



## Original article

# *Theobroma cacao* L. compounds: Theoretical study and molecular modeling as inhibitors of main SARS-CoV-2 protease

Oswaldo Yañez<sup>a,b,c</sup>, Manuel Isaías Osorio<sup>d,e</sup>, Carlos Areche<sup>f</sup>, Alejandro Vasquez-Espinal<sup>a</sup>, Jessica Bravo<sup>d</sup>, Angélica Sandoval-Aldana<sup>g</sup>, José M. Pérez-Donoso<sup>e</sup>, Fernando González-Nilo<sup>e</sup>, Maria João Matos<sup>h</sup>, Edison Osorio<sup>i</sup>, Olimpo García-Beltrán<sup>i,j,\*</sup>, William Tiznado<sup>a,\*\*</sup>

<sup>a</sup> Computational and Theoretical Chemistry Group, Departamento de Ciencias Químicas, Facultad de Ciencias Exactas, Universidad Andres Bello, República 498, Santiago, Chile

<sup>b</sup> Center of New Drugs for Hypertension (CENDHY), Santiago, Chile

<sup>c</sup> Department of Pharmaceutical Science and Technology, School of Chemical and Pharmaceutical Sciences, Universidad de Chile, Santiago, Chile

<sup>d</sup> Facultad de Medicina, Centro de Investigación Biomédica, Universidad Diego Portales, Ejército 141, Santiago 837007, Chile

<sup>e</sup> Center for Bioinformatics and Integrative Biology (CBIB), Facultad de Ciencias de la Vida, Universidad Andres Bello, Av. República 330, Santiago 8370146, Chile

<sup>f</sup> Departamento de Química, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Nuñoa, Santiago 7800024, Chile

<sup>g</sup> Grupo Interdisciplinario de Investigación en Fruticultura Tropical, Facultad de Ingeniería Agronómica, Universidad del Tolima, Ibagué 730006, Colombia

<sup>h</sup> CIQUP/Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal

<sup>i</sup> Facultad de Ciencias Naturales y Matemáticas, Universidad de Ibagué, Carrera 22 calle 67, Ibagué 730002, Colombia

<sup>j</sup> Universidad Bernardo O'Higgins, Centro Integrativo de Biología y Química Aplicada (CIBQA), General Gana 1702, Santiago, Chile, 8370854



## ARTICLE INFO

## Keywords:

*Theobroma cacao*

Antioxidant

Bioflavonoids

SARS-CoV-2

Molecular dynamics

DFT

## ABSTRACT

Cocoa beans contain antioxidant molecules with the potential to inhibit type 2 coronavirus (SARS-CoV-2), which causes a severe acute respiratory syndrome (COVID-19). In particular, protease. Therefore, using *in silico* tests, 30 molecules obtained from cocoa were evaluated. Using molecular docking and quantum mechanics calculations, the chemical properties and binding efficiency of each ligand was evaluated, which allowed the selection of 5 compounds of this series. The ability of amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin to bind to the main viral protease was studied by means of free energy calculations and structural analysis performed from molecular dynamics simulations of the enzyme/inhibitor complex. Isorhoifolin and rutin stand out, presenting a more negative binding  $\Delta G$  than the reference inhibitor N-[(5-methylisoxazol-3-yl)carbonyl]alanyl-L-valyl-N~1~-((1R,2Z)-4-(benzyloxy)-4-oxo-1-[[3R]-2-oxopyrrolidin-3-yl)methyl]but-2-enyl)-L-leucinamide (N3). These results are consistent with high affinities of these molecules for the major SARS-CoV-2. The results presented in this paper are a solid starting point for future *in vitro* and *in vivo* experiments aiming to validate these molecules and /or test similar substances as inhibitors of SARS-CoV-2 protease.

## 1. Introduction

The cocoa tree is known since ancient times in Mesoamerica, as drinks obtained from its seeds can be traced through archeological and chemical studies that show the presence, in pottery containers dating from approximately 1000 B.C., of compounds originating from the secondary metabolism of this species [1,2]. Shortly after the arrival of the Spanish in the new world, this product became known in Europe, and in 1737 Carl Nilsson Linnaeus designated it taxonomically as *Theobroma cacao* L. [3]. Currently the genus *Theobroma* comprises 22

species and is classified within the family Malvaceae, and 9 of them are native to the Amazon forest [4–6]. Cocoa beans are a rich source of phenolic compounds (10–12%, dry weight), specifically catechins (flavan-3-ols) and procyanidins, which are a large and diverse group of metabolites acting as antioxidants [7,8]. According to the literature reviewed, 2 main genetic groups are considered, known as *Criollo* and *Forastero*, based on their geographic origins and morphological characteristics. In addition, there is a third group known as *Trinitario*, which is considered a hybrid product of the previous varieties [9]. The phenol content depends mainly on the genetics of the trees and can also vary

\* Corresponding author at: Facultad de Ciencias Naturales y Matemáticas, Universidad de Ibagué, Carrera 22 calle 67, Ibagué 730002, Colombia.

\*\* Corresponding author.

E-mail addresses: [jose.garcia@unibague.edu.co](mailto:jose.garcia@unibague.edu.co) (O. García-Beltrán), [wtiznado@unab.cl](mailto:wtiznado@unab.cl) (W. Tiznado).

<https://doi.org/10.1016/j.bioph.2021.111764>

Received 29 March 2021; Received in revised form 7 May 2021; Accepted 20 May 2021

Available online 24 May 2021

0753-3322/© 2021 The Author(s).

Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

according to the area of cultivation, fruit maturity, climatic conditions, harvest time and storage time of the cobs until they are opened [10–12]. The 3 main groups of polyphenols are present in cocoa beans – catechins, anthocyanins and proanthocyanidins – representing 35%, 4% and 58%, respectively [13]. It has been shown that during the different stages of processing and manufacturing (fermentation, drying, roasting and refining), the chemical composition of cocoa beans changes drastically, affecting the content and structure of polyphenols, and resulting in a reduction of their antioxidant activity [2,14,15]. The high content of antioxidant compounds makes cocoa a very interesting product because it can protect against various diseases generated by reactive oxygen species (ROS), such as neurodegenerative diseases [16,17], cardiometabolic diseases and cancers [18–20].

Quite recently, compounds have been postulated for the inhibition of the main protease of the virus with a specific interaction in the potential treatment of infected patients [21,22]. However, secondary metabolites are an unexplored source of chemical resources to control SARS-CoV-2 proliferation. Among natural products there are substances such as nitrogenous, terpenic and phenolic compounds, and some of them exhibit antiviral properties [23,24]. In the polyphenol family we find flavonoids, and literature reports show the ability of some of them, such as formononetin and penduletin, to inhibit virus replication. Both are reported to be effective *in vivo* models against enterovirus infections that cause hand, foot and mouth disease [25]. In addition, baicalin inhibits replication of the Dengue virus in an *in vitro* model [26]. The increased effect of the biflavone amentoflavone compared to flavones and biflavonoid derivatives with methoxy groups [27], and the enzyme inhibiting activity of pectolarin, herbacetin and rhoifolin [28], have been reported to inhibit SARS-CoV 3CLpro. The flavonoids hesperidin and naringin, which are abundant in citrus fruits, exhibit potential in the inhibition of the proteins responsible for the replication and propagation of SARS-CoV-2 [29]. *In silico* studies show that caflanone, hesperetin, myricetin and 5'-chloroquercetin can bind with high affinity to the spike protein, helicase, and protease sites on the ACE2 receptor [30], and penimethavone, a molecule mined from microorganisms, presents a high potential to modulate/inhibit the SARS-CoV-2 M<sup>Pro</sup> active site [31].

In this work, we present an *in silico* investigation of the antioxidant capacity and of the SARS-CoV-2 M<sup>Pro</sup> inhibitory activity of 30 polyphenolic molecules derived from *Theobroma cacao*: flavonoids, hydroxybenzoic acids, hydroxycinnamic acids and N-phenylpropanoyl-L-amido acids. The computational study started by evaluating some reactivity descriptors, defined in the framework of density functional theory (DFT), in order to inspect the reactive nature of these molecules. This methodology has been tested and applied in the literature by several research groups, proving to be very useful to rationalize the reactivity patterns of molecular systems.

Subsequently, the evaluation of 30 flavonoids by molecular docking showed promising results on the anti-SARS-CoV-2 activity of 5 compounds: Amentoflavone, isorhoifolin, nicotiflorin, naringin, and rutin. These compounds were subjected to more specific computational analysis to establish their potential inhibitory activity. A set of molecular dynamics simulations and free energy calculations, using MMGBSA and quantum mechanical calculations, established that all the selected compounds have binding energies very close to that of N3, a reference molecule with a high inhibitory activity against SARS-CoV-2.

We emphasize that isorhoifolin and rutin possess more negative binding energies than N3, approximately by 5 kcal mol<sup>-1</sup>. These results can be associated with the high proximity and formation of hydrogen bonds with Glu166, Cys145 and His41 residues, fundamental amino acids for the activity of M<sup>Pro</sup>. These promising results make it possible to propose these selected compounds for future experimental tests.

## 2. Computational methods

### 2.1. Density functional theory (DFT) calculations

Thirty polyphenolic compounds selected for their possible capability to inhibit SARS-CoV-2 protease M<sup>Pro</sup>, and N3 (reference compound) obtained from the Protein Data Bank (PDB) [32] (PDB id: 6LU7), were drawn using Discovery Studio [33] 3.1 (Accelrys, CA) and geometries were optimized using the M06–2X-D3 method [34,35] in conjunction with the 6–31G(d,p) basis set. M06–2X-D3 is the best dispersion-corrected meta-GGA hybrid functional on the GMTKN30 database [36], and it is implemented in the Gaussian16 software program suite [37]. The water was simulated as a solvent using the SMD parametrization of the IEF-PCM.

Some DFT-based global reactivity descriptors (Table 1), such as electronegativity ( $\chi$ ), global hardness ( $\eta$ ), electrophilicity ( $\omega$ ), electrodonating ( $\omega^-$ ), electroaccepting ( $\omega^+$ ) and net electrophilicity ( $\Delta\omega^\pm$ ), were calculated to better understand the molecules reactivity.

### 2.2. Molecular docking

The 30 polyphenolic compounds and N3 were docked in the binding cavity of the main protease M<sup>Pro</sup> of SARS-CoV-2 using the AutoDock 4.0 [49] suite. The crystal structure of SARS-CoV-2 M<sup>Pro</sup> protein (2.16 Å resolution) [32] (PDB Code: 6LU7, <https://www.rcsb.org/structure/6LU7>) was downloaded from the PDB [50]. The SARS-CoV-2 M<sup>Pro</sup> model was modified with the Schrödinger [51] Protein Preparation Wizard [52]. Polar hydrogen atoms were added, non-polar hydrogen atoms were merged, and charges were assigned, in addition to delimiting the binding region of possible inhibitors of SARS-CoV-2 M<sup>Pro</sup> [21, 53–58]. The binding pocket of SARS-CoV-2 M<sup>Pro</sup> was established as the center of mass between the Cys145 and His41 residues of the catalytic site. In general, the grid maps were calculated using the AutoGrid 4.0 option and the volume chosen for the grid maps was made up of 60×60×60 points, with a grid-point spacing of 0.375 Å. The author's option was used to define the rotating bond in the ligand. In the Lamarckian genetic algorithm (LGA) dockings, an initial population of random individuals with a population size of 150, a maximum number of 2.5 × 10<sup>7</sup> energy evaluations, a maximum number of generations of 27,000, a mutation rate of 0.02 and crossover rate of 0.80 were employed. Each complex was built using the lowest docked-energy binding positions. The van der Waals interactions were computed by means of a smoothed 12–6 Lennard Jones potential, while the hydrogen bonding interactions were evaluated by a 12–10 function which incorporated a directionality term. That is, interactions which deviate from ideal hydrogen bonding geometries were progressively weighted.

The partial charges of each ligand were determined with the PM6-D3H4 semi-empirical method [59,60] implemented in the MOPAC2016 software [61]. PM6-D3H4 introduces dispersion and hydrogen-bonded corrections to the PM6 method. 3D representations of the docking results were analyzed using the VMD molecular graphics system [62].

**Table 1**

Equations for global reactivity indexes calculated in TAFF [38] pipeline.

	Koopmans' theorem	Reference
Global hardness ( $\eta$ )	$\eta = \frac{1}{2}(\epsilon_L - \epsilon_H)$	[39–44]
Electronegativity ( $\chi$ )	$\chi = -\frac{1}{2}(\epsilon_L + \epsilon_H)$	[40,45,46]
Electrophilicity ( $\omega$ )	$\omega = \frac{\mu^2}{2\eta} = \frac{(\epsilon_L + \epsilon_H)^2}{2(\epsilon_L - \epsilon_H)}$	[47]
Electron acceptor ( $\omega^+$ )	$\omega^+ = \frac{(\epsilon_L + 3\epsilon_H)^2}{16(\epsilon_L - \epsilon_H)}$	[47]
Electron donator ( $\omega^-$ )	$\omega^- = \frac{(3\epsilon_L + \epsilon_H)^2}{16(\epsilon_L - \epsilon_H)}$	[47]
Net electrophilicity ( $\Delta\omega^\pm$ )	$\Delta\omega^\pm = \omega^+ + \omega^-$	[48]

### 2.3. Ligand efficiency approach

Ligand efficiency (*LE*) calculations were performed using one parameter:  $K_d$ . The  $K_d$  parameter corresponds to the dissociation constant of a ligand/protein complex, and its value indicates the bond strength between the ligand and protein [63–65]. Low values indicate strong binding of the molecule to the protein.  $K_d$  calculations were done using the following equations:

$$\Delta G^0 = -2.303RT \log(K_d) \quad (1)$$

$$K_d = 10^{\frac{\Delta G^0}{2.303RT}} \quad (2)$$

where  $\Delta G^0$  corresponds to binding energy ( $\text{kcal mol}^{-1}$ ) obtained from docking computations,  $R$  is the gas constant whose value is  $1.987207 \text{ cal mol}^{-1} \text{ K}^{-1}$ , and  $T$  is the temperature in Kelvin at standard conditions of aqueous solution at 298.15 K, neutral pH and remaining concentrations of 1 M. The ligand efficiency (*LE*) allows us to compare molecules according to their average binding energy [65,66]. It is determined as the binding energy per non-hydrogen atom, as follows [63–65,67]:

$$LE = -\frac{2.303RT}{HAC} \log(K_d) \quad (3)$$

where  $K_d$  is obtained from Eq. 2 and *HAC* denotes the heavy atom count (i.e., number of non-hydrogen atoms) in a ligand.

### 2.4. Molecular dynamics simulations

MD calculations were performed for the five systems with the lowest binding energy according to the docking calculations and also for compound N3, which is our reference ligand. Each model was confined inside a periodic simulation box. The compounds were bound to SARS-CoV-2 M<sup>pro</sup> protein [32] (PDB ID: 6LU7) in aqueous solution with an explicit solvent TIP3P water model [68] ( $\approx 20,000$  water molecules). Protonation states of ionizable residues corresponding to pH 7.0 were determined by the H++ web interface that computes pK values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment [69]. The N3, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin molecules were parameterized using the GAFF Force Field for organic molecules [70, 71], using the Antechamber module in AmberTools18.

The partial charges of each compound were determined by the restrained electrostatic potential (RESP) model [72] M06–2X-D3/6–31G (d,p) level. MD simulations were carried out using the modeled ff14SB [73] force field [74,75] within the AMBER-GPU Implementations18 [76]. The simulations were carried out using a standard MD protocol: i) Minimization and structural relaxation of water molecules with 2000 steps of minimization and MD simulation with an NPT (300 K) assembly for 1000 ps using harmonic restrictions of  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  for protein and ligand; ii) minimization of the complete structure considering 6500 steps of conjugate gradient minimization; iii) the minimized systems were progressively heated to 300 K, with harmonic restrictions of  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  in the backbone protein and ligand during 0.5 ns; iv) the system was then equilibrated for 0.5 ns maintaining the restrictions and then for 5 ns without restrictions at 300 K in a canonical assembly (NVT); v) finally, the total duration of simulation was approximately 75 ns for each system. During the MD simulations, motion equations were integrated with a 2 fs time step in the NPT ensemble at a pressure of 1 atm. The SHAKE algorithm was applied to all hydrogen atoms, and the van der Waals cutoff was set to 12 Å. The temperature was maintained at 310 K, employing the Langevin thermostat method with a relaxation time of 1 ps. The Berendsen barostat was used to control the pressure at 1 atm. Long-range electrostatic forces were taken into account by means of the particle-mesh Ewald (PME) approach. Data were collected every 1 ps during the MD runs. Molecular visualization of the systems and MD

trajectory analysis were carried out with the VMD software package [62].

### 2.5. Free energy calculation

The molecular MM-GBSA method was employed to estimate the binding free energy of the SARS-CoV-2 M<sup>pro</sup>/ligand complexes. From calculations for a total of 75 ns of MD, the last 70 ns were extracted for analysis, and the explicit water molecules and ions were removed. The MM-GBSA analysis was performed on three subsets of each system: the protein alone, the ligand alone, and the complex (protein-ligand). For each of these subsets, the total free energy ( $\Delta G_{tot}$ ) was calculated as follows:

$$\Delta G_{tot} = E_{MM} + G_{solv} - T\Delta S_{conf} \quad (4)$$

where  $E_{MM}$  is the bonded (bond, angle and dihedral) and non-bonded (electrostatics and Lennard-Jones) terms;  $G_{solv}$  is the polar contribution of solvation energy and non-polar contribution to the solvation energy;  $T$  is the temperature; and  $\Delta S_{conf}$  is the conformational entropy [77]. Both  $E_{MM}$  and  $G_{solv}$  were calculated using AMBER-GPU Implementations18 [76] software with the generalized Born implicit solvent model [78,79].  $\Delta G_{tot}$  was calculated as a linear function of the solvent-accessible surface area, which was calculated with a probe radius of 1.4 Å [80]. The binding free energies of SARS-CoV-2 M<sup>pro</sup> and ligand complexes ( $\Delta G_{bind}$ ) were calculated as differences where  $G_{tot}$  values are the average values over the simulation.

$$\Delta G_{bind} = G_{tot}(complex) - G_{tot}(protein) - G_{tot}(ligand) \quad (5)$$

### 2.6. Non-covalent interactions

The principal clusters of the main component analyses of trajectory were analyzed with the non-covalent interaction index (NCI) [81,82] using the NCIPLOT program to identify and map non-covalent interactions, such as hydrogen bonds, steric repulsion, and van der Waals interactions, a M06–2X-D3/6–31G(d,p). For the N3 and the antioxidants and for the protein-ligand complexes we used promolecular densities ( $\rho^{pro}$ ), computed as the sum of all atomic contributions. The NCI is based on the electron density ( $\rho$ ), its derivatives, and the reduced density gradient ( $s$ ). The reduced density gradient is given by:

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{\nabla\rho}{\rho^{4/3}} \quad (6)$$

These interactions are local and manifest in real space as low-gradient isosurfaces with low densities which are interpreted and colored according to the corresponding values of  $\text{sign}(\lambda_2)\rho$ , where  $\lambda_2$  represents the second eigenvalue of the Hessian matrix. The surfaces are colored on a blue-green-red scale according to the strength and type of interaction. Blue indicates strong attractive interactions, green indicates weak van der Waals interactions, and red indicates a strong non-bonded overlap.

## 3. Results and discussion

### 3.1. Global reactivity molecular descriptors

The global reactivity indices, approximated using the Koopmans' theorem, are evaluated to predict the ligand molecules reactivity against and biological receptors. The electronegativity ( $\chi$ ), global hardness ( $\eta$ ) electrophilicity ( $\omega$ ), electrodonating ( $\omega^-$ ), electroaccepting ( $\omega^+$ ) and net electrophilicity ( $\Delta\omega^\pm$ ) of antioxidants from *Theobroma cacao* L. and the reference molecule N3 are presented in Table 2. In general terms we can see that amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin compounds have electronegativity values very close to N3's. Now, systems with high electronegativity values tend to accept H-bonds, and the

**Table 2**

Global reactivity descriptors for the ligands in SARS-CoV-2 M<sup>PRO</sup> complexes, calculated with the M06–2X-D3 density functional. All units measured in eV.

Compound	$\chi$	$\eta$	$\omega$	$\omega^-$	$\omega^+$	$\Delta\omega^\pm$
N3	4.45	3.90	2.54	5.26	0.80	6.06
Amentoflavone	4.41	3.10	3.13	5.72	1.31	7.03
Naringin	4.45	3.38	2.93	5.58	1.12	6.71
Isorhoifolin	4.48	3.15	3.19	5.82	1.34	7.17
Rutin	4.26	3.13	2.89	5.42	1.15	6.57
Nicotiflorin	4.40	3.07	3.15	5.74	1.33	7.07
Kaempferol-7-O-neohesperidoside	4.20	2.98	2.96	5.44	1.23	6.68
Prunin	4.31	3.46	2.68	5.27	0.95	6.23
Kaempferol 3-O- $\beta$ -D-glucoside	4.25	3.20	2.82	5.35	1.09	6.44
Isoquercetin	4.23	3.12	2.86	5.37	1.13	6.51
Luteolin 7-O- $\beta$ -D-glucoside	4.34	3.18	2.96	5.53	1.19	6.73
Apigenin 7-O-glucoside	4.34	3.25	2.90	5.48	1.14	6.62
Hyperoside	4.36	3.07	3.09	5.66	1.30	6.96
Quercitrin	4.34	3.04	3.10	5.65	1.31	6.96
Isoorientin	4.36	3.12	3.04	5.61	1.25	6.87
Chlorogenic acid	4.42	3.06	3.19	5.79	1.36	7.16
Luteolin	4.28	3.18	2.88	5.41	1.13	6.55
Quercetin	4.28	2.87	3.19	5.69	1.41	7.11
Orientin	4.41	3.09	3.15	5.74	1.33	7.08
Epicatechin	3.70	3.80	1.80	4.13	0.42	4.56
(+)-Catechin	3.63	3.83	1.72	4.02	0.38	4.40
(-)-Catechin	3.68	3.75	1.81	4.12	0.43	4.56
Apigenin	4.44	3.16	3.12	5.74	1.29	7.04
Naringenin	4.24	3.54	2.54	5.10	0.86	5.97
Isorhamnetin	4.13	2.91	2.92	5.35	1.22	6.57
Isovitexin	4.38	3.19	3.00	5.60	1.21	6.81
Caffeic acid	4.35	3.07	3.08	5.64	1.28	6.93
Ferulic acid	4.33	3.06	3.05	5.60	1.27	6.88
Gallic acid	4.21	3.59	2.47	5.03	0.81	5.84
Protocatechuic acid	4.25	3.63	2.49	5.08	0.82	5.90
Coumaric acid	4.62	3.20	3.33	6.05	1.42	7.47

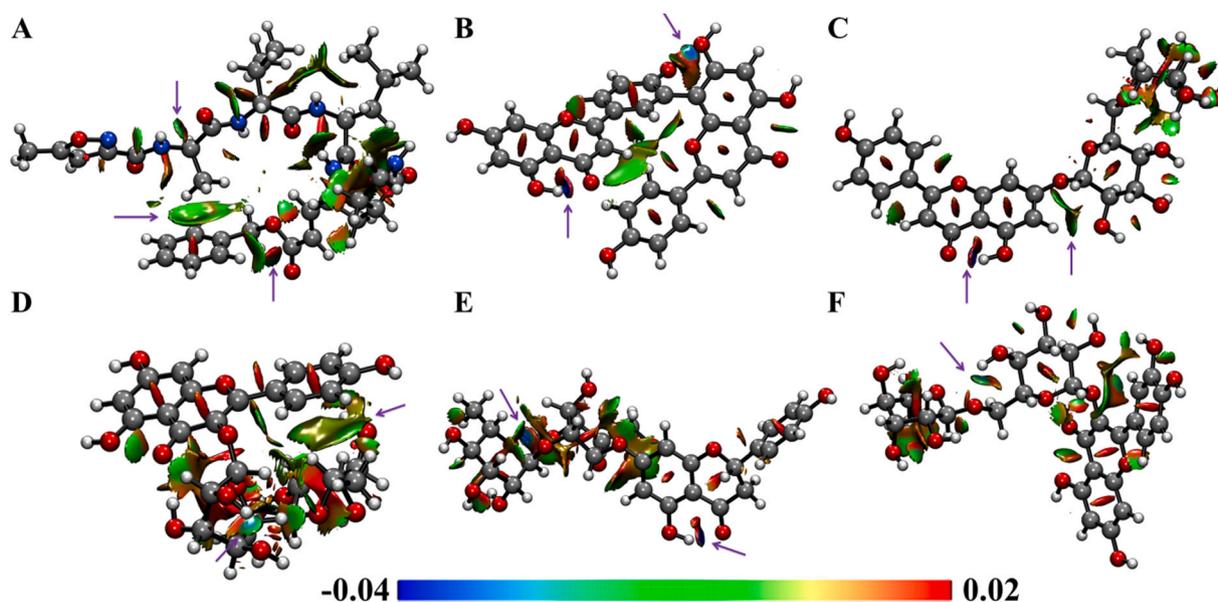
bond strength should increase with the electronegativity of the two bonded atoms, i.e., hydrogen bonding interactions could form in the active site. On the other hand, these molecules have lower hardness values than the reference molecule N3, therefore they are less averse to the arrival of electrons, and thus they are more reactive systems. These results contribute to proposals in the literature that compounds with phenol groups are more susceptible to electron donation [83]. In the

case of electrophilicity values, there is an increase relative to the value of N3, indicating that these molecules are more electrophilic.

NCI plots at DFT geometric optimization of N3, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin at the M06–2X-D3/6–31G (d,p) level reveal different intramolecular interactions (Fig. 1) in terms of polar groups, which can contribute to the stability in the active site of a protein through the formation of non-covalent interactions such as hydrogen bonds and ionic interactions. Note that the regions corresponding to the hydrogen bridges (blue) are tiny and surrounded by repulsive interactions (red).

### 3.2. Molecular docking analysis

M<sup>PRO</sup> is a homodimeric cysteine protease and plays an important role in the SARS virus replication and transcription. When the mRNA of the virus is translated into polyproteins, M<sup>PRO</sup> is first self-cleaved to become a mature enzyme, which in turn cleaves all of the 11 remaining downstream nonstructural proteins of the polyproteins into polypeptides which are required for the replication process of the virus [84]. Consequently, M<sup>PRO</sup> is a key drug target for the inhibition of SARS-CoV-2. The binding site of M<sup>PRO</sup> is made up of subsites S1, S2, S3, S4, and S1', which are represented based on the binding position of the substrate polyprotein [57]. The binding site is located in the groove between a His41 and Cys145 catalytic dyad. Cys145 is involved in the covalent binding of some M<sup>PRO</sup> inhibitors, specifically N3. Tung Ngo et al. reported in recent studies that Glu166 also has a prominent and important role in binding ligands to M<sup>PRO</sup> [85]. In order to investigate the possible mechanism by which selected flavonoids act, molecular docking analysis of all the flavonoids was done in the active site of M<sup>PRO</sup>. The flavonoid names and docking scores are displayed in Table 3. Five prominent ligands with the highest affinities for the active site are amentoflavone, naringin, isorhoifolin, rutin and nicotiflorin, with binding energies of  $-10.0$ ,  $-9.0$ ,  $-8.8$ ,  $-8.7$  and  $-8.5$  kcal mol<sup>-1</sup>, respectively. These values are more negative than those obtained for N3 (7.5 kcal mol<sup>-1</sup>). Furthermore, it is clearly shown that glycosylated flavonoids show the highest scores, with a maximum value of 10.0 kcal mol<sup>-1</sup>, which is complemented by the study carried out by Cherrak et al. [86]. Amentoflavone showed hydrogen bond interactions with Asn142, Glu166 and Thr190. Naringin showed hydrogen bonds with His163, His164 and Phe140. Isorhoifolin



**Fig. 1.** NCIplot of the non-covalent interaction regions with isosurface gradient (0.7 au) for A) N3, B) Amentoflavone, C) Isorhoifolin, D) Nicotiflorin, E) Naringin and F) Rutin molecules. The arrows indicate potential intramolecular interactions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

Molecular docking study between selected ligands and SARS-CoV-2M<sup>PRO</sup>. Intermolecular docking values, presented with their interaction energy ( $\Delta E_{\text{binding}}$ ), H-bond residues, interacting residues, and Ligand efficiency calculation for SARS-CoV-2M<sup>PRO</sup> complexes are shown.

Compound	Docking results		Ligand efficiency	
	$\Delta E_{\text{binding}}$ (kcal mol <sup>-1</sup> ) <sup>a</sup>	Residue interactions <sup>b</sup>	$K_d$	LE kcal mol <sup>-1</sup>
N3 <sup>c</sup>	-7.5	ARG188, ASN142, ASP187, <b>CYS145</b> , GLN189, GLU166, GLY143, HIS164, HIS163, <b>HIS41</b> , LEU141, LEU167, MET165, MET49, PHE140, PRO168, SER144, SER305, SER46, THR25	$3.18 \times 10^{-6}$	0.15
Amentoflavone	-10.0	ASN142, CYS44, GLN189, GLU166, <b>HIS41</b> , MET165, MET49, PRO168, SER46, THR190, THR25	$4.69 \times 10^{-8}$	0.25
Naringin	-9.0	ASN142, <b>CYS145</b> , GLU166, GLY143, HIS164, HIS163, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144, SER305, SER46, THR24, THR25, THR26, THR45	$2.53 \times 10^{-7}$	0.21
Isorhoifolin	-8.8	ASN142, <b>CYS145</b> , GLN189, GLU166, GLY143, HIS163, HIS172, MET165, MET49, PHE140, SER144, SER305, SER46, THR24, THR25, THR45	$3.55 \times 10^{-7}$	0.21
Rutin	-8.7	<b>CYS145</b> , GLN189, GLU166, GLY143, HIS163, HIS172, LEU167, MET165, MET49, PHE140, PRO168, SER144, SER305, THR25	$4.20 \times 10^{-7}$	0.20
Nicotiflorin	-8.5	ARG188, ASN142, <b>CYS145</b> , GLN189, GLU166, GLY143, HIS164, HIS163, <b>HIS41</b> , LEU141, LEU27, MET165, MET49, PHE140, SER144, SER305, SER46, THR25	$5.89 \times 10^{-7}$	0.20
Kaempferol-7-O-neohesperidoside	-8.2	ASN142, <b>CYS145</b> , CYS44, GLU166, GLY143, HIS164, HIS163, LEU141, MET165, PHE140, SER144, SER305, SER46, THR24, THR25, THR26, THR45	$9.78 \times 10^{-7}$	0.19
Prunin	-8.2	<b>CYS145</b> , CYS44, GLU166, HIS163, MET165, MET49, PHE140, SER144, SER46, THR25, THR26, THR45	$9.78 \times 10^{-7}$	0.26
Astragalin	-8.1	ARG188, ASN142, CYS44, GLN189, GLU166, HIS164, HIS163, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144, SER305, SER46, THR25, THR45	$1.15 \times 10^{-6}$	0.25
Isoquercetin	-8.1	ARG188, ASN142, <b>CYS145</b> , GLN189, GLU166, GLY143, HIS163, LEU141, MET49, PHE140, SER144, SER305, THR190, THR25	$1.15 \times 10^{-6}$	0.24
Luteolin 7-O- $\beta$ -D-glucoside	-8.1	<b>CYS145</b> , GLU166, HIS163, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144, SER305, SER46, THR24, THR25, THR45	$1.15 \times 10^{-6}$	0.25
Apigenin 7-O-glucoside	-7.9	<b>CYS145</b> , GLU166, HIS164, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144, SER305, SER46, THR24, THR25, THR26, THR45	$1.62 \times 10^{-6}$	0.25
Hyperoside	-7.7	ASN142, <b>CYS145</b> , GLN189, GLU166, GLY143, HIS163, HIS172, LEU141, MET49, PHE140, SER144, SER46, THR25, THR26	$2.27 \times 10^{-6}$	0.23
Quercitrin	-7.7	ASN142, <b>CYS145</b> , GLN189, GLU166, GLY143, HIS164, HIS163, HIS172, LEU141, MET165, MET49, PHE140, SER144, SER305, THR25	$2.27 \times 10^{-6}$	0.24
Isorientin	-7.6	ARG188, ASN142, <b>CYS145</b> , GLN189, GLU166, GLY143, <b>HIS41</b> , MET165, MET49, SER46, THR24, THR25	$2.69 \times 10^{-6}$	0.23
Chlorogenic acid	-7.4	ASN142, <b>CYS145</b> , CYS44, GLU166, HIS163, LEU141, MET165, MET49, PHE140, SER144, THR25, THR45	$3.77 \times 10^{-6}$	0.29
Luteolin	-7.4	ARG188, GLN189, GLU166, HIS164, HIS163, LEU141, MET165, MET49, PHE140, SER144	$3.77 \times 10^{-6}$	0.35
Quercetin	-7.3	<b>CYS145</b> , GLU166, HIS164, HIS163, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144, SER305, THR25	$4.46 \times 10^{-6}$	0.33
Orientin	-7.3	ALA191, ASN142, GLN189, GLU166, HIS164, <b>HIS41</b> , MET165, MET49, PRO168, THR190	$4.46 \times 10^{-6}$	0.22
Epicatechin	-7.2	ARG188, GLN189, GLU166, HIS163, LEU141, MET165, MET49, PHE140, SER144	$5.28 \times 10^{-6}$	0.34
(+)-Catechin	-7.1	ARG188, <b>CYS145</b> , GLN189, GLU166, HIS164, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER305	$6.26 \times 10^{-6}$	0.33
(-)-Catechin	-7.1	ASN142, GLN189, GLU166, HIS164, HIS163, HIS172, <b>HIS41</b> , MET165, MET49, PHE140, SER305	$6.26 \times 10^{-6}$	0.33
Apigenin	-7.1	ARG188, GLN189, GLU166, HIS163, LEU141, MET165, MET49, PHE140, SER144	$6.26 \times 10^{-6}$	0.35
Naringenin	-6.9	ARG188, GLN189, GLU166, HIS163, LEU141, MET165, MET49, PHE140, SER144	$8.77 \times 10^{-6}$	0.34
Isorhamnetin	-6.9	<b>CYS145</b> , GLU166, HIS163, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144, SER305, THR25	$8.77 \times 10^{-6}$	0.30
Isovitexin	-6.9	ARG188, GLN189, GLU166, HIS164, <b>HIS41</b> , MET165, MET49, SER46, THR24, THR25, THR45	$8.77 \times 10^{-6}$	0.22
Caffeic acid	-5.4	<b>CYS145</b> , GLU166, HIS164, HIS163, LEU141, MET165, MET49, PHE140, SER144, SER305	$1.10 \times 10^{-4}$	0.41
Ferulic acid	-5.3	<b>CYS145</b> , GLU166, HIS164, HIS163, <b>HIS41</b> , LEU141, MET165, PHE140, SER305	$1.30 \times 10^{-4}$	0.37
Galic acid	-5.2	ASN518, GLY170, GLY306, LEU586, LYS137, PHE307, SER305, THR169, VAL171	$1.54 \times 10^{-4}$	0.43
Protocatechuic acid	-5.2	ARG308, GLY138, GLY170, GLY306, LEU586, LYS137, PHE307, THR169, VAL171	$1.54 \times 10^{-4}$	0.47
Coumaric acid	-5.2	ASN142, GLU166, HIS163, <b>HIS41</b> , LEU141, MET165, MET49, PHE140, SER144	$1.54 \times 10^{-4}$	0.43

\*Residues His41 and Cys145 of the catalytic site are highlighted with bold font.

<sup>a</sup> In each site, the energy was calculated to see which bound more tightly to the ligand.

<sup>b</sup> A cutoff distance of 3 Å was chosen to consider that a hydrogen bond was established.

<sup>c</sup> N3 is our reference ligand and the structure of its complex with M<sup>pro</sup> was obtained from PDB (id: 6LU7).

showed hydrogen bonds with Ser305, Cys145 and Thr24. Rutin showed hydrogen bonds with Glu166 and Gln189. Nicotiflorin showed hydrogen bonds with Asn142, Glu166 and Cys44. The reference ligand N3 shows four hydrogen bonds with Arg188, Gln189, Cys145 and Gly143 residues. This allows us to conclude that the hydrogen bonds and the hydrophobic interactions dominate formation of these complexes.

### 3.3. Ligand efficiency analysis

The parameters  $K_d$  and  $LE$  were used to compare the affinity of the molecules bound to SARS-CoV-2 M<sup>pro</sup> in this study.  $K_d$  corresponds to the dissociation constant of a ligand-protease complex. Therefore, very low values indicate that the compound binds tightly to the protein.  $LE$  represents the average binding energy per non-hydrogen atom, where tolerable values of  $LE$  for inhibitor candidates are  $LE > 0.3$  kcal mol<sup>-1</sup> (see Table 2). The five best ligands, amentoflavone, isorhoifolin, nicotiflorin, naringin, rutin and the reference molecule N3, as obtained from the docking, exhibit low  $K_d$  values, which means that these complexes are the most stable in the series. Although the  $LE$  values are close to 0.3 kcal mol<sup>-1</sup>, this does not affect binding efficiency to the active site of the protease. The results are consistent with those obtained in the molecular docking in which these complexes were the most stable according to their  $\Delta E_{binding}$  values. Therefore, these molecules are excellent prospects for testing as SARS-CoV-2 M<sup>pro</sup> inhibitors.

### 3.4. Molecular dynamics simulation and MM/GBSA analysis

Molecular dynamics simulations were performed for 75 ns to analyze the steady nature and conformational stability of SARS-CoV-2 M<sup>pro</sup>

bound to N3, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin.

The RMSD results along the MD trajectories show high stability for the SARS-CoV-2 M<sup>pro</sup> dimer, as illustrated in Fig. 2. After a molecular dynamics simulation of 75 ns the N3, amentoflavone, isorhoifolin, nicotiflorin and rutin molecules remain within parameters that agree with the system being in equilibrium. Therefore, no complex suffered structural destabilization during the simulation, except for the naringin complex with the SARS-CoV-2 M<sup>pro</sup>. Instability is observed from 42ns where the RMSD values oscillate by more than 3.0 Å (see Fig. 2-E). Deviations with a maximum difference of 3.0 Å of RMSD [87] indicate that the system is in equilibrium. Also, the RMSD curves for N3 and amentoflavone are remarkably more stable than for isorhoifolin, nicotiflorin and rutin, with RMSD values differing by about 1.0 Å between the above-mentioned complexes.

The Radius of Gyration ( $R_{Gyr}$ ) study complements the RMSD analysis.  $R_{Gyr}$  was computed for the same runs. The results shown in the boxplot (see Fig. 3) illustrate the distribution of the data within  $R_{Gyr}$  as a function of time. The trajectories show that the  $R_{Gyr}$  of N3 and amentoflavone have compactly distributed interquartile range values, showing uniform fluctuations of  $R_{Gyr}$  as a function of the molecular trajectory. Their mean values are 4.97 and 4.74 Å. In the case of isorhoifolin, rutin, nicotiflorin and naringin, the Interquartile Range values are widely distributed, meaning that their  $R_{Gyr}$  values fluctuate more depending on the molecular trajectory. Their mean values are 5.31, 4.51, 4.25 and 4.84 Å. Although the fluctuation of the interquartile range of  $R_{Gyr}$  for these cases is higher, these never exceeded the N3 and amentoflavone values by as much as 0.6 Å, meaning that the ligands remain quite stable during the 75 ns simulation for  $R_{Gyr}$ , indicating good

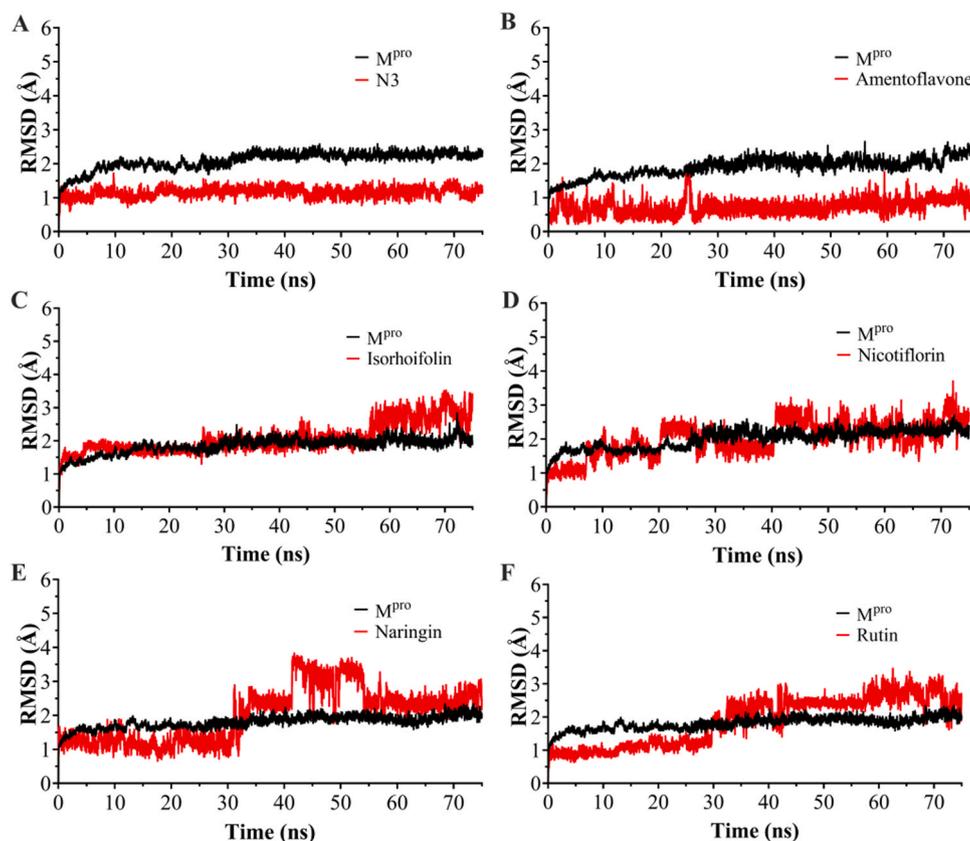


Fig. 2. Root Mean Square Deviation (RMSD) as a function of simulation times for the complexes formed between SARS-CoV-2 M<sup>pro</sup> and A) N3, B) Amentoflavone, C) Isorhoifolin, D) Nicotiflorin, E) Naringin and F) Rutin.

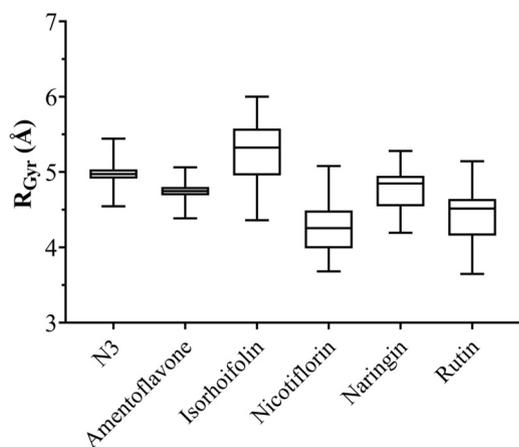


Fig. 3. Box-plot displaying radii of gyration for the SARS-CoV-2 M<sup>pro</sup> in complex with N3, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin.

thermodynamic stability of their M<sup>pro</sup> binding, with only minor conformational changes.

After verifying that SARS-CoV-2 M<sup>pro</sup> maintains its folding when bound to the ligands, it is desirable to perform a quantitative analysis of the structural fluctuation at the residue level using the C<sub>α</sub> atom of each amino acid. To compare in detail the differences between the complexes a suitable measure is Normal Mode Analysis (NMA) and Root Mean Square Fluctuation (RMSF) values, which indicate the fluctuation of the structure with respect to the average conformation in the 75 ns simulation. In this way, the conformational stability of the structure is quantified. When overlaying the RMSF profiles of SARS-CoV-2 M<sup>pro</sup> with different ligands (top Fig. 4), the M<sup>pro</sup> complex present with naringin

showed the greatest deviation of its residues. In contrast, the other complexes show less fluctuation, which could mean greater stability of the backbone in the M<sup>pro</sup> complexes with amentoflavone, isorhoifolin, nicotiflorin and rutin. The most notable differences are found in the Tyr54, Glu55, Ala193, Ala194 and Gly195 residues, which show higher RMSF values. In the active site of the protease, the fluctuation values of the main residues [21] (His41, His163, His164, Phe140 and Cys145) were similar among the complexes with the five selected molecules. In the case of NMA profiles (bottom Fig. 4), they showed similar trends for M<sup>pro</sup> bound to isorhoifolin, nicotiflorin and rutin, and different trends with amentoflavone and naringin. The amentoflavone-M<sup>pro</sup> and naringin-M<sup>pro</sup> complexes showed a comparatively larger fluctuation of residues near Asn51 and the terminal residues of M<sup>pro</sup>. This larger displacement in their normal modes did not affect the flexibility of the protein throughout the simulations. These results indicate that isorhoifolin and rutin are very likely to have the same capability to inhibit the SARS-CoV-2 M<sup>pro</sup> as N3, but reversibly.

The analyses of intermolecular interactions performed during the molecular simulation show that N3, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin preserve hydrogen bonds with active site residues of the SARS-CoV-2 M<sup>pro</sup>. However, the number of hydrogen bonds formed was different for each ligand (Figs. 5–6). N3 formed 3 hydrogen bonds between the residues Thr190, Gln192 and Glu166, highlighting the participation of the residues Gln189, Met49, His41 and Cys145. Amentoflavone formed 4 hydrogen bonds with Cys44, Glu166, His41 and Thr190, highlighting the participation of Asn142, Ser46 and Met49. In the case of isorhoifolin, 3 hydrogen bonds between es Asp187, Glu166 and Leu141 were determined, highlighting the participation of Gln189 and Phe140. Nicotiflorin formed 4 hydrogen bonds with Asp187, Glu166, Asn142 and Gln189, highlighting the participation of Arg188, His164, Ser46, Cys145 and His41. Naringin formed 2 hydrogen bonds with Glu166 and Asn142, highlighting the participation of ues Glu47, Gln192 and Gln189. Three hydrogen bonds are formed between rutin and Glu166, Thr25 and Asp187, highlighting the participation of Cys145, Gln189, His163 and His41. Finally, these residues (see Figs. 5–6) are consistent with previous theoretical-experimental studies carried out by Dai et al. [55], in which they detail the interactions that some of the synthesized compounds have with the active site of the M<sup>pro</sup>.

The binding mode of the molecules in the active site of the SARS-CoV-2 M<sup>pro</sup> at equilibrium is displayed in Fig. 7. The NCIPLOT analysis labeled all the hydrogen-bonding and van der Waals interactions in total agreement with the classical molecular dynamics simulations, providing a qualitative confirmation of these interactions, using a topological and visual analysis of a scalar field related to the electron density and reduced density gradient (Fig. 7-III). These results suggest that the improved character of the inhibitors is due to direct interactions with the protease.

The binding free energy (MM-GBSA) was calculated after MD simulation. The last 70 ns were collected for all complexes, and the results are shown in Table 4. Isorhoifolin and rutin showed better binding free energies with values of  $-43.91$  and  $-44.91$  kcal·mol<sup>-1</sup>, rutin being the molecule with the lowest binding free energy, while amentoflavone, nicotiflorin and naringin showed less negative binding free energies with values of  $-33.77$ ,  $-34.43$  and  $-19.10$  kcal·mol<sup>-1</sup>, respectively. Furthermore, N3 showed a binding free energy value of  $-38.87$  kcal·mol<sup>-1</sup> with M<sup>pro</sup> of SARS-CoV-2, this being the third best value compared to isorhoifolin and rutin, with energy differences of 5.04 and 6.04 kcal·mol<sup>-1</sup>. Although 3N is a synthetic molecule with M<sup>pro</sup> inhibitory activity for SARS-CoV-2, the binding free energy results showed two possible candidates, isorhoifolin and rutin, that are also likely to exhibit M<sup>pro</sup> inhibitory activity. Experiments performed by Huynh and col. and Rahman and col. mention that rutin can be used as a potential M<sup>pro</sup> inhibitor of COVID-19 [88,89], based on its well-documented significant antibacterial and antiviral properties [28, 90,91].

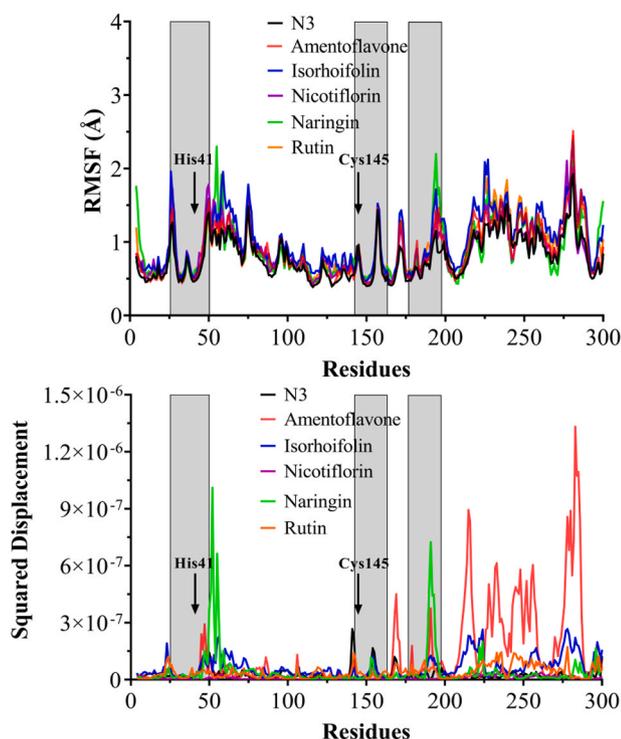
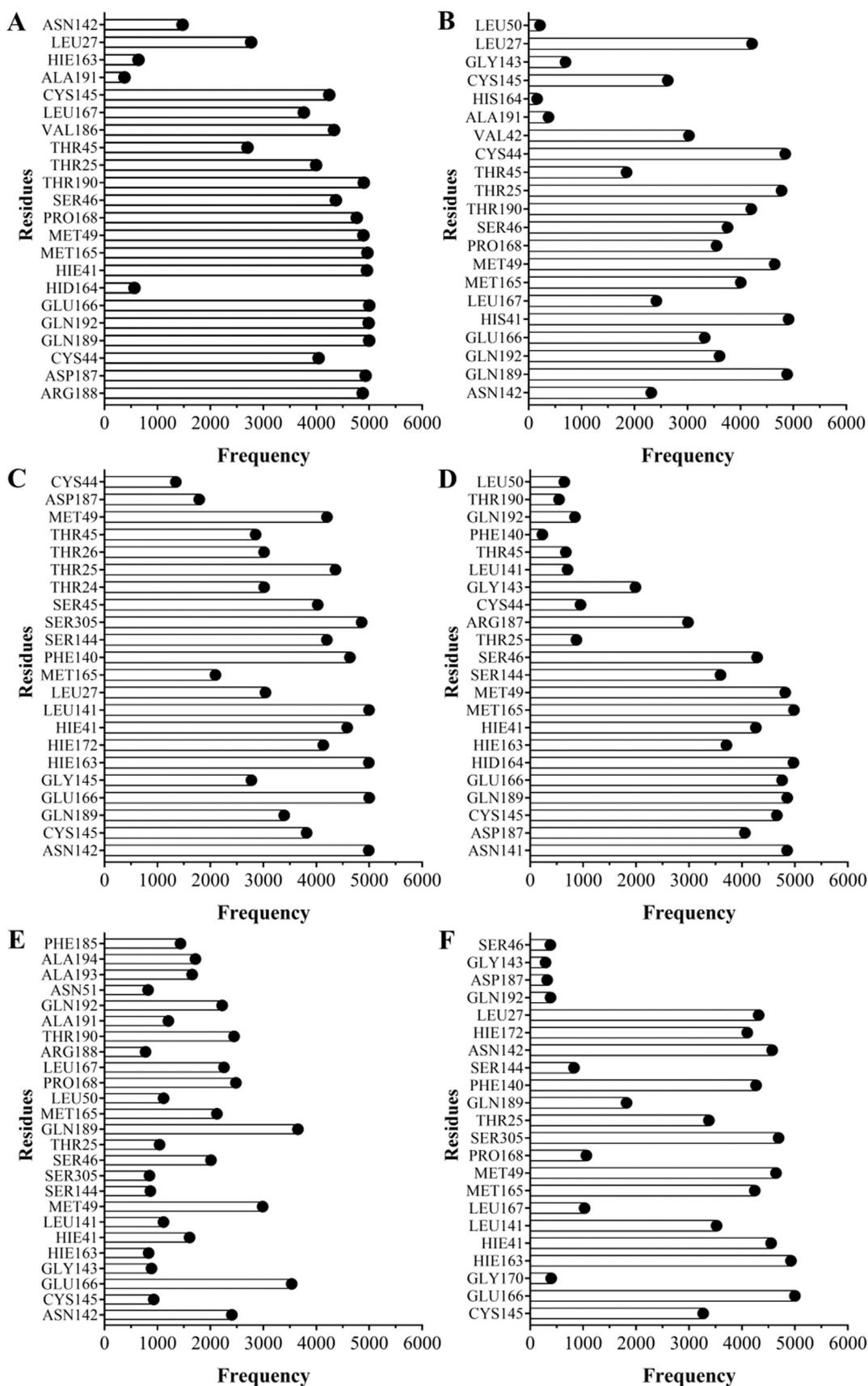
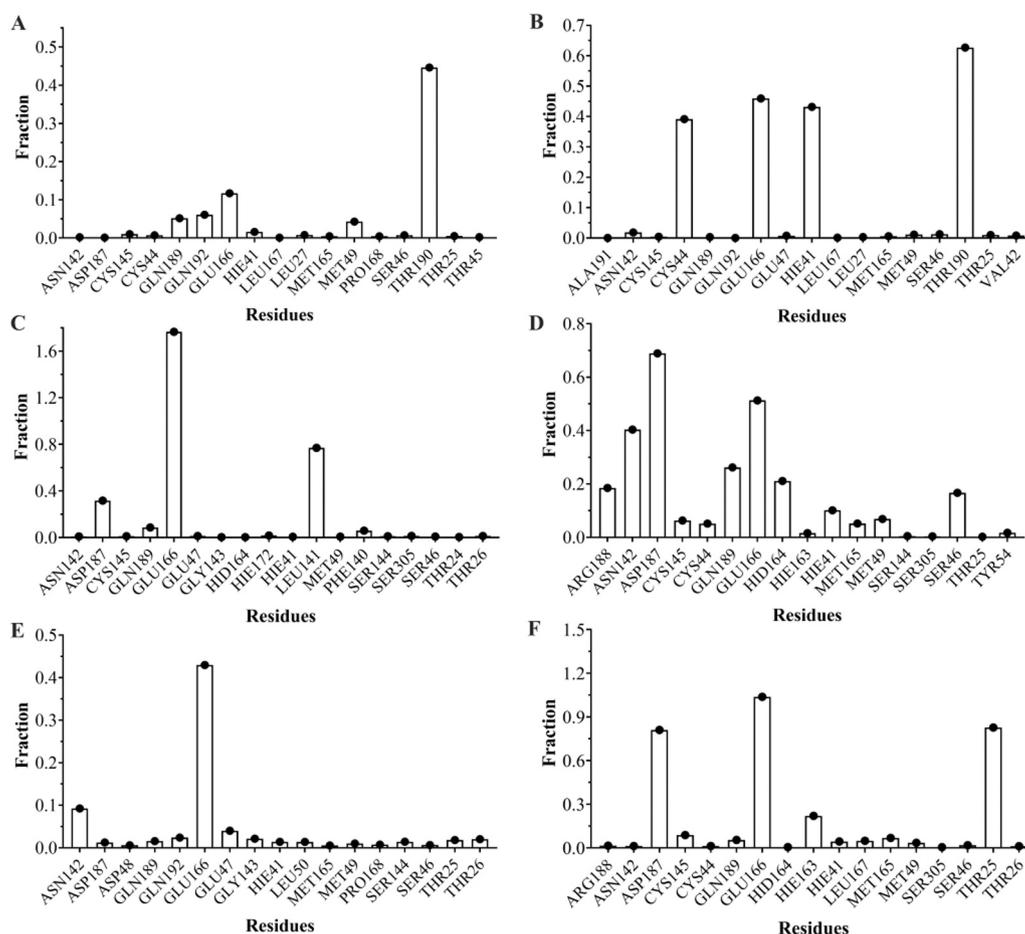


Fig. 4. NMA and RMSF of the  $\alpha$ -carbons. A principal component analysis was carried out using the 75 ns trajectories, and the main normal mode of movement was obtained. The displacement was plotted for each residue of SARS-CoV-2 M<sup>pro</sup> in complex with N3, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin. Pocket site residues are distinguished in gray boxes. His41 and Cys145 residues of the catalytic site are highlighted with bold font.



**Fig. 5.** Frequency of the appearance of residues at a distance of 3.0 Å or less from ligands A) N3, B) Amentoflavone, C) Isorhoifolin, D) Nicotiflorin, E) Naringin and F) Rutin calculated using MD procedures.



**Fig. 6.** Fraction of intermolecular hydrogen bonds for SARS-CoV-2 M<sup>pro</sup> interacting with A) N3, B) Amentoflavone, C) Isorhoifolin, D) Nicotiflorin, E) Naringin and F) Rutin. The graph bar shows the most common hydrogen bonds formed between the residues in the binding pocket and the inhibitors. Values obtained from CPPTRAJ script in AMBER.

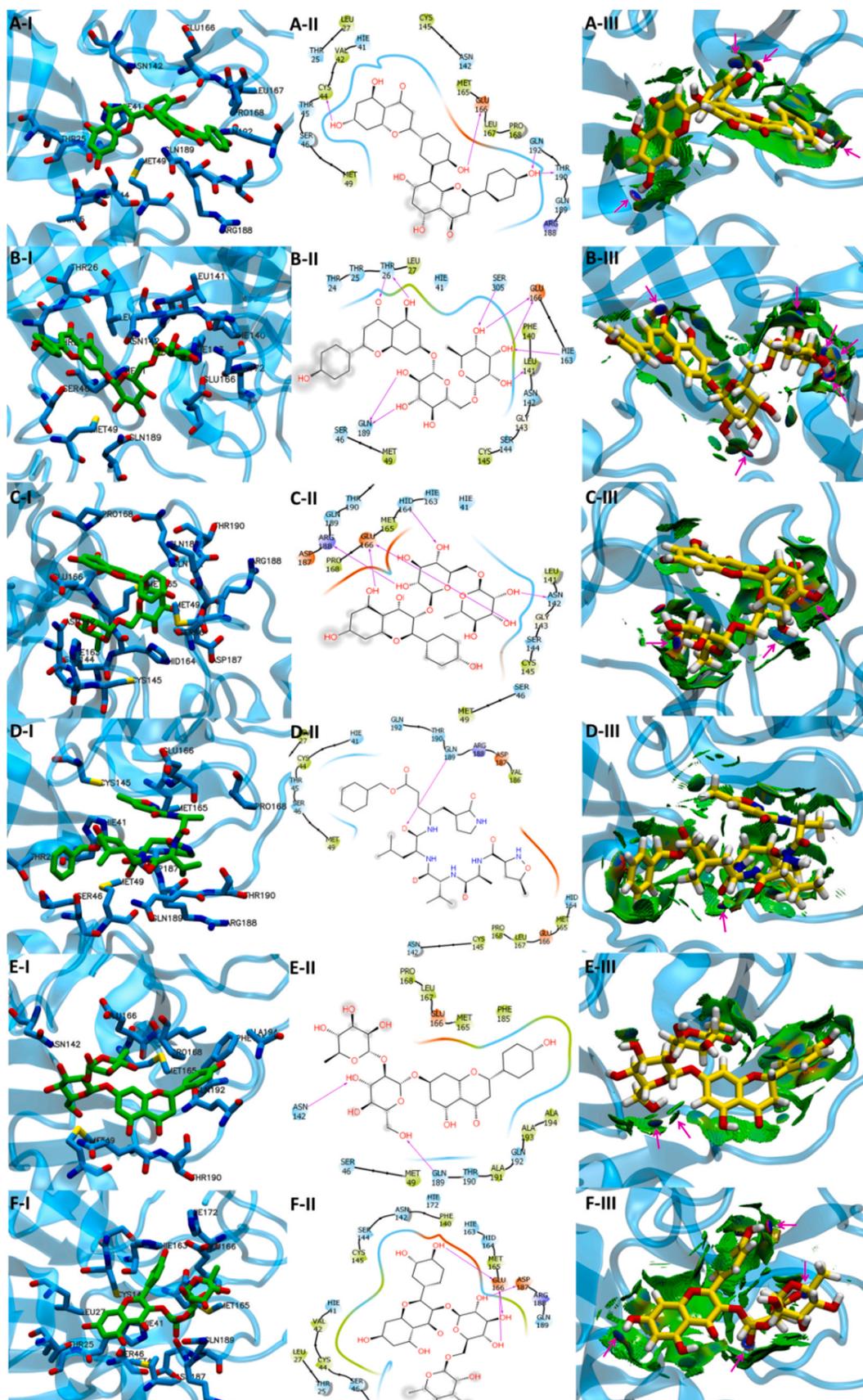
### 3.5. N3, rutin and isorhoifolin as antiviral molecules

Consistent with the results obtained in this work, rutin has been reported to show antimicrobial activity [92] and, through *in silico* studies, possible inhibitory activity of several proteins that are essential for SARS-CoV-2 to complete its viral cycle, [86,88,89,93,94]. However, its antiviral spectrum is broader and it is being tested experimentally as an antiviral agent against retroviruses, orthomyxoviruses, herpes viruses, hepatitis B and C viruses, and the H1N1 influenza virus. In addition to showing antiviral activity, some studies demonstrate its activity associated with anti-inflammatory processes, which is very important in patients with COVID-19 [95]. On the other hand, no antiviral activity of isorhoifolin has been shown, but anti-acaricidal activity has been reported, with an LC<sub>50</sub> = 0.65 mg ml<sup>-1</sup> [96], and potential inhibition of  $\beta$ -amyloid protein aggregation. These compounds interact with the Glu166 residue, reaching an occupancy greater than 0.6. This interaction may be relevant for its antiviral activity since Glu166 has been shown to stabilize the M<sup>pro</sup> dimer and its mutation drastically reduces its activity [85]. Furthermore, these ligands could form hydrogen bonds with the catalytic residues Cys145 and His41, an interaction that could explain the very favorable free energy of binding of these compounds. In this sense, rutin and isorhoifolin form interactions similar to those described for the reversible stage of N3 binding [97] (Fig. 8). As we have reported in this research, amentoflavone, isorhoifolin, nicotiflorin, naringin and rutin, all present in *Criollo*, *Trinitario* and *Forastero* cocoa beans [9] (Table 5), can interact with various residues in the active site of M<sup>pro</sup>, including catalytic residues, showing excellent potential to inhibit the replication of SARS-CoV-2.

Therefore, this research is a starting point for *in vitro* studies and the development of pharmacological therapies against this virus, using bioproducts developed from cocoa rich in bioflavonoids such as rutin and isorhoifolin.

## 4. Conclusions

*Theobroma cacao* contains polyphenolic compounds, secondary metabolites that are widely available in natural medicines. The identification of these polyphenolic compounds is a remarkable step towards the discovery of drugs targeting SARS-CoV-2 M<sup>pro</sup>, and may help in the development of new drugs. Results suggest that the flavonoids isorhoifolin and rutin should display strong inhibitory activities against SARS-CoV-2 M<sup>pro</sup>. Isorhoifolin and rutin seem to bind more strongly than N3 co-crystallized inhibitor. The estimated free energies of binding of these two flavonoids are -43.91 and -44.91 kcal mol<sup>-1</sup>, respectively. The binding of the flavonoids at the protease active site is structure dependent, proffering the spontaneous and energetically favored production of the protein ligand complex. The interactions of flavonoids with active site residues mainly occur through the hydroxyl groups in their structure. These results represent an important step in the exploration of natural products from *Theobroma cacao* for structure-based design of anti-SARS-CoV-2 drugs. Furthermore, they encourage further *in vitro* research and also boost the traditional preventive use of isorhoifolin and rutin. We anticipate that the insights obtained in the present study may prove valuable for researching and developing novel anti-SARS-CoV-2 coronavirus therapeutic agents in the future.

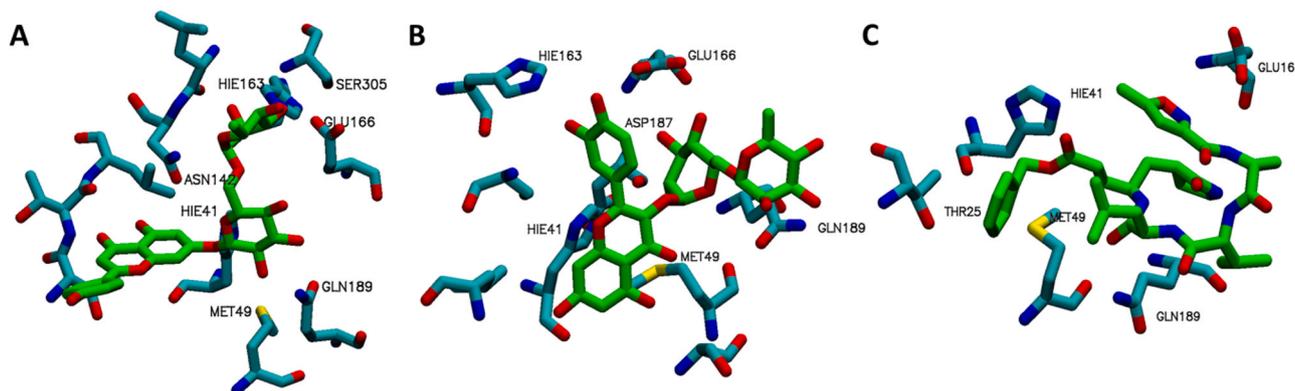


**Fig. 7.** Schematic representations of principal component analysis of the respective production runs for A) Amentoflavone, B) Isorhoifolin, C) Nicotiflorin, D) N3, E) Naringin and F) Rutin bound to SARS-CoV-2 M<sup>pro</sup>. (I) Representative amino acid residues surrounding ligands in the binding pocket of M<sup>pro</sup>. (II) Two-dimensional interaction map of ligands and M<sup>pro</sup>. (III) NCIPLOT isosurface gradient (0.5 au) of ligands in the structure of M<sup>pro</sup>. The arrows indicate potential interactions between amino acid residues and ligands.

**Table 4**

Predicted binding free energies (kcal mol<sup>-1</sup>) and individual energy terms calculated from molecular dynamics simulation following the MM-GBSA protocol for M<sup>PTO</sup> complexes.

Compounds	Calculated free energy of decomposition (kcal mol <sup>-1</sup> )				
	$\Delta G_{binding}$	$\Delta E_{vdW}$	$\Delta E_{elect}$	$\Delta G_{gas}$	$\Delta G_{solv}$
N3	-38.87 ± 0.04	-54.57 ± 0.05	-17.74 ± 0.06	-72.32 ± 0.08	33.44 ± 0.05
Amentoflavone	-33.77 ± 0.09	-41.23 ± 0.08	-25.78 ± 0.11	-67.01 ± 0.15	33.24 ± 0.08
Isorhoifolin	-43.91 ± 0.09	-45.54 ± 0.08	-62.18 ± 0.16	-107.72 ± 0.17	63.81 ± 0.13
Nicotiflorin	-34.43 ± 0.10	-43.99 ± 0.07	-41.23 ± 0.22	-85.23 ± 0.22	50.80 ± 0.14
Naringin	-19.10 ± 0.12	-30.97 ± 0.13	-14.64 ± 0.19	-45.62 ± 0.25	26.51 ± 0.15
Rutin	-44.91 ± 0.08	-55.18 ± 0.06	-42.78 ± 0.19	-97.97 ± 0.18	53.06 ± 0.11



**Fig. 8.** Representative amino acid residues surrounding ligands in the binding pocket of M<sup>PTO</sup> for A) Isorhoifolin, B) Rutin and C) N3.

**Table 5**

Concentrations of compounds that showed the highest inhibitory activity of M<sup>PTO</sup> protease in varieties of *Theobroma cacao*.

No.	Compound	Class	Polyphenols (mg/kg)		
			Criollo	Trinitario	Forastero
1	Amentoflavone	Flavones	n.d.	0.24	5.8
2	Naringin	Flavanones	n.d.	n.d.	n.d.
3	Isorhoifolin	Flavones	37.5	95	44.7
4	Rutin	Flavonols	50	20	10
5	Nicotiflorin	Flavonols	12.1	7.4	8.5
6	Kaempferol-7-O-neohesperidoside	Flavonols	9.9	2.85	3.9
7	Prunin	Flavanones	33.8	54.65	31.6
8	Astragalin	Flavanols	0.5	0.5	0.4
9	Isoquercetin	Flavonols	n.d.	n.d.	n.d.
10	Luteolin 7-O- $\beta$ -D-glucoside	Flavones	1.2	0.55	1.0

### CRedit authorship contribution statement

Oswaldo Yañez, Olimpo García-Beltrán, Carlos Areche, William Tiznado and Angelica Sandoval-Aldana contributed to the conception and design of the study; Oswaldo Yañez, Alejandro Vasquez-Espinal, Edison Osorio, Manuel Isaías Osorio and Jessica Bravo performed the theoretical calculations; Oswaldo Yañez, José M. Pérez-Donoso, Fernando González-Nilo, Manuel Isaías Osorio and Angelica Sandoval-Aldana organized the database; Oswaldo Yañez, William Tiznado, Carlos Areche and Olimpo García-Beltrán wrote the first draft of the manuscript; Oswaldo Yañez, Olimpo García-Beltrán and Maria João Matos wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

### Declaration of Competing Interest

No conflict of interest.

### Acknowledgments

Olimpo García-Beltrán acknowledges funding from the Ministry of Science, Technology and Innovation, the Ministry of Education, the Ministry of Industry, Commerce and Tourism, and ICETEX, Programme Ecosistema Científico-Colombia Científica, from the Francisco José de Caldas Fund, Grant RC-FP44842-212-2018 (Colombia), Angelica Sandoval-Aldana and Olimpo García-Beltrán, acknowledge funding from FONTAGRO (ATN-RF 16109 RG, Maria João Matos would like to thank Fundação para a Ciência e Tecnologia (FCT, CEECIND/02423/2018 and UIDB/00081/2020) (Portugal), William Tiznado acknowledges funding from the National Agency for Research and Development through ECOS170045 and FONDECYT projects 1211128 (Chile) and Osvaldo Yañez acknowledges funding from the ANID/PIA ACT192144 (Chile).

### References

- [1] J.S. Henderson, R.A. Joyce, G.R. Hall, W.J. Hurst, P.E. McGovern, Chemical and archaeological evidence for the earliest cacao beverages, *Proc. Natl. Acad. Sci. USA* 104 (2007) 18937–18940, <https://doi.org/10.1073/pnas.0708815104>.
- [2] M. Rusconi, A. Conti, *Theobroma cacao* L., the food of the gods: a scientific approach beyond myths and claims, *Pharmacol. Res.* 61 (2010) 5–13, <https://doi.org/10.1016/j.phrs.2009.08.008>.
- [3] T.L. Dillinger, P. Barriga, S. Escárcega, M. Jimenez, D.S. Lowe, L.E. Grivetti, Food of the gods: cure for humanity? A cultural history of the medicinal and ritual use of chocolate, *J. Nutr.* 130 (2000) 2057S–2072S, <https://doi.org/10.1093/jn/130.8.2057S>.
- [4] W. Pérez-Mora, J.V. Jorrián-Novo, L.M. Melgarejo, Substantial equivalence analysis in fruits from three *Theobroma* species through chemical composition and protein profiling, *Food Chem.* 240 (2018) 496–504, <https://doi.org/10.1016/j.foodchem.2017.07.128>.
- [5] R. Campos-Vega, K.H. Nieto-Figueroa, B.D. Oomah, Cocoa (*Theobroma cacao* L.) pod husk: renewable source of bioactive compounds, *Trends Food Sci. Technol.* 81 (2018) 172–184, <https://doi.org/10.1016/j.tifs.2018.09.022>.
- [6] M. Arlorio, M. Locatelli, F. Travaglia, J.D. Coisson, E. Del Grosso, A. Minassi, G. Appendino, A. Martelli, Roasting impact on the contents of clovamide (N-caffeoyl-L-DOPA) and the antioxidant activity of cocoa beans (*Theobroma cacao* L.), *Food Chem.* 106 (2008) 967–975, <https://doi.org/10.1016/j.foodchem.2007.07.009>.

- [7] A. Othman, A. Ismail, N. Abdul Ghani, I. Adenan, Antioxidant capacity and phenolic content of cocoa beans, *Food Chem.* 100 (2007) 1523–1530, <https://doi.org/10.1016/j.foodchem.2005.12.021>.
- [8] R.M. Lamuela-Raventós, A.I. Romero-Pérez, C. Andrés-Lacueva, A. Tornero, Review: health effects of cocoa flavonoids, *Food Sci. Technol. Int.* 11 (2005) 159–176, <https://doi.org/10.1177/1082013205054498>.
- [9] Y. Rodríguez-Carrasco, A. Gaspari, G. Graziani, A. Santini, A. Ritieni, Fast analysis of polyphenols and alkaloids in cocoa-based products by ultra-high performance liquid chromatography and Orbitrap high resolution mass spectrometry (UHPLC-Q-Orbitrap-MS/MS), *Food Res. Int.* 111 (2018) 229–236, <https://doi.org/10.1016/j.foodres.2018.05.032>.
- [10] M. del R. Brunetto, L. Gutiérrez, Y. Delgado, M. Gallignani, A. Zambrano, Á. Gómez, G. Ramos, C. Romero, Determination of theobromine, theophylline and caffeine in cocoa samples by a high-performance liquid chromatographic method with on-line sample cleanup in a switching-column system, *Food Chem.* 100 (2007) 459–467, <https://doi.org/10.1016/j.foodchem.2005.10.007>.
- [11] I. Ramirez-Sanchez, L. Maya, G. Ceballos, F. Villarreal, Fluorescent detection of (-)-epicatechin in microsomes from cacao seeds and cocoa products: comparison with Folin-Ciocalteu method, *J. Food Compos. Anal.* 23 (2010) 790–793, <https://doi.org/10.1016/j.jfca.2010.03.014>.
- [12] L.C. Carrillo, J. Londoño-Londoño, A. Gil, Comparison of polyphenol, methylxanthines and antioxidant activity in Theobroma cacao beans from different cocoa-growing areas in Colombia, *Food Res. Int.* 60 (2014) 273–280, <https://doi.org/10.1016/j.foodres.2013.06.019>.
- [13] H. Kim, P.G. Keeney, Method of analysis for (-)-epicatechin in cocoa beans by high performance liquid chromatography, *J. Food Sci.* 48 (1983) 548–551, <https://doi.org/10.1111/j.1365-2621.1983.tb10787.x>.
- [14] J. Wollgast, E. Anklam, Review on polyphenols in Theobroma cacao: changes in composition during the manufacture of chocolate and methodology for identification and quantification, *Food Res. Int.* 33 (2000) 423–447, [https://doi.org/10.1016/S0963-9969\(00\)00068-5](https://doi.org/10.1016/S0963-9969(00)00068-5).
- [15] N. Ortega, J. Reguant, M.P. Romero, A. Macià, M.J. Motilva, Effect of fat content on the digestibility and bioaccessibility of cocoa polyphenol by an in vitro digestion model, *J. Agric. Food Chem.* 57 (2009) 5743–5749, <https://doi.org/10.1021/jf900591q>.
- [16] O. García-Beltrán, N.P. Mena, P. Aguirre, G. Barriga-Gonzalez, A. Galdamez, E. Nagles, T. Adasme, C. Hidalgo, M.T. Nuñez, Development of an iron-selective antioxidant probe with protective effects on neuronal function, *PLoS One* 12 (2017) 1–18, <https://doi.org/10.1371/journal.pone.0189043>.
- [17] P. Aguirre, O. García-Beltrán, V. Tapia, Y. Muñoz, B.K. Cassels, M.T. Nuñez, Neuroprotective effect of a new 7,8-dihydroxycoumarin-based Fe<sup>2+</sup>/Cu<sup>2+</sup> chelator in cell and animal models of parkinson's disease, *ACS Chem. Neurosci.* 8 (2017) 178–185, <https://doi.org/10.1021/acschemneuro.6b00309>.
- [18] J. Oracz, E. Nebesny, D. Zyzewicz, G. Budryn, B. Luzak, Bioavailability and metabolism of selected cocoa bioactive compounds: a comprehensive review, *Crit. Rev. Food Sci. Nutr.* 60 (2020) 1947–1985, <https://doi.org/10.1080/10408398.2019.1619160>.
- [19] M.S. Beg, S. Ahmad, K. Jan, K. Bashir, Status, supply chain and processing of cocoa - a review, *Trends Food Sci. Technol.* 66 (2017) 108–116, <https://doi.org/10.1016/j.tifs.2017.06.007>.
- [20] S. Yuan, X. Li, Y. Jin, J. Lu, Chocolate consumption and risk of coronary heart disease, stroke, and diabetes: a meta-analysis of prospective studies, *Nutrients* 9 (2017), <https://doi.org/10.3390/nu9070688>.
- [21] L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. Von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. De Wilde, E.J. Snijder, H. Liu, R. Hilgenfeld,  $\alpha$ -Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment, *J. Med. Chem.* (2020), <https://doi.org/10.1021/acs.jmedchem.9b01828>.
- [22] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors, *Science* 368 (80-) (2020) 409–412, <https://doi.org/10.1126/science.abb3405>.
- [23] M. Wink, Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective, *Phytochemistry* 64 (2003) 3–19, [https://doi.org/10.1016/S0031-9422\(03\)00300-5](https://doi.org/10.1016/S0031-9422(03)00300-5).
- [24] S. Ben-Shabat, L. Yarmolinsky, D. Porat, A. Dahan, Antiviral effect of phytochemicals from medicinal plants: applications and drug delivery strategies, *Drug Deliv. Transl. Res.* 10 (2020) 354–367, <https://doi.org/10.1007/s13346-019-00691-6>.
- [25] W. Dai, J. Bi, F. Li, S. Wang, X. Huang, X. Meng, B. Sun, D. Wang, W. Kong, C. Jiang, W. Su, Antiviral efficacy of flavonoids against enterovirus 71 infection in vitro and in newborn mice, *Viruses* 11 (2019), <https://doi.org/10.3390/v11070625>.
- [26] E. Moghaddam, B.T. Teoh, S.S. Sam, R. Lani, P. Hassandarvish, Z. Chik, A. Yueh, S. Abubakar, K. Zandi, Baicalin, a metabolite of baicalin with antiviral activity against dengue virus, *Sci. Rep.* 4 (2014) 1–8, <https://doi.org/10.1038/srep05452>.
- [27] Y.B. Ryu, H.J. Jeong, J.H. Kim, Y.M. Kim, J.Y. Park, D. Kim, T.T.H. Nguyen, S. J. Park, J.S. Chang, K.H. Park, M.C. Rho, W.S. Lee, Biflavonoids from *Torreya nucifera* displaying SARS-CoV 3CLpro inhibition, *Bioorg. Med. Chem.* 18 (2010) 7940–7947, <https://doi.org/10.1016/j.bmc.2010.09.035>.
- [28] S. Jo, S. Kim, D.H. Shin, M.S. Kim, Inhibition of SARS-CoV 3CL protease by flavonoids, *J. Enzym. Inhib. Med. Chem.* 35 (2020) 145–151, <https://doi.org/10.1080/14756366.2019.1690480>.
- [29] F. Meneguzzo, R. Ciriminna, F. Zabini, M. Pagliaro, Review of evidence available on hesperidin-rich products as potential tools against COVID-19 and hydrodynamic cavitation-based extraction as a method of increasing their production, *Processes* 8 (2020) 1–18, <https://doi.org/10.3390/PR8050549>.
- [30] W. Ngwa, R. Kumar, D. Thompson, W. Lyerly, R. Moore, T.E. Reid, H. Lowe, N. Toyang, Potential of flavonoid-inspired phytochemicals against COVID-19, *Molecules* 25 (2020) 1–10, <https://doi.org/10.3390/molecules25112707>.
- [31] A.M. Sayed, H.A. Alhadrami, A.O. El-Gendy, Y.I. Shamikh, L. Belbahri, H. M. Hassan, U.R. Abdelmohsen, M.E. Rateb, Microbial natural products as potential inhibitors of SARS-CoV-2 main protease (Mpro), *Microorganisms* 8 (2020) 970, <https://doi.org/10.3390/microorganisms8070970>.
- [32] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [33] Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, 2017.
- [34] Y. Zhao, N.E. Schultz, D.G. Truhlar, Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions, *J. Chem. Theory Comput.* 2 (2006) 364–382, <https://doi.org/10.1021/ct0502763>.
- [35] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, *J. Comput. Chem.* 32 (2011) 1456–1465, <https://doi.org/10.1002/jcc.21759>.
- [36] L. Goerigk, S. Grimme, A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions, *Phys. Chem. Chem. Phys.* 13 (2011) 6670–6688, <https://doi.org/10.1039/c0cp02984j>.
- [37] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J. V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian16 Revision B.01, 2016.
- [38] R. Pino-Ríos, O. Yañez, D. Inostroza, L. Ruiz, C. Cardenas, P. Fuentealba, W. Tiznado, Proposal of a simple and effective local reactivity descriptor through a topological analysis of an orbital-weighted fukui function, *J. Comput. Chem.* 38 (2017) 481–488, <https://doi.org/10.1002/jcc.24699>.
- [39] T. Koopmans, Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms, *Physica* 1 (1934) 104–113, [https://doi.org/10.1016/S0031-8914\(34\)90011-2](https://doi.org/10.1016/S0031-8914(34)90011-2).
- [40] R.G. Parr, R.G. Pearson, Absolute hardness: companion parameter to absolute electronegativity, *J. Am. Chem. Soc.* 105 (1983) 7512–7516, <https://doi.org/10.1021/ja00364a005>.
- [41] E.G. Lewars, Computational Chemistry-Introduction to the Theory and Applications of Molecular and Quantum Mechanics, Springer Netherlands, New York, 2011, <https://doi.org/10.1007/978-90-481-3862-3>.
- [42] D.C. Young, Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems, John Wiley & Sons, New York, 2002, <https://doi.org/10.1002/0471220655>.
- [43] F. Jensen, Introduction to Computational Chemistry, 2nd ed., John Wiley & Sons, New York, 2006.
- [44] C.J. Cramer, Essentials of Computational Chemistry: Theories and Models, 2nd ed., John Wiley & Sons, New York, 2004.
- [45] R.G. Pearson, Hard and soft acids and bases, *J. Am. Chem. Soc.* 85 (1963) 3533–3539, <https://doi.org/10.1021/ja00905a001>.
- [46] R.G. Pearson, Chemical Hardness, Wiley, New York, 2005, <https://doi.org/10.1002/3527606173>.
- [47] J.L. Gázquez, A. Cedillo, A. Vela, Electrodonating and electroaccepting powers, *J. Phys. Chem. A* 111 (2007) 1966–1970, <https://doi.org/10.1021/jp065459f>.
- [48] P.K. Chattaraj, A. Chakraborty, S. Giri, Net electrophilicity, *J. Phys. Chem. A* 113 (2009) 10068–10074, <https://doi.org/10.1021/jp904674x>.
- [49] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* 19 (1998) 1639–1662, [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B).
- [50] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The Protein Data Bank, *Nucleic Acids Res.* 28 (2000) 235–242, <https://doi.org/10.1093/nar/28.1.235>.
- [51] L. Schrödinger, Maestro, Schrödinger, (2020).
- [52] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, *J. Comput. Aided Mol. Des.* 27 (2013) 221–234, <https://doi.org/10.1007/s10822-013-9644-8>.
- [53] Z. Xu, C. Peng, Y. Shi, Z. Zhu, K. Mu, X. Wang, W. Zhu, Nelfinavir was predicted to be a potential inhibitor of 2019-nCoV main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation, *BioRxiv* (2020), <https://doi.org/10.1101/2020.01.27.921627>.
- [54] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors, *Science* 3405 (80-) (2020) eabb3405, <https://doi.org/10.1126/science.abb3405>.

- [55] W. Dai, B. Zhang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang, H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 4489 (80-) (2020) 1–10, <https://doi.org/10.1126/science.abb4489>.
- [56] M.T. ul Qamar, S.M. Alqahtani, M.A. Alamri, L.-L. Chen, Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants, *J. Pharm. Anal.* 10 (2020) 313–319, <https://doi.org/10.1016/j.jpha.2020.03.009>.
- [57] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [58] O. Yañez, M.I. Osorio, E. Uriarte, C. Areche, W. Tiznado, J.M. Pérez-Donoso, O. García-Beltrán, F. González-Nilo, In silico study of coumarins and quinolines derivatives as potent inhibitors of SARS-CoV-2 main protease, *Front. Chem.* 8 (2021) 1273, <https://doi.org/10.3389/fchem.2020.595097>.
- [59] J.J.P. Stewart, Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *J. Mol. Model.* 13 (2007) 1173–1213, <https://doi.org/10.1007/s00894-007-0233-4>.
- [60] J. Rezáč, P. Hobza, Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods, *J. Chem. Theory Comput.* 8 (2012) 141–151, <https://doi.org/10.1021/ct200751e>.
- [61] J.J.P. Stewart, MOPAC, (2016).
- [62] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38, [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [63] C. Abad-Zapatero, Ligand efficiency indices for effective drug discovery, *Expert Opin. Drug Discov.* 2 (2007) 469–488, <https://doi.org/10.1517/17460441.2.4.469>.
- [64] C. Abad-Zapatero, O. Perišić, J. Wass, A.P. Bento, J. Overington, B. Al-Lazikani, M. E. Johnson, Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, *Drug Discov. Today* 15 (2010) 804–811, <https://doi.org/10.1016/J.DRUDIS.2010.08.004>.
- [65] C. Abad-Zapatero, Ligand Efficiency Indices for Drug Discovery, Academic Press, San Diego, 2013, <https://doi.org/10.1016/B978-0-12-404635-1.00009-8>.
- [66] C.H. Reynolds, B.A. Tounge, S.D. Bembek, Ligand binding efficiency: trends, physical basis, and implications, *J. Med. Chem.* 51 (2008) 2432–2438, <https://doi.org/10.1021/jm701255b>.
- [67] M.M. Cavalluzzi, G.F. Mangiatordi, O. Nicolotti, G. Lentini, Ligand efficiency metrics in drug discovery: the pros and cons from a practical perspective, *Expert Opin. Drug Discov.* 12 (2017) 1087–1104, <https://doi.org/10.1080/17460441.2017.1365056>.
- [68] E. Neria, S. Fisher, M. Karplus, Simulation of activation free energies in molecular systems, *J. Chem. Phys.* 105 (1996) 1902–1921, <https://doi.org/10.1063/1.472061>.
- [69] R. Anandakrishnan, B. Aguilar, A.V. Onufriev, H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, *Nucleic Acids Res.* 40 (2012) W537–W541, <https://doi.org/10.1093/nar/gks375>.
- [70] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174, <https://doi.org/10.1002/jcc.20035>.
- [71] G.A. Özplnar, W. Peukert, T. Clark, An improved generalized AMBER force field (GAFF) for urea, *J. Mol. Model.* 16 (2010) 1427–1440, <https://doi.org/10.1007/s00894-010-0650-7>.
- [72] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model, *J. Phys. Chem.* 97 (1993) 10269–10280, <https://doi.org/10.1021/j100142a004>.
- [73] R. Salomon-Ferrer, D.A. Case, R.C. Walker, An overview of the Amber biomolecular simulation package, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3 (2013) 198–210, <https://doi.org/10.1002/wcms.1121>.
- [74] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkerich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (1998) 3586–3616, <https://doi.org/10.1021/jp973084f>.
- [75] J. Huang, A.D. MacKerell Jr, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data, *J. Comput. Chem.* 34 (2013) 2135–2145, <https://doi.org/10.1002/jcc.23354>.
- [76] D.J. Mermelstein, C. Lin, G. Nelson, R. Kretsch, J.A. McCammon, R.C. Walker, Fast and flexible gpu accelerated binding free energy calculations within the amber molecular dynamics package, *J. Comput. Chem.* 39 (2018) 1354–1358, <https://doi.org/10.1002/jcc.25187>.
- [77] J.M. Hayes, G. Archontis, MM-GB(PB)SA calculations of protein-ligand binding free energies, in: L. Wang (Ed.), *Molecular Dynamics*, IntechOpen, Rijeka, 2012, <https://doi.org/10.5772/37107>.
- [78] L. Song, T.-S. Lee, C. Zhu, D.M. York, K.M.M. Jr, Validation of AMBER/GAFF for Relative Free Energy Calculations, (2019). (<https://doi.org/10.26434/chemrxiv.7653434.v1>).
- [79] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born, *J. Chem. Theory Comput.* 8 (2012) 1542–1555, <https://doi.org/10.1021/ct200909j>.
- [80] H. Abroshan, H. Akbarzadeh, G.A. Parsafar, Molecular dynamics simulation and MM-PBSA calculations of sickle cell hemoglobin in dimer form with Val, Trp, or Phe at the lateral contact, *J. Phys. Org. Chem.* 23 (2010) 866–877, <https://doi.org/10.1002/poc.1679>.
- [81] J. Contreras-García, E.R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang, NCIPLLOT: a program for plotting noncovalent interaction regions, *J. Chem. Theory Comput.* 7 (2011) 625–632, <https://doi.org/10.1021/ct100641a>.
- [82] E.R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A.J. Cohen, W. Yang, Revealing noncovalent interactions, *J. Am. Chem. Soc.* 132 (2010) 6498–6506, <https://doi.org/10.1021/ja100936w>.
- [83] C. Ortega-Moo, J. Garza, R. Vargas, The substituent effect on the antioxidant capacity of catechols and resorcinols, *Theor. Chem. Acc.* 135 (2016) 177, <https://doi.org/10.1007/s00214-016-1932-7>.
- [84] D.S. Hui, E.I. Azhar, T.A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z.A. Memish, C. Drosten, A. Zumla, E. Petersen, The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health; the latest 2019 novel coronavirus outbreak in Wuhan, China, *Int. J. Infect. Dis.* 91 (2020) 264–266, <https://doi.org/10.1016/j.ijid.2020.01.009>.
- [85] S.T. Ngo, N. Quynh Anh Pham, L. Thi Le, D.-H. Pham, V.V. Vu, Computational determination of potential inhibitors of SARS-CoV-2 main protease, *J. Chem. Inf. Model.* 60 (2020) 5771–5780, <https://doi.org/10.1021/acs.jcim.0c00491>.
- [86] S.A. Cherrak, H. Merzouk, N. Mokhtari-Soulimane, Potential bioactive glycosylated flavonoids as SARS-CoV-2 main protease inhibitors: a molecular docking and simulation studies, *PLoS One* 15 (2020), 0240653, <https://doi.org/10.1371/journal.pone.0240653>.
- [87] O. Carugo, How root-mean-square distance (r.m.s.d.) values depend on the resolution of protein structures that are compared, *J. Appl. Crystallogr.* 36 (2003) 125–128, <https://doi.org/10.1107/S002188902020502>.
- [88] F. Rahman, S. Tabrez, R. Ali, A.S. Alqahtani, M.Z. Ahmed, A. Rub, Molecular docking analysis of rutin reveals possible inhibition of SARS-CoV-2 vital proteins, *J. Tradit. Complement. Med.* 11 (2021) 173–179, <https://doi.org/10.1016/j.jtcm.2021.01.006>.
- [89] T. Huynh, H. Wang, B. Luan, Structure-based lead optimization of herbal medicine rutin for inhibiting SARS-CoV-2's main protease, *Phys. Chem. Chem. Phys.* 22 (2020) 25335–25343, <https://doi.org/10.1039/D0CP03867A>.
- [90] N. Chéron, C. Yu, A.O. Kolawole, E.I. Shakhnovich, C.E. Wobus, Repurposing of rutin for the inhibition of norovirus replication, *Arch. Virol.* 160 (2015) 2353–2358, <https://doi.org/10.1007/s00705-015-2495-y>.
- [91] J.-H. Choi, D.-W. Kim, S.-E. Park, H.-J. Lee, K.-M. Kim, K.-J. Kim, M.-K. Kim, S.-J. Kim, S. Kim, Anti-thrombotic effect of rutin isolated from *Dendropanax moribifera* Leveille, *J. Biosci. Bioeng.* 120 (2015) 181–186, <https://doi.org/10.1016/j.jbiosc.2014.12.012>.
- [92] A. Ganeshpurkar, A.K. Saluja, The pharmacological potential of rutin, *Saudi Pharm. J.* 25 (2017) 149–164, <https://doi.org/10.1016/j.jpsps.2016.04.025>.
- [93] A.A. Al-Zahrani, Rutin as a promising inhibitor of main protease and other protein targets of COVID-19: in silico study, *Nat. Prod. Commun.* 15 (2020), <https://doi.org/10.1177/1934578x20953951>.
- [94] H. Puttaswamy, H.G. Gowtham, M.D. Ojha, A. Yadav, G. Choudhri, V. Raguraman, B. Kongkham, K. Selvaraju, S. Shareef, P. Gehlot, F. Ahamed, L. Chauhan, In silico studies evidenced the role of structurally diverse plant secondary metabolites in reducing SARS-CoV-2 pathogenesis, *Sci. Rep.* 10 (2020) 1–24, <https://doi.org/10.1038/s41598-020-77602-0>.
- [95] S.A. Adefegha, D.B.R. Leal, J.S. de Oliveira, A.G. Manzoni, J.M. Bremm, Modulation of reactive oxygen species production, apoptosis and cell cycle in pleural exudate cells of carrageenan-induced acute inflammation in rats by rutin, *Food Funct.* 8 (2017) 4459–4468, <https://doi.org/10.1039/C7FO01008G>.
- [96] O.T. Adenubi, M. Ali Abdalla, A.S. Ahmed, E.M. Njoya, L.J. McGaw, J.N. Eloff, V. Naidoo, Isolation and characterization of two acaricidal compounds from *Calpurnia aurea* subsp. *aurea* (Fabaceae) leaf extract, *Exp. Appl. Acarol.* 75 (2018) 345–354, <https://doi.org/10.1007/s10493-018-0269-4>.
- [97] K. Arafet, N. Serrano-Aparicio, A. Lodola, A.J. Mulholland, F.V. González, K. Świderek, V. Moliner, Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity, *Chem. Sci.* 12 (2021) 1433–1444, <https://doi.org/10.1039/D0SC06195F>.