Flavonoids as potential agents for development of multi-target drugs for covid-19 treatment: An in silico study

Nguyen Thu Hang, Nguyen Van Phuong*

Department of Pharmacognosy, Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Phan Chu Trinh, Hoan Kiem, Hanoi 10000, Viet Nam

Submitted August 17, 2021; Revised October 27, 2021; Accepted November 6, 2021

Abstract

COVID-19 is an infectious disease caused by SARS-CoV-2 that is spreading in many countries around the world. In attempts to discover compounds that have an effect on SARS-CoV-2, many important molecular targets have been identified, involved in viral infection and replication including spike protein, main protease, capthesin L, helicase, 2-O-methyltransferase, endoRNAse. In this study, we would like to identify potential flavonoids that could simultaneously inhibit 3CLP, capthesin L, endoRNAse, 2-O-methyltransferase, and PLP from a 4389-flavonoid database using molecular docking, molecular dynamics simulation, pharmacokinetic and toxicity prediction. Out of 4389 compounds, 79 potential flavonoids that could simultaneously inhibit five COVID-19 molecular targets were identified. Pharmacokinetic and toxicity prediction showed that these compounds were well absorbed from the gastrointestinal tract and safe for human use. These potential compounds were noteworthy during drug research and development for SARS-CoV-2 treatment.

Keywords. SARS-CoV-2, flavonoid, molecular docking, ADME prediction, molecular dynamics simulation.

1. INTRODUCTION

The COVID-19 pandemic, which was caused by Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2), first appeared in Wuhan, China in 2019 and has rapidly spread around the world. As of October 22nd, 2021, there has been a total of 242,348,657 cases of infection, of which nearly 5 million people died.[1] Even though much effort has been made, to date, there is no effective and specific treatment for COVID-19. Therefore, it is necessary to carry out research and development of drugs that have an effect on SARS-CoV-2.

In attempts to discover compounds that inhibit SARS-CoV-2, many important molecular targets have been identified, involved in viral infection and replication including spike protein, main protease, capthesin L, helicase, 2-O-methyltransferase, endoRNAse.[2] Numerous studies have shown that a multi-target drug that could simultaneously inhibit different enzymes at the same time is expected to bring greater efficiency as well as safety for users much more than a single-target drug.[3]

Main protease (Mpro) or 3-chymotrypsin-like protease (3CLP) and papain-like protease (PLP) are two proteases that cleave apo-polyprotein to functional components, therefore were vital to the replication of SARS-CoV-2. Many studies showed that these two enzymes could be potential targets for the treatment of COVID-19.[4,5] Besides, 2-O-methyltransferase and EndoRNAse also play a very important role in viral RNA replication in the human body. On the other hand, the presence of capthesin is essential for the coronavirus to enter the human cell. Some recent studies have shown that it is not capthesin B but capthesin L that involves in the host cell entry process of SARS-CoV-2. Recognizing the important role of cathepsin L in the SARS-CoV-2 infection cycle, many new drug development studies have focused on cathepsin L as an important molecular target in the treatment of COVID-19.[2]

Natural compounds are considered a large and diverse library for the research and development of new drugs.[5] Among them, flavonoids are an important group of compounds, which is widely distributed in nature with many valuable biological activities such as antioxidant, anti-inflammatory, analgesic, antiviral.[6] Some studies have shown that flavonoids were potential agents for the prevention and treatment of COVID-19 because of their ability to inhibit SARS-CoV-2 main protease, NTPase/helicase, and N proteins.[7] However, out of nearly 10,000 known flavonoids, the number of flavonoids screened is still quite limited.
Drug research and development is a strenuous process, often lasting 10-15 years and costing about $1 billion with a very high failure rate. Many methods have been introduced to shorten the time, cost as well as improve the success rate. One of these solutions is in silico methods including molecular docking, molecular dynamics simulation, pharmacokinetic and toxicity prediction. In this study, our object was to apply these techniques to search for the most potential flavonoids which simultaneously inhibit SARS-CoV-2 main protease, Capthesin L, 2-O-methyltransferase, endoRNAse and have suitable pharmacokinetic parameters and toxicity profiles.

2. MATERIALS AND METHODS

2.1. Protein preparation

In this study, COVID-19 molecular targets including 3CLP (ID: 7LKT), capthesin L (ID: 3HHA), endoRNAse (ID: 6WXC), 2-O-methyltransferase (ID: 6WKQ) and papain-like protease (ID: 7JIW) were selected and downloaded from protein data bank RCSB. To be able to evaluate the molecular docking protocol and ensure the most accurate results, proteins with low resolution and co-crystal ligands were preferred. However, since the protein structures from RCSB did not include hydrogen atoms, in the next step, hydrogen atoms were added followed by removing water molecules which were far away from the ligands using MOE 2009 software. Finally, all the atoms were assigned charges based on Amber99 forcefield.

2.2. Ligand preparation

A database containing structures of 4389 flavonoids was collected from the Dictionary of Flavonoids. These structures were drawn using ChemDraw 17 software, optimized and assigned MMFF94x force field using MOE 2009 software. The structure information of these compounds was saved as .mdb files.

2.3. Molecular docking

Before performing molecular docking to identify potential flavonoids, the method was validated by redocking co-ligands including Y7M (3CLP), AZ12878478 (Capthesin L), tipiracil (EndoRNAse), sinefungin (2-O-methyltransferase) and VBY (PLP) into the active sites of proteins, using Triangle matcher algorithm with a scoring function of London dG. After that, RMSD (Root Mean Square Deviation of atomic positions) between the re-docked and experimental confirmations of ligands was calculated. The molecular docking protocol is considered accurate if RMSD is less than 2 Angstroms.

After evaluating the docking method, potential flavonoids were screened by docking 4389 compounds into the active sites of 5 molecular targets, respectively then these flavonoids were arranged in order of increasing binding energy. The positive control used were the co-ligands including Y7M (3CLP), AZ12878478 (Capthesin L), tipiracil (EndoRNAse), sinefungin (2-O-methyltransferase) and VBY (PLP), respectively. The structures of these reference compounds were shown in figure 1. Flavonoids were considered potential if their binding energy with the enzyme was less than or equal to that of the corresponding positive control.

2.4. Prediction of pharmacokinetic parameters, drug-like properties and toxicity

After the structures of potential flavonoids were converted to SMILES, drug-like properties of these compounds were predicted by SwissAdme server based on Lipinski’s rule of 5, including the number of hydrogen bond donors ≤ 5, number of hydrogen bond acceptors ≤ 10, molecular mass < 500 and octanol/water partition coefficient ≤ 5. After that, based on pre-built QSARS models on the SwissAdme server, pharmacokinetic parameters including: gastrointestinal tract (GI) absorption, blood-brain barrier (BBB) permeability, inhibition of the hepatic enzyme and P-glycoprotein (P-gp) substrate identification of compounds which did not violate any rule were determined to identify potential flavonoids with suitable pharmacokinetic properties to develop drugs for the treatment of COVID-19. The toxicity of these potential compounds includes the following properties: alerts for S. typhimurium mutagenicity, negative for carcinogenicity and DNA binding alerts was also predicted using Toxtree 3.1.0 software.

2.5. Molecular dynamics simulation

Molecular docking often uses some approximate assumptions to reduce resources and time of calculation. However, this approach could lead to inaccuracies in which compounds predicted to have low protein-ligand binding energies are unlikely to inhibit the enzyme well in actuality. To overcome this problem, molecular dynamics simulation has been proposed. In this study, after the prediction of pharmacokinetic properties, drug likeliness and
toxicity, 1000-ps molecular dynamics simulations of potential flavonoid enzyme complexes were performed using MOE software. The temperature was set at 310 K and Nosé-Poincaré-Andesen algorithm was used. Other parameters were set by default. The free energy and the positions of the atoms of the systems were recorded every 0.5 ps. The results were analyzed based on the change of free energy, root mean square deviation (RMSD) over time and the root mean square fluctuation (RMSF) of the alpha carbon.

![Flavonoids as potential agents for...](image)

© 2022 Vietnam Academy of Science and Technology, Hanoi & Wiley-VCH GmbH www.vjc.wiley-vch.de 283

3. RESULTS AND DISCUSSION

3.1. Molecular docking

The traditional drug discovery process is often based on the “one target - one drug” paradigm which has successfully explored many drugs that could effectively fight several diseases. However, in the case of diseases with complex pathogenesis such as cancer, Alzheimer’s, Parkinson’s, this approach proved to be an inefficient method. Therefore, the multi-target drug discovery paradigm was proposed as a solution to overcome this shortcoming.[18] This approach shows many advantages compared to the traditional approach such as increasing treatment efficiency, reduction of drug dose and toxicity, prevention of drug resistance. Based on this principle, in this study, we would like to identify potential flavonoids that could simultaneously inhibit 3CLP, capthensin L, endoRNAse, 2-O-methyltransferase and PLP. Proteins were selected from the protein database bank based on the following criteria: high resolution to ensure the structures of enzymes were accurate and containing ligand(s) so that the molecular docking protocol could be validated. Therefore, five proteins: 3CL protease (ID: 7LKT), Capthensin L (ID: 3HHA), endoRNAse (ID: 6WXC), 2-O-methyltransferase (ID: 6WKQ) and papain-like protease (ID: 7JIW) were chosen.

Before molecular docking was performed to identify potential flavonoids that had effects on SARS-CoV-2, the protocol was validated by redocking the co-ligand into the active site of the enzyme and the theoretical and experimental data were compared. The validation results of the docking method were illustrated in figure 2.

The validation results showed that the molecular docking protocol was highly accurate with RMSD values of all five ligands less than 0.2 nm. Therefore, this method was used to screen for potential flavonoids. After being screened, the binding energies of flavonoids with enzymes were calculated and arranged in ascending order of binding affinity. The lower the energy, the stronger the predicted inhibitory effect of flavonoids, and vice versa. Flavonoids were considered potential if
the binding energy of the compound was less than that of positive control.

Figure 2: Validation of molecular docking protocol. Green: native ligands; Gray: redocked ligands
In this study, positive controls were chosen as compounds with known corresponding enzyme inhibitory effects. For 3CLP, capthesin L, endoRNAse, 2-O-methyltransferase and PLP, the positive controls selected were Y7M, AZ12878478, tipiracil, sinefungin and VBY, respectively. The results of the screening of 4389 flavonoids in the database were presented in figure 3.

3.1.1. 3-Chymotrypsin like protease (3CLP)

The results of molecular docking of flavonoids with 3CLP showed that the binding energy of flavonoids with 3CLP ranged from -1.86 to -12.01 kcal/mol, compared with the positive control, Y7M, was -6.08 kcal/mol. Of the 4389 screened flavonoids, 1240 potential compounds had the binding affinity to 3CLP less than or equal to Y7M, accounting for 28.25%. Interaction between 3CLP and one of 1240 flavonoids, compound 2275 was illustrated in figure 4. The results revealed that compound 2275 interacted with SARS-CoV-2 3CLP by the formation of 4 hydrogen bonds between 2275 and...
Figure 4: Interaction between 2275 with SARS-CoV-2 3CLP, Capthesin L, endoRNAse, 2-O-methyltransferase and PLP
Gly143, Cys145, Glu166 and Arg188 residues. In which, Gly143 and Glu166 were also two residues that interact with the positive control, Y7M in the 7LKT complex. These interactions might play an important role that should be considered in the research and development of treatment for COVID-19. This result was also completely consistent with some previous studies on the complex of 3CLP with flavonoids.\(^{19,20}\)

### 3.1.2. Capthesin L

Similar to 3CLP, the molecular docking results showed that the binding energy of flavonoids with Capthesin L ranged from -1.60 to -9.27 kcal/mol, compared with the affinity of positive control of -3.96 kcal/mol. The number of flavonoids considered as potential Capthesin L inhibitors was 2200 compounds, equivalent to more than 50% of total flavonoids in the database. However, unlike the positive control AZ12878478, the docking result of compound 2275 with Capthesin L suggested that Glu159 may play a key role in the Capthesin L inhibitory effect of flavonoids and not Gly68 or Asp162 as for AZ12878478.

### 3.1.3. EndoRNAse

The screening results showed that most flavonoids had lower binding energy with endoRNAse than tipiracil did. The results of the interaction analysis between compound 2275 and EndoRNAse revealed that this compound bound to the protein through hydrogen bonding with residue Lys345 which was similar to tipiracil. In addition, compound 2275 also formed a hydrogen bond between the OH group with Val262 residue and a \(\pi-\pi\) interaction between the A-ring of the flavonoid and Tyr343. This result may partly explain the low protein binding energy value as well as the antiviral activity of many flavonoids.

### 3.1.4. 2-O-methyltransferase

The screening results showed that the binding energy between flavonoids and 2-O-methyltransferases ranged from -2.56 to -15.57 kcal/mol, compared with the positive control (sinefungin) of -5.45 kcal/mol. Thus, out of the 4389 screened flavonoids, 2490 compounds had a lower energy binding to the enzyme than sinefungin. These compounds were predicted to be potent 2-O-methyltransferase inhibitors and could continue to develop as a treatment for COVID-19. The results of molecular docking between the flavonoid (compound 2275), sinefungin and proteins showed that Asn6899, Gly6860 and Asp6912 could be important residues related to the 2-O-methyltransferase inhibitory ability of these compounds.

#### 3.1.5. Papain like protease (PLP)

The molecular docking results between 4389 flavonoids and SARS-CoV-2 PLP showed that the binding energy of flavonoids with this enzyme ranged from -2.52 to -13.95 kcal/mol, compared with the positive control (VBY) of -5.50 kcal/mol. Among them, the number of flavonoids considered to be potential was 3251 compounds. The interaction analysis revealed that two residues Leu162 and Thr301 were involved in forming hydrogen bonds with flavonoids. This result was quite consistent with the previous research conducted by D. Li et al.\(^{21}\)

From the molecular docking results, 684 compounds that simultaneously inhibit 5 molecular targets of COVID-19 were identified. In the next step, information on the drug-like properties, pharmacokinetics and toxicity of these 684 flavonoids would be determined to identify the most potent compound, leading to the development of a therapeutic drug for SARS-CoV-2.

### 3.2. Prediction of drug-like properties, pharmacokinetic parameters and toxicity

Using the SwissADME server, the drug-like properties of 684 potential flavonoids were determined based on Lipinski's Rule of Five, including the number of hydrogen bond donors \(\leq 5\), the number of hydrogen bond acceptors \(\leq 10\), the molecular mass \(< 500\) and \(\log P \leq 5\). The results were shown in figure 5.

The results revealed that out of 684 potential flavonoids, 287 compounds violated 3/4 of the rule, 79 compounds violated 2 rules, 124 compounds violated a rule, and only 194 compounds did not violate any rule. In other words, these 194 flavonoids were the most drug-like and had a high chance of success in developing into COVID-19 drugs. The results of predicting pharmacokinetic parameters of these 194 compounds showed that there were 84 potential compounds with the capacity of well absorbed through the gastrointestinal tract. Then, the toxicity of these 84 compounds continued to be investigated using Toxtree 3.1 software. The results showed that there were five compounds predicted to have at least one structural alert for...
carcinogenicity. Therefore, these compounds were eliminated. Thus, the remaining 79 substances were notable compounds for research and development of the drug for treatment SARS-CoV-2 from flavonoids. The SMILES, pharmacokinetics properties of these flavonoids were presented in table 1. The toxicity of these 79 compounds was presented in table 2.

3.3. Molecular dynamics simulation

To verify molecular docking results as well as have a further look at the interaction between the flavonoids and the proteins of SARS-CoV-2, molecular dynamics simulation was performed between 2275 and 3CLP, Capthesin L, endoRNAse, 2-O-methyltransferase and PLP, using MOE 2009 software. The free energy and RMSD of the complex over time were shown in figures 6 and 7. The RMSF of the alpha carbon was also illustrated in figure 7.

The results revealed that the models were stable with free energy U reaching equilibrium after about 50ps. This result suggested that the predicted molecular docking results were stable and consistent with the experimental data. However, it is necessary to perform molecular dynamics simulation over a longer time to obtain more accurate results.

4. CONCLUSION

In this study, 79 potential flavonoids that could simultaneously inhibit 3CLP, Capthesin L, endoRNAse, helicase, 2-O-methyltransferase and PLP were identified from a database of 4389 flavonoids. Pharmacokinetic and toxicity prediction showed that these 79 compounds were well absorbed from the gastrointestinal tract and were safe for human use. These potential compounds were noteworthy during drug research and development for SARS-CoV-2 treatment.

**Supporting information.** The structure and docking scores of 4389 screened flavonoids against five COVID-19 molecular targets.

**REFERENCES**

1. World Health Organization, WHO Coronavirus (COVID-19) Dashboard.

2. Shagufta I. Ahmad. The race to treat COVID-19: Potential therapeutic agents for the prevention and treatment of SARS-CoV-2, *Eur. J. Med. Chem.*, 2021, 213, 113157.

3. J. -J. Lu, W. Pan, Y. -J. Hu, Y. -T. Wang. Multi-target drugs: the trend of drug research and development, *PLoS One*, 2012, 7(6), e40262.

4. Faheem, B. K. Kumar, K. V. G. C. Sekhar, S. Kunjiappan, J. Jamalis, R. Balaña-Fouce, B. L Tekwani, M. Sankaranarayanan. Druggable targets of SARS-CoV-2 and treatment opportunities for COVID-19, *Bioorg. Chem.*, 2020, 104, 104269.

5. M. S. Butler. The role of natural product chemistry in drug discovery, *J. Nat. Prod.*, 2004, 67(12), 2141-2153.

6. D. Atmani, N. Chaher, D. Atmani, M. Berboucha, N. Debbache, and H. Boudaoud. Flavonoids in human health: from structure to biological activity, *Curr. Nutr. Food Sci.*, 2009, 5(4), 225-237.

7. M. Russo, S. Moccia, C. Spagnuolo, I. Tedesco, and G. L. Russo. Roles of flavonoids against coronavirus infection, *Chem. Biol. Interact.*, 2020, 109211.

8. J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott. Principles of early drug discovery, *Br. J. Pharmacol.*, 2011, 162(6), 1239-1249.
9. I. M. Kapetanovic. Computer-aided drug discovery and development (CADD): *in silico*-chemico-biological approach. *Chem. Biol. Interact.*, **2008**, 171(2), 165-176.

10. P. W. Rose et al. The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.*, **2012**, 41(D1), D475-D482.

11. S. Vilar, G. Cozza, and S. Moro. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery, *Curr. Top. Med. Chem.*, **2008**, 8(18), 1555-1572.

12. J. Buckingham and V. R. N. Munasinghe. *Dictionary of Flavonoids with CD-ROM*, **2015**, CRC Press.

13. C. S. de Magalhães, H. J. C. Barbosa, and L. E. Dardenne. A genetic algorithm for the ligand-protein docking problem, *Genet. Mol. Biol.*, **2015**, 27(4), 605-610.

14. A. Daina, O. Michielin, and V. Zoete. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci. Rep.*, **2017**, 7(1), 1-13.

15. C. A. Lipinski. Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.*, **2004**, 1(4), 337-341.

16. A. A. Parikesit and R. Nurdiansyah. Natural products repurposing of the H5N1-based lead compounds for the most fit inhibitors against 3C-like protease of SARS-CoV-2, *J. Pharm. Pharmacogn. Res.*, **2021**, 9(5), 730-745.

17. D. Gioia, M. Bertazzo, M. Recanatini, M. Masetti, and A. Cavalli. Dynamic docking: A paradigm shift in computational drug discovery, *Molecules*, **2017**, 22(11), 2029.

*Figure 6:* The dependence of free energy of the enzyme-ligand complexes on time.
Figure 7: Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) during molecular dynamics simulation

18. J. L. Medina-Franco, M. A. Giulianotti, G. S. Welmaker, and R. A. Houghten. Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov. Today*, 2013, 18(9-10), 495-501.

19. H. Liu et al. *Scutellaria baicalensis* extract and baicalein inhibit replication of SARS-CoV-2 and its 3C-like protease in vitro. *J. Enzyme Inhib. Med. Chem.*, 2021, 36(1), 497-503.

20. H. Su et al. Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients, *Acta Pharmacol. Sin.*, 2020, 41(9), 1167-1177.

21. D. Li, J. Luan, and L. Zhang. Molecular docking of potential SARS-CoV-2 papain-like protease inhibitors, *Biochem. Biophys. Res. Commun.*, 2021, 538, 72-79.

*Corresponding author: Nguyen Van Phuong*  
Department of Pharmacognosy, Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Phan Chu Trinh, Hoan Kiem, Hanoi, 11000, Viet Nam  
E-mail: phuongnv@hup.edu.vn
### Table 1: SMILES and pharmacokinetic parameters of 79 potential flavonoids

| Comp | Pubchem ID | SMILES                                                                 | GI absorption | BBB permeant | Pgp substrate | CYP inhibitions         |
|------|------------|------------------------------------------------------------------------|---------------|--------------|---------------|--------------------------|
| 55   | 46216999   | O[C@@H]1Ce2e(O)cc(c3e2[C@@H][C@@H]1)Ce1e3oe2e(c1=O)(O)ce1e2C=C(C=C(0)(C)C)O | High          | No           | No            | CYP2C9                  |
| 74   | 44258658   | CC(=C[C@@H]1Ce2e(=O)ce3e(O)ce4e(c3oe2[C@@H][23][C@@H]1)(O2)(C(=C3 =O)O)Ce=CC(C(=O)(C)(C)C) | High          | No           | No            | CYP2C9, CYP3A4           |
| 259  | 11486856   | COc1el(C(=O)Ce2eccc2e)Ce(c2e1=C(Ce1eccc1)C(=O)O2)O                    | High          | No           | No            | CYP2C9, CYP3A4           |
| 304  | 42607594   | OC1C[C@@H]1Ce2eccc2e2(O1)(C(=C(c1ccc(C1=C(C(=O)CC(=O2)C(=CC3=C(=O)O)Ce=CC(C(=O)(C)(C)C)C) | High          | No           | No            | CYP2C9                  |
| 344  | 44259049   | OC1Ce2eccc1e(ce3e(e3=O)O)Ce1eccc(e1)O                                  | High          | No           | No            | CYP2C9                  |
| 369  | 124350926  | CC(=CCel(O)ce(e2e1e3ecco(e30)(C=([@H](c1e2=O)C=C(C=C(C)O)O)C) | High          | No           | No            | CYP2C9, CYP3A4           |
| 597  | 5464381    | COc1ecco(e2e1e3ecco(e30)(C=C(=O)(C)C)C)                              | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 726  | 123684269  | O=C1oce2e(O)oce2e1Ce1eccc(e1)O                                      | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 748  | 11036802   | Oc1ecco(e1e1e2e3ecco(e30)(Oe1e10)(O)CC(O)C                        | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 962  | 10087098   | Oc1ecco(e2e1e3ecco(e30)(Oe1e10)(O)CC(O)C                        | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 981  | 44130318   | COc1ecco(Ce2ecco(e30)(O)cc10)O                                       | High          | Yes          | No            | CYP1A2, CYP2C9, CYP2D6   |
| 1035 | 10221174   | Oc1ecco(e1e1e3ecco(e30)(O)cc10)O                                     | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4   |
| 1189 | 10070028   | C(C=CCecco(e1e10)Ce1ecco(e30)Ce1c1ecco(e30)OCCC=C(C)C                | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4   |
| 1228 | 71713861   | C=C(Cecco(e10)OCC1Cecco(e20)Cecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP2C9, CYP3A4           |
| 1322 | 10070990   | O=C(C=CCecco(e10)Ce1ecco(e20)OCC1Cecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP2C9, CYP3A4           |
| 1390 | 6450959    | OC(C=CCecco(e10)Ce1ecco(e20)OCC1Cecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP2C9, CYP2D6           |
| 1715 | 71728402   | Oc1ecco(e1e1Cecco(e30)Ce1ecco(e10)O                                 | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4   |
| 1796 | 21637735   | CC=C(Cecco(CCC2ecco(e20)O)cc10e1e10ecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP2C9                  |
| 1811 | 14539881   | C(C=C(Cecco(C=CC(C=CCecco(e20)Occ10ecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4   |
| 1815 | 10342292   | C(C=C(Cecco(C=CC(C=CCecco(e20)Occ10ecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4   |
| 1858 | 11303195   | Oc1ecco(e1e1ecco(e20)Ce1ecco(e10)OCC1Cecco(e10)OCCC=C(C)C                | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6   |
| Comp | Pubchem ID | SMILES | GI absorption | BBB permeant | Pgp substrate | CYP inhibitions |
|------|------------|--------|---------------|--------------|---------------|----------------|
| 1860 | 50411994   | CC(=C(C=Cc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)O) | High          | No           | No            | CYP2C9, CYP2D6  |
| 1865 | 14237664   | CC(=C(Cc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)C(=O)(C(=O)O)C) | High          | No           | No            | CYP2C9         |
| 1916 | -          | Oc1c1(c1)cc(c1c(O)c1(O)c1c1cc(c1)C(=O)(C2)CC(=C(c2)O)O)CC=O(C2)=O/cc1ccc1 | High          | No           | No            | No             |
| 1976 | 21721840   | CC(=C(Cc1cc(c1O)O)C1CC(=C(=O)c1c(c2O)O)C)CC=C(C)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 1991 | 131751488  | Oc1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2042 | 5481948    | Oc1c1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2262 | 102334441  | C(=C(Cc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)C)CC=C(C)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2275 | 73950872   | CC(=C(Cc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)C)CC=C(C)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2333 | 25016411   | C(=CC(=C(CCc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)O)O)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2373 | -          | C(=CC(=C(CCc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)O)O)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2393 | 44139000   | C(=CC(=C(CCc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)O)O)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2475 | -          | C(=CC(=C(CCc1c(O)cc2c(c1O)c(=O)c1c(2)c21cc(c2)O)O)O)C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2534 | 5320451    | Oc1c1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/C | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2567 | 11142766   | OCc1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/O | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2669 | 5381322    | Oc1c1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/O | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2680 | 5319492    | Oc1c1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/O | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2688 | 85981660   | Oc1c1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/O | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| 2699 | 25087106   | Oc1c1c1cc(c1)C=C=O(c1)O c1(O)c1c1cc(c1)O)O/O | High          | No           | No            | CYP1A2, CYP2C9, CYP3A4 |
| Comp | Pubchem ID | SMILES | GI absorption | BBB permeant | Pgp substrate | CYP inhibitions |
|------|------------|--------|---------------|--------------|---------------|----------------|
| 2726 | -          | O[C@@H]1Oc2e(2c(1)=O)[C@@H]1c1ceccc1O)c1O)|O)O | High | No | No | No | CYP1A2, CYP2D6, CYP3A4 |
| 2752 | 91665876   | Oe1ce(1)oe2(1)oecc(2=O)oe1ce(2)oe(2=O) | High | No | No | No | CYP1A2, CYP2D6, CYP3A4 |
| 2815 | -          | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2D6 |
| 2826 | 44259962   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9, CYP2D6 |
| 2837 | -          | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9, CYP2D6 |
| 2858 | 44257530   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9 |
| 2919 | 12096317   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C19, CYP2C9 |
| 3003 | 5281294    | Oe1ce(1)O)c1O)O | High | No | No | No | CYP1A2, CYP2C9, CYP3A4 |
| 3008 | 71614261   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP3A4 |
| 3035 | 71713860   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP3A4 |
| 3038 | 5481948    | Oe1ce(1)O)c1O)O | High | No | No | No | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3084 | 44562556   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP3A4 |
| 3124 | 480772     | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP2D6, CYP3A4 |
| 3126 | 14542252   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP2D6, CYP3A4 |
| 3127 | 42607994   | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP3A4 |
| 3146 | 137637878  | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP3A4 |
| 3147 | 134751536  | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP2C9, CYP3A4 |
| 3291 | 52951750   | Oe1ce(1)O)e2c(1=O)e(2=O)e(1=O)e(1=O)e1ce(1)O)O | High | No | No | No | CYP1A2, CYP2D6 |
| 3342 | 5377945    | Oe1ce(1)O)e2c(1=O)e(2=O)e(1=O)e(1=O)e1ce(1)O)O | High | No | No | No | CYP1A2, CYP2D6, CYP3A4 |
| 3384 | -          | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3412 | 73205      | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3429 | -          | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9, CYP2D6 |
| 3436 | -          | CC(=CC(2c(1=O)ce(2)=O)2c1ce(1)O)O)C | High | No | No | No | CYP1A2, CYP2C9, CYP2D6 |
| Comp | Pubchem ID | SMILES | GI absorption | BBB permeant | Pgp substrate | CYP inhibitions |
|------|------------|---------|---------------|--------------|---------------|----------------|
| 3540 | 3540       | Oc1cc(O)c2e(c10)oe1c(e2=O)OCc2e1ccc1OCc2e1ccc2 | High          | No           | No            | CYP1A2, CYP2D6, CYP3A4 |
| 3560 | 5320945    | CCo1cc(e1c(c2=O)OCc2e1ccc1)OC | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3612 | 10251762   | CC=C(e1e(e10)OC2eCl=O)Ce2eCl=OOCc2e1ccc1OCc2e1ccc2 | High          | No           | No            | CYP2C9, CYP3A4 |
| 3680 | 131834785  | Oc1cc(O)c2e(c10)oe1c(e2=O)oe2e1ccc1OCc2e1ccc2 | High          | No           | No            | CYP1A2, CYP2D6 |
| 3702 | 148856     | CCo1cc(e1c(c3=O)OCc3e1ccc1)OC | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3714 | 13873661   | O=C(c1e(O)oce1e10)Ce1ecccc1OCc1ecccc1 | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3732 | -          | Oe1cc(O)oce1e10)Ce1ecccc10)CCc1ecccc1 | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6, CYP3A4 |
| 3765 | 13964266   | CC=C(e1e(O)oce1e10)oce2e1cc(O)oce2e1cc1cc1O | High          | No           | No            | CYP1A2, CYP2C9 |
| 3900 | 25208435   | CC=C(e1e(O)oce1e10)oce2e1cc2Oce2e1cocc2e1cc2Oc | High          | No           | No            | CYP1A2, CYP2C9, CYP2D6 |
| 4163 | 15227646   | Oc1cc(O)oce1e10)Ce1ecccc1CCc1ecccc1OCc1ecccc1 | High          | No           | No            | CYP2C9, CYP3A4 |
| 4215 | 10021410   | COc1cc(O)oce1e10)Ce1ecccc1OCc1ecccc10)OCc1ecccc1 | High          | Yes          | No            | CYP1A2, CYP2C19, CYP2C9, CYP3A4 |
| 4255 | -          | Oc1ecccc1(c1)ecccc1=O)oce2e1ccc10)ccoc1e1c(c2)O | High          | No           | No            | CYP1A2, CYP2D6, CYP3A4 |
| 4272 | 10090416   | CC=C(e1e(O)oce1e10)oce2e1cc1cc10)OCc1ecccc1OCc1ecccc1 | High          | No           | No            | CYP2C9 |
| 4276 | 5324358    | CC=C(e1e(O)oce1e10)oce2e1cc1cc10)OCc1ecccc1OCc1ecccc1 | High          | No           | No            | CYP2C9 |
| 4322 | -          | CC=C(e1e(O)oce1e10)oce2e1cc1cc10)OCc1ecccc1OCc1ecccc1 | High          | No           | No            | CYP2C9, CYP2D6 |
| 4347 | 10980660   | COc1ee1e(O)oce1e10)Ce1ecccc1(C=c=O)Ce1ecccc1 | High          | Yes          | No            | CYP2C19, CYP2C9, CYP3A4 |
Table 2: Toxicity of 79 potential flavonoids predicted by Toxtree

| Compound | No alerts for S. typhimurium mutagenicity | Crammer rules | Negative for carcinogenicity | No DNA binding alerts identified |
|----------|------------------------------------------|--------------|-----------------------------|-------------------------------|
| 55       | No                                       | High (Class III) | Yes                         | No                            |
| 74       | No                                       | High (Class III) | Yes                         | No                            |
| 259      | Yes                                      | High (Class III) | Yes                         | No                            |
| 304      | No                                       | High (Class III) | Yes                         | No                            |
| 344      | Yes                                      | High (Class III) | Yes                         | No                            |
| 369      | No                                       | High (Class III) | Yes                         | No                            |
| 597      | Yes                                      | High (Class III) | Yes                         | No                            |
| 726      | No                                       | High (Class III) | Yes                         | No                            |
| 748      | No                                       | High (Class III) | Yes                         | No                            |
| 962      | Yes                                      | High (Class III) | Yes                         | No                            |
| 981      | Yes                                      | High (Class III) | Yes                         | No                            |
| 1035     | Yes                                      | High (Class III) | Yes                         | No                            |
| 1189     | No                                       | High (Class III) | Yes                         | No                            |
| 1228     | No                                       | High (Class III) | Yes                         | No                            |
| 1322     | Yes                                      | High (Class III) | Yes                         | No                            |
| 1390     | No                                       | High (Class III) | Yes                         | No                            |
| 1715     | Yes                                      | High (Class III) | Yes                         | No                            |
| 1796     | No                                       | High (Class III) | Yes                         | No                            |
| 1811     | No                                       | High (Class III) | Yes                         | No                            |
| 1815     | No                                       | High (Class III) | Yes                         | No                            |
| 1858     | No                                       | High (Class III) | Yes                         | No                            |
| 1860     | No                                       | High (Class III) | Yes                         | No                            |
| 1865     | No                                       | High (Class III) | Yes                         | No                            |
| 1916     | Yes                                      | High (Class III) | Yes                         | No                            |
| 1976     | No                                       | High (Class III) | Yes                         | No                            |
| 1991     | No                                       | High (Class III) | Yes                         | No                            |
| 2042     | Yes                                      | High (Class III) | Yes                         | No                            |
| 2262     | No                                       | High (Class III) | Yes                         | No                            |
| 2275     | No                                       | High (Class III) | Yes                         | No                            |
| 2333     | No                                       | High (Class III) | Yes                         | No                            |
| 2373     | No                                       | High (Class III) | Yes                         | No                            |
| 2393     | No                                       | High (Class III) | Yes                         | No                            |
| 2475     | No                                       | High (Class III) | Yes                         | No                            |
| 2534     | No                                       | High (Class III) | Yes                         | No                            |
| 2567     | No                                       | High (Class III) | Yes                         | No                            |
| 2669     | Yes                                      | High (Class III) | Yes                         | No                            |
| 2680     | Yes                                      | High (Class III) | Yes                         | No                            |
| 2688     | Yes                                      | High (Class III) | Yes                         | No                            |
| 2699     | Yes                                      | High (Class III) | Yes                         | No                            |
| 2726     | Yes                                      | High (Class III) | Yes                         | No                            |
| Compound | No alerts for *S. typhimurium* mutagenicity | Crammer rules | Negative for carcinogenicity | No DNA binding alerts identified |
|----------|------------------------------------------|---------------|-----------------------------|---------------------------------|
| 2752     | Yes                                      | High (Class III) | Yes                         | No                              |
| 2815     | Yes                                      | High (Class III) | Yes                         | No                              |
| 2826     | No                                       | High (Class III) | Yes                         | No                              |
| 2837     | No                                       | High (Class III) | Yes                         | No                              |
| 2858     | No                                       | High (Class III) | Yes                         | No                              |
| 2919     | No                                       | High (Class III) | Yes                         | No                              |
| 3003     | No                                       | High (Class III) | Yes                         | No                              |
| 3008     | No                                       | High (Class III) | Yes                         | No                              |
| 3035     | No                                       | High (Class III) | Yes                         | No                              |
| 3038     | Yes                                      | High (Class III) | Yes                         | No                              |
| 3084     | No                                       | High (Class III) | Yes                         | No                              |
| 3124     | No                                       | High (Class III) | Yes                         | No                              |
| 3126     | No                                       | High (Class III) | Yes                         | No                              |
| 3127     | No                                       | High (Class III) | Yes                         | No                              |
| 3146     | No                                       | High (Class III) | Yes                         | No                              |
| 3147     | No                                       | High (Class III) | Yes                         | No                              |
| 3291     | No                                       | High (Class III) | Yes                         | No                              |
| 3342     | Yes                                      | High (Class III) | Yes                         | No                              |
| 3384     | No                                       | High (Class III) | Yes                         | No                              |
| 3412     | No                                       | High (Class III) | Yes                         | No                              |
| 3429     | No                                       | High (Class III) | Yes                         | No                              |
| 3436     | Yes                                      | High (Class III) | Yes                         | No                              |
| 3540     | No                                       | High (Class III) | Yes                         | No                              |
| 3560     | No                                       | High (Class III) | Yes                         | No                              |
| 3612     | No                                       | High (Class III) | Yes                         | No                              |
| 3680     | No                                       | High (Class III) | Yes                         | No                              |
| 3702     | No                                       | High (Class III) | Yes                         | No                              |
| 3714     | Yes                                      | High (Class III) | Yes                         | No                              |
| 3732     | Yes                                      | High (Class III) | Yes                         | No                              |
| 3765     | No                                       | High (Class III) | Yes                         | No                              |
| 3841     | No                                       | High (Class III) | Yes                         | No                              |
| 3900     | No                                       | High (Class III) | Yes                         | No                              |
| 4163     | No                                       | High (Class III) | Yes                         | No                              |
| 4215     | No                                       | High (Class III) | Yes                         | No                              |
| 4255     | No                                       | High (Class III) | Yes                         | No                              |
| 4272     | No                                       | High (Class III) | Yes                         | No                              |
| 4276     | No                                       | High (Class III) | Yes                         | No                              |
| 4322     | No                                       | High (Class III) | Yes                         | No                              |
| 4347     | Yes                                      | High (Class III) | Yes                         | No                              |