International Journal of *Plant Biology*

MDPI

Technical Note

Theoretical Prediction of Gastrointestinal Absorption of Phytochemicals

Luis A. Vélez [1], Yamixa Delgado [1,*], Yancy Ferrer-Acosta [2], Ivette J. Suárez-Arroyo [2], Priscilla Rodríguez [1] and Daraishka Pérez [1]

[1] Department of Biochemistry and Pharmacology, San Juan Bautista School of Medicine, Urb. Turabo Gardens PR-172, Caguas, PR 00726, USA; luisvl@sanjuanbautista.edu (L.A.V.); priscillamrt@sanjuanbautista.edu (P.R.); daraishkapc@sanjuanbautista.edu (D.P.)

[2] School of Medicine, Universidad Central del Caribe, Ave. Laurel 100, Sta. Juanita, Bayamón, PR 00960, USA; yancy.ferrer@uccaribe.edu (Y.F.-A.); ivette.suarez@uccaribe.edu (I.J.S.-A.)

* Correspondence: ydelgado@sanjuanbautista.edu; Tel.: +1-787-743-3038 (ext. 3220)

Abstract: The discovery of bioactive compounds for non-invasive therapy has been the goal of research groups focused on pharmacotherapy. Phytonutrients have always been attractive for researchers because they are a significant source of bioactive phytochemicals. Still, it is challenging to determine which components show high biomedical activity and bioavailability after administration. However, based on the chemical structure of these phytochemicals, their physicochemical properties can be calculated to predict the probability of gastrointestinal (GI) absorption after oral administration. Indeed, different researchers have proposed several rules (e.g., Lipinski's, Veber's, Ghose's, and Muegge's rules) to attain these predictions, but only for synthetic compounds. Most phytochemicals do not fully comply with these rules even though they show high bioactivity and high GI absorption experimentally. Here, we propose a detailed methodology using scientifically validated web-based platforms to determine the physicochemical properties of five phytochemicals found in ginger, echinacea, and tobacco. Furthermore, we analyzed the calculated data and established a protocol based on the integration of these classical rules, plus other extended parameters, that we called the Phytochemical Rule, to obtain a more reliable prediction of the GI absorption of natural compounds. This methodology can help evaluate bioactive phytochemicals as potential drug candidates and predict their oral bioavailability in patients.

Keywords: physicochemical properties; gastrointestinal absorption; Lipinski's rules; Veber's rules; Ghose's rules; Muegge's rules; phytochemicals

Citation: Vélez, L.A.; Delgado, Y.; Ferrer-Acosta, Y.; Suárez-Arroyo, I.J.; Rodríguez, P.; Pérez, D. Theoretical Prediction of Gastrointestinal Absorption of Phytochemicals. *Int. J. Plant Biol.* **2022**, *13*, 163–179. https://doi.org/10.3390/ijpb13020016

Academic Editor: Adriano Sofo

Received: 27 May 2022
Accepted: 15 June 2022
Published: 17 June 2022

1. Introduction

The physicochemical properties of a drug candidate can be used to understand and predict its physiological absorption and, therefore, determine the chances of having a biological effect after oral consumption [1]. These properties can be theoretically calculated to improve and facilitate the drug design process. Also, they can help us create guidelines to understand the behavior of our drug candidates in a biological environment such as the gastrointestinal (GI) tract. We can identify potential orally absorbable and non-absorbable bioactive compounds using these guidelines. Several researchers have created different pharmacokinetic rules to aid in predicting whether a compound is likely to be absorbed and is readily permeable to the GI tract. Among these classical rules is Lipinski's rule of five (L-Ro5), Ghose filter (GF), Veber's rule (VR), and Muegge's rule (MR). These rules are mainly focused on lipophilicity, electronic distribution, hydrogen bonding, molecule size, and structural flexibility. Since their creation, up to the present, these rules have helped predict the absorption of molecules in the GI tract and define compound drug-likeness. A typical drug-like synthetic compound is a molecule that falls into the proposed ranges of these rules. Pharmaceutical companies have made wide use of

these theoretical rules to create groups of active molecules with favorable physicochemical properties to develop novel drugs [2]. Based on these rules, a large number of active synthetic and natural compounds are rejected because they fall out of the established ranges. However, these rules (the selected physicochemical properties and molecular ranges of the compounds) were established to evaluate only synthetic compounds. Nevertheless, there is little knowledge regarding rules specific to predicting the GI absorption of compounds derived from plants. In this area, we found studies that incorporate the additional property of molecular complexity (Cm) [3] and the extension of specific ranges to the previously proposed classical rules [4].

Plants rich in phytochemicals have been used for centuries in traditional medicine. Because of their well-known bioactivity, there is continuous study within evidence-based medicine of the chemical compositions and biological activity of plants and their phytochemicals (as isolated bioactive compounds) [5]. Phytochemicals are divided into several categories based on their structural features i.e., phenolics, terpenoids, alkaloids, and organosulfur compounds [6]. This review article systematically analyzes five phytochemicals found within three plants: echinacea, ginger, and tobacco. We selected phytochemicals from echinacea and ginger due to the recent increase in their consumption to strengthen the immune system as we face the COVID-19 pandemic [7,8]. On the other hand, we selected tobacco phytochemicals due to the well-known functions of its bioactive phytochemicals, such as nicotine [9] and cembranoids [10], which can have very different biological activities, yet are found within the same plant. In addition, we included in our analyses a well-known natural compound as theoretical control (ascorbic acid (vitamin C)) to compare and validate our methodology and GI absorption predictions for natural products.

Our previous review article [6], explained the biomedical effect of phytochemicals in several plants. From this article, we understood that the ranges for these theoretical rules did not consider phytochemicals, which could exclude their study in drug discovery. The objective of this systematic review is to clearly explain our methodology using the scientifically validated web-based platforms PubChem, SwissADME, and ChemSpider-ACD/Labs. To validate the calculations from these platforms, we include published experimental values for the lipophilicity (LogP) of each phytochemical. We propose inclusion of the Cm property and increasing the ranges of the classical rules to develop the best approximation of GI absorption prediction for natural products. Thus, this study combines the knowledge of several established properties and ranges [3,4], to make a final prediction, which we named the Phytochemical Rule (PR) to study natural compounds in drug discovery.

It is important to mention that obtaining these calculations is free of charge, and the theoretical and experimental information was acquired from trusted scientific platforms (PubMed, SwissADME, ChemSpider). The presented data and methodology can be instrumental in evaluating bioactive natural compounds as potential drug candidates and predicting their bioavailability in patients. As important as these data are, obtaining them is relatively fast and can be readily done at the initial stages of drug discovery assessment.

2. The Classical Rules and the Phytochemical Rule (PR)

L-Ro5 is the main set of rules created to predict the drug-likeness of small synthetic compounds. This rule states that poor absorption and permeation are more likely when the molecular weight (MW) is over 500 Da, lipophilicity (LogP) and hydrogen-bond donors (HBD) are more than five, and there are more than ten hydrogen-bond acceptors (HBA) [11]. The authors concluded that the selected ranges for these properties were delimited based on synthetic compounds, but they never claim that these properties can only be used for synthetic compounds. Later, Lipinski understood that the range limits in his rules must be different to evaluate natural compounds [12]. After his conclusion, several research groups have worked to include other properties and adjust the ranges to study small natural compounds. The Ghose filter (GF) attempts to improve prediction by stating that high absorption is likely with the following criteria: MW of 160 to 480 Da, a LogP of −0.4 to 5.6, a molar refractivity (A, cm^3) of 40 to 130, and a total number of atoms (TNA) of 20 to

70 [13]. Veber's rule (VR) further increases the criteria for bioavailability with less than ten rotatable bonds (RB) and a polar surface area (PSA, Å^2) no greater than 140 [14]. Muegge's rule extends the ranges of several properties and included other parameters to differentiate between drug-like and nondrug-like compounds by stricter rules. These are: MW 200–600, LogP from −2 to 5, PSA $\leq$ 150, number of rings (NR) $\leq$ 7, number of carbons (NC) > 4, number of heteroatoms (NH) > 1, RB $\leq$ 15, HBD $\leq$ 5, HBA $\leq$ 10 [15].

Lipophilicity (LogP) is one of the most important properties of all these rules, defined as the partition coefficient ratio of a compound between the hydrophobic and hydrophilic phases [16]. Other researchers have proposed that the lipophilicity of the ionizable groups at pH 7.4, called LogD, is much more critical for physiological absorption [17]. However, recent literature has shown that the determination of LogD is not easy because the calculated pKa and hence, LogD values are, in some cases, very different from those found experimentally [18]. Consequently, we have included LogD values, but they will not be used for predictions.

The other properties in these rules mainly focus on the molecules' interactions with themselves, the solvent, and additional molecules around. On the other hand, the molecular complexity (Cm) is another property considered important to predict GI absorption that is a rough estimate of how complicated the structure is, seen from the point of view of both the elements contained and the displayed structural features, including symmetry [19]. In general, larger compounds display greater complexity than smaller ones, but large symmetrical compounds and large compounds with low diversity of atoms are considered less complex. A recent study has investigated whether Cm can be a useful property in medicinal chemistry by calculating Cm values for approved drugs of different major classes of synthetic and natural antibiotics. The results demonstrate that a Cm of 100–900 had favorable outcomes for absorption and permeation for synthetic and natural compounds [3].

Interestingly, other researchers, in studying hundreds of clinical orally administered drugs, concluded that a larger LogP, MW, PSA, and HBA could be allowed, especially in natural products. Stratton et al. showed that some structural features and properties in synthetic products could be successfully extrapolated into natural products but display greater chemical diversity and flexibility [20]. Croy et al., argued that to be an orally-bioavailable compound (synthetic or natural), these properties need to be balanced depending on its chemical features [4]. Thus, they studied these classical rules (L-Ro5, GF, VR, MR) in natural compounds, and, after their analysis, proposed to increase the ranges of these properties to effectively apply them to natural compounds. The ranges from this study are: MW $\leq$ 800 Da, TNA $\leq$ 80, −2 $\leq$ LogP $\leq$ 7, HB $\leq$ 6, HBA $\leq$ 15, PSA $\leq$ 250 Å^2, and RB $\leq$ 20.

Based on these studies that used natural compounds, we developed the Phytochemical Rule (PR) that includes the Cm property and the extension of the ranges for LogP, MW, PSA, and HBA to predict the drug-likeness by GI absorption of phytochemicals. It is also important to mention that the predictions of all these rules are established on molecules passively transported into the cells. Thus, L-Ro5, GF, VR, MR, Cm, and ER do not consider actively transported substrates by biological transporters (e.g., endocytosis) [21]. Furthermore, we evaluated the phytochemicals from these plants as isolated compounds because the GI absorption of phytonutrient extracts needs the evaluation of additional effects. For example, the synergistic effect between the different metabolites in the extract that influences the phytochemicals' absorption [22], and the formation of emulsions/suspensions in aqueous plant extracts also affects the absorption after oral administration [23].

For this work, we obtained the values for the molecular formula, the molecular structure, MW (Da), TNA, HBA, HBD, RB, LogP, LogD, Cm, PSA, and A for the theoretical predictions, using the web-based platforms PubChem, ChemSpider/ACD Labs, and SwissADME.

3. Methodology

3.1. Data Source Platform: PubChem

The primary data source was obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov; accessed on 10 February 2022). First, the name of each phytochemical was typed into the database's search engine. Then, the program calculated and provided the values of different physicochemical properties of the searched compound (Figure 1).

i. Search for the common compound name on the PubChem engine.
ii. This engine will provide the structure, molecular formula, molecular weight, LogP, HBD, HBA, RB, PSA, A and Cm of the chosen compound. We included these parameters in Tables 1 and 2 for each phytochemical.
iii. To determine the total number of atoms (TNA) for each compound, we manually added the number of atoms in the molecular formula.

PUBCHEM > ASCORBIC ACID > COMPUTED PROPERTIES

CID 54670067

Ascorbic acid

Computed Properties

Showing 1 of 1

Property Name	Property Value	Reference
Molecular Weight	176.12	Computed by PubChem 2.2 (PubChem release 2021.10.14)
XLogP3	-1.6	Computed by XLogP3 3.0 (PubChem release 2021.10.14)
Hydrogen Bond Donor Count	4	Computed by Cactvs 3.4.8.18 (PubChem release 2021.10.14)
Hydrogen Bond Acceptor Count	6	Computed by Cactvs 3.4.8.18 (PubChem release 2021.10.14)
Rotatable Bond Count	2	Computed by Cactvs 3.4.8.18 (PubChem release 2021.10.14)
Exact Mass	176.03208797	Computed by PubChem 2.2 (PubChem release 2021.10.14)
Monoisotopic Mass	176.03208797	Computed by PubChem 2.2 (PubChem release 2021.10.14)
Topological Polar Surface Area	107 Å²	Computed by Cactvs 3.4.8.18 (PubChem release 2021.10.14)
Heavy Atom Count	12	Computed by PubChem
Formal Charge	0	Computed by PubChem
Complexity	232	Computed by Cactvs 3.4.8.18 (PubChem release 2021.10.14)
Isotope Atom Count	0	Computed by PubChem
Defined Atom Stereocenter Count	2	Computed by PubChem
Undefined Atom Stereocenter Count	0	Computed by PubChem
Defined Bond Stereocenter Count	0	Computed by PubChem
Undefined Bond Stereocenter Count	0	Computed by PubChem
Covalently-Bonded Unit Count	1	Computed by PubChem
Compound Is Canonicalized	Yes	Computed by PubChem (release 2021.10.14)

Figure 1. PubChem results on ascorbic acid properties. We selected MW, HBA, HBD, RB, LogP, PSA, and Cm from these values. They can be found following the instructions mentioned above in the data source PubChem (https://pubchem.ncbi.nlm.nih.gov; National Center for Biotechnology Information; accessed on 10 February 2022) and are summarized in Tables 1 and 2.

Table 1. Structural and physicochemical properties of the main phytochemicals in echinacea, tobacco, and ginger.

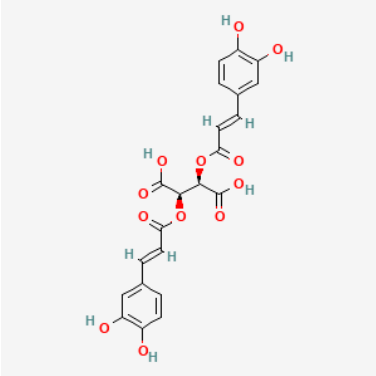
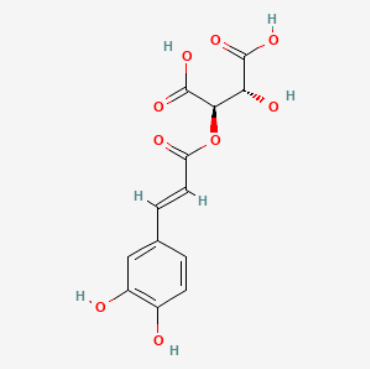
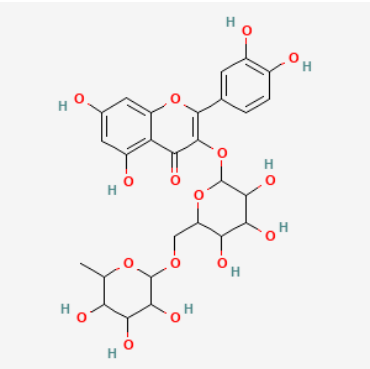
Name/ Category	Structure/Molecular Formula	MW (Da)	TNA/HBA/HBD/RB	A (cm^3)	PSA ($Å^2$)	Cm
		Echinacea				
Cichoric acid		474.4	52/12/6/11	114	208	740
Phenolic acid	$C_{22}H_{18}O_{12}$					
Caftaric acid		312.2	34/9/5/7	70.6	162	458
Phenolic acid	$C_{13}H_{12}O_9$					
Quercetin-3-O-rutinoside		610.5	73/16/10/6	141.4	269.4	1020
Phenolic Flavonoid	$C_{27}H_{30}O_{16}$					

Table 1. *Cont.*

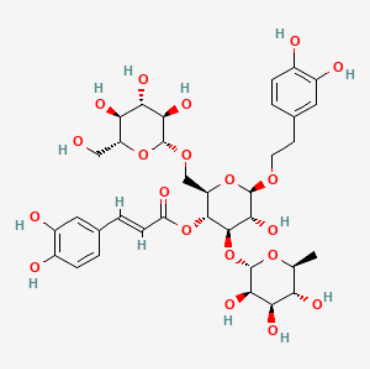
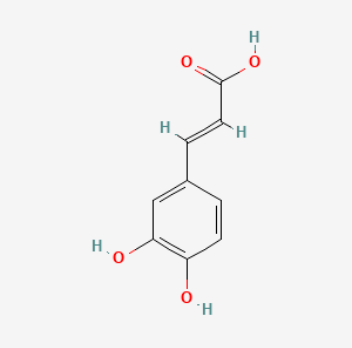
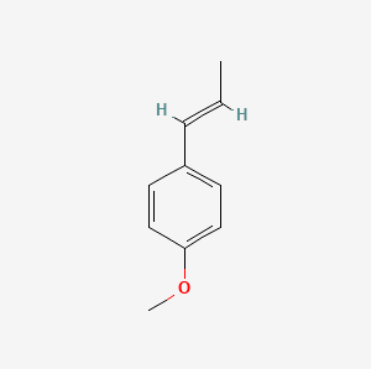
Name/ Category	Structure/Molecular Formula	MW (Da)	TNA/HBA/HBD/RB	A (cm^3)	PSA ($Å^2$)	Cm
Echinacoside		786.7	101/20/12/14	180.8	324	1230
Phenolic glycoside	$C_{35}H_{46}O_{20}$					
Caffeic acid		180.2	21/4/3/2	47.2	77.8	212
Phenolic acid	$C_9H_8O_4$					
	Tobacco					
Anethole		148.2	23/1/0/2	47.8	9.2	121
Phenolic stilbene	$C_{10}H_{12}O$					

Table 1. *Cont.*

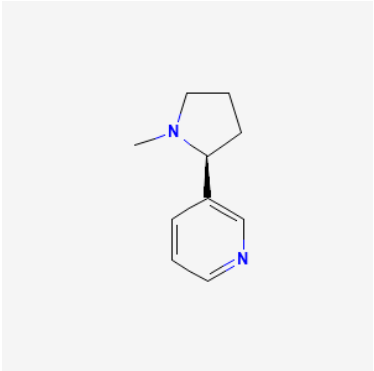
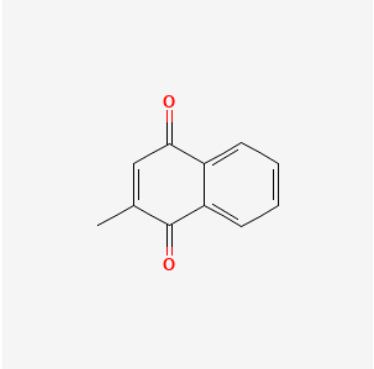
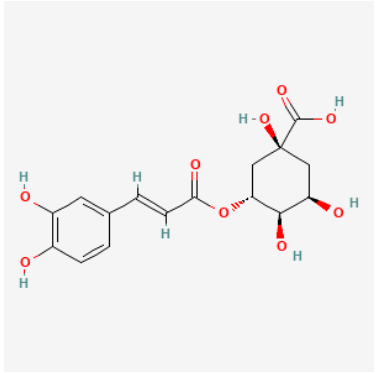
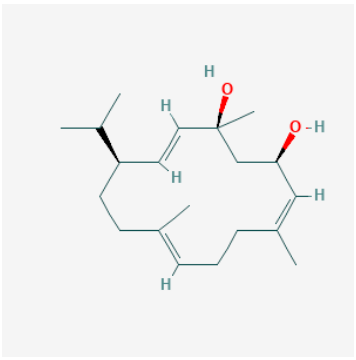
Name/ Category	Structure/Molecular Formula	MW (Da)	TNA/HBA/HBD/RB	A (cm^3)	PSA ($Å^2$)	Cm
Nicotine		162.2	26/2/0/1	47.8	9.2	147
Alkaloid	$C_{10}H_{14}N_2$					
Menadione		172.2	21/2/0/0	53.1	16.1	289
Phenolic Flavonoid	$C_{11}H_8O_2$					
Chlorogenic acid		354.3	43/9/6/5	49.1	34.1	534
Phenolic acid	$C_{16}H_{18}O_9$					
Cembra-2,7,11-triene-4,6-diol		306.5	56/2/2/1	83.5	164.8	431
Cyclic terpene	$C_{20}H_{34}O_2$					

Table 1. *Cont.*

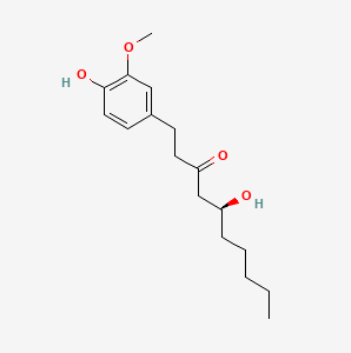
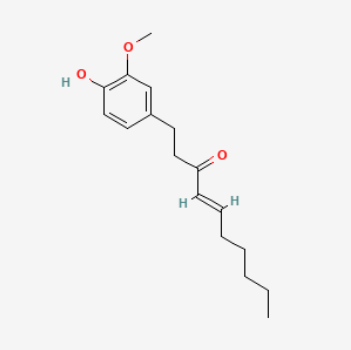
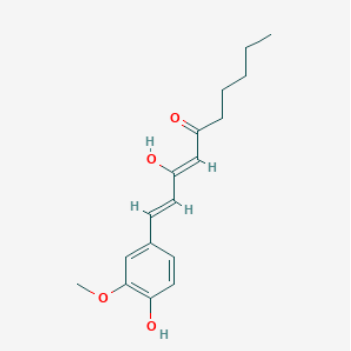
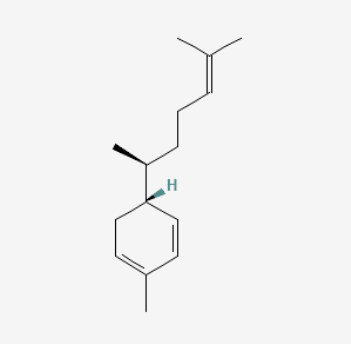
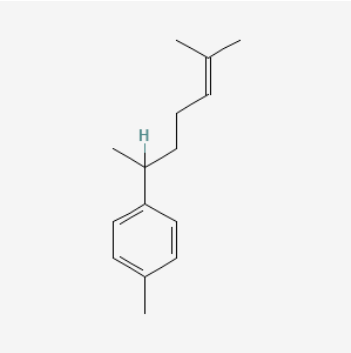
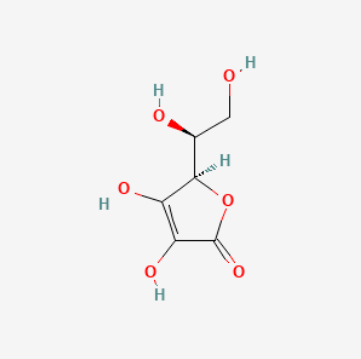
Name/ Category	Structure/Molecular Formula	MW (Da)	TNA/HBA/HBD/RB	A (cm^3)	PSA ($Å^2$)	Cm
		Ginger				
6-Gingerol		294.4	47/4/2/10	84.6	66.8	293
Polyphenol	$C_{17}H_{26}O_4$					
6-Shogaol		276.4	44/3/1/9	82.9	46.5	299
Polyphenol	$C_{17}H_{24}O_3$					
6-Dehydro gingerdione		290.4	43/4/2/8	84.8	66.8	373
Polyphenol	$C_{17}H_{22}O_4$					
Zingiberene		204.4	39/0/0/4	70.68	0	274
Sesquiterpene	$C_{15}H_{24}$					

Table 1. *Cont.*

Name/ Category	Structure/Molecular Formula	MW (Da)	TNA/HBA/HBD/RB	A (cm^3)	PSA ($Å^2$)	Cm
α-Curcumene		202.3	37/0/0/4	69.55	0	190
Sesquiterpene	$C_{15}H_{22}$					
	Control					
Ascorbic acid		176.1	20/6/4/2	35.1	107.2	232
Phenolic acid	$C_6H_8O_6$					

MW: molecular weight; A: molar refractivity; HBD: hydrogen bond donors, HBA: hydrogen bond acceptors; RB: rotatable bonds; PSA: polar surface area; TNA: total number of atoms; Cm: molecular complexity. TNA was determined manually by adding the number of atoms in the molecular formula. Cm was determined using PubChem. The remaining values in this table were determined using PubChem, ChemSpider and SwissADME, and there are almost no differences (<±0.5) when comparing across the platforms.

Table 2. Lipophilicity of the main phytochemicals of the selected plants, obtained from PubChem, ChemSpider, and SwissADME databases.

Name	LogP %/$/&/@	LogD *
	Echinacea	
Cichoric acid	2.00/3.81/1.01/1.18	−2.56
Caftaric acid	0.10/1.14/−0.23/−0.53	−4.39
Quercetin -3-O-rutinoside	−1.30/1.76/−1.12/−0.74	−1.75
Echinacoside	−2.10/0.14/−2.08/−1.82	−1.06
Caffeic acid	1.20/1.42/0.93/0.89	−1.74
	Tobacco	
Anethole	3.30/3.17/2.79/2.65	3.08
Nicotine	1.20/0.72/1.50/ND	−0.37
Menadione	2.2/2.38/1.98/0.67	2.02
Chlorogenic acid	−0.4/−0.36/−0.38/−0.7	−3.91
Cembra-2,7,11-triene-4,6-diol	4.00/6.26/3.93/ND	5.34

Table 2. *Cont.*

Name	LogP %/$/&/@	LogD *
	Ginger	
6-Gingerol	2.5/2.48/3.13/2.44	2.88
6-Shogaol	3.70/3.85/3.76/3.78	4.15
6-Dehydro gingerdione	4.20/3.05/3.45/ND	3.17
Zingiberene	5.20/6.60/4.47/ND	5.63
α-Curcumene	5.40/6.22/4.86/5.76	5.20
	Control	
Ascorbic acid	−1.6/−2.41/−1.28/−1.85	−4.99

LogP: lipophilicity; LogD: lipophilicity considering ionizable groups at pH 7.4; Cm: molecular complexity; ND: not determined. %/$/&/@ determined using PubChem/ChemSpider-ACDLabs/Consensus LogP from SwissADME/Experimental LogP from ChemSpider. * determined using ChemSpider.

3.2. Data Source Platform: SwissADME

Our second data source was SwissADME (http://www.swissadme.ch/index.php/; accessed on 13 February 2022). This database requires the input of the Simplified Molecular Input Line System (SMILES) of the compound of interest, which is a chemical notation that allows a user to represent a chemical structure in a way that the computer can use. This notation allows the computation of physicochemical descriptors and predicts small-molecule pharmacokinetics and drug-likeliness to support drug discovery [24]. The program provides the results of several physicochemical properties and pharmacokinetics of the searched compound (Figure 2).

i. Search for the common compound name on the PubChem engine.
ii. Identify the Canonical SMILES in the category of Computed Descriptors.
iii. Go to the SwissADME program and write the SMILES (from PubChem) in the space indicating "Enter a list of SMILES here" and click "Run."
iv. This program will provide the user with the compound's RB, HBD, HBA, A, PSA, Consensus LogP, GI absorption, blood-brain barrier (BBB) permeability, and P-glycoprotein (P-gp) substrate. The values of these parameters are shown in Tables 1–3.

Table 3. Pharmacokinetics of the main bioactive compounds of the selected plants from the SwissADME database.

Name	BBB Permeant	P-gp Substrate	GI Absorption
	Echinacea		
Cichoric acid	No	Yes	Low
Caftaric acid	No	No	Low
Quercetin-3-O-rutinoside	No	Yes	Low
Echinacoside	No	No	Low
Caffeic acid	No	No	High
	Tobacco		
Anethole	Yes	No	High
Nicotine	Yes	No	High
Menadione	Yes	No	High
Chlorogenic acid	No	No	Low

Table 3. *Cont.*

Name	BBB Permeant	P-gp Substrate	GI Absorption
Cembra-2,7,11-triene-4,6-diol	Yes	No	High
Ginger			
6-Gingerol	Yes	No	High
6-Shogaol	Yes	No	High
6-Dehydro gingerdione	Yes	No	High
Zingiberene	No	No	Low
α-Curcumene	No	No	Low
Control			
Ascorbic acid	No	No	High

BBB permeant: blood-brain barrier permeant; Pgp substrate: P-glycoprotein substrate; GI absorption: gastrointestinal absorption predicted by the program for synthetic compounds using the BOILED-Egg algorithm model.

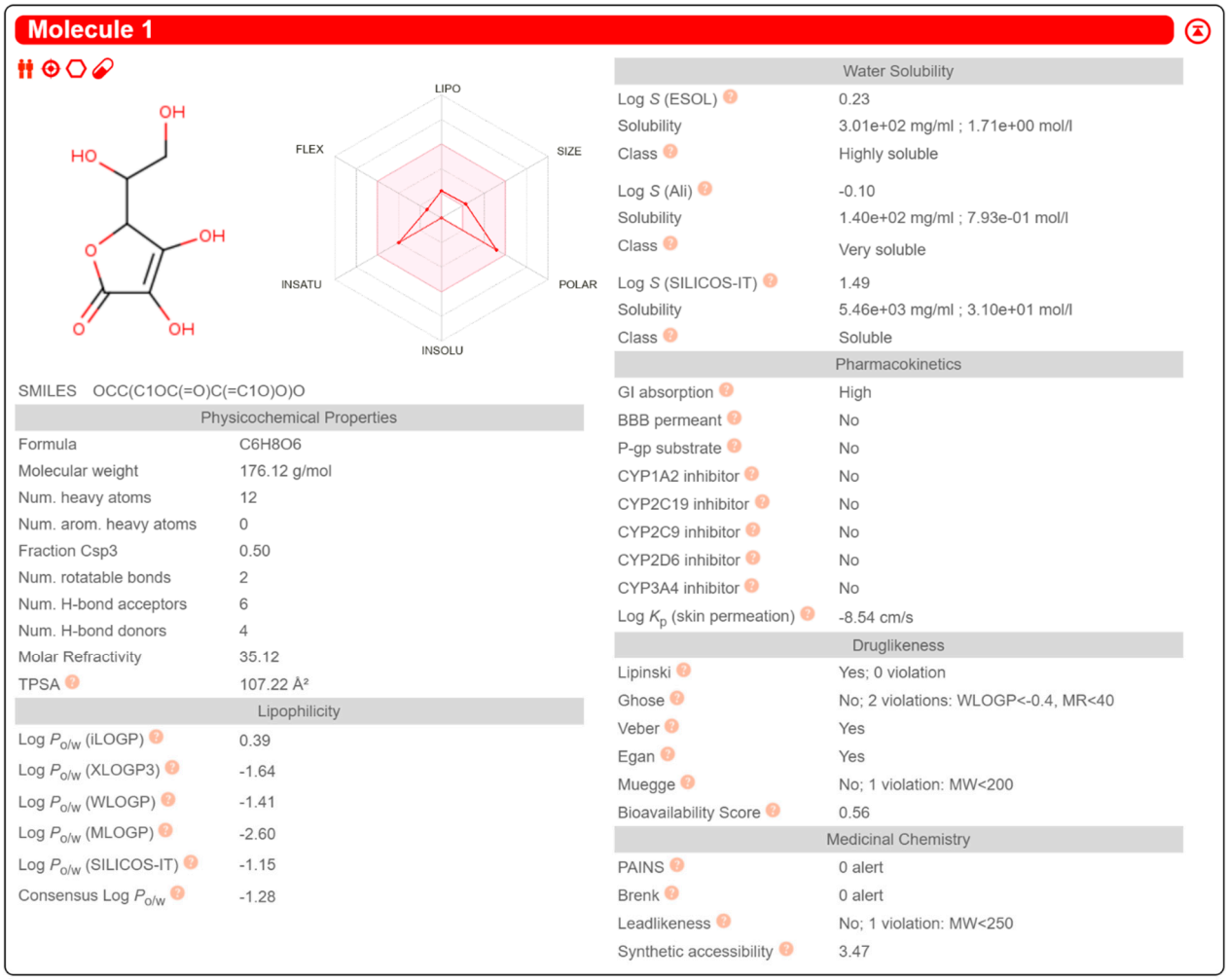


Figure 2. SwissADME results on ascorbic acid properties. We selected RB, HBD, HBA, A, PSA, Consensus LogP, GI absorption, BBB permeant, and P-gp substrate from these values included in Tables 1–3. They can be calculated following the instructions above using the data source SwissADME (http://www.swissadme.ch/index.php/; Swiss Institute of Bioinformatics; accessed on 13 February 2022).

3.3. Data Source Platform: ChemSpider

Another data source used to obtain phytochemical's parameters was ChemSpider (http://www.chemspider.com; accessed on 15 February 2022). First, the name of each phytochemical was typed into the search engine of the database. Then, the program provides results for the values of different physicochemical properties (Figure 3).

i. Search for the common compound name on the ChemSpider engine.
ii. Click on the "Properties" Table
iii. Click on the "Predicted—ACD/Labs" sub-tab.
iv. Look for the parameters ACD/LogP, ACD/LogD(pH 7.4), HBA, HBD, RB, PSA. We included these parameters in Tables 1 and 2 for each phytochemical.

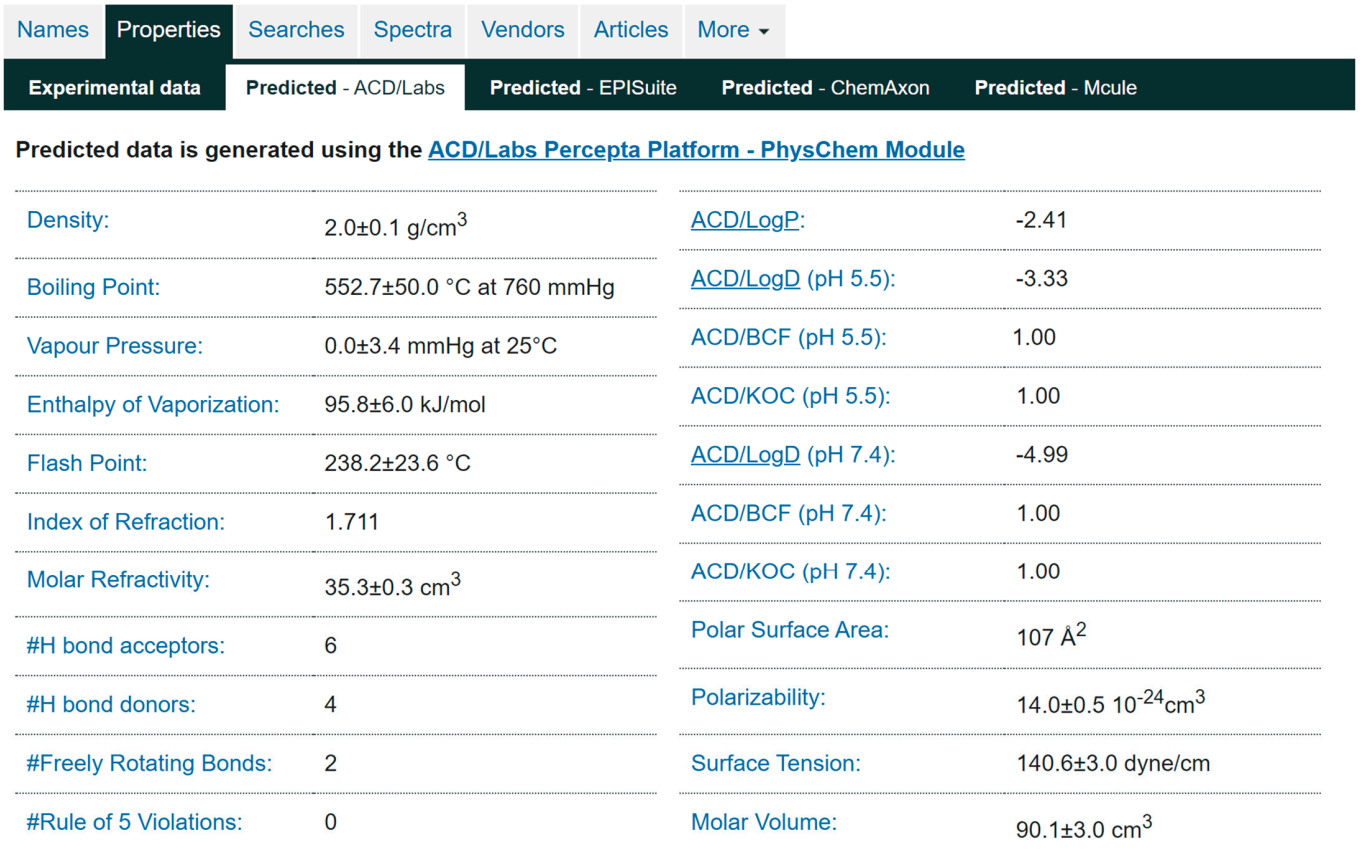


Figure 3. ChemSpider results on ascorbic acid properties. From these values, we selected the LogP (ACD/LogP) and ACD/LogD$_{7.4}$, RB, HBD, HBA, A, PSA included in Tables 1 and 2. These values were found following the instructions mentioned above in ChemSpider's data source (http://www.chemspider.com; Royal Society of Chemistry; accessed on 15 February 2022).

4. Results and Discussion

4.1. Determination of Physicochemical Properties

We calculated and analyzed the physicochemical properties of each phytochemical as an isolated compound. Tables 1–3 summarize the most important parameters that compose the physicochemical properties of these natural compounds. It is important to mention that most of these calculations from the different platforms may have only minor variations (<±0.5) for the same property. For this reason, in Table 1 we grouped the properties with equal or similar values across the platforms. These properties are mostly focused on the structural features (common compound name, phytochemical category, structure, molecular formula, MW, TNA, HBA, HBD, RB, PSA, Cm) for the selected plants' top five bioactive compounds. Of these 15 phytochemicals, 11 are phenolics (biggest phytochemical category), 3 are terpenoids, 1 is an alkaloid, and MW is ~140–790 Da. In contrast, we

found major differences (~±1) in some of the calculated LogP from PubChem, ChemSpider, and SwissADME, primarily due to differences in the algorithms used. Considering these variabilities, Table 2 includes all the lipophilic properties (LogP's and LogD's) from the three platforms and the available experimental LogP for the selected phytochemicals from ginger, tobacco, and echinacea. As we mentioned before, LogD values were added as supplementary information but were not used for further predictions. The experimental LogP was not available for menadione, cembra-2,7,11-triene-4,6-diol, zingiberene, and 6-dehydrogingerdione. From the LogP comparison, Consensus LogP (SwissADME) showed more similarities (~<1) among the phytochemicals with the experimental LogP. Of all the phytochemicals, nicotine showed the most variability in the experimental LogP within the platforms. The most significant limitation for the theoretical calculation of LogP is when the molecule structure has a combined low MW, high polarity, and high acidic properties [25] From these tables, we also want to show that different phytochemicals from the plant and/or from the same phytochemical category can exhibit very diverse physicochemical properties.

In addition to the properties in the classical rules (L-Ro5, GF, VR, MR), in Table 3, we included the theoretical calculations for blood-brain barrier (BBB) permeation, targeting the P-glycoprotein (P-gp), and GI absorption determined by SwissADME. The P-gp transporter is expressed in the intestinal epithelium and cancer cells, decreasing cellular uptake of its substrates [26]. This property is also important to study cancer because P-gp is the key efflux pump of chemotherapeutic drugs and the inductor of chemoresistance [27]. Furthermore, the permeability of a drug to the BBB is significant for laboratories working on brain therapies because brain-targeted drugs must have the capacity to cross this barrier to target neurological disorders of the central nervous system. Interestingly, SwissADME predicts that ascorbic acid is highly absorbable even when it shows violations of the classical rules. This program mainly adjusts the GI absorption using the BOILED-Egg algorithm model [28]. This model makes predictions by constructing two ellipses using the coordinates of only two properties: PSA (0–142.1 $Å^2$) and LogP (−2.3–6.8).

4.2. Prediction of GI Absorption for Phytochemicals Using the Phytochemical Rule (PR)

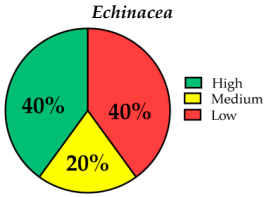
Considering the values of the physicochemical properties determined and summarized in previous tables, we identified in Table 4 any violation of the classical rules, L-Ro5, VR, GF, and MR. Because these rules mainly apply to the study of synthetic compounds, we included in Table 4 the proposed PR that includes the Cm and the extension of the ranges in the classical rules. SwissADME also predicts the probability of being absorbed through the GI, as we showed in Table 3. However, this program was also developed for synthetic compounds. In Table 4, we used the following ranges for these rules:

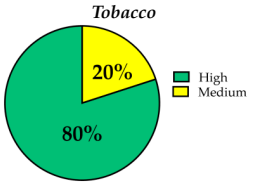
a. L-Ro5: HBD $\leq$ 5, HBA $\leq$ 10, MW $\leq$ 500, logP $\leq$ 5 [11];
b. GF: $-0.4 \leq$ logP $\leq$ 5.6, A (40–130), MW (160–480), TNA (20–70) [13];
c. VR: RB $\leq$ 10, PSA $\leq$ 140 [14];
d. MR: MW (200–600), $-2 \leq$ logP $\leq$ 5, PSA $\leq$ 150, NR $\leq$ 7, NC> 4, NH > 1, RB $\leq$ 15, HBD $\leq$ 5, HBA $\leq$ 10 [15];
e. PR: MW $\leq$ 800 Da, TNA $\leq$ 80, $-2 \leq$ LogP $\leq$ 7, HBD $\leq$ 6, HBA $\leq$ 15, PSA $\leq$ 250 $Å^2$, RB $\leq$ 20 [4], and 100 $\leq$ Cm $\leq$ 900 [3].

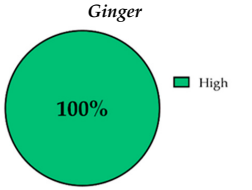
Violations of these rules affect GI absorption. The GI predictions for phytochemicals were manually determined by the combination of all the rules above as follows: High: The compound fully complies with all the rules or has up to 3 violations in the L-Ro5, GF, VR, or MR, covered by the PR.; Medium: The compound fully complies with the PR but has >3 violations to any of the other rules. Low: The compound does not comply with the PR, and, therefore, neither with the other rules.

Table 4. Combination of the classical rules and the PR to predict the drug-likeness and GI absorption of phytochemicals.

Name	L-Ro5	GF	VR	MR	PR	Predicted GI Absorption [#]	
						Phytochemical	Plant
Caffeic acid	✔	✔	✔	1/MW < 200	✔	High	
Caftaric acid	✔	1/LogP < −0.4	1/PSA > 140	1/PSA > 150	✔	High	
Cichoric acid	2/HBA > 10 HBD > 5	✔	2/RB > 10 PSA > 140	3/PSA > 150 HBA > 10 HBD > 5	✔	Medium	
Quercetin-3-O-rutinoside	3/MW > 500 HBA > 10 HBD > 5	4/MW > 480 LogP < −0.4 A > 130 TNA > 70	1/PSA > 140	4/MW > 600 PSA > 150 HBA > 10 HBD > 5	4/PSA > 250 HBA > 10 HBD > 5 Cm > 900	Low	
Echinacoside	3/MW > 500 HBA > 10 HBD > 5	4/MW > 480 LogP < −0.4 A > 130 TNA > 70	2/RB > 10 PSA > 140	5/MW > 600 LogP < −2 PSA > 150 HBA > 10 HBD > 5	6/PSA > 250 LogP < −2 TNA > 80 HBA > 10 HBD > 5 Cm > 900	Low	
Nicotine	✔	✔	✔	1/MW < 200	✔	High	
Menadione	✔	✔	✔	1/MW < 200	✔	High	
Cembra-2,7,11-triene-4,6-diol	✔	✔	✔	✔	✔	High	
Anethole	✔	1/MW < 160	✔	2/MW < 200 NH < 2	✔	High	
Chlorogenic acid	1/HBD > 5	1/LogP < −0.4	1/PSA > 140	2/PSA > 150 HBD > 5	✔	Medium	
6-Gingerol	✔	✔	✔	✔	✔	High	
6-Shogaol	✔	✔	✔	✔	✔	High	
6-Dehydrogingerdione	✔	✔	✔	✔	✔	High	
Zingiberene	✔	✔	✔	1/NH < 2	✔	High	
α-Curcumene	✔	✔	✔	1/NH < 2	✔	High	
Ascorbic acid	✔	2/LogP < −0.4 A < 40	✔	1/MW < 200	✔	High	







L-Ro5: Lipinski's rule of five; GF: Ghose filter; VR: Veber's rule; MR: Muegge's rule; ER: extended rules; Cm: molecular complexity; LogP: Consensus LogP (lipophilicity); A: molar refractivity; HBD: hydrogen bond donors, HBA: hydrogen bond acceptors; RB: rotatable bonds; PSA: polar surface area; TNA: total number of atoms; NR: number of rings; NH: number of heteroatoms; NC: number of carbons; GI: gastrointestinal. ✔: complies with all the rules. L-Ro5: HBD ≤ 5, HBA ≤ 10, MW ≤ 500, logP ≤ 5; GF: logP (−0.4–5.6), A (40–130), MW (160–480), TNA (20–70); VR: RB ≤ 10, PSA ≤ 140; MR: MW (200–600), logP (−2–5), PSA ≤ 150, NR ≤ 7, NC > 4, NH > 1, RB ≤ 15, HBD ≤ 5, HBA ≤ 10; PR: MW ≤ 800 Da, TNA ≤ 80, LogP (−2–7), Cm (100–900), HBD ≤ 6, HBA ≤ 15, PSA ≤ 250 $Å^2$, and RB ≤ 20. [#] The GI predictions of phytochemicals were manually determined as follows: High: ≤3 violations covered by the PR; Medium: >3 violations covered by the PR; Low: any violation to the PR.

We want to clarify that the graphs shown in the right panel of Table 4 summarize the GI results for each plant (in %) based on the PR of the 5 analyzed isolated phytochemicals from the same plant (e.g., if all 5 phytochemicals of the specific plant have high GI, then 100% are "high"). These graphs do not consider the synergism for the GI absorption of the whole phytonutrient (e.g., any whole plant preparation for consumption as a solid or liquid extract). On the other hand, phytochemicals with predicted low absorption could still be interesting to study in phytonutrient extracts where the synergistic effect could increase their GI absorption. In addition, we must always take into account that these rules do not consider the active transport of molecules [21]. We consider these as the limitations of our study.

From echinacea phytochemicals, we predict that only caffeic acid and caftaric acid will have a high GI absorption with 1 and 3 violations, respectively, and comply with the PR. Cichoric acid is predicted to have a medium GI absorption because it shows 7 violations, but it complies with the PR. In contrast, quercetin-3-O-rutinoside and echinacoside are predicted to have a low GI absorption because they show 16 and 20 violations, respectively, and show no compliance with the PR. Based on the five analyzed phytochemicals, echinacea is predicted to show partially (high 40%/medium 40%) GI absorption.

Tobacco has four out of five phytochemicals predicted to have high GI absorption. These are the following: nicotine and menadione, which have 1 violation; anethole which has three violations; and cembra-2,7,11-triene-4,6-diol with no violations. Chlorogenic acid is predicted to have a medium absorption because it shows five violations but still complies with the PR. As a result, according to these five analyzed phytochemicals, tobacco is predicted to show mostly high (80%) GI absorption.

In our prediction of ginger's phytochemicals, all of them (6-gingerol, 6-shogaol, 6-dehydrogingerdione, zingiberene, and α-curcumene) are predicted to have high GI absorption. Only zingiberene, and α-curcumene showed 1 violation. According to these five analyzed phytochemicals, ginger is predicted to show a high GI absorption. Moreover, ascorbic acid, the well-known vitamin C, was analyzed as a theoretical control. Although vitamin C is an established orally absorbed compound [29], it shows three violations of 2 classical rules, two from GF and one from MR, further supporting our analysis that natural compounds may have wider ranges than those proposed in the classical rules. Based on these findings for vitamin C, we expanded the limit to 3 violations to the classical rules for high absorption while complying to the PR.

Comparing the GI predictions in these 16 phytochemicals, caftaric acid, zingiberene and α-curcumene are the compounds that show the greatest differences from our predictions (high) vs. the SwissADME (low). For cichoric acid and chlorogenic acid, we predicted medium absorption while SwisADME, low. In an in vivo study using rats, researchers found that the caftaric acid was rapidly absorbed from the stomach to the plasma, and excreted as fertaric acid by the kidneys [30]. We also found in a study with humans that ~ 33% of orally administered chlorogenic acid was absorbed through the GI and found in the blood circulation [31]. Furthermore, in a study administering ginger oil by oral gavage in rats, zingiberene (the component at the highest concentration) was absorbed and detected in serum [32]. For α-curcumene and cichoric acid, we did not find any recent experimental GI study. On the other hand, it is known that some synthetic drugs for oral administration also fall out of the ranges of these classical rules. For example, Selpercatinib, a recently FDA-approved oral drug for lung cancer, has three violations to the classical rules (1 violation of L-Ro5 and 2 violations of GF) [33]. These results support our methodology for GI predictions by combining the classical rules with our theoretical calculations using the PR to evaluate natural compounds as potential drug candidates.

5. Conclusions

This study proposes a detailed methodology using scientifically validated web-based platforms to determine the physicochemical properties of five phytochemicals found in ginger, echinacea, and tobacco. Furthermore, we developed a filter called the Phytochemical

Rule (PR) based on integrating the classical rules with other extended parameters to obtain a more reliable prediction of the GI absorption of natural compounds. This methodology can help evaluate bioactive phytochemicals as potential drug candidates. For an initial analysis of oral bioavailability and drug-likeness of phytochemicals, the PR proved to be excellent in predicting their drug-relevant properties. Nevertheless, further in vivo and clinical studies should be conducted to confirm the predicted GI absorption.

Author Contributions: Coordination, Y.D.; conceptualization, Y.D.; formal analysis, L.A.V., Y.D. and I.J.S.-A.; theoretical calculations, L.A.V., P.R. and D.P.; writing—original draft preparation, Y.D. and L.A.V.; writing—review and editing, Y.D., L.A.V., Y.F.-A. and I.J.S.-A.; supervision, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Career Development Grant (Y.D., Y.F.A.) from Sloan Scholars Mentoring Network, Biomedical Fellowships from Fundación Intellectus (D.P., Y.D.), and National Institute of General Medical Sciences (NIGMS) National Institutes of Health (NIH) under the Award Number U54GM133807 (Y.D., Y.F.A.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the other funding agencies.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This publication was made possible by the support of the San Juan Bautista School of Medicine (SJBSM) and Universidad Central del Caribe (UCC). The authors thank Estela Estapé (Director of the SJBSM Research Center and senior advisor), for her outstanding dedication and support in the writing process of this article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study's design, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Lin, L.; Wong, H. Predicting Oral Drug Absorption: Mini Review on Physiologically-Based Pharmacokinetic Models. *Pharmaceutics* **2017**, *9*, 41. [CrossRef]
2. Benet, L.Z.; Hosey, C.M.; Ursu, O.; Oprea, T. BDDCS, the Rule of 5 and drugability. *Adv. Drug Deliv. Rev.* **2016**, *101*, 89–98. [CrossRef] [PubMed]
3. Böttcher, T. An Additive Definition of Molecular Complexity. *J. Chem. Inf. Model.* **2016**, *56*, 462–470. [CrossRef]
4. Croy, B.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates. *Chem. Biol.* **2014**, *21*, 1115–1142. [CrossRef]
5. Yoo, S.; Kim, K.; Nam, H.; Lee, D. Discovering Health Benefits of Phytochemicals with Integrated Analysis of the Molecular Network, Chemical Properties and Ethnopharmacological Evidence. *Nutrients* **2018**, *10*, 1042. [CrossRef] [PubMed]
6. Delgado, Y.; Cassé, C.; Ferrer-Acosta, Y.; Suárez-Arroyo, I.J.; Rodríguez-Zayas, J.; Torres, A.; Torres-Martínez, Z.; Pérez, D.; González, M.J.; Velázquez-Aponte, R.A.; et al. Biomedical Effects of the Phytonutrients Turmeric, Garlic, Cinnamon, Graviola, and Oregano: A Comprehensive Review. *Appl. Sci.* **2021**, *11*, 8477. [CrossRef]
7. Aucoin, M.; Cooley, K.; Saunders, P.R.; Carè, J.; Anheyer, D.; Medina, D.N.; Cardozo, V.; Remy, D.; Hannan, N.; Garber, A. The effect of Echinacea spp. on the prevention or treatment of COVID-19 and other respiratory tract infections in humans: A rapid review. *Adv. Integr. Med.* **2020**, *7*, 203–217. [CrossRef]
8. Jafarzadeh, A.; Jafarzadeh, S.; Nemati, M. Therapeutic potential of ginger against COVID-19: Is there enough evidence? *J. Tradit. Chin. Med Sci.* **2021**, *8*, 267–279. [CrossRef]
9. Mishra, A.; Chaturvedi, P.; Datta, S.; Sinukumar, S.; Joshi, P.; Garg, A. Harmful effects of nicotine. *Indian J. Med Paediatr. Oncol.* **2015**, *36*, 24–31. [CrossRef]
10. Rojas-Colón, L.A.; Dash, P.K.; Morales-Vías, F.A.; Lebrón-Dávila, M.; Ferchmin, P.A.; Redell, J.B.; Maldonado-Martínez, G.; Vélez-Torres, W.I. 4R-cembranoid confers neuroprotection against LPS-induced hippocampal inflammation in mice. *J. Neuroinflammation* **2021**, *18*, 95. [CrossRef]
11. Pollastri, M.P. Overview on the Rule of Five. *Curr. Protoc. Pharmacol.* **2010**, *49*, 9.12.1–9.12.8. [CrossRef] [PubMed]
12. Lipinski, C. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. *Adv. Drug Deliv. Rev.* **2016**, *101*, 34–41. [CrossRef] [PubMed]

13. Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1*, 55–68. [CrossRef] [PubMed]
14. Veber, D.F.; Johnson, S.R.; Cheng, H.-Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623. [CrossRef]
15. Muegge, I.; Heald, S.L.; Brittelli, D. Simple Selection Criteria for Drug-like Chemical Matter. *J. Med. Chem.* **2001**, *44*, 1841–1846. [CrossRef]
16. Rutkowska, E.; Pajak, K.; Jóźwiak, K. Lipophilicity—Methods of determination and its role in medicinal chemistry. *Acta Pol. Pharm.* **2013**, *70*, 3–18.
17. Tinworth, C.P.; Young, R.J. Facts, Patterns, and Principles in Drug Discovery: Appraising the Rule of 5 with Measured Physicochemical Data. *J. Med. Chem.* **2020**, *63*, 10091–10108. [CrossRef]
18. Winiwarter, S.; Ridderström, M.; Ungell, A.; Andersson, T.; Zamora, I. 5.22—Use of molecular descriptors for absorp-tion, distribution, metabolism, and excretion predictions. In *Comprehensive Medicinal Chemistry II*; Taylor, J., Triggle, D., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; Volume 5, pp. 531–554.
19. Kermen, F.; Chakirian, A.; Sezille, C.; Joussain, P.; Le Goff, G.; Ziessel, A.; Chastrette, M.; Mandairon, N.; Didier, A.; Rouby, C.; et al. Molecular complexity determines the number of olfactory notes and the pleasantness of smells. *Sci. Rep.* **2011**, *1*, 206. [CrossRef]
20. Stratton, C.F.; Newman, D.J.; Tan, D.S. Cheminformatic comparison of approved drugs from natural product versus synthetic origins. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4802–4807. [CrossRef]
21. Bickerton, G.R.; Paolini, G.V.; Besnard, J.; Muresan, S.; Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4*, 90–98. [CrossRef]
22. Silva, D.M.; Da Costa, P.A.; Ribon, A.O.; Purgato, G.A.; Gaspar, D.-M.; Diaz, G. Plant Extracts Display Synergism with Different Classes of Antibiotics. *Acad Bras. Cienc.* **2019**, *91*, e20180117. [CrossRef] [PubMed]
23. Date, A.A.; Desai, N.; Dixit, R.; Nagarsenker, M. Self-nanoemulsifying drug delivery systems: Formulation insights, applications and advances. *Nanomedicine* **2010**, *5*, 1595–1616. [CrossRef] [PubMed]
24. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef]
25. Sobańska, A.W.; Robertson, J.; Brzezińska, E. Application of RP-18 TLC Retention Data to the Prediction of the Transdermal Absorption of Drugs. *Pharmaceuticals* **2021**, *14*, 147. [CrossRef] [PubMed]
26. Alqahtani, M.S.; Kazi, M.; Alsenaidy, M.A.; Ahmad, M.Z. Advances in Oral Drug Delivery. *Front. Pharmacol.* **2021**, *12*, 618411. [CrossRef]
27. Ughachukwu, P.; Unekwe, P. Efflux pump-mediated resistance in chemotherapy. *Ann. Med. Health Sci. Res.* **2012**, *2*, 191–198. [CrossRef]
28. Daina, A.; Zoete, V. A BOILED-Egg to Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **2016**, *11*, 1117–1121. [CrossRef]
29. Lykkesfeldt, J.; Tveden-Nyborg, P. The Pharmacokinetics of Vitamin C. *Nutrients* **2019**, *11*, 2412. [CrossRef]
30. Vanzo, A.; Cecotti, R.; Vrhovsek, U.; Torres, A.M.; Mattivi, F.; Passamonti, S. The Fate of *trans*-Caftaric Acid Administered into the Rat Stomach. *J. Agric. Food Chem.* **2007**, *55*, 1604–1611. [CrossRef]
31. Olthof, M.R.; Hollman, P.C.H.; Katan, M.B. Chlorogenic Acid and Caffeic Acid Are Absorbed in Humans 1. *J. Nutr.* **2001**, *131*, 66–71. [CrossRef]
32. Jeena, K.; Liju, V.B.; Kuttan, R. A Preliminary 13-Week Oral Toxicity Study of Ginger Oil in Male and Female Wistar Rats. *Int. J. Toxicol.* **2011**, *30*, 662–670. [CrossRef] [PubMed]
33. Pathania, S.; Singh, P.K. Analyzing FDA-approved drugs for compliance of pharmacokinetic principles: Should there be a critical screening parameter in drug designing protocols? *Expert Opin. Drug Metab. Toxicol.* **2020**, *17*, 351–354. [CrossRef] [PubMed]