Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Virtual screening of plant-derived compounds against SARS-CoV-2 viral proteins using computational tools

María Antonela Zígolo a,b, Matías Rivero Goytia c,d, Hugo Ramiro Poma a, Verónica Beatriz Rajala e,f,⁎ Verónica Patricia Irazusta a,b

a Instituto de Investigaciones para la Industria Química (INIQUI), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) - Universidad Nacional de Salta (UNSa), Av. Bolivia 5150, 4400 Salta, Argentina
b Facultad de Ciencias Naturales, UNSa, Salta, Argentina
c Silentium Apps, Salta, Argentina
d Facultad de Economía y Administración, Universidad Cathólica de Salta (UCASAL), Salta, Argentina
e Facultad de Ingeniería, UNSa, Salta, Argentina
f Singapore Centre for Environmental Life Sciences Engineering (SCELSE), Nanyang Technological University, Singapore, Singapore

HIGHLIGHTS

• The viral spike (S) protein and the main protease Mpro were the molecular targets.
• Plant-derived compounds were assessed as potential SARS-CoV-2 inhibitors by docking.
• Synthetic antiviral drugs were used as controls for both proteins.
• Best candidates for S protein were glycyrrhizin, gitoxin, dicumarol, and diosgenin.
• Best candidates for Mpro: spirostan, N-(3-acetylglycyrrhetinoyl)-2-amino-propanol.

GRAPHICAL ABSTRACT

ABSTRACT

The new SARS-CoV-2, responsible for the COVID-19 pandemic, has been threatening public health worldwide for more than a year. The aim of this work was to evaluate compounds of natural origin, mainly from medicinal plants, as potential SARS-CoV-2 inhibitors through docking studies. The viral spike (S) glycoprotein and the main protease Mpro, involved in the recognition of virus by host cells and in viral replication, respectively, were the main molecular targets in this study. Molecular docking was performed using AutoDock, which allowed us to select the plant actives with the highest affinity towards the viral targets and to identify the interaction molecular sites with the SARS-CoV2 proteins. The best energy binding values for S protein were, in kcal/mol: −19.22 for glycyrrhizin, −17.84 for gitoxin, −12.05 for dicumarol, −10.75 for diosgenin, and −8.12 for delphinidin. For Mpro were, in kcal/mol: −9.36 for spirostan, −8.75 for N-(3-acetylglycyrrhetinoyl)-2-amino-propanol, −8.41 for α-amyrin, −8.35 for oleanane, −8.11 for taraxasterol, and −8.03 for glycyrrhetic acid. In addition, the synthetic drugs umifenovir, chloroquine, and hydroxychloroquine were used as controls for S protein, while atazanavir and nelfinavir were used for Mpro. Key hydrogen bonds and hydrophobic interactions between natural
1. Introduction

New strains of viruses are constantly emerging and represent a challenge to public health. Among the recent emerging viruses were the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) in 2003 influenza A (H1N1) virus in 2009, Middle East Respiratory Syndrome coronavirus (MERS-CoV) in 2012, and Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2), which causes the 2019-nCoV infection (COVID-19) that appeared at the end of 2019 in China (Gorbatenya et al., 2020). The disease started in Wuhan City, central China, mainly linked to workers from the Wuhan South China Wholesale Seafood Market, which sells, among other products, various types of live and dead exotic animals. After 14 months of the outbreak, the global acceleration in case incidence has slowed down, with around 3 million new cases reported; however, death rates continue to increase with over 88,000 new deaths reported last week (World Health Organization, 2020). As of February 9, there have been almost 105 million cases and over 2.3 million deaths reported globally since the beginning of the pandemic. These data demonstrate the urgent need to find quick solutions to the problem.

To date, very vaccines against SARS-CoV-2 have been developed showing good percentages of effectiveness in clinical trials; however, the world population has not been massively vaccinated and the period of immunity that these vaccines will generate is yet unknown (Detoc et al., 2020; Kaur & Gupta, 2020; Malik et al., 2020). On the other hand, it is important to mention that therapies with monoclonal antibodies proved to be effective as modulators of the immune response against SARS-CoV-2 (Jahanshahi & Rezaei, 2020; Pecetta et al., 2020; Zhao, 2020). Some monoclonal antibodies such as infliximab are already in the stage of clinical trials in patients with COVID-19 (Clinical Trials, 2020). Also, many synthetic antiviral molecules, already existing in the market were evaluated at different levels, many of which reached the stage of clinical trials. However, none of them showed to be effective

Compounds known for their effectiveness against other viruses are being investigated in search of a solution to the rapid and lethal advance of the new SARS-CoV-2. Examples of that are some protease inhibitors, already existing in the stage of clinical trials in patients with COVID-19 (Clinical Trials, 2020). Also, many synthetic antiviral molecules, already existing in the market were evaluated at different levels, many of which reached the stage of clinical trials. However, none of them showed to be effective enough against SARS-CoV-2 (Clinical (Clinical Trials, 2020; Fredisjaysah et al., 2021; Whang et al., 2020)). In this context, the search for new therapeutic agents is imperative as another possible alternative for the treatment of COVID-19; indeed, great effort and important resources for extensive experimental work are scarce. Considering that in the last months many SARS-CoV-2 macromolecules have been crystallized, more effective simulations may be performed cross-out if model proteins from other closely related viruses have to be used (Jin et al., 2020; Wrapp et al., 2020; Zhang et al., 2020).

In the initial stage of the SARS-CoV-2 replication cycle, the attachment of the virion to the host cell is initiated by interactions between the S protein (spike) and human angiotensin-converting enzyme (ACE2 receptor) (Hoffmann et al., 2020). The S protein–receptor interaction is the primary determinant for a coronavirus to infect a host species and governs the tissue tropism of the virus. Structurally, the coronavirus spike is a trimer arranged asymmetrically. Each monomeric unit has three segments: an extensive ectodomain, a single-pass transmembrane segment, and a small intracellular tail (Walls et al., 2016). The ectodomain contains two subunits; the S1 subunit allows binding to the ACE2 receptor while the S2 allows the fusion of viral and host membranes. Two major domains in S1, N-terminal domain (S1-NTD) and C-terminal domain (S1-CTD), have been identified. The S1-CTD, also known as the receptor-binding domain (RBD), is responsible for recognizing protein receptors ACE2 (Liu et al., 2015). The next step in the infection cycle is the translation of the replicase gene from the viral genome (positive polarity RNA). Replicate gene encodes two overlapping polypeptides, pp1a and pp1ab, required for viral replication and transcription (Baranov et al., 2005). Polypeptides pp1a and pp1ab contain the non-structural proteins (nsps) 1–11 and 1–16, respectively. These polypeptides are subsequently cleaved into the individual nsps (Ziebuhr et al., 2000). SARS-CoV-2 encodes two proteases that cleave the polypeptides; the papain-like proteases, PLpro (encoded by nsp3) and chymotrypsin-like cysteine protease known as 3C-like protease (3CLpro). The 3CLpro enzyme, also called main protease (Mpro), hydrolyses the polypeptide at no less than 11 conserved sites between nsp4 and nsp16, starting with the autolytic cleavage of this enzyme itself (Hegyi & Ziebuhr, 2002).

The aims of this work were to evaluate by molecular docking selected compounds obtained from natural sources for antiviral activity against SARS-CoV-2. Synthetic drugs that are currently being evaluated for the treatment of COVID-19 were used as controls. Two viral proteins with different and essential functions were tested: the SARS-CoV-2 main protease (Mpro) and the S protein (spike) in its prefusion conformation.
2. Materials and methods

2.1. Viral proteins

SARS-CoV-2 Mpro (PDB ID: 6LU7) and S protein (PDB ID: 6VSB) were used in this study. Three-dimensional structure are shown in Table 1.

2.2. Antiviral compounds

Secondary metabolites with different structures were analyzed to select the most promising to assess a possible relationship between structure and activity. Many of them already had proven antiviral activity against other viruses (Alonso & Desmarchelier, 2015; Cecil et al., 2011; Enkhtaivan et al., 2018; Shin et al., 2015).

Sixteen plant-derived compounds were assessed for the interaction with Mpro:

- α-amyрин (CID 73170)
- берберин (CID 2353)
- гераниал (CID 638011)
- гликерритиновая кислота (CID 10114)
- лапаченол (CID 174859)
- лапачол (CID 3884)
- лимонен (CID 22311)
- майтансин (CID 5281828)
- мелацин (Andrei et al., 1990; Hassan et al., 2015)
- ментол (CID 1254)
- рутин (CID 91434)
- олеанан (CID 9548717)
- санбунгин (CID 91434)
- сауроксин (CID 5462447)
- спиростан (CID 6857439)
- тараксестерол (CID 115250)
- теофyllин (CID 2153).

The semisynthetic compound analyzed was the N-(3-acetylglycyrrhetinoyl)-2-amino-propanol (Zígolo et al., 2018). The synthetic nelﬁnavir (CID 64143) and atazanavir (CID 148192), were used as controls for the interaction with Mpro.

Regarding the interaction with the S protein (PDB: 6VSB), seven natural compounds: алоин (CID 9866696), дигосгенин (CID 99474), амьдалгин (CID 34751), гитоксин (CID 91540), гликерритин (CID 14982), дельфинидин (CID 68245), and дикумарол (CID 54676038), and also three synthetic compounds: хлорохинон (CID 2719), гидроксихлорохинон (CID 3652), and умиеновир (CID 131411), were assessed.

2.3. Docking analysis

Docking calculations were carried out with Autodock 4.2 program, which uses Lamarckian genetic algorithm (LGA) for calculations (Morris et al., 2009). Number of genetic algorithm (GA) runs was set to 200 for each case analyzed and binding affinities were determined and reported in kcal/mol.

The Autodock 4.2 program was applied considering all rotatable bonds for ligands and the whole protein as a rigid structure. For the location and extent of the 3D area the search space was defined by specifying a center, the number of points at each dimension and the points between spaces to focus the search space in the enzyme active site or putative binding region.

For protease, the grid box center coordinates corresponded to the Cα atom of the histidine residue 163: x (-19.055), y (16.637), and z (64.100). Dimensions of the grid box were 90 × 90 × 90 and point spacing was 0.492 Å. For the S protein, the grid box center coordinates corresponded to the oxygen atom of the alanine 419 residue for chain B: x (232.020), y (245.760), and z (265.659). Dimensions of the grid box were 124 × 124 × 124 and point spacing was 0.869 Å.

Prior to docking Mpro, the inhibitor N3 and water molecules were removed from the protein structure. Polar hydrogen atoms were added, and Gasteiger atom charges were assigned to protein atoms. Other parameters were set to default values.

The 200 conformers found for each compound with the Autodock program were grouped in clusters that were ordered according to a ranking, which was used to evaluate such similarity was the residue mean quadratic root deviation (RMSD). The residues were obtained through the difference between the atom coordinates of a given conformer respect to the cluster to which the most stable conformer belongs. Each cluster grouped conformers with RMSD lower or equal to 2.0 Å.

For visualization of protein-ligand complexes the software Visual Molecular Dynamics 1.9.1 (VMD) was used (Theoretical and Computational Biophysics Group, University of Illinois).

2.4. Data processing using data mining and statistics

The frequency distribution (number in cluster) of plant-derived compounds were first plotted vs. Gibbs free energy of conformers
for a specific target or protein. Then, the best compounds for each protein were selected considering the following criteria for each compound: i) clusters with highest frequencies, ii) clusters with lowest Gibbs free energies, iii) conformational dispersion associated with the number of clusters (the higher the number of clusters was, there was more conformational dispersion for the interaction studied), iv) intra-cluster dispersion given by the variation around the mean binding energy, and v) intra-cluster dispersion given by the percentile of energy values close to the minimum binding energy (more negative free energy). Statistical software R (https://cran.r-project.org/) with the Rstudio IDE (https://rstudio.com/) was used for processing.

Different clusters were ordered according to the energy of the most stable conformer within each group: the ranking was ordered according to the increasing stability of the lowest energy conformers in each cluster.

3. Results

3.1. Selection of conformers

From the 200 docking runs for all the evaluated compounds, both from natural origin and the synthetic controls, the binding energies for each conformer were obtained which, in turn, were grouped into

![Cluster analysis of conformations](image)

**Fig. 1.** A) Cluster analysis of conformations. Frequency (number of conformers in clusters) of each compound after 200 runs considering the binding energy of conformations corresponding to the docking of α-amyrin, berberine, geranial, glycyrrhetic acid, lapachenole, lapachol, limonene, meitansine, meliacine, mentors, nelnavir, oleanone, sambunigrin, sauroxine, spirostan, taraxasterol, and theophylline with SARS-CoV-2 main protease (Mpro). B) Frequency (number of conformers in clusters) and binding energy of the most promising compounds regarding affinity with SARS-CoV-2 main protease (Mpro): α-amyrin, N-(3-acetylcyrrheticinoyl)-2-amino-propanol, oleanane, spirostan, and nelnavir and atazanavir as controls.
clusters. The binding energy for each conformer was plotted as a function of the frequencies (number of conformers in each cluster) for the Mpro protein (Fig. 1, A and B; S1A) and for S glycoprotein (Fig. 2; Fig. S2).

The interaction of all the natural compounds evaluated with each viral protein showed, in general, a low number of clusters although they were highly frequent, suggesting that there was a high probability for them to happen. Conversely, the synthetic antiviral nelﬁnavir and atazanavir used as controls for the Mpro protein and umifenovir, chloroquine, and hydroxychloroquine for the spike protein, showed broad conformational dispersion or high number of clusters (Figs. 1B and 2, S1A).

Each compound was represented by the most populated clusters, which in some cases were also those with the lowest binding energies. Then, statistical criteria were applied to select the best clusters when

---

**Table 2**

Compounds evaluated by docking with Mpro. Ranking; frequency; lowest and mean and standard deviation (SD) binding energies (ΔG); variation coefficient (CV); Lowest 1% ΔG and Lowest 1% perc, for selected clusters. Synthetic antivirals used as controls are in bold. Lowest 1% perc is an intra-cluster dispersion given by the 1% percentile of the Gibbs free energy close to the lowest binding energy.

| Compound                                      | Ranking | Frequency (Num in cls) | Lowest ΔG (kcal/mol) | Mean ΔG (kcal/mol) | SD ΔG (kcal/mol) | CV (%) | Lowest 1% ΔG (kcal/mol) | Lowest 1% perc |
|-----------------------------------------------|---------|------------------------|----------------------|-------------------|-----------------|--------|------------------------|----------------|
| α-Amyrin                                       | 1       | 125                    | −8.41                | −8.40             | 0.05            | 0.65   | −8.33                  | 96             |
| Atazanavir                                     | 2       | 13                     | −10.56               | −9.47             | 0.91            | 9.60   | −10.45                 | 15             |
| Berberine                                      | 1       | 66                     | −7.35                | −7.29             | 0.04            | 0.49   | −7.28                  | 80             |
| Geranial                                       | 1       | 108                    | −5.60                | −5.20             | 0.19            | 3.75   | −5.54                  | 5              |
| Glycyrrhetic acid acid                        | 1       | 46                     | −8.15                | −8.03             | 0.07            | 0.91   | −8.07                  | 43             |
| Lapachenone                                    | 1       | 79                     | −7.60                | −7.45             | 0.09            | 1.27   | −7.52                  | 18             |
| Lapachenone                                    | 3       | 77                     | −6.37                | −6.35             | 0.03            | 0.50   | −6.31                  | 96             |
| Lapachol                                       | 3       | 97                     | −6.91                | −6.71             | 0.12            | 1.83   | −6.84                  | 12             |
| Lapachol                                       | 1       | 61                     | −8.17                | −7.65             | 0.43            | 5.60   | −8.09                  | 8              |
| Limonene                                       | 1       | 141                    | −6.31                | −6.21             | 0.11            | 1.84   | −6.25                  | 55             |
| Meitansine                                     | 7       | 18                     | −6.23                | −5.46             | 0.40            | 7.24   | −6.17                  | 6              |
| Melacine                                       | 5       | 39                     | −7.36                | −7.05             | 0.10            | 1.27   | −7.29                  | 3              |
| Menthol                                        | 4       | 64                     | −5.38                | −5.19             | 0.13            | 2.55   | −5.33                  | 20             |
| Menthol                                        | 1       | 47                     | −6.55                | −6.33             | 0.35            | 5.54   | −6.48                  | 66             |
| N-(3-Acetylglycyrrhetinoyl)-2-amino-propanol    | 19      | 14                     | −7.58                | −7.38             | 0.22            | 2.97   | −7.50                  | 43             |
| N-(3-Acetylglycyrrhetinoyl)-2-amino-propanol    | 3       | 10                     | −8.37                | −7.68             | 0.54            | 7.06   | −8.29                  | 10             |
| N-(3-Acetylglycyrrhetinoyl)-2-amino-propanol    | 1       | 7                      | −8.75                | −8.23             | 0.43            | 5.27   | −8.66                  | 29             |
| Nelfinavir                                     | 1       | 6                      | −9.66                | −8.25             | 0.85            | 10.26  | −9.56                  | 17             |
| Oleanane                                       | 4       | 5                      | −7.83                | −7.18             | 0.41            | 5.66   | −7.75                  | 20             |
| Rutin                                          | 2       | 59                     | −8.35                | −8.34             | 0.01            | 0.10   | −8.27                  | 100            |
| Sambunigrin                                    | 6       | 2                      | −4.65                | −4.30             | 0.49            | 11.51  | −4.60                  | 50             |
| Saffronine                                     | 6       | 20                     | −5.33                | −4.59             | 0.48            | 10.54  | −5.30                  | 10             |
| Siprostan                                      | 4       | 71                     | −8.50                | −8.50             | 0.01            | 0.04   | −8.42                  | 100            |
| Siprostan                                      | 1       | 59                     | −9.36                | −9.35             | 0.01            | 0.08   | −9.27                  | 100            |
| Taraxasterol                                   | 6       | 72                     | −8.21                | −8.11             | 0.04            | 0.59   | −8.13                  | 38             |
| Theophylline                                   | 5       | 56                     | −4.69                | −4.48             | 0.01            | 0.11   | −4.45                  | 100            |
| Theophylline                                   | 1       | 32                     | −5.04                | −5.02             | 0.02            | 0.32   | −4.98                  | 97             |
more than one was considered for one compound, and the binding energies associated with their interactions with the viral proteins. Three different statistics were calculated: a) the standard deviation of the Gibbs free energy of all the cluster’s conformers (SD binding energy), b) the variation coefficient of the Gibbs free energy of all the cluster’s conformers (CV), and c) the intra-cluster dispersion given by the 1% percentile of the Gibbs free energy close to the lowest binding energy (Lowest 1% perc). These metrics were applied in order to assign to each compound the Gibbs free energy corresponding to the interaction with the viral proteins based on the following criteria: i) the Mean binding energy was assigned when CV was lower than or equal to 10% and the value of the Lowest 1% perc was not higher than 60%; ii) the Lowest binding energy was assigned when the Lowest 1% perc reached at least 60%, despite the CV value was lower than or equal to 10% (Table 2 for Mpro and Table 3 for S protein).

When two or more clusters were initially selected for the same compound (for having similar frequencies and binding energies), the following criteria were applied: i) if the clusters compared had Lowest 1% perc value equal to or greater than 60%, the one with the highest Lowest 1% perc value was chosen, and ii) when the clusters compared had a Lowest 1% perc value lower than 60%, the CVs were compared and that with the lowest CV value was chosen.

Following the statistical criteria for the docking results, the binding energies for the interactions of each compound with the corresponding protein were selected (Table 4).

When M<sup>pro</sup> protein interacted with natural compounds, similar values than those observed with the synthetic antivirals atazanavir and nelfinavir, were found. However, for synthetics control compounds greater dispersion and lower frequencies were observed (Fig. 1A and B; S1A). The strongest affinities for spirostan and for N-(3-acetylglycyrrhetinoyl)-2-amino-propanol, with ΔG of −9.36 and −8.75 kcal/mol, were obtained, respectively. Other compounds showed interesting affinities for M<sup>pro</sup>, for example: α-amyrin, oleanane, taraxasterol, and glycyrrhetic acid (Table 4).

On the other hand, the plant-derived natural compounds evaluated for S protein, yielded very high affinity values compared to controls, except for aloin (−5.94 kcal/mol). The highest affinities were observed for glycyrrhizin, gitoxin, dicumarol, diosgenin, and delphinidin, which also showed low dispersion and high frequencies (Fig. 2; S2). The most interesting results were observed for glycyrrhizin and gitoxin (−19.22 and −17.84 kcal/mol, respectively), which presented more than twice the affinity for the viral S protein compared to the synthetic antivirals umifenovir, chloroquine, and hydroxychloroquine (−7.47, −7.25, and −5.30 kcal/mol, respectively) (Table 4).

### 3.2. Analysis of the molecular interactions of candidate compounds with the proteins evaluated

Key interactions of possible candidate drugs with M<sup>pro</sup> were analyzed: α-amyrin, spirostan, taraxasterol, oleanane, lapachol, glycyrrhetic acid,
and N-(3-acyethylglycyrrhetinoyl)-2-amino-propanol (Table 5). All the natural and semisynthetic active compounds bound to the enzyme active site, as well as the controls nelfinavir and atazanavir. Only interactions of nelfinavir with Mpro with lower binding energy were shown (Table 5). The semi-synthetic compound N-(3-acyethylglycyrrhetinoyl)-2-amino-propanol presented five hydrogen bond interactions, four with the residues of the active site and one with the tyrosine residue 54. It also showed hydrophobic interactions with six residues of the protease in the vicinity of the catalytic site. All these interactions increased the binding affinity with respect to the control compound, which presented only two hydrogen bond interactions with the active site residues and hydrophobic interactions with other four protease residues. Glycyrrhetic acid, a precursor of the above-mentioned semisynthetic compound, had four hydrogen bond interactions with catalytic residues and hydrophobic interactions with leucine residue 4.

It is important to note that five of the six candidate compounds (Table 5) have in common pentacyclic triterpene structures: taraxasterol, N-(3-acyethylglycyrrhetinoyl)-2-amino-propanol, glycyrrhetic acid, α-amyrin, and oleanane. All of them but oleanane presented hydrogen bond interactions; however, its high affinity, comparable with the compounds that showed hydrogen bond interactions highlighted the importance of hydrophobic interactions that could help stabilize oleanane binding with the main protease.

Spirostan interacted with the Mpro mainly by hydrophobic interactions, presenting only one hydrogen bond interaction with the serine 46 catalytic residue. Similarly, taraxasterol only had one hydrogen bond interaction.

In addition to glycyrrhizin and gitoxin, the interaction of dicumarol with the spike protein showed one of the most favorable Gibbs free energies (−12.05 kcal/mol) (Table 6). Glycyrrhizin, with the highest affinity for the glycoprotein, interacted with a great number of residues from A and C chains. There was a total of four hydrogen bonds with residues of the A chain (PHE855, VAL976 and ASP979), and another two with the B chain (ASP571 and ASP574). Hydrophobic interactions occurred with the remaining of the 14 residues.

4. Discussion

Plants have a long evolutionary history of developing resistance against viruses. This adaptation, together with the ability to produce secondary metabolities, has generated a wide range of possible sources for antiviral drugs. Herbal infusions and essential oils from plants have been used ancestrally in different civilizations. Their low toxicity and minimal adverse effects at the doses used have long been known, but with remarkably interesting biological activity (Krajcovicova et al., 2004; Zígolo et al., 2018). Further, it would be also important to evaluate more compounds of plant origin with common steroid structures due to the high value found in this work for spirostan. Other compounds in this family with steroidal structures and potent antiviral activity have already been reported against poliovirus, measles virus, influenza virus, herpes simplex virus, human Immunodeficiency virus, and other viral diseases (Kalipaperumal et al., 2020).

Hydrophobic interactions played a particularly important role for the Mpro protein. In the case of oleanane, which showed a good binding energy value, the only type of interactions that presented were those of hydrophobic nature with the protease residues. Furthermore, for all the other compounds evaluated, despite forming hydrogen bonds there were also a significant number of hydrophobic interactions with protease residues. Khaerunnisa et al. (Khaerunnisa et al., 2020) also showed the importance of hydrophobic interactions for a series of compounds of plant origin in the interaction with the viral protease Mpro. It is known that spirostan is the most abundant steroidal saponin, being the triterpene and steroid glycosides of important biological and medicinal applications (Pérez-Labrada et al., 2012). Several of these molecules have exhibited potent antiviral activities, making them promising candidates for drug development according to the results of this work and other studies (Cecil et al., 2011; Enkhaïtaïv et al., 2018; Shin et al., 2015).

In the molecular docking for Mpro protein of SARS-CoV-2, the semisynthetic compound N-(3-acyethylglycyrrhetinoyl)-2-amino-propanol performed better than glycyrrhizin acid. The former, obtained by biocatalytic reaction of aminolysis from glycyrrhetic acid (derived from the plant Glycyrrhiza glabra L.), already showed powerful antiviral activity against the herpes virus types I and II, including strains resistant to acyclovir (De Clercq, 2000; Zígolo et al., 2018). In addition, Zígolo et al. (Zígolo et al., 2018) showed low toxicity of the N-(3-acyethylglycyrrhetinoyl)-2-amino-propanol in Vero cell cultures, making this compound a promising candidate for COVID-19 treatment. In a molecular docking study of natural products that are commonly used in Indian and other cuisines against SARS-CoV-2 main protease (PDB: 6LU7 and PDB: 6Y2E), it was shown that several of these compounds had high binding affinity against Mpro, and they were comparable with a known anti-HIV drug, saquinavir. In our study, we report plant-derived active compounds with much higher binding affinities than other natural compounds reported as possible antivirals for SARS-CoV-2. Glycyrrhizin (binding energy of −8.40 kcal/mol) was one of the best potential candidates assessed. It is remarkable that a diammonium salt of glycyrrhizin was recommended for the cure of COVID-19 in China. The binding affinity of the positive control, saquinavir, was −7.90 kcal/mol (Sampangi-Ramaiah et al., 2020).

In a recent study, several compounds, mainly flavonoids (kaempferol, quercetin, luteolin-7-glucoside, demethoxycurcumin, naringenin, apigenin-7-glucoside, oleuropein, curcumin, catechin,
epicatechingallate, zingerol, gingerol, and allicin) and the synthetic compounds nelfinavir and lopinavir were analyzed as potential inhibitors of the SARS-CoV-2 M\textsuperscript{pro} (PDB: 6LU7) by docking studies (Khaerunnisa et al., 2020). In our hands, the binding energy found (−8.25 kcal/mol) for the interaction of nelfinavir with M\textsuperscript{pro} protein was similar to that reported by Khaerunnisa et al. (Khaerunnisa et al., 2020).

Table 5
Three-dimensional (3D) molecular interaction of candidate compounds with M\textsuperscript{pro}.

| Compound         | 3D Interactions | Hydrogen bond donor/acceptor (black dotted lines, indicated by red arrows) | Hydrophobic interactions (spheres) |
|------------------|-----------------|---------------------------------------------------------------------------|------------------------------------|
| Nelfinavir       |                 | GLU166 (grey)                                                             | PRO168 (red)                       |
| \(C_{32}H_{43}N_{3}O_{5}S\) |                 | HIS172 (green)                                                            | HIS141 (blue)                      |
| \(C_{30}H_{59}O_{3}\) |                 |                                                                           | HIS41 (light blue)                 |
| \(\alpha\)-Amyrin |                 |                                                                           | MET49 (yellow)                     |
| \(C_{30}H_{59}O_{3}\) |                 | THR26 (violet)                                                            | ASP48 (black)                      |
| Spirostan        |                 |                                                                           | CYS44 (green)                      |
| \(C_{27}H_{44}O_{2}\) |                 |                                                                           | HIS41 (light blue)                 |

In our studies, we also analyzed the interactions of the synthetic compounds nelfinavir and lopinavir with M\textsuperscript{pro} protein. The table above summarizes the hydrogen bond and hydrophobic interactions observed for each compound.
Interestingly, despite the similarity in the ΔG obtained in both studies, differences in the hydrogen bonds formed were observed (Khaerunnisa et al., 2020). While residues Glu166, Gln189, and Gln192 were used according to Khaerunnisa et al. (Khaerunnisa et al., 2020), we only found Glu166 and His172 in the interaction between nefinavir and Mpro protein by docking (Khaerunnisa et al., 2020). Differences observed would be consequences of the parameters considered by different authors, for example, the rotatable links in the coupling. Furthermore, this difference may also be due to the selection of different interaction conformations made by the authors. In this work, rigorous statistical criteria to make the best decision in choosing the correct cluster (detailed in Section 3.1), were applied. In this sense, nefinavir presented a broad conformational dispersion (number of clusters) and similar frequencies in the clusters with better binding energies.

All of the compounds evaluated against S glycoprotein showed very favorable binding energies, higher than those for the synthetic antiviral compounds, umifenovir (−7.47 kcal/mol), chloroquine (−7.25 kcal/mol), and hydroxychloroquine (−5.30 kcal/mol), directed to that viral target. The most favorable binding energies were obtained for glycyrrhizin (−19.22 kcal/mol), gitoxin (−17.84 kcal/mol), dicumarol (−12.05 kcal/mol), diosgenin (−10.75 kcal/mol), and delphinidin (−8.12 kcal/mol). These results are very promising compared to those obtained in recent works, where the coupling of natural and synthetic compounds with viral glycoprotein S, were analyzed. Thus, the best binding energies obtained for S protein were: −11.55 kcal/mol to coenzyme A, −11.089 kcal/mol to flavin adenine dinucleotide, and −9.36 kcal/mol to tiludronate (Hall & Ji, 2020).

In agreement with the low Gibbs free energy, numerous molecular interactions were observed for the glycyrrhizin: six hydrogen bonds and many hydrophobic interactions, all of them with residues from chains A and C from the glycoprotein. This is despite no interactions with the RBD region (recognition of ACE human receptors), chain B, were observed. Therefore, glycyrrhizin would be a good candidate to be evaluated experimentally to learn how it would affect the glycoprotein activity. Interestingly, a dietary supplement of Glycyrrhiza glabra L. (SoriaNatural®) is already in the market demonstrating the safety of consuming this compound since it has already been approved by the corresponding organizations.

Gitoxin is a secondary cardiac glycoside from Digitalis purpurea and Digitalis lanata. This compound and others from Digitalis sp. have demonstrated action on heart muscle (Haustein et al., 1970). The aglycone that is the active part of the compound has a cyclopentanepreydrophenanthrene structure, which is common in steroids. This hydrophobic portion of the molecule established many hydrophobic interactions with the spike protein residues. The glycone portion of the gitoxin interacted with two N-acetyl-α-glucosamine residues from spike protein. It also formed hydrogen
bonds. Many of these interactions of different nature were established in the RBD region of S glycoprotein. For this reason, we consider that it is the most promising compound with potential to inhibit spike protein.

Dicumarol is a derivative from cumarin and was successfully used as an inhibitor against influenza A (An et al., 2014). Moreover, it could be a particularly good candidate for inhibiting the S glycoprotein by blocking the RBD region, which contains human ACE2 receptors, as the compound interacted with several residues in that region. One important aspect to consider regarding dicumarol and other cumarins, is that they have already been approved in the pharmaceutical market for their effectiveness as: antidiabetic (Dwivedi et al., 2008), antiviral (Yu et al., 2003), anti-inflammatory (Kontogiorgis & Hadjipavlou-Litina, 2004), and anticancer and antileukemia activities (Stefanachi et al., 2011; Yang et al., 2011); thus, they are considered safe compounds for human oral consumption. Considering that dicumarol and coumarins have powerful anticoagulant effects, their use in certain patients must be under medical control. In fact, being derived from medicinal plants with ancestral use proven for oral consumption is an important advantage of some natural origin compounds. The medicinal use of phytocomplexes allows not only increasing the desired action of the main active compounds but also decreasing adverse effects than if any of the pure active compounds were used for the same treatment. Phytocomplexes are plant extracts or essential oils that contain more than one main active compounds and other substances that can help decreasing the adverse effects of the main compounds, contributing to the better absorption and solubility of the main compounds. This important advantage would facilitate the production and reduce the cost because fewer downstream operations would be needed as there is no need to obtain the high purity product (Cortés Gallardo et al., 2004; Duraffourd et al., 1987; Hostettmann & Marston, 1995; Pellecuer, 1993; Rawat et al., 2020; Sharapin et al., 2000).

5. Conclusions

COVID-19 is a global health problem. Despite all the efforts dedicated to stop its pandemic dissemination there is no conclusive treatment to date.

Based on the binding energies obtained and on the analysis of the frequency and dispersion of clusters, interaction conformations between the proposed natural compounds and Mpro and spike proteins are highly possible which means an effective and strong bond. This could translate into a high possibility that these active compounds are efficient inhibitors of viral proteins, which must be corroborated experimentally. In addition, the tested natural compounds performed better than the synthetic antivirals used as controls: nelfinavir and atazanavir, for protease Mpro and umifenovir, chloroquine and hydroxichloroquine for S protein. Thus, the compounds evaluated as inhibitors for the viral
Table 6
Three-dimensional (3D) molecular interaction of candidate compounds with S protein.

| Compound   | 3D interactions | Interactions         |
|------------|-----------------|----------------------|
| Glycyrrhizin | Chain C         | LEU546               |
| C_{6}H_{12}O_{16} |              | THR547               |
|            |                 | GLY548               |
|            |                 | THR549               |
|            |                 | ILE587               |
|            |                 | THR588               |
|            |                 | PRO589               |
|            |                 | THR49                |
|            |                 | ASP571 (H-bond 3.00 Å) |
|            |                 | THR572               |
|            |                 | THR573               |
|            |                 | ASP574 (H-bond 2.79 Å) |
|            |                 | ARG567               |
|            |                 | ASP568               |
| Gitoxin    | Chain A         | PHE855 (2H-bond 2.29/3.10 Å) |
| C_{6}H_{12}O_{14} |           | ASN856               |
|            |                 | VAL976 (H-bond 3.11 Å) |
|            |                 | ASN978               |
|            |                 | ASP979 (H-bond 2.20 Å) |
|            | Chain B (RBD domain) | ARG466             |
|            |                 | PHE464 (H-bond 3.50 Å) |
|            |                 | ARG355               |
|            |                 | LYS356               |
|            |                 | TYR396 (H-bond 2.75 Å) |
|            |                 | ARG357               |
|            |                 | LYS462 (H-bond 2.64 Å) |
|            | Chain C         | PRO230               |
|            |                 | TYR200               |
|            |                 | THR167               |
|            |                 | PHE168               |
|            |                 | ASP198 (2H-bond 3.10/3.15 Å) |
|            |                 | ILE231               |
|            |                 | GLY199               |
The S protein residues that interacted by hydrogen bonds are represented with lines and with balls and sticks those that interacted hydrophobically with the compounds.
promote and glycoprotein S could be potential active agents for treatment of COVID-19, either alone or as adjunct therapies with other medications demonstrated have antiviral activity in vivo.

This study also contains valuable information to increase the knowledge of certain structural families of compounds with great affinities by Mpro protease and SARS-CoV-2 spike proteins. Furthermore, the present study provides molecular details that set the basis to further propose structural modifications of some compounds to make the interaction between them and proteins even more effective.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.146400.

CRediT authorship contribution statement

María Antonela Zígolo: Conceptualization, Methodology, Formal analysis, Writing — original draft. Matías Rivero Goytia: Formal analysis, Data curation. Hugo Ramiro Poma: Methodology, Writing — review & editing. Verónica Beatriz Rajal: Visualization, Writing — review & editing, Supervision. Verónica Patricia Irazust: Visualization, Writing — review & editing, Supervision.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by Project #2361 from Consejo de Investigaciones de la Universidad Nacional de Salta (CIUNSA, Argentina). María Antonela Zígolo is a recipient of a postdoctoral fellowship from Investigaciones de la Universidad Nacional de Salta (CIUNSa, Argentina).

The authors would like to thank Dr. Jerold Last for his kind help in the preparation of the English manuscript.

References

Abbas Zaidi, S., Jameel, S., Baric, R.S., Groot, R.J.D., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, S.W., Peng, D., Perlman, S., Poon, L.L.M., Samborskyi, D.V., Sidorov, L.A., Sola, I., Ziebuhr, J. 2020. The species severe acute respiratory syndrome-related coronavirus. Nature https://doi.org/10.1038/s41586-020-2223-y.

Kalyaperumal, S., Periyasamy, K., Balakrishnan, U., Panalivel, P., Eghbala, C., 2020. Chapter 15 - antiviral phytochemicals for drug development: a data mining studies. In: Eghbala, C., Kumar, S., Ilbene, J.C., Ezzat, S.M., Kalyaperumal, S. (Eds.), Phytochemicals as Lead Compounds for New Drug Discovery. Elsevier, Amsterdam, pp. 239–244.

Kaur, S.P., Gupta, V., 2020. COVID-19 Vaccine: a comprehensive status report. Virus Res. https://doi.org/10.1016/j.virusres.2020.104111.

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., Soetjipto, S., 2020. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. https://doi.org/10.20944/preprints202003.0226.v1.

Kontogiorgis, C.A., Hadjadjov-Lutina, D.J., 2004. Synthesis and biological evaluation of novel coumarin derivatives with a 7-azomethine linkage. published correction appears in, Bioorg. Med. Chem. 14 (3), 611–614.

Krajčovec, Z., Krajčová, A., Vychádková, K., 2004. Horváthová. Taraxasterol and 2700. 13. Kaliyaperumal, S., Periyasamy, K., Balakrishnan, U., Panalivel, P., Eghbala, C., 2020. Chapter 15 - antiviral phytochemicals for drug development: a data mining studies. In: Eghbala, C., Kumar, S., Ilbene, J.C., Ezzat, S.M., Kalyaperumal, S. (Eds.), Phytochemicals as Lead Compounds for New Drug Discovery. Elsevier, Amsterdam, pp. 239–244.

Kaur, S.P., Gupta, V., 2020. COVID-19 Vaccine: a comprehensive status report. Virus Res. https://doi.org/10.1016/j.virusres.2020.104111.

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., Soetjipto, S., 2020. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. https://doi.org/10.20944/preprints202003.0226.v1.

Kontogiorgis, C.A., Hadjadjov-Lutina, D.J., 2004. Synthesis and biological evaluation of novel coumarin derivatives with a 7-azomethine linkage. published correction appears in, Bioorg. Med. Chem. 14 (3), 611–614.

Krajčovec, Z., Krajčová, A., Vychádková, K., 2004. Horváthová. Taraxasterol and 2700. 13. Kaliyaperumal, S., Periyasamy, K., Balakrishnan, U., Panalivel, P., Eghbala, C., 2020. Chapter 15 - antiviral phytochemicals for drug development: a data mining studies. In: Eghbala, C., Kumar, S., Ilbene, J.C., Ezzat, S.M., Kalyaperumal, S. (Eds.), Phytochemicals as Lead Compounds for New Drug Discovery. Elsevier, Amsterdam, pp. 239–244.

Kaur, S.P., Gupta, V., 2020. COVID-19 Vaccine: a comprehensive status report. Virus Res. https://doi.org/10.1016/j.virusres.2020.104111.

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., Soetjipto, S., 2020. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. https://doi.org/10.20944/preprints202003.0226.v1.

Kontogiorgis, C.A., Hadjadjov-Lutina, D.J., 2004. Synthesis and biological evaluation of novel coumarin derivatives with a 7-azomethine linkage. published correction appears in, Bioorg. Med. Chem. 14 (3), 611–614.

Krajčovec, Z., Krajčová, A., Vychádková, K., 2004. Horváthová. Taraxasterol and 2700. 13. Kaliyaperumal, S., Periyasamy, K., Balakrishnan, U., Panalivel, P., Eghbala, C., 2020. Chapter 15 - antiviral phytochemicals for drug development: a data mining studies. In: Eghbala, C., Kumar, S., Ilbene, J.C., Ezzat, S.M., Kalyaperumal, S. (Eds.), Phytochemicals as Lead Compounds for New Drug Discovery. Elsevier, Amsterdam, pp. 239–244.

Kaur, S.P., Gupta, V., 2020. COVID-19 Vaccine: a comprehensive status report. Virus Res. https://doi.org/10.1016/j.virusres.2020.104111.

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., Soetjipto, S., 2020. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. https://doi.org/10.20944/preprints202003.0226.v1.

Kontogiorgis, C.A., Hadjadjov-Lutina, D.J., 2004. Synthesis and biological evaluation of novel coumarin derivatives with a 7-azomethine linkage. published correction appears in, Bioorg. Med. Chem. 14 (3), 611–614.

Krajčovec, Z., Krajčová, A., Vychádková, K., 2004. Horváthová. Taraxasterol and 2700. 13. Kaliyaperumal, S., Periyasamy, K., Balakrishnan, U., Panalivel, P., Eghbala, C., 2020. Chapter 15 - antiviral phytochemicals for drug development: a data mining studies. In: Eghbala, C., Kumar, S., Ilbene, J.C., Ezzat, S.M., Kalyaperumal, S. (Eds.), Phytochemicals as Lead Compounds for New Drug Discovery. Elsevier, Amsterdam, pp. 239–244.

Kaur, S.P., Gupta, V., 2020. COVID-19 Vaccine: a comprehensive status report. Virus Res. https://doi.org/10.1016/j.virusres.2020.104111.

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., Soetjipto, S., 2020. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. https://doi.org/10.20944/preprints202003.0226.v1.

Kontogiorgis, C.A., Hadjadjov-Lutina, D.J., 2004. Synthesis and biological evaluation of novel coumarin derivatives with a 7-azomethine linkage. published correction appears in, Bioorg. Med. Chem. 14 (3), 611–614.

Krajčovec, Z., Krajčová, A., Vychádková, K., 2004. Horváthová. Taraxasterol and 2700. 13. Kaliyaperumal, S., Periyasamy, K., Balakrishnan, U., Panalivel, P., Eghbala, C., 2020. Chapter 15 - antiviral phytochemicals for drug development: a data mining studies. In: Eghbala, C., Kumar, S., Ilbene, J.C., Ezzat, S.M., Kalyaperumal, S. (Eds.), Phytochemicals as Lead Compounds for New Drug Discovery. Elsevier, Amsterdam, pp. 239–244.

Kaur, S.P., Gupta, V., 2020. COVID-19 Vaccine: a comprehensive status report. Virus Res. https://doi.org/10.1016/j.virusres.2020.104111.

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., Soetjipto, S., 2020. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. https://doi.org/10.20944/preprints202003.0226.v1.
Sampangi-Ramaiah, M.H., Vishwakarma, R., Shaanker, U., 2020. Molecular docking analysis of selected natural products from plants for inhibition of SARS-CoV-2 main protease. Curr. Sci. 118 (7), 1087–1092.

Sharapin, N., Machado, L., Souza, E., de Albuquerque, E., Valverde, E., López, J.M., 2000. Fundamentos de tecnología de productos fitoterapéuticos. Santa Fe de Bogotá: Convención Andrés Bello y Red Iberoamericana de Productos Fitofarmacéuticos (RIPROFITA) del Subprograma X de CYTED, p. 17.

Shin, H.B., Choi, M.S., Yi, C.M., Lee, J., Kim, N.J., Inn, K.S., 2015. Inhibition of respiratory syncytial virus replication and virus-induced p38 kinase activity by berberine. Int. Immunopharmacol. 27, 65–68.

Stefanachi, A., Favia, A.D., Nicolotti, O., et al., 2011. Design, synthesis, and biological evaluation of imidazolyl derivatives of 4, 7-disubstituted coumarins as aromatase inhibitors selective over 17-α-hydroxylase/C17-20 lyase. J. Med. Chem. 54 (6), 1613–1625.

Walls, A.C., Tortorici, M.A., Bosch, B.J., Frenz, B., Rottier, P.J., et al., 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature 531, 114–117.

Yang, L., Wen, K.S., Ruan, X., Zhao, Y.X., Wei, F., Wang, Q., 2018. Response of plant secondary metabolites to environmental factors. Molecules 23 (4), 1–26.

Yu, D., Suzuki, M., Xie, L., Morris-Natschke, S.L., Lee, K.H., 2003. Recent progress in the development of coumarin derivatives as potent anti-HIV agents. Med. Res. Rev. 23 (3), 322–345.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., Hilgenfeld, R., 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Nature https://doi.org/10.1038/s41586-020-2223.

Zhao, M., 2020. Cytokine storm and immunomodulatory therapy in COVID-19: Role of chloroquine and anti-IL-6 monoclonal antibodies. Int. J. Antimicrob. Agents 55, 105982. https://doi.org/10.1016/j.ijantimicag.2020.105982.