No	CYP3A4	No	No	2.218
17	Vanillic acid	No	No	No	No	No	2.454
18	4-hydroxybenzoic acid	Yes	No	No	No	No	2.255
19	Syringic acid	No	No	No	No	No	2.157
20	Isovanillic acid	No	No	No	No	No	2.487
21	Benzoic acid	Yes	No	No	No	No	2.170
22	Salicylic acid	Yes	No	No	No	No	2.282

¹ BBB: Blood Brain Barrier

² P-gp: P-glycoprotein substrate

³ CYP: Cytochrome P

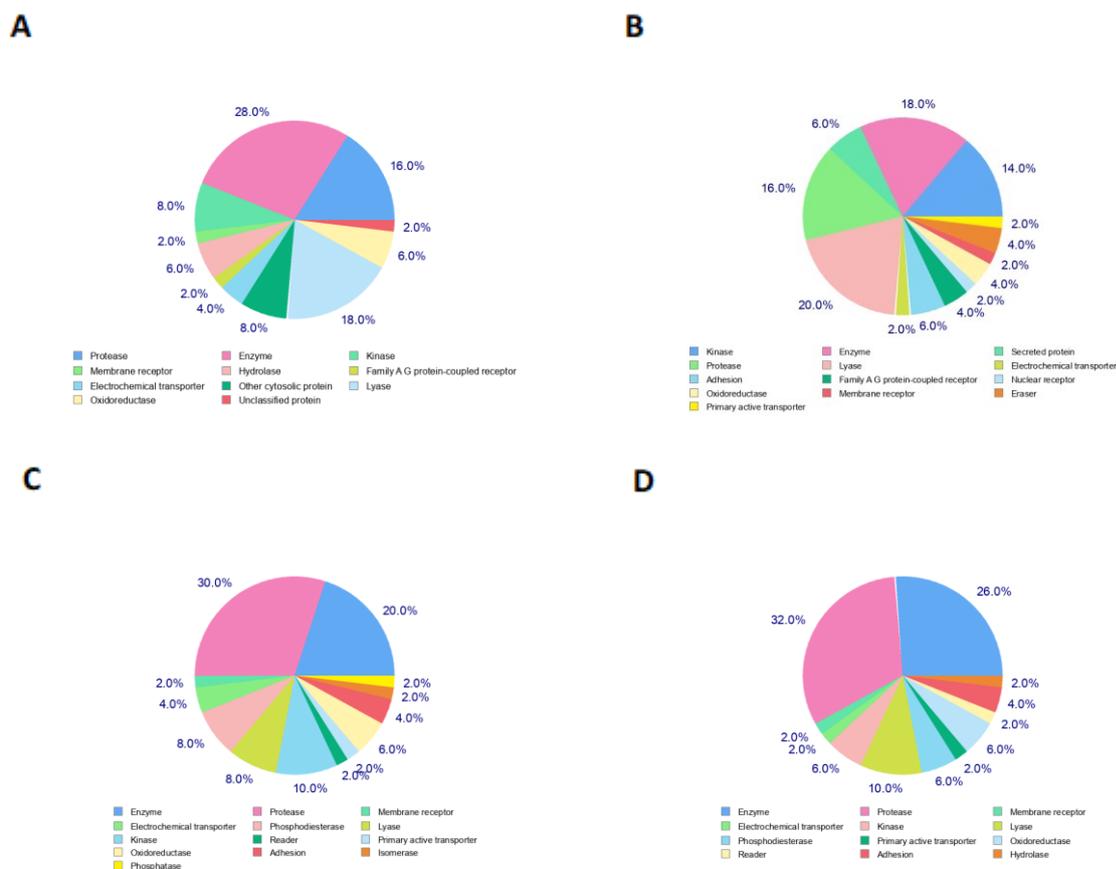
⁴ <http://biosig.unimelb.edu.au/pkcsim/prediction>

* <https://www.swissadme.ch>

Figure 7B shows that possible intracellular targets of rosmarinic acid comprise lyases (20%), various enzymes (18%), proteases (16%), and kinases (14%). Quite impressively, rosmarinic acid is likely to interact with each of these targets at 96% ($p = 0.96$), a statistically very significant value. This finding suggests that rosmarinic acid exerts possible inhibitory/activator effect on these enzyme groups. Most of the intracellular targets of 3-*p*-coumaroylquinic acid consist of proteases (30%), various enzymes (20%) and kinases (10%) (Figure 7C). 3-*p*-coumaroylquinic interacts with these targets at a statistically insignificant probability level ($p = 0.10 - 0.41$).

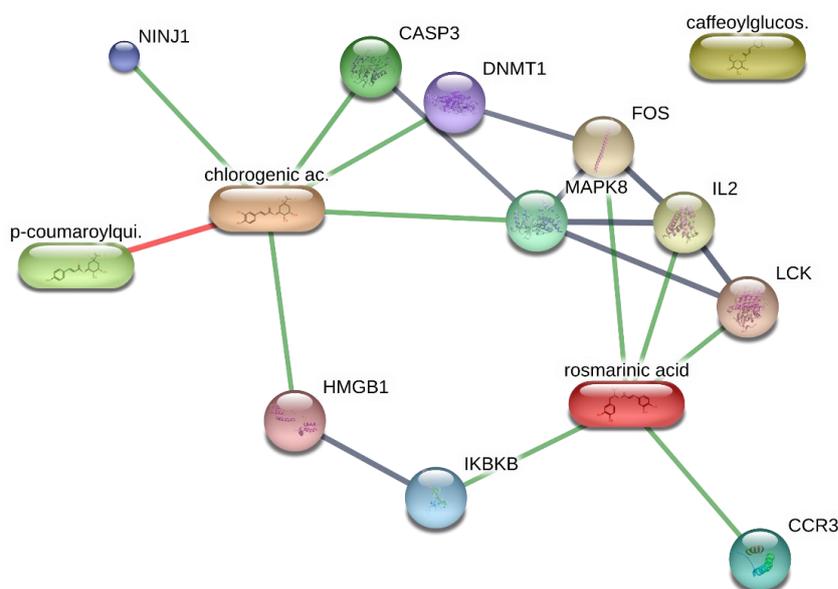
Intracellular targets of chlorogenic acid generally consist of proteases (32%), various enzymes (26%) and lyases (10%) (Figure 7D). Although chlorogenic acid shows a statistically insignificant level of probability ($p = 0.09 - 0.33$) in terms of interaction with the proteases and lyases, its interaction with various intracellular enzymes appears slightly significant ($p = 0.77 - 0.80$).

Figure 7. Target prediction of A - 1-caffeoyl- β -D-glucose B - rosmarinic acid, C - 3-*p*-coumaroylquinic acid, D - chlorogenic acid.



The STITCH platform was used to predict putative targets of hit phenolic acids in the human proteome. Thus, the direct interactions of 1-caffeoyl- β -D-glucose, rosmarinic acid, 3-*p*-coumaroylquinic acid and chlorogenic acid with different proteins were mapped (Figure 8). Prior to mapping target-component interactions, the minimum required interaction score was set to a high confidence score which was ≥ 0.7 . A high confidence score indicates a strong interaction between hit phytochemicals and protein(s). According to the targets-components interaction network in Figure 8, it can be observed that 2 ligands other than rosmarinic acid and chlorogenic acid do not directly interact with the human proteome. Rosmarinic acid interacts with FOS, IL2, LCK, CCR3 and IKBKB proteins, while chlorogenic acid interacts with NINJ1, CASP3, DNMT1, MAPK8 and HMGB1 proteins. For rosmarinic acid, it is noteworthy that kinases (LCK, IKBKB) play a central role among protein targets predicted by both SwissTargetPrediction and STITCH platforms (Figures 7A, 7B, 7C, 7D and Figure 8).

Figure 8. Targets-components analysis (chemical-protein interactions) of top 4 hit phytochemicals performed via the STITCH platform (<http://stitch.embl.de>). Note the explicit role of rosmarinic acid in the targets-components network.



4. DISCUSSION and CONCLUSION

Although the novel coronavirus disease (SARS-CoV-2), which emerged in Wuhan in 2019, affected the whole world as a pandemic in a short time, an effective drug molecule has not yet been found in the treatment of this virus. Therefore, a safe orally administered pharmaceutical would be a breakthrough in the treatment of SARS-CoV-2. In this context, organic molecules of plant origin always maintain their potential among possible treatment options against viral diseases (Chavez *et al.*, 2006; Ruibo *et al.*, 2017; Taguchi *et al.*, 2017; Wu *et al.*, 2017; Jahan & Onay, 2020; Piccolella *et al.*, 2020; Cano-Avenidaño *et al.*, 2021). On the other hand, although herbal compounds are biologically diverse and accessible, drug discovery is an arduous process and molecular docking-based bioinformatics approach has become an increasingly important tool in this field in recent years (Meng *et al.*, 2011).

In the present study, the molecular interactions of twenty-two polyphenolic compounds, derivatives of hydroxycinnamic acid and hydroxybenzoic acid (Table 1, Figure 1), with the receptor binding domain (RBD) of SARS-CoV-2 spike glycoprotein, TMPRSS2, CatB and CatL were analyzed. According to the RBCI analysis we applied, it was determined that 1-Caffeoyl- β -D-glucose was the phytochemical with the highest affinity (binding free energy) against all target proteins (Table 1). There is no docking or molecular dynamic study of 1-Caffeoyl- β -D-glucose with the four target proteins (spike RBD, TMPRSS2, CatB and CatL) in the literature. Therefore, our results regarding this ligand constitute the first record in terms of literature. In our study, the second top-ranked ligand, rosmarinic acid, showed a high binding affinity for spike, TMPRSS2, CatB and CatL proteins (-6,13, -5,10, -5,44 and -6,21 kcal/mol, respectively) (Table 1). Although there is no other study showing the binding affinity of rosmarinic acid against spike protein, this ligand was found to show a strong binding (MolDock score of -89.17) against the host cell furin (a protein convertase) enzyme that the spike protein of SARS-CoV-2 uses to enter the cell using the Molegro Virtual Docker program (Kumar Verma *et al.*, 2021). On the other hand, in support of our study, it was reported in a different study using the Glide XP docking program that rosmarinic acid showed a reasonably high binding affinity (-5.6 kcal/mol) for the TMPRSS2 enzyme (Coban *et al.*, 2021). However, the

molecular docking study of rosmarinic acid regarding CatB and CatL enzymes has not been found in the literature. Regarding 3-*p*-coumaroylquinic acid (Table 1), the third promising ligand in this study, no molecular docking studies on spike, TMPRSS2, CatB or CatL enzymes were found in the literature. However, it has been reported that 3-*p*-coumaroylquinic acid showed a very favorable binding affinity (-7.2 kcal/mol) against non-structural protein 10 (nsp10) of SARS-CoV-2 in molecular docking simulation (Mohammad *et al.*, 2021). This result supports our finding we obtained with SARS-CoV-2 spike glycoprotein. In a molecular docking study conducted with the post fusion core of the SARS-CoV-2 spike S2 subunit, spike glycoprotein open state and spike glycoprotein closed state structures, chlorogenic acid, our 4th top-ranked ligand (Table 1), has been reported to bind to these different spike protein states with high affinities (-101.663, -108,993 and -92,121 MolDock scores) (Adem *et al.*, 2021). These results are also in agreement with our study.

In addition, there are molecular docking studies reporting that gallic acid, *p*-coumaric acid, ferulic acid, chicoric acid, cinnamic acid and 4-hydroxybenzoic acid, the other polyphenolic phytochemicals in our study, show high binding affinity for spike, TMPRSS2, CatB and CatL enzymes (Georgousaki *et al.*, 2020; Adem *et al.*, 2021; Guler *et al.*, 2021; Srivastava *et al.*, 2021; Surucic *et al.*, 2021).

In this present study, 1-caffeoyl- β -D-glucose, rosmarinic acid, 3-*p*-coumaroylquinic acid, and chlorogenic acid, which we determined as the first four top-ranked ligands, have drug-likeness properties according to Lipinski's Rule of Five (1-caffeoyl- β -D-glucose 1 violation [NH or OH>5]; chlorogenic acid 1 violation [NH or OH>5]). 6-tri-*o*-caffeoyl- β -D-glucopyranose (TCGP), an isomer of 1-caffeoyl- β -D-glucose isomer, has been reported to effectively inhibit the entry of HIV-1 Env pseudovirus into host cells (IC₅₀ = 5.5 μ g/mL) (Dong *et al.*, 2013). Therefore, although the spatial arrangement of the atoms of these two isomers are different, it can be expected that these two close isomers will show similar antiviral activity. It has been also shown that rosmarinic acid and chlorogenic acid have wide-ranging pharmacological effects on the inflammatory response, tumor occurrence and development, antioxidant, antimicrobial, anti-inflammatory and hepatoprotective effects in the host organism (Maalik *et al.*, 2016; Guan *et al.*, 2021). In general, it is obvious that the ligands in our study have various pharmacological effects (Table 2). However, considering the ADMET profiles, 1-caffeoyl- β -D-glucose was positive in the bacterial gene mutation test (AMES test), indicating that 1-caffeoyl- β -D-glucose may be a potential mutagen (Table 3).

Although our top-four ligands generally interact with cellular enzymes, lyases, proteases and kinases according to SwissTarget analysis (Figures 7A, 7B, 7C and 7D), based on probability analysis, only rosmarinic acid showed potential for interaction ($p = 0.96$) with its putative targets (lyases, various enzymes, proteases and kinases). Interestingly, it should be noted that rosmarinic acid showed an anti-inflammatory effect by inhibiting the PI3K-Akt pathway stimulated by IL2 in mice gastric cells (Nam *et al.*, 2020) and by inhibiting the kinase IKK- β (inhibitor of nuclear factor kappa-B kinase subunit beta) activity in human dermal fibroblasts (Lee *et al.*, 2006). This reported that inhibitory potential of rosmarinic acid on interleukins and kinases in the inflammatory pathway is in high agreement with the 'targets-components analysis' we performed on the STITCH public server. Based on this analysis, putative targets of rosmarinic acid include IKK- β (IKBKB), CCR3 and IL2, which play central roles in the inflammatory pathway (Figure 8).

In this study, it was investigated whether 22 polyphenolic phytochemicals (hydroxycinnamic acid and hydroxybenzoic acid derivatives) have effective pharmaceutical properties in the fight against SARS-CoV-2 by determining their binding affinities against the spike glycoprotein, TMPRSS2, CatB and CatL via molecular docking method. Considering the molecular docking scores in combination with the drug-likeness and intracellular target predictions, it can be

suggested that rosmarinic acid may be the most promising ligand among these 22 phytochemicals. As known, SARS-CoV-2 viral replication causes an aggressive inflammation (cytokine storm) in patients infected with this virus, and increased plasma levels of IL-2, IL-7, IL-10, GCSF, IP-10, MCP-1, MIP- 1A and TNF- α are frequently observed in affected individuals (Fu *et al.*, 2020). Therefore, based on its additional anti-inflammatory effects, we believe that further molecular optimization and investigation of the efficacy of rosmarinic acid in pre-clinical in vitro and in vivo studies may be promising in drug development efforts against SARS-CoV-2.

Acknowledgments

The authors would like to thank the Research Council of Kilis 7 Aralik University for the financial support (Grant number: 2028MAP2).

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research and publishing ethics. The scientific and legal responsibility for manuscripts published in IJSM belongs to the authors.

Authorship Contribution Statement

Erman Salih Istifli: Methodology, Resources, Visualization, Software, Formal Analysis. **Arzuhan Sihoglu Tepe:** Investigation, Resources, Validation, and Writing -original draft. **Cengiz Sarikurkcu:** Methodology, Formal Analysis, Software. **Bektas Tepe:** Investigation, Resources, Validation, and Writing -original draft.

Orcid

Erman Salih ISTIFLI  <https://orcid.org/0000-0003-2189-0703>

Arzuhan SIHOGLU TEPE  <https://orcid.org/0000-0001-8290-9880>

Cengiz SARIKURKCU  <https://orcid.org/0000-0001-5094-2520>

Bektas TEPE  <https://orcid.org/0000-0001-8982-5188>

5. REFERENCES

- Adem, Ş., Eyupoglu, V., Sarfraz, I., Rasul, A., Zahoor, A.F., Ali, M., Abdalla, M., Ibrahim, I.M., & Elfiky, A.A. (2021). Caffeic acid derivatives (CAFDs) as inhibitors of SARS-CoV-2: CAFDs-based functional foods as a potential alternative approach to combat COVID-19. *Phytomedicine*, 85, 153310.
- 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, 450–452.
- Cano-Avenidaño, B.A., Carmona-Hernandez, J.C., Rodriguez, R.E., Taborda-Ocampo, G., & González-Correa, C.H. (2021). Chemical properties of polyphenols: a review focused on anti-inflammatory and anti-viral medical application. *Biomedicine*, 41(1), 3-8.
- Chavez, J.H., Leal, P.C., Yunes, R.A., Nunes, R.J., Barardi, C.R., Pinto, A.R., Simoes, C.M., & Zanetti, C.R. (2006). Evaluation of antiviral activity of phenolic compounds and derivatives against rabies virus. *Veterinary Microbiology*, 116(1-3), 53-59.
- Coban, M.A., Morrison, J., Maharjan, S., Hernandez Medina, D.H., Li, W., Zhang, Y.S., Freeman, W.D., Radisky, E.S., Le Roch, K.G., & Weisend, C.M. (2021). Attacking COVID-19 progression using multi-drug therapy for synergetic target engagement. *Biomolecules*, 11(6), 787.
- Dong, Y., Tang, D., Zhang, N., Li, Y., Zhang, C., Li, L., & Li, M. (2013). Phytochemicals and biological studies of plants in genus *Hedysarum*. *Chemistry Central Journal*, 7(1), 1-13.

- Fu, Y., Cheng, Y., & Wu, Y. (2020). Understanding SARS-CoV-2-mediated inflammatory responses: from mechanisms to potential therapeutic tools. *Virologica Sinica*, 35(3), 266-271.
- Georgousaki, K., Tsafantakis, N., Gumeni, S., Lambrinidis, G., González-Menéndez, V., Tormo, J.R., Genilloud, O., Trougakos, I.P., & Fokialakis, N. (2020). Biological evaluation and in silico study of benzoic acid derivatives from *Bjerkandera adusta* targeting proteostasis network modules. *Molecules*, 25(3), 666.
- Guan, M., Guo, L., Ma, H., Wu, H., & Fan, X. (2021). Network pharmacology and molecular docking suggest the mechanism for biological activity of rosmarinic acid. *Evidence-Based Complementary and Alternative Medicine*, 2021.
- Guler, H.I., Fulya, A., Zehra, C., Yakup, K., Belduz, A.O., Canakci, S., & Kolayli, S. (2021). Targeting CoV-2 Spike RBD and ACE-2 Interaction with Flavonoids of Anatolian Propolis by *in silico* and *in vitro* Studies in terms of possible COVID-19 therapeutics. *BioRxiv*, <https://doi.org/10.1101/2021.02.22.432207>.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., & Nitsche, A. (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.
- Huang, I.-C., Bosch, B.J., Li, F., Li, W., Lee, K.H., Ghiran, S., Vasilieva, N., Dermody, T.S., Harrison, S.C., & Dormitzer, P.R. (2006). SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. *Journal of Biological Chemistry*, 281(6), 3198-3203.
- Jahan, I., & Onay, A. (2020). Potentials of plant-based substance to inhabit and probable cure for the COVID-19. *Turkish Journal of Biology*, 44(3), 228-241.
- Kumar Verma, A., Kumar, V., Singh, S., Goswami, B.C., Camps, I., Sekar, A., Yoon, S., & Lee, K.W. (2021). Repurposing potential of Ayurvedic medicinal plants derived active principles against SARS-CoV-2 associated target proteins revealed by molecular docking, molecular dynamics and MM-PBSA studies. *Biomedicine & Pharmacotherapy*, 137, 111356.
- Lee, J., Jung, E., Kim, Y., Lee, J., Park, J., Hong, S., Hyun, C.-G., Park, D., & Kim, Y.S. (2006). Rosmarinic acid as a downstream inhibitor of IKK- β in TNF- α -induced upregulation of CCL11 and CCR3. *British Journal of Pharmacology*, 148(3), 366-375.
- 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.
- Li, F., Li, W., Farzan, M., & Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, 309(5742), 1864-1868.
- 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. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965), 450-454.
- 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.
- Maalik, A., Bukhari, S.M., Zaidi, A., Shah, K.H., & Khan, F.A. (2016). Chlorogenic acid: a pharmacologically potent molecule. *Acta Poloniae Pharmaceutica*, 73(4), 851-854.
- Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, 7(2), 146-157.
- Mohammad, A., Alshawaf, E., Marafie, S.K., Abu-Farha, M., Al-Mulla, F., & Abubaker, J. (2021). Molecular Simulation-Based Investigation of Highly Potent Natural Products to

- Abrogate Formation of the nsp10-nsp16 Complex of SARS-CoV-2. *Biomolecules*, 11(4), <https://doi.org/10.3390/biom11040573>.
- Nam, H.-H., Kim, J.S., Lee, J., Seo, Y.H., Kim, H.S., Ryu, S.M., Choi, G., Moon, B.C., & Lee, A.Y. (2020). Pharmacological Effects of *Agastache rugosa* against Gastritis Using a Network Pharmacology Approach. *Biomolecules*, 10(9), 1298.
- Piccolella, S., Crescente, G., Faramarzi, S., Formato, M., Pecoraro, M.T., & Pacifico, S. (2020). Polyphenols vs. coronaviruses: how far has research moved forward? *Molecules*, 25(18), 4103.
- Ruibo, L., Narita, R., Nishimura, H., Marumoto, S., Yamamoto, S., Ouda, R., Yatagai, M., Fujita, T., & Watanabe, T. (2017). Antiviral Activity of Phenolic Derivatives in Pyrolytic Acid from Hardwood, Softwood, and Bamboo. *Sustainable Chemistry & Engineering*, 6(1), 119-126.
- Srivastava, N., Garg, P., Srivastava, P., & Seth, P.K. (2021). A molecular dynamics simulation study of the ACE2 receptor with screened natural inhibitors to identify novel drug candidate against COVID-19. *PeerJ*, 9, e11171.
- Sudhan, D.R., & Siemann, D.W. (2015). Cathepsin L targeting in cancer treatment. *Pharmacology & Therapeutics*, 155, 105-116.
- Surucic, R., Tubic, B., Stojiljkovic, M.P., Djuric, D.M., Travar, M., Grabez, M., Savikin, K., & Skrbic, R. (2021). Computational study of pomegranate peel extract polyphenols as potential inhibitors of SARS-CoV-2 virus internalization. *Molecular and Cellular Biochemistry*, 476(2), 1179-1193.
- Taguchi, R., Hatayama, K., Takahashi, T., Hayashi, T., Sato, Y., Sato, D., Ohta, K., Nakano, H., Seki, C., & Endo, Y. (2017). Structure–activity relations of rosmarinic acid derivatives for the amyloid β aggregation inhibition and antioxidant properties. *European Journal of Medicinal Chemistry*, 138, 1066-1075.
- Worldometers.info. (2021). COVID-19 Coronavirus Pandemic Retrieved 20.06.2021
- Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., & Wang, J. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host & Microbe*, 27, 325-328.
- 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. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269.
- Wu, Y.H., Zhang, B.Y., Qiu, L.P., Guan, R.F., Ye, Z.H., & Yu, X.P. (2017). Structure properties and mechanisms of action of naturally originated phenolic acids and their derivatives against human viral infections. *Current Medicinal Chemistry*, 24(38), 4279-4302.
- Zakaryan, H., Arabyan, E., Oo, A., & Zandi, K. (2017). Flavonoids: promising natural compounds against viral infections. *Archives of Virology*, 162(9), 2539-2551.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., & Huang, C.-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. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382, 727-733.

6. APPENDIX: SUPPLEMENTARY

6.1. Structural optimization of ligands

The Automatic Topology Builder (ATB) online server was utilized for geometry optimization of all the ligands using density functional theory DFT/B3LYP/6-31G* basis set. Following geometry optimization, the energetically minimized 3D ligand structures generated by the ATB server was exported in pdb file format for further use in molecular docking analyzes (Malde *et al.*, 2011).

6.2. Homology modeling of TMPRSS2

The crystallographic structure of human TMPRSS2 enzyme has not been resolved until today, therefore, a homology model was generated for 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 (Remmert *et al.*, 2012).

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 was 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 indicated that the model was 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 S1) using the PROCHECK web server (Laskowski *et al.*, 1993). Also, ERRAT web-based tool (Figure S2) was also deployed to calculate the overall quality factor (OQF) for non-bonded atomic interactions (Colovos and Yeates, 1993).

6.3. Protein retrieval and energy minimization of the Spike, TMPRSS2, CatB and CatL proteins using Nanoscale Molecular Dynamics (NAMD)

The purpose of performing energy minimization (optimization) of proteins is the physical significance of the obtained 3D structures: the optimized structures often resemble to their native conformation as they found in nature. Thus, this process searches to find the lowest energy conformation of the proteins in question. The spike glycoprotein was retrieved by removing the ACE2 subunit from the angiotensin-converting enzyme 2 - 2019nCoV RBD complex in Discovery Studio Visualizer v16. This model was downloaded from "<https://swissmodel.expasy.org/interactive/HLkhkP/models/03>" (PDB ID: model_03.pdb) (Camacho *et al.*, 2009; Remmert *et al.*, 2012). Since, the crystallographic structure of human TMPRSS2 enzyme has not still been defined, using SWISS-MODEL, we generated a homology model of this enzyme to utilize in further molecular docking analyses. The details of the homology modeling procedure can be found in the study of Istifli *et al.* (2020). The PROCHECK web server was utilized whether our generated model has an energetically favorable conformation (Figure S1) and the ERRAT web-based tool was also employed to calculate the protein overall quality factor (OQF) (Figure S2) (Colovos & Yeates, 1993; Laskowski *et al.*, 1996). The crystal structures of CatB (PDB ID: 1GMY) and CatL (PDB ID: 2YJ9) enzymes were retrieved from Protein Data Bank.

Water molecules, co-crystallized ligands and non-interacting ions were removed from proteins before energy minimization. During the protein energy minimization step, the atom types and electrical charges of the spike glycoprotein, TMPRSS2, CatB and CatL were fixed using CHARMM22_PROT force field and Gasteiger-Marsili charges in the Vega ZZ software (Pedretti *et al.*, 2004). Next, for the energy minimization of all the proteins using NAMD, the parameters were 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 structures corresponding to the last minimization step of all proteins were saved as the lowest energy conformation.

6.4. Molecular docking analyses of phenolic acids

Molecular docking analyses between spike glycoprotein, TMPRSS2, CatB and CatL and the selected phenolic acids were performed using AutoDock 4.2.6. Correspondingly, the docking scores (binding free energy) of the ligands with the spike receptor binding motif, TMPRSS2, CatB (PDB ID: 1GMY) and CatL (PDB ID: 2YJ9) were determined. AutoDockTools-1.5.6 was used to prepare the targets and ligands as well as the parameters prior to initiating molecular docking using AutoDock 4.2.6 (Sanner, 1999). The grid box coordinates determined in molecular docking studies were adjusted to allow the analyzed phytochemicals to interact with the catalytic amino acid residues (active sites) of the target proteins. Accordingly, the grid box sizes were adjusted as: a) $80 \times 90 \times 40$ Å points (x: -34.42, y: 27.69, z: 5.37) for the spike glycoprotein; b) $60 \times 110 \times 86$ Å points (x: 14.67, y: -3.01, z: 6.88) for TMPRSS2; c) $86 \times 84 \times 44$ Å points (x: 22.86, y: 5.73, z: 27.96) for CatB; and d) $54 \times 52 \times 60$ Å points (x: 5.31, y: 6.05, z: 0.12) for CatL.

Docking calculations were performed using 100 genetic algorithm (GA) runs, an initial population of 150 individuals, max. number of 3.000.000 energy evaluations, and a max. number of 27.000 generations. The mutation and crossover rates were set as default values, 0.02 and 0.8, respectively. After 100 independent docking runs, all the ligand binding modes (conformations) were clustered and ranked on the basis of the most negative free energy of binding (kcal/mol). The best poses of receptor-ligand pre-reactive complexes obtained by AutoDock 4.2.6 were visualized and examined with BIOVIA Discovery Studio Visualizer v16.

6.5. Calculation of relative binding capacity index (RBCI)

In this study, the relative binding capacity index (RBCI) was applied to statistically rank the activity potentials of phytochemicals using the binding free energy values obtained from the binding analysis (Figure 2 and Table S1) (Istifli *et al.*, 2020). Using RBCI, it is possible to compare statistically relevant data with different scientific meanings. Since the binding energies of ligands are different for each protein, phytochemicals can only be ranked in terms of their potential at this parameter if they are performed in the light of their binding energies only to one protein. However, sequencing based on only one of these proteins cannot represent the full activity potential of these molecules. The most common method used to calculate the interaction between each receptor and ligand is the "central bias" in which components are ranked according to the mean value of each component.

If the values (binding free energy) in each data set are converted into standard scores, it is possible to compare them with each other. Arithmetic mean and standard deviation values were calculated for each protein by using the binding free energies of the ligands. Raw standard scores were obtained by subtracting the binding free energies of each protein for each ligand from this arithmetic mean and then dividing by the standard deviation value (see equation given below) (Sharma, 1995). The RBCI values of each phytochemical were then calculated by taking the average of these standard scores obtained separately for each protein target.

$$\text{Standard score} = \frac{(x - \mu)}{\sigma}$$

where 'x' is the raw data, 'μ' is the mean, and 'σ' is the standard deviation.

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

6.6. Drug-likeness, ADMET profile and target prediction

The determination of drug-likeness, ADMET and target profiles of promising hit compounds in structure-based drug design studies is important in terms of reducing their side effects on the target organism. In this study, web-based SwissTargetPrediction and pkCSM tools were used to investigate such effects of analyzed phenolic acids (Pires *et al.*, 2015; Daina *et al.*, 2019).

6.7. Network pharmacology analysis

The one-drug/one-target approach in drug discovery has some deficiencies in terms of safety and efficacy (Chandran *et al.*, 2015). Therefore, analyzing the mutual interactions of small molecules with the protein network in discovering the possible side effects of hit or lead compounds or the elucidation of novel therapeutic effects requires the application of a network pharmacology approach. In our study, the targets-components analysis of identified hit phytochemicals was performed by selecting the target organism as '*Homo sapiens*' through the STITCH (<http://stitch.embl.de/>) public database.

Table S1. RBCI values of phenolic acids.

Compound	Total RBCI
1-Caffeoyl-β-D-glucose	-1.47
Rosmarinic acid	-1.30
3-p-Coumaroylquinic acid	-1.18
Chlorogenic acid	-1.16
6-O-p-Coumaroyl-D-glucose	-1.03
3-Feruloylquinic acid	-0.54
p-Coumaroyltartaric acid	-0.47
N-Caffeoyl-L-aspartic acid	-0.25
Ferulic acid	-0.16
Caffeic acid	-0.13
1-O-Feruloyl-β-D-glucose	0.00
Chicoric acid	0.06
p-Coumaric acid	0.12
Protocatechuic acid	0.43
Cinnamic acid	0.56
Gallic acid	0.60
Vanillic acid	0.76
4-hydroxybenzoic acid	0.77
Syringic acid	0.94
Isovanillic acid	0.98
Benzoic acid	1.20
Salicylic acid	1.31

Table S2. Molecular interactions between the phenolic acids and RBD of the spike glycoprotein of 2019-nCoV.

No	Compound	Classical H-bond	Van der Waals	Non-Classical H-bond (C-H, Pi-Donor)	Hydrophobic interaction		Electrostatic	Miscellaneous (Lone pair/Pi- sulphur)
					π - π interaction	Mixed π /Alkyl		
1	1-caffeoyl- β -D-glucose	Arg403, Gln409, Lys417, Tyr453, Gly496, Tyr505 ¹	Glu406, Gly416, Ile418, Gln493 ¹ , Tyr495, Ala419, Gln493 ¹ , Phe497, Asn501 ¹	Arg403, Ser494 ¹	-	Lys417	-	-
2	Rosmarinic acid	Tyr449, Ser494 ¹ , Gly496, Gln498, Tyr505 ¹	Tyr453, Gln493 ¹ , Tyr495, Phe497	-	-	Arg403	-	-
3	3-p-Coumaroylquinic acid	Arg403, Gln409, Tyr453, Ser494 ¹ , Asn501 ¹ , Tyr505 ¹	Gly416, Ile418, Ala419, Gln493 ¹ , Tyr495, Gly496, Phe497, Gln506	Arg403	-	Lys417	Glu406	-
4	Chlorogenic acid	Glu406, Gln409, Lys417, Ser494 ¹ , Gly496	Arg403, Asp405, Gly416, Ile418, Tyr453, Gln493 ¹ , Phe497, Asn501 ¹ , Tyr505 ¹	-	Tyr495	-	-	-

¹Amino acid residues involved in binding to ACE2 in the RBM of 2019-nCoV (Leu455, Phe486, Gln493, Ser494, Asn501, and Tyr505)

Table S3. Molecular interactions between the phenolic acids and Tmprss2.

Compound	Classical H-bond	Van der Waals	Non-Classical H-bond (C-H, Pi- Donor)	Hydrophobic interaction		Electrostatic	Miscellaneous (Lone pair/Pi- sulphur)
				π - π interaction	Mixed π /Alkyl		
1-caffeoyl- β -D-glucose	Val280, His296 ¹ , Ser394, Gly439, Ser441 ¹	His279, Cys281, Cys297, Gln317, Trp384, Glu395, Gln438, Ser460	-	-	-	Thr393	-
Rosmarinic acid	His279, Ser460, Gly462	Val278, Cys281, Gln438, Gly439, Trp461, Ser463, Gly464	Val280, Thr393, Ser441 ¹	His296 ¹	Thr393	-	-
3-p-Coumaroylquinic acid	Val280, Cys281, His296 ¹	His279, Cys297, Trp308, Thr393, Ser436, Cys437, Gly439, Ser460, Trp461, Gly462, Ser463, Gly464, Cys465	Ser441 ¹	-	-	-	-
Chlorogenic acid	His279, His296 ¹ , Ser394, Gly439, Ser441 ¹	Cys281, Cys297, Gln317, Trp384, Cys437	Val280	-	-	-	-

¹The active amino acid residues of Tmprss2 (His296, Asp345, Ser441)

Table S4. Molecular interactions between the phenolic acids and CatB.

No	Compound	Classical H-bond	Van der Waals	Non-Classical H-bond (C-H, Pi- Donor)	Hydrophobic interaction		Electrostatic	Miscellaneous (Lone pair/Pi- sulphur)
					π - π interaction	Mixed π /Alkyl		
1	1-caffeoyl- β -D-glucose	Gln23 ¹ , Gly24 ¹ , Gly27 ¹ , Cys29 ¹ , Gly72, His197	Ser25, Cys26 ¹ , Ser28 ¹ , Trp30 ¹ , Gly71, Pro74 ¹ , His108, Cys117, Thr118, Glu120, Ala198	His197	-	-	-	-
2	Rosmarinic acid	Gln23 ¹ , Cys29 ¹ , Asn70, His108, Gly119, Glu120, Gly196, His197	Gly24 ¹ , Ser25, Gly27 ¹ , Cys26 ¹ , Ser28 ¹ , Gly71, His109, Cys117, Val174, Trp219	Gly24 ¹	-	-	-	-
3	3-p-Coumaroylquinic acid	Gln23 ¹ , Cys29 ¹ , Gly72, Gly119, Glu120	Gly24 ¹ , Ser25, Gly27 ¹ , Ser28 ¹ , Asn70, Gly71, His108, Cys117, Gly196, His197, Trp219	Trp30 ¹	-	Cys26 ¹ , Gly27 ¹	-	-
4	Chlorogenic acid	Gln23 ¹ , Gly24 ¹ , Cys26 ¹ , Cys29 ¹ , Gly119, Gly196, His197	Ser25, Gly27 ¹ , Ser28 ¹ , His109, Thr118, Glu120, Val174, Leu179, Met194, Gly195, Trp219 Gln23 ¹ , Ser25, Gly27 ¹ , Cys69, Gly71, His108, Cys117, Gly119,	His108, His197	His108	Cys117	-	-

¹The active amino acid residues of CatB (Gln23, Gly24, Cys26, Gly27, Ser28, Cys29, Trp30, Gly73, Pro74, His110, His111, His199, Trp221).

Table S5. Molecular interactions between the phenolic acids and CatL.

No	Compound	Classical H-bond	Van der Waals	Non-Classical H-bond (C-H, Pi-Donor)	Hydrophobic interaction		Electrostatic	Miscellaneous (Lone pair/Pi-sulphur)
					π - π interaction	Mixed π /Alkyl		
1	1-caffeoyl- β -D-glucose	Met70 ¹ , Asp71, Met161 ¹ , Asp114, Asp162 ¹ , Ala215	Trp26 ¹ , Gly68 ¹ , Leu69 ¹ , Phe112, Ser133, His163, Gly164, Ser216, Ala214	-	-	Cys25 ¹ , Ala135 ¹	-	Met70 ¹
2	Rosmarinic acid	Gly68 ¹ , Met70 ¹ , Lys117, Ser213	Cys25 ¹ , Trp26 ¹ , Phe74, Asp114, Ser133, Met161 ¹ , Asp162 ¹ , His163, Gly164, Ser216	-	-	Leu69 ¹ , Ala135 ¹ , Ala214, Ala215	Asp71	Met70 ¹
3	3-p-Coumaroylquinic acid	Cys25 ¹ , Gly68 ¹ , Met70 ¹ , Asp71, Asp114, Ser216	Gly23 ¹ , Trp26 ¹ , Ser133, Gly164, Ala214, Ala215	Gly67 ¹ , Leu69 ¹	Asp162 ¹ , His163	Cys25 ¹ , Ala135 ¹	-	Gly68 ¹
4	Chlorogenic acid	Asp71, Asp114, Lys117, Ser213, Ser216	Trp26 ¹ , Gly68 ¹ , Leu69 ¹ , Tyr72, Phe112, Met161 ¹ , Asp162 ¹ , His163, Gly164, Ala214	-	-	Ala135 ¹	-	Cys25 ¹ , Met70 ¹

¹The active amino acid residues of CatL (Gln19, Gly20, Gln21, Cys22, Gly23, Ser24, Cys25, Trp26, Gly61, Asn66, Gly67, Gly68, Leu69, Met70, Ala135, Met161, Asp162, Trp189)

Figure S1. Ramachandran plot of Tmprss2 model.

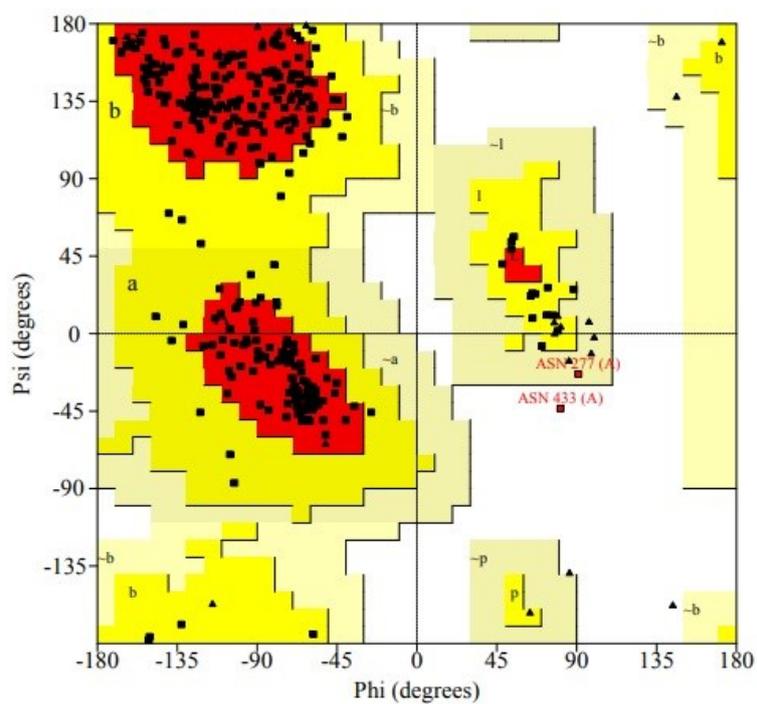
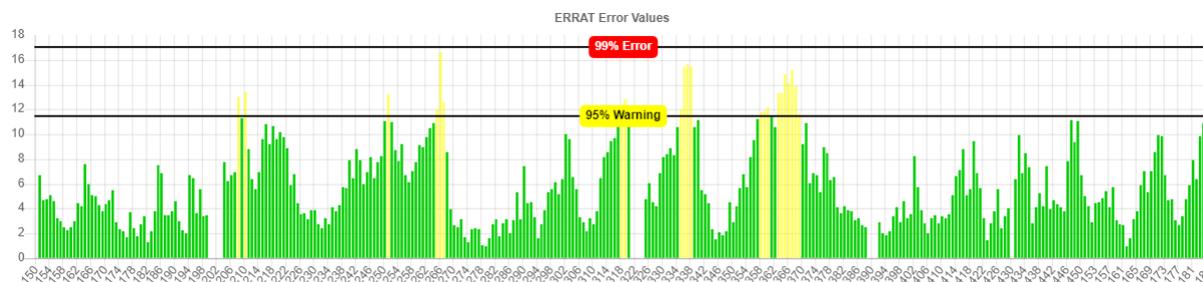


Figure S2. ERRAT error values for Tmprss2 model.



(Protein regions show misfolding at 95% confidence level were indicated with yellow bars. Green bars, on the other hand, point to regions that show correct folding)

6.8. References

- 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.
- Chandran, U., Mehendale, N., Tillu, G., & Patwardhan, B. (2015). Network Pharmacology of Ayurveda Formulation Triphala with Special Reference to Anti-Cancer Property. *Combinatorial Chemistry & High Throughput Screening*, 18(9), 846-854.
- Colovos, C., & Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, 2(9), 1511-1519.
- Daina, A., Michielin, O., & Zoete, V. (2019). SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Research*, 47(W1), W357-W364.
- Istifli, E.S., Netz, P.A., Sihoglu Tepe, A., Husunet, M.T., Sarikurkcü, C., & Tepe, B. (2020). 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. *Journal of Biomolecular Structure and Dynamics*, <https://doi.org/10.1080/07391102.2020.1840444>.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., & Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, 8(4), 477-486.
- Malde, A.K., Zuo, L., Breeze, M., Stroet, M., Poger, D., Nair, P.C., Oostenbrink, C., & Mark, A.E. (2011). An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of Chemical Theory and Computation*, 7(12), 4026-4037.
- Pedretti, A., Villa, L., & Vistoli, G. (2004). VEGA—an open platform to develop chemo-bioinformatics applications, using plug-in architecture and script programming. *Journal of Computer-Aided Molecular Design*, 18(3), 167-173.
- 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.
- 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.
- Sanner, M.F. (1999). Python: a programming language for software integration and development. *Journal of Molecular Graphics and Modelling*, 17(1), 57-61.
- Sharma, S. (1995). *Applied multivariate techniques*. New York, United States: John Wiley & Sons, Inc.