Unlocking the potential of phytochemicals in inhibiting SARS-CoV-2 M Pro protein - An in-silico and cell-based approach

Khushboo Singh (khushboo.singh@amway.com)
Amway (United States)
J. J. Patten
Boston University
Andrea Dimet
The University of Texas Medical Branch at Galveston
Robert A. Davey
Boston University
Stanley J. Watowich
The University of Texas Medical Branch at Galveston
Amit Chandra
Amway (United States)
Jesse Leverett
Amway (United States)

Article

Keywords:

Posted Date: January 29th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3888947/v1

Additional Declarations: No competing interests reported.

Abstract

The main protease (M^{Pro}) of SARS-CoV-2 plays a crucial role in viral replication and is a prime target for therapeutic interventions. Phytochemicals, known for their antiviral properties, have been previously identified as potential M^{Pro} inhibitors in several in silico studies. However, the efficacy of these remains in question owing to the inherent flexibility of the M^{Pro} binding site, posing challenges in selecting suitable protein structures for virtual screening. In this study, we conducted an extensive analysis of the M^{Pro} binding pocket, utilizing molecular dynamics (MD) simulations to explore its conformational diversity. Based on pocket volume and shape-based clustering, five representative protein conformations were selected for virtual screening. Virtual screening of a library of ~ 48,000 phytochemicals suggested 39 phytochemicals as potential M^{Pro} inhibitors. Based on subsequent MM-GBSA binding energy calculations and ADMET property predictions, five compounds were advanced to cell-based viral replication inhibition assays, with three compounds (demethoxycurcumin, shikonin, and withaferin A) exhibiting significant (EC50 < 10 uM) inhibition of SARS-CoV-2 replication. Our study provides an understanding of the binding interactions between these phytochemicals and M^{Pro}, contributing significantly to the identification of promising M^{Pro} inhibitors. Furthermore, beyond its impact on therapeutic development against SARS-CoV-2, this research highlights a crucial role of proper nutrition in the fight against viral infections.

Introduction

As the world marks four years since the onset of the COVID-19 pandemic, we continue to grapple with the devastating impact of this global health crisis caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first reported in Wuhan, China, in late 2019[1,2] As per the latest data from World Health Organization, (https://data.who.int/dashboards/covid19/cases?n=c, as of 22 January 2024), the global death toll from COVID-19 has now exceeded 7 million, and the disease continues to claim hundreds of lives daily. The onset of 2024 marked a deadly surge in COVID-19 cases, with over 1 million new cases and above 8000 deaths reported in the past month alone, highlighting the persistent threat posed by the virus. Although there are several highly efficient vaccines authorized for SARS-CoV-2 in the USA and other nations, the number of approved small-molecule COVID-19 drugs remain limited to remdesivir (e.g., Veklury), baricitinib (e.g., Olumiant), and nirmatrelvir/ritonavir (e.g., Paxlovid).

In the pursuit of effective drug development strategies against SARS-CoV-2, researchers are exploring a diverse range of approaches, including investigating plant-derived compounds that have a rich history of use in traditional medicine for combating viral infections[5,6,7,8,9,10]. Phytochemicals, chemicals synthesized by plants demonstrate vast scaffold diversity and structural complexity, which enable these molecules to interact with high specificity to a wide range of biological macromolecules. Unique chemical features in phytochemicals can provide distinct advantages for their effectiveness as interventions. For instance, high rigidity due to a large number of chiral centers, often enable phytochemicals to disrupt protein-protein interactions[11]. However, identifying the bioactive component within plant extracts is a cumbersome process. Usually, plant-derived natural product screens involve a library of botanical extracts containing complex mixtures of molecules; this complexity is often unsuitable for traditional target-based efficacy studies[12]. Recognizing this challenge, application of cheminformatics has emerged as a highly promising strategy. Computational techniques offer a powerful toolset for identifying potential molecular targets, synergistic effects, partial agonist/antagonist activities, off-target interactions, and the intricate mechanisms[13] of action associated with phytochemicals in extracts. Thereby, adding a layer of convenience in facilitating experimental validation and potentially streamlining the process of discovering novel therapeutics.

The application of cheminformatics to pinpoint active phytochemicals within complex botanical blends hold immense potential for advancing nutritional product development. This newfound insight could guide designing extraction strategies to ensure the presence and abundance of key bioactive markers within the final extracts. Large phytochemical databases, combined with cutting-edge computational methods, offer a robust toolkit for achieving this, ultimately fine-tuning the production of nutritionally efficacious products.

The protein M^{Pro} is a homodimer of two protomers, with each protomer comprising domain I (residues 8-101), domain II (residues 102–184), and domain III (residues 201–303)[17,18,19,20,21] (Fig. 1). Domains I and II form an active site consisting of S1', S1, S2, and S4 subsites[22]. The catalytic residues His41 and Cys145 form a portion of S2 and S1 pockets, respectively. His41 forms a crucial hydrogen bond with a water molecule, facilitating interactions with the side chains of Asp187 and His164[23]. Asp187 is further stabilized through a salt-bridge to Arg40[23]. This intricate arrangement allows His41 to function as a base, extracting a proton from the catalytic Cys145 side chain and activating it for a nucleophilic attack and cleavage of the overlapping polyproteins pp1a and pp1ab into 16 non-structural proteins. This cleavage releases the functional polypeptide for viral replication and transcription[15,16].

The pursuit of M^{Pro} inhibitors, whether synthetic or plant-derived, as potential antiviral agents against the virus, has been the focus of extensive research. One of the primary challenges in targeting M^{Pro} stems from the inherent plasticity of its active site. Molecular dynamics (MD) simulations of SARS-CoV-2 and SARS-CoV M^{Pro} proteins have identified differences in shape and size[24,25,] of the binding site, despite the active

site residues being identical. This conformational flexibility arises largely from 12 distant residues that differ between SARS-CoV-2 M^{Pro} and SARS-CoV M^{Pro} [26] ; SARS-CoV-2 M^{Pro} exhibit increased structural flexibility and plasticity[27]. These remote residues play a pivotal role in the dynamic plasticity of the binding site, exerting substantial influence on inhibitor binding affinity and specificity. Thus, delineating the conformational changes of the binding site is important for selecting suitable target structures for virtual screening. Therefore, prior to virtual screening, we sampled ensembles of binding site pockets produced from MD simulations and compared pocket volumes and shapes. These comparisons identified novel, pharmacologically relevant binding-pocket conformations (that may not be readily apparent in experimental structures) for virtual screening. In this study, structure-based virtual screening of phytochemical libraries against diverse M^{Pro} binding site pockets was employed to identify potential phytochemical inhibitors. We utilized the CMAUP[14] database containing ~ 48K small-molecule phytochemicals for screening, which (to the best of our knowledge) is largest library of plant-derived molecules to be screened for M^{Pro} inhibitors to date.

Results

Stability Analysis of M^{Pro} Protein during MD Simulation

We measured the stability of the M^{Pro} protein by calculating the RMSD (Root Mean Square Deviation) of the Cα atoms during MD simulations. As shown in Fig. 2A, the RMSD plot demonstrated high stability of the protein with deviations ranging between 0.05 and 0.08 nm. The RMSF (Root Mean Square Fluctuations) analyses of the backbone, backbone + Cβ, and sidechain atoms of M^{Pro} protein was subsequently performed to reveal the flexibility of the residues during simulations. As shown in Fig. 2B, a significant difference in RMSF values was observed between the backbone and sidechain atoms, as well as between the backbone + Cβ and sidechain atoms. The sidechain atoms showed fluctuations between 0.04 and 0.20 nm, while the backbone and backbone + Cβ atoms displayed more limited fluctuations in the range of 0.03 to 0.10 nm. Among the subsites forming residues, relatively higher fluctuations were observed in the sidechain atoms of Thr45, Asn119, Asn142, Arg188, and Gln189, which exhibited RMSF values in the range of 0.15 nm to 0.19 nm. The highest fluctuations of the sidechain chain atoms were observed for Thr45 forming the S2 site with RMSF value of 0.19 nm. Asn142, Arg188, and Gln189 have previously been reported[28,29,30] to play a crucial role in ligand interactions. The high flexibility of these residues could allow the binding site to adapt to different ligand structures and sizes, facilitating the recognition and binding of various inhibitors.

Pocket Dynamics Analysis

To optimize the selection of protein conformations for virtual screening, we employed pocket-based clustering of the ensembles obtained from simulations. POVME[31,32] was utilized to calculate the binding site volumes of 60 distinct frames of M^{Pro} were obtained from each simulation. These frames were sampled at 5 ns intervals during the 300 ns simulation. Active site pocket volumes ranged from 190 to 498 $Å^3$, with surface areas varying between 188 to 380 $Å^2$ during the MD simulations. The active site volumes in our study were slightly greater than previously reported[33,34] due to the widening of subsite pockets and the formation of new pockets in the active site.

The POVME clustering workflow classified the binding sites within the sampled frames into five clusters, each representing frequently observed pocket shapes. The representative structure from each cluster is shown in Fig. 3. Differences between the binding site shape in each cluster originate from the opening or closing of regions that constitute the active site. For Clusters 1 through 4, a gradual widening of the S1 pocket is observed; however, this widening coincides with a concurrent reduction in the pocket's depth, resulting in a shallower S1 pocket configuration (shown in Fig. 3A-D). In Cluster 5, the shape and size of the S1 subsite is decreased significantly compared to the other clusters (shown in Fig. 3E). The S2 subsite, on the other hand, maintained its small size in all conformations, with a portion of it being deeply embedded and thus inaccessible for ligand binding. New subsites, termed the S2’ subsite, emerged adjacent to the S1’ subsite in structures corresponding to Clusters 1, 3, 4, and 5 (shown in Fig. 3A,C-E). The S1’ pocket showed the least divergence when compared to the average pocket geometry by retaining a similar shape in Clusters 1–3 but was notably diminished in size and volume in the structures of Cluster 4 and 5. Whereas S4 exhibited the most pronounced dissimilarity in shape across all five clusters.

Molecular Docking

The CMAUP database[14] was screened to identify ligands capable of binding to each of the five active site conformations of M^{Pro}. We identified nearly 500 phytochemicals with docking scores above − 7.5 kcal/mol across five protein conformations. The criteria for the docking score cutoff was based on the docking scores observed for reported small-molecule M^{Pro} inhibitors[35,36,37]. Our analysis revealed a preference for the binding site within Cluster 3, as approximately 200 compounds bound to this conformation with scores above the set threshold. In contrast, only 25 compounds docked to the active site of Cluster 5 with docking scores above the threshold.

Table 1 lists the 20 phytochemicals with high docking scores across all five conformations. Additional details of the phytochemicals, including structure, PubChem ID, and the plant sources are listed in Supplementary Table 1. Of these 20 phytochemicals, four were identified as aglycones, while the remaining compounds were glycosides. When discussing phytochemical, the aglycone is of particular interest because it typically holds the pharmacological or biological activity. Whereas the glycosidic bond could affect the compound's solubility, stability, and bioavailability.

Table 1

Top phytochemicals with high docking scores (in kcal/mol) across all five M^{Pro} conformations.

Phytochemical Name	Docking scores				
	Conformation 1	Conformation 2	Conformation 3	Conformation 4	Conformation 5
1,3,6-Tri-O-Galloyl-Beta-D-Glucose	-7.6	-8.8	-11.1	-7.5	-10.1
2'-Acetylacteoside	-8.6	-9.5	-12.1	-13.4	-7.6
2"-O-Acetylrutin	-10.3	-9.6	-12.2	-10.8	-10.4
*AHDPH	-8.1	-9.0	-11.6	-7.5	-9.0
Balanophotannin E	-7.5	-11.0	-12.9	-9.9	-8.4
**DDHHG	-9.7	-8.3	-10.9	-11.3	-8.6
***DHMMP-TRTH-TMMO-Chr-One	-9.7	-10.5	-10.7	-9.1	-9.9
Eriodictyol 7-O-Sophoroside	-12.6	-9.3	-10.0	-11.1	-10.0
Forsythiaside	-10.3	-12.6	-14.3	-14.6	-9.2
Hyperin 6"-[glucosyl-(1->3)-rhamnoside]	-9.7	-10.9	-15.9	-12.1	-11.9
Kaempferol 3-(3R-glucosylrutinoside)	-10.0	-10.6	-12.0	-11.1	-8.5
Luteolin 7-rutinoside	-9.8	-9.4	-14.4	-12.0	-9.9
Narcissin	-9.7	-10.5	-10.7	-9.1	-9.9
Pectolinarin	-8.9	-7.7	-13.9	-8.5	-7.5
Plantagineoside C	-9.4	-10.3	-13.3	-10.4	-9.3
Quercetin 3-glucoside2"-gallate	-7.8	-9.2	-12.1	-10.6	-7.5
Quercetin-3-o-rutinose	-12.2	-11.0	-11.1	-11.5	-11.4
Salvianolic Acid L (SAL)	-9.1	-8.2	-13.3	-11.3	-7.6
Shikonin	-8.1	-8.4	-8.6	-8.9	-9.5
Shimobashiric Acid C (SAC)	-8.2	-8.7	-10.5	-9.6	-10.2
*AHDPH = (3R,5R)-3-Acetoxy-5-Hydroxy-1,7-Bis(3,4-Dihydroxyphenyl)Heptane.					
**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.					
***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.					

The four aglycones included shimobashiric acid C (SAC), salvianolic acid L (SAL), AHDPH, and shikonin. The binding modes of these phytochemicals were then ranked for the most favorable binding site conformation by docking score (Fig. 4). SAC, SAL, and AHDPH exhibited a highest binding affinity of when bound to the third binding site conformation with a docking score of -10.5 kcal/mol, -13.3 kcal/mol, and − 11.6 kcal/mol. SAC is a complex molecule featuring a cyclobutane core with multiple hydroxyphenyl groups. As shown in Fig. 4A-C the S2 pocket is predicted to accommodate one of the four hydroxyphenyl groups linked to the cyclobutene core. The three other hydroxy groups from the phenyl moieties interact with the peripheral residues of the S1, S4, and newly formed S2’ pocket. The catalytic residue, His41 is engaged in a pi-pi interaction with one of the hydroxyphenyl groups and residues Cys44, Glu166, and Asp187 established hydrogen bonds with the hydroxy groups. Furthermore, residues Gly143, Ser144, Cys145, and Gln189 formed hydrogen bonds with the oxygen atoms of the two propanoic acid units of the phytochemical. SAL consists of a naphthalene molecule with three 3,4-dihydroxyphenyl substituents and a 3-(3,4-dihydroxyphenyl) propanoic acid moiety connected via a carboxyethyl linker. As depicted in Fig. 4D-F the three 3,4-dihydroxyphenyl groups occupied the S1’, S1, and S4 pockets. The naphthalene ring interacted with S2’ pocket. Six hydrogen bonds were formed between the molecule and protein residues,

including Cys44, Gly143, Cys145, His164, Glu166, Arg188, and Gln192. In the case of AHDPH, S1 and S2 pockets remained unoccupied, while S1' and S4 accommodated the two dihydroxyphenyl groups of the compound. Three residues, Thr26, Arg188, and Gln192 formed hydrogen bonds with the hydroxy groups of the phenyl while Glu 166 formed a hydrogen bond with the hydroxy group of the carbon chain of the phytochemical (Fig. 4G-I). For shikonin, the best binding was achieved with the fifth protein conformation, with docking scores of -9.5 kcal/mol, respectively. The naphthalene ring of shikonin occupied the narrow S4 pocket, with its hydroxy groups interacting with charged pocket residues —Glu166, Asp187, Arg188, and Gln189. Additionally, the phytochemical's pentyl chain interacted with the hydrophobic residues forming the S2 pocket (Fig. 4J-L).

Considering the metabolic cleavage of glycosides to aglycones in the body we investigated docking of such metabolites (Table 2). In each case, aglycones exhibited moderate binding capabilities and displayed a preference for the active site of a specific cluster conformation, except for quercetin that showed good binding to all conformations with a docking score > 7.0 kcal/mol. Dihydrocaffeic acid, brevifolincarboxylic acid, DDHH, pinoresinol, kaempferol, luteolin, isoharmnetin, pectolinarigenin, and secoisolariciresinol preferentially bound to at least two conformations with a docking score > = 7.0 kcal/mol. Hesperetin and eriodictyol bounded favorably with the first conformation, while hydroxytyrosol preferred the third conformation. Gallic acid, and caffeic acid exhibited a low docking score across all conformations compared to its glycoside. Some of these aglycones have been previously reported for their anti-SARS-Cov-2 activity[38,39,40,41,42].

Table 2
Docking Scores (in kcal/mol) of metabolites for the top phytochemicals.

Glycoside Name	Metabolite	Docking scores				
		Conformation 1	Conformation 2	Conformation 3	Conformation 4	Conformation 5
1,3,6-Tri-O-Galloyl-Beta-D-Glucose	Gallic acid	-5.8	-5.3	-5.8	-6.3	-5.4
2'-Acetylacteoside/Forsythiaside	Caffeic acid	-4.9	-4.3	-4.0	-5.5	-4.4
2'-Acetylacteoside/Forsythiaside	Dihydrocaffeic acid	-5.3	-6.0	-8.2	-8.0	-6.3
2'-Acetylacteoside/Forsythiaside	Hydroxytyrosol	-4.8	-4.3	-7.5	-6.7	-5.0
2"-O-Acetylrutin/ Hyperin 6"-[glucosyl-(1->3)-rhamnoside]/ Quercetin 3-glucoside2"-gallate/ Quercetin-3-o-rutinose	Quercetin	-7.8	-7.2	-7.3	-7.6	-7.0
Balanophotannin E	Brevifolincarboxylic acid	-6.4	-6.4	-8.0	-7.1	-6.7
Balanophotannin E	Gallic acid	-5.8	-5.3	-5.8	-6.3	-5.4
**DDHHG	****DDHH	-7.5	-6.5	-10.2	-6.8	-6.5
***DHMMP-TRTH-TMMO-Chr-One	Hesperetin	-7.0	-5.7	-6.5	-5.2	-5.3
Eriodictyol 7-O-Sophoroside	Eriodictyol	-7.0	-5.6	-6.4	-6.7	-6.7
Forsythiaside	Pinoresinol	-5.0	-6.2	-8.7	-7.8	-6.0
Kaempferol 3-(3R glucosylrutinoside)	Kaempferol	-7.0	-5.4	-7.9	-5.1	-5.5
Luteolin 7-rutinoside	Luteolin	-7.3	-6.6	-8.0	-6.9	-5.4
Narcissin	Isoharmnetin	-7.3	-5.9	-7.0	-5.9	-6.1
Pectolinarin	Pectolinarigenin	-7.0	-5.8	-7.6	-5.4	-3.7
Plantagineoside C	Secoisolariciresinol	-7.8	-5.5	-8.8	-7.3	-6.7
**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.						
***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.						
****DDHH = 3,5-dihydroxy-1-(3,4-dihydroxyphenyl)-7-(4-hydroxyphenyl)-heptane.						

In addition to examining the top 20 phytochemicals and their respective metabolites, we extended our analysis to include four phytochemicals that exhibited high docking scores for at least one protein conformation. These included cynarin, demethoxycurcumin, hexahydrocurcumin, and withaferin A (Table 3).

Table 3

Phytochemicals with high docking scores (in kcal/mol) for at least one M^{Pro} conformation.

Phytochemical Name	Docking scores				
	Conformation 1	Conformation 2	Conformation 3	Conformation 4	Conformation 5
Cynarin	-10.7	-5.7	-9.4	-3.9	-6.9
Demethoxycurcumin	-5.9	-8.0	-6.9	-5.8	-4.9
Hexahydrocurcumin	-3.4	-4.7	-9.9	-7.1	-5.6
Withaferin A	-3.9	-4.8	-7.4	-4.6	-8.4

Cynarin had moderate to low docking scores across all binding site conformations, except for the conformations representing Cluster 1 and 3, where its docking scores exceeded − 9.0 kcal/mol. As shown in Fig. 5A-C, the core cyclic structure of dihydrocyclohexane occupied the S1 pocket, while the carboxylic acid group attached to the cyclohexane extended towards the S1’ site. One of the two hydroxyphenyl groups occupied the S4 pocket, while the other hydroxyphenyl group, although not positioned within the S1’ pocket, interacted with the residues forming the pocket. The phytochemical establishes hydrogen bonds with five key residues—Thr26, Asn142, Gly143, Glu166, and Thr190.

Demethoxycurcumin exhibits a docking score of approximately − 8.0 kcal/mol when binding to the representative protein conformation of Cluster 2, the highest among other protein conformations. The hydroxyphenyl group of demethoxycurcumin occupies the region between the S2 and S1’ pockets, with the S2 pocket accommodating the hepta-1,6-diene-3,5-dione bridge. Simultaneously, the hydroxy-3-methoxy group of the phytochemical effectively occupies the S1 pocket. This binding arrangement leads to the formation of hydrogen bonds with four significant residues—Thr26, Leu141, Ser144, and Gly143 (refer to Fig. 5. D-F).

Hexahydrocurcumin had its highest binding score (~ 10.0 kcal/mol) when interacting with the representative conformation of Cluster 3. One of the two hydroxy methoxyphenyl groups of the phytochemical occupy the S2-S4 pocket, interacting with Glu166, Asp187, and Thr190. The other hydroxy methoxy phenyl group interacts with the residues Gly143 and Cys 145, fitting into the S1 pocket (refer to Fig. 5G-I). Interestingly, despite the structural similarity between the two curcumin derivatives, their predicted affinities toward specific protein conformations and their adopted binding orientations are distinct.

Withaferin A preferred the representative structure of Cluster 5 conformation, binding with a docking score of -8.4 kcal/mol. In this binding orientation the dihydropyran ring and the hydroxymethyl substituent occupied the S4 pocket. The main cyclic structure of withaferin A (oxapentacyclooctadec-4-en-3-one) interacted with the S2 and S1’ sites of the binding pocket. Crucial interactions also included M^{Pro} residuesThr26, Asn119, and Glu166 interacting with the hydroxy groups of withaferin A (Fig. 5J-L).

MM-GBSA Prediction

To improve the accuracy of our inhibitor binding predictions, the initially identified 39 phytochemicals were rescored based on binding energies calculated with Prime/MM-GBSA. For these binding energy calculations, the protein-ligand conformation with the highest docking score was selected. Table 4 summarizes the MM-GBSA energies for the selected 20 phytochemicals (Table 1), their corresponding aglycones (Table 2), and phytochemicals exhibiting strong docking scores for at least one protein conformation (Table 3), a total of 39 phytochemicals. The MM-GBSA energy calculations revealed that the binding mode of all the phytochemicals was primarily driven by negative ΔG values for $\Delta G_{coulomb}$, ΔG_{hbond}, ΔG_{lipo}, $\Delta G_{packing}$, and ΔG_{vdw}. These scores indicate the presence of attractive Coulombic interactions, hydrogen bonding, and strong hydrophobic interactions, including lipophilic and van der Waals forces. However, the positive $\Delta G_{covalent}$ and $\Delta G_{solv-GB}$ scores suggest that covalent bond formation and binding of the phytochemicals in an aqueous environment are energetically unfavorable.

Table 4
Prime MM_GBSA energies (in kcal/mol) of phytochemicals.

Phytochemical Name	ΔG_{bind}	$\Delta G_{coulomb}$	$\Delta G_{covalent}$	ΔG_{hbond}	ΔG_{lipo}	$\Delta G_{packing}$	ΔG_{solv_GB}	ΔG_{vdW}
1,3,6-Tri-O-Galloyl-Beta-D-Glucose	-63.8	-45.1	7.4	-5.3	-15.3	-2.2	48.3	-51.5
2'-Acetylacteoside	-59.0	-30.9	8.2	-4.4	-20.8	-2.2	41.9	-50.9
2"-O-Acetylrutin	-53.3	-43.3	14.5	-5.7	-11.7	-3.1	47.1	-51.3
*AHDPH	-58.3	-27.4	3.9	-5.0	-15.1	-2.2	29.9	-42.5
Balanophotannin E	-70.9	-57.1	17.4	-5.5	-15.9	-1.5	43.0	-51.3
Brevifolincarboxylic acid	-40.0	-14.3	0.0	-2.3	-9.5	-2.3	18.2	-29.7
Caffeic acid	-26.8	11.3	3.8	-2.2	-8.1	-0.2	-13.5	-17.9
Cynarin	-34.1	-2.6	7.6	-5.2	-13.1	-1.8	18.2	-37.1
****DDHH	-49.9	-27.4	1.9	-3.4	-11.1	-2.2	22.2	-29.9
**DDHHG	-70.2	-51.9	6.1	-5.8	-12.4	-1.5	32.8	-37.4
Demethoxycurcumin	-57.2	-39.5	6.0	-2.8	-12.3	-0.3	27.6	-35.9
***DHMMP-TRTH-TMMO-Chr-One	-54.9	-32.1	5.9	-4.0	-11.0	-3.0	24.5	-35.2
Dihydrocaffeic acid	-41.1	-26.1	1.0	-2.7	-9.6	-0.8	17.3	-20.2
Eriodictyol	-35.5	-32.2	3.6	-3.2	-4.6	-1.3	29.1	-26.9
Eriodictyol 7-O-Sophoroside	-60.6	-50.5	5.8	-6.7	-13.1	-1.4	44.9	-39.7
Forsythiaside	-82.2	-55.0	6.9	-6.0	-18.4	-0.8	39.4	-48.2
Gallic acid	-24.5	11.8	4.3	-2.8	-6.1	-0.3	-15.7	-15.7
Hesperetin	-43.1	-33.1	3.5	-3.2	-5.9	-1.3	26.2	-29.3
Hexahydrocurcumin	-67.6	-40.4	3.9	-3.8	-16.4	-1.7	27.5	-36.7
Hydroxytyrosol	-54.9	-35.2	1.4	-3.7	-14.7	-0.9	15.9	-17.7
Hyperin 6"-[glucosyl-(1->3)-rhamnoside]	-76.6	-59.1	8.4	-8.8	-15.0	-1.9	36.0	-36.1
Isoharmnetin	-47.2	-22.8	0.7	-2.4	-7.5	-2.8	21.7	-34.1
Kaempferol	-52.6	-34.1	1.1	-2.9	-4.7	-2.3	24.8	-34.5
Kaempferol 3-(3R-glucosylrutinoside)	-35.4	-21.1	14.6	-5.5	-12.8	-0.9	27.7	-37.3
Luteolin	-39.4	-33.7	2.8	-3.3	-4.0	-1.4	26.6	-26.5
Luteolin 7-rutinoside	-80.2	-58.1	7.0	-6.6	-12.5	-2.5	39.0	-46.5
Narcissin	-56.0	-42.0	20.4	-3.0	-15.4	-2.4	34.7	-48.4
Pectolinarigenin	-45.1	-27.2	6.3	-3.1	-6.9	-2.8	21.2	-32.5
Pectolinarin	-71.1	-45.6	8.0	-5.4	-14.1	-2.9	37.3	-48.2
Pinoresinol	-59.1	-34.4	4.7	-4.2	-13.7	-2.8	22.2	-31.0
Plantagineoside C	-62.3	-41.1	5.5	-6.3	-17.5	-2.3	32.1	-32.8
Quercetin	-36.4	-32.8	2.9	-3.3	-4.5	-1.3	28.9	-26.3

*AHDPH = (3R,5R)-3-Acetoxy-5-Hydroxy-1,7-Bis(3,4-Dihydroxyphenyl)Heptane.

**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.

***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.

****DDHH = 3,5-dihydroxy-1-(3,4-dihydroxyphenyl)-7-(4-hydroxyphenyl)-heptane.

Phytochemical Name	ΔG_{bind}	$\Delta G_{coulomb}$	$\Delta G_{covalent}$	ΔG_{hbond}	ΔG_{lipo}	$\Delta G_{packing}$	ΔG_{solv_GB}	ΔG_{vdW}
Quercetin 3-glucoside2"-gallate	-76.7	-61.6	13.2	-6.9	-11.9	-3.0	51.1	-57.5
Quercetin-3-o-rutinose	-45.3	-38.8	3.1	-5.8	-8.4	-1.9	44.4	-37.8
Salvianolic Acid L (SAL)	-42.1	25.3	11.3	-5.2	-24.2	-2.4	2.7	-49.6
Secoisolariciresinol	-53.8	-32.6	6.9	-4.0	-11.0	-3.0	25.5	-35.5
Shikonin	-52.9	-25.8	6.9	-2.8	-15.1	0.0	19.8	-35.9
Shimobashiric Acid C (SAC)	-39.0	2.0	8.3	-7.2	-5.0	-1.5	-1.3	-34.4
Withaferin A	-51.8	-19.8	4.0	-1.7	-14.1	0.0	22.6	-42.8
*AHDPH = (3R,5R)-3-Acetoxy-5-Hydroxy-1,7-Bis(3,4-Dihydroxyphenyl)Heptane.								
**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.								
***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.								
****DDHH = 3,5-dihydroxy-1-(3,4-dihydroxyphenyl)-7-(4-hydroxyphenyl)-heptane.								

Among the selected phytochemicals, Forsythiaside had the best calculated binding energy, ΔG_{bind} = -82.2 kcal/mol. Luteolin 7-rutinoside ranked second with a ΔG_{bind} value of -80.2 kcal/mol, followed by quercetin 3-glucoside 2”-gallate and hyperin 6’-[glucosyl-(1 > 3)-rhamnoside] with a ΔG_{bind} value of approximately − 76.0 kcal/mol. All of these molecules also showcased high docking scores above − 12.0. kcal/mol.

Pinoresinol, one of the metabolites of forsythiaside exhibited the highest binding energy of ~ -59.0 kcal/mol among all the aglycones studied. It also showcased a significant docking score of -8.7 kcal/mol. Secoisolariciresinol was the second-best aglycone with a binding energy of -53.8 kcal/mol, followed by kaempferol. Overall, the binding energy calculations aligned well with the computed docking scores of top 20 phytochemicals (Table 1), except for SAC and kaempferol 3-(3R-glucosylrutinoside), which deviated with relatively low binding energies (ΔG_{bind} < -40 kcal/mol), despite securing high docking scores across all protein conformations. Cynarin, while displaying a significantly high docking score for the first protein conformation, exhibited a low binding energy value (ΔGbind = -36.1 kcal/mol). Additionally, the aglycones luteolin and quercetin demonstrated lower binding energies (ΔG_{bind} = -39.4 and − 36.4 kcal/mol, respectively), although they showcased moderate docking scores in at least one protein conformation.

We studied the protein-ligand interaction analysis of the top ten phytochemicals based on docking scores and MM-GBSA energies. The heatmap illustrated in Fig. 6 highlights the significance of specific binding site residues, including His41, Asn142, Gly143, Cys145, Met165, Glu166, Arg188, and Glu189, in stabilizing the phytochemicals within the binding site. These residues formed van der Waals and Coulombic interactions with at least five out of the ten phytochemicals, emphasizing their crucial role in ligand binding.

Bioavailability Prediction

To gain insights into the bioavailability of the selected phytochemicals, we conducted an in silico ADMET study using ADMET Predictor by Simulations Plus[43]. This calculation generates a ADMET Risk score that reflects potential compound liabilities based on 20 rules within three Risk models (Absn_Risk, CYP_Risk, and Tox_Risk) of their ADMET Predictor. Additionally, the ADMET Risk score models pharmacokinetic properties termed fraction unbound (fu) and high-steady-state volume of distribution (Vd). Each Risk score is paired with a mnemonic Code list that identifies the rules that have been violated. Of the 39 phytochemicals predicted to bind strongly to M^{Pro}, ten had ADMET Risk scores of < = 1 (Table 5). On the other hand, six phytochemicals, including 1,3,6-Tri-O-galloyl-beta-D-glucose, 2-acetylacteoside, quercetin 3-glucoside2" -gallate, balanophotannin E, SAL, and SAC had problematic ADEMT Risk scores exceeding 6.

Table 5
ADMET Prediction of phytochemicals.

Phytochemical ID	ADMET Risk	ADMET Code	Liver Microsomes		Hepatocytes		Systemic		THalf_hum-100.0	S + MDCK-LE
			%Fa	%Fb	%Fa	%Fb	%Fa	%Fb		
1,3,6-Tri-O-Galloyl-Beta-D-Glucose	7.0	Size; HBD; HBA; ch; Kow-; Peff+; Sw-; fu-; Vd-; CL+	28.3	23.4	28.3	23.4	28.2	24.9	1.2	Low
2-Acetylacteoside	6.5	Size; RotB; HBD; HBA; ch; Peff; CL+	40.1	34.4	40.1	34.4	40.0	36.9	0.8	Low
2"-O-Acetylrutin	5.5	Size; HBD; HBA; ch; Peff; CL-	24.0	19.1	24.0	19.1	23.9	21.2	4.8	Low
*AHDPH	3.4	RotB; HBD; ch; CL	99.9	93.7	99.9	93.7	99.8	99.8	1.8	Low
Balanophotannin E	7.0	Size; HBD; HBA; ch; Kow-; Peff+; Sw-; fu-; Vd-; CL+	5.1	4.5	5.1	4.5	5.1	4.8	1.3	Low
Brevifolincarboxylic acid	1.5	HBD; ch	95.0	82.9	95.0	82.9	95.8	95.1	1.6	Low
Caffeic acid	1.0	HBD	99.5	86.9	99.3	87.2	97.6	84.0	1.0	High (96%)
Cynarin	5.0	Size; HBD; HBA; ch; Peff	44.6	40.1	44.6	40.1	44.4	42.0	3.8	Low
****DDHH	2.2	HBD; CL	99.8	82.2	99.8	82.2	99.6	99.6	3.5	Low
**DDHHG	5.0	Size; RotB; HBD; HBA; ch	54.9	48.3	54.7	48.2	53.3	47.6	3.5	Low
Demethoxycurcumin	1.1	fu; CL	97.7	87.0	97.7	87.0	97.6	97.6	10.6	High (99%)
***DHMMP-TRTH-TMMO-Chr-One	4.5	Size; HBD; HBA; ch; CL-	22.9	18.7	22.9	18.7	22.9	20.4	4.5	Low
Dihydrocaffeic Acid	0.0		99.6	83.6	99.6	83.6	99.8	99.7	0.9	High
Eriodictyol	1.3	HBD; CL	99.9	90.1	99.9	90.1	99.6	99.5	1.9	High (82%)
Eriodictyol 7-O-Sophoroside	6.0	Size; HBD; HBA; ch; Kow-; Peff; CL-	17.1	14.7	17.1	14.7	17.1	15.8	1.0	Low
Forsythiaside	6.0	Size; RotB; HBD; HBA; ch; Peff; CL-	33.5	28.4	33.5	28.4	33.4	30.2	1.0	Low
Gallic Acid	0.5	HBD	99.0	78.1	99.0	78.1	92.0	91.0	0.4	High (85%)
Hesperetin	0.9	CL	99.9	92.7	99.9	92.7	92.5	91.9	5.7	High (93%)
Hexahydrocurcumin	2.0	RotB; CL	99.9	82.0	99.9	82.0	99.9	99.9	4.6	High (99%)

Red hues signify improved predictive scores, and blue hues represent poorer predictive scores for three clearance models.

RotB = rotatable bonds; HBD = H-bond donors; HBA = H-bond acceptors; ch = charge; Kow = lipophilicity; Peff = permeability; Sw = water solubility; fu = fraction unbound; Vd = volume of distribution; CL = high microsomal clearance.

%Fa = fraction absorbed; %Fb = fraction bioavailable.

*AHDPH = (3R,5R)-3-Acetoxy-5-Hydroxy-1,7-Bis(3,4-Dihydroxyphenyl)Heptane.

**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.

***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.

****DDHH = 3,5-dihydroxy-1-(3,4-dihydroxyphenyl)-7-(4-hydroxyphenyl)-heptane.

Phytochemical ID	ADMET Risk	ADMET Code	Liver Microsomes		Hepatocytes		Systemic		THalf_hum-100.0	S + MDCK-LE
			%Fa	%Fb	%Fa	%Fb	%Fa	%Fb		
Hydroxytyrosol	0.0		99.9	63.2	99.9	63.2	98.6	98.2	1.0	High (99%)
Hyperin 6"-[glucosyl-(1->3)-rhamnoside]	6.0	Size; HBD; HBA; ch; Kow-; Peff; CL-	4.9	4.0	4.9	4.0	4.9	4.4	1.0	Low
Isorhamnetin	0.5	HBD; ch	98.8	90.3	98.8	90.3	98.6	98.5	2.1	High (88%)
Kaempferol	0.5	HBD	99.9	92.0	99.9	92.0	99.7	99.6	2.2	High (93%)
Kaempferol 3-(3R-glucosylrutinoside)	6.0	Size; HBD; HBA; ch; Kow-; Peff; CL-	5.3	4.2	5.3	4.2	5.3	4.7	1.1	Low
Luteolin	0.5	HBD; CL	93.9	87.5	93.9	87.5	93.7	93.7	3.8	High (93%)
Luteolin 7-rutinoside	5.5	Size; HBD; HBA; ch; Peff; CL-	20.8	16.9	20.8	16.9	20.8	18.3	4.1	Low
Narcissin	5.5	Size; HBD; HBA; ch; Peff; CL-	22.9	18.7	22.9	18.7	22.9	20.4	1.0	Low
Pectolinarigenin	2.0	Sw; CL	23.6	21.7	23.6	21.7	23.6	23.6	5.2	High (99%)
Pectolinarin	5.5	Size; HBD; HBA; ch; Peff; CL-	33.4	26.7	33.4	26.7	33.3	29.9	1.6	Low
Pinoresinol	1.0	CL	99.9	76.2	99.9	76.2	99.8	99.8	3.7	High (99%)
Plantagineoside C	5.5	Size; HBD; HBA; ch; Peff; CL-	70.9	64.8	70.9	64.8	70.2	69.0	0.8	Low
Quercetin	2.0	HBD; ch; CL	98.1	88.3	98.1	88.3	97.8	97.7	1.5	High (76%)
Quercetin 3-glucoside2"-gallate	6.5	Size; HBD; HBA; ch; Kow-; Peff; Sw-; CL-	24.2	21.6	24.2	21.6	24.1	22.9	1.6	Low
Quercetin-3-o-rutinose	6.0	Size; HBD; HBA; ch; Kow-; Peff; CL-	16.9	13.9	16.9	13.9	16.9	15.2	1.1	Low
Salvianolic acid L (SAL)	8.0	Size; RotB; HBD; HBA; ch; Kow-; Peff; Sw-; fu-; CL+	48.5	43.9	48.5	43.9	48.3	45.6	1.2	Low
Secoisolariciresinol	1.3	RotB; HBD; CL	99.9	80.4	99.9	80.4	99.8	99.8	3.7	Low
Shikonin	0.0		99.9	93.7	99.9	93.7	99.9	99.9	7.7	High (88%)

Red hues signify improved predictive scores, and blue hues represent poorer predictive scores for three clearance models.

RotB = rotatable bonds; HBD = H-bond donors; HBA = H-bond acceptors; ch = charge; Kow = lipophilicity; Peff = permeability; Sw = water solubility; fu = fraction unbound; Vd = volume of distribution; CL = high microsomal clearance.

%Fa = fraction absorbed; %Fb = fraction bioavailable.

*AHDPH = (3R,5R)-3-Acetoxy-5-Hydroxy-1,7-Bis(3,4-Dihydroxyphenyl)Heptane.

**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.

***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.

****DDHH = 3,5-dihydroxy-1-(3,4-dihydroxyphenyl)-7-(4-hydroxyphenyl)-heptane.

Phytochemical ID	ADMET Risk	ADMET Code	Liver Microsomes		Hepatocytes		Systemic		THalf_hum-100.0	S + MDCK-LE
			%Fa	%Fb	%Fa	%Fb	%Fa	%Fb		
Shimobashiric Acid C (SAC)	8.0	Size; RotB; HBD; HBA; ch; Kow-; Peff; Sw-; fu-; CL+	45.8	43.4	45.8	43.4	45.8	43.4	14.0	Low
Withaferin A	1.5	Size; ch	80.4	62.3	80.4	62.3	80.3	80.3	11.7	High (99%)
Red hues signify improved predictive scores, and blue hues represent poorer predictive scores for three clearance models.										
RotB = rotatable bonds; HBD = H-bond donors; HBA = H-bond acceptors; ch = charge; Kow = lipophilicity; Peff = permeability; Sw = water solubility; fu = fraction unbound; Vd = volume of distribution; CL = high microsomal clearance.										
%Fa = fraction absorbed; %Fb = fraction bioavailable.										
*AHDPH = (3R,5R)-3-Acetoxy-5-Hydroxy-1,7-Bis(3,4-Dihydroxyphenyl)Heptane.										
**DDHHG = (3R,5R)-3,5-Dihydroxy-1-(3,4-Dihydroxyphenyl)-7-(4-Hydroxyphenyl)-Heptane 3-O-Beta-D-Glucopyranoside.										
***DHMMP-TRTH-TMMO-Chr-One = 5,7-Dihydroxy-2-(4-Hydroxy-3-Methoxyphenyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-Trihydroxy-6-[[(3R,4R,5R,6S)-3,4,5-Trihydroxy-6-Methyloxan-2-Yl]Oxymethyl]Oxan-2-Yl]Oxychromen-4-One.										
****DDHH = 3,5-dihydroxy-1-(3,4-dihydroxyphenyl)-7-(4-hydroxyphenyl)-heptane.										

We calculated absorption and bioavailability characteristics of these phytochemicals using three human clearance models (liver microsomes, hepatocytes, and systemic) at a dose of 100 mg for an immediate release tablet and observed significant variability in the calculated pharmacokinetics parameters (Table 5). Out of 39 phytochemicals examined, 20 exhibited high absorption (%Fa) and minimal metabolism or elimination in the liver during the first-pass effect (%Fb). A slight decrease in %Fb values were observed in the hepatocytes for some of these phytochemicals, suggesting that a portion of the compound is subject to metabolism.

In addition to absorption and bioavailability estimations, we calculated phytochemical plasma half-life (T-half) in humans and cell permeability (Madin-Darby canine kidney; MDCK), the latter serving as an estimate for intestinal epithelial absorption (Table 5). Demethoxycurcumin, shikonin, SAC, and withaferin, demonstrated an extended plasma half-life, surpassing 7 hours. Phytochemicals with low ADMET risks also exhibited high MDCK permeability, except for secoisolariciresinol and DDHH.

Cytotoxicity and Viral Replication Assay

We performed cell-based antiviral assay to evaluate the inhibitory effects of five highly performing phytochemicals (demethoxycurcumin, hydroxytyrosol, kaempferol, shikonin, and withaferin A) on SARS-CoV-2 replication. These compounds were selected based on their high overall performance in terms of docking score (above − 7.5 kcal/mol for at least one protein conformation), MM-GBSA binding energy (ΔG_{bind} > -50 kcal/mol), and ADMET properties. Glycosides were excluded from the viral replication assay as their metabolic modification in the body would render the results less relevant. Although pinoresinol and hexahydrocurcumin fulfilled the criteria for top-performing molecules, meeting the specified factors, we opted not to include them in the study due to their current unavailability for immediate testing.

Viral assay demonstrated the SARS-Cov-2 replication inhibitory activity of three compounds, shikonin (EC50 = ~ 10 µM), demethoxycurcumin (EC50 = ~ 8.8 µM), and withaferin,(EC50 = ~ 2.8 µM) (Fig. 7). Importantly, in contrast to shikonin and withaferin, demethoxycurcumin showed no apparent cytotoxicity.

The inhibitory effects of these three compounds against M^{Pro} were previously speculated [44, 45, 46, 47, 48,49,50,51], however, the referenced studies primarily relied on computational or enzyme-based assays. Our study presents a robust cell-based antiviral data, providing a more direct and clinically relevant perspective on the inhibitory potential of these compounds against SARS-CoV-2 replication.

While our findings contribute to the ongoing discussion on effective viral inhibition potential of demethoxycurcumin, shikonin, and withaferin A, we also recognize the necessity for further research to elucidate the intricate mechanisms underlying this inhibition. Our work adds valuable insights to this evolving dialogue, underscoring the importance of continued investigation in this area.

Discussion

Plant-derived natural product reservoir harbors a distinctive wealth of 'bioactive' molecules, spanning broader chemical diversity than synthetic libraries [52]. Interestingly, phytochemicals are structurally optimized by evolution, to serve specific biological functions. Among these functions is

the regulation of endogenous defense mechanisms against microorganisms such as viruses and bacteria[53]. Additionally, the utilization of phytochemicals in traditional medicine could offer valuable insights into their effectiveness and safety. These characteristics elevates the importance of phytochemicals for antiviral and antibacterial interventions, revealing fresh avenues for exploration. Our study aimed to identify phytochemicals capable of inhibiting SARS-CoV-2 M^{Pro} protein, using a phytochemical database comprising ~ 45K compounds.

M^{Pro} is a challenging target due to its inherent plasticity mediated by allosteric residues. The high flexibility of M^{Pro} alters the characteristics of its binding pocket subsites, leading to diverse packing modes for inhibitors. This variation is evident from over 200 experimental structures of M^{Pro} deposited in the protein data bank[54]. To investigate the ensemble of binding site conformations adopted by the protein we integrated MD simulations with pocket shape-based clustering approach for conducting virtual screening. Since performing docking on MD simulations generated hundreds of thousands of protein conformations is computationally demanding, employing pocket-based filtering allows the selection of representative protein structures that capture the full range of conformational diversity observed in the binding site.

Our pocket-based analysis revealed significant differences in subsite geometry across the five clusters, indicating the ability of these conformers to bind dissimilar ligands. The representative structures of the clusters highlighted a dual pattern: the S2 and S4 pockets demonstrated shifts—progressing from two distinct pockets to a single, larger pocket—and alternatively, the area and the volume of S4 pocket increased. Furthermore, opening of the S2 pocket and formation of new accessible subsites were observed. Some of the structural features of the binding sites that we observed in our analysis have been previously captured in experimental structures of the M^{Pro} [55,56,57]. While our analysis provided valuable insights, a potential avenue for further enhancement would be to extend the length of the MD simulation and expand the number of protein conformations used for pocket-based clustering. Enriching the dataset could yield more comprehensive understanding of the intricate dynamics governing the M^{Pro} binding site, thereby facilitating an improved selection of potential drug candidates.

In line with the variations observed during pocket-based clustering analysis, virtual screening against the representative structures yielded a striking divergence in the landscape of high-performing phytochemicals; around 1% of the phytochemicals in the dataset showed potential binding across all examined M^{Pro} pocket conformers. A majority of molecules with high docking scores belonged to the class of flavonoid glycosides. This observation was consistent with previous reports[58] on insilico studies of natural products, although the structural basis for their potency compared to other phytochemical classes remain unclear. In this study, we assessed 20 phytochemicals with high docking scores across all protein conformations and four phytochemicals that displayed high docking scores for at least one conformation. Given the likelihood of metabolic transformations of glycosides within the body, we examined the binding capabilities of the aglycones of these glycosides. This study showed a consistent superiority of glycosides over their aglycone counterparts in terms of docking scores, signaling a pivotal role for glycosylation in increasing the score of binding calculations. Moreover, this observation accentuates the importance of investigating phytochemical metabolites to ensure efficacy, especially in the case where parent compounds that are prone to biotransformation. By examining metabolite-protein interactions, the likelihood of overlooking potential therapeutic benefits originating from these modified forms can be minimized, concurrently highlighting potential reductions in bioactivity among the resultant metabolites.

Among all the phytochemicals identified in the virtual screening process, seven are reported here for the first time for their M^{Pro} binding tendency. These include hyperin 6’-[glucosyl-(1 > 3)-rhamnoside], balanophotannin E, plantagineoside C, SAC, AHDPH, brevifolincarboxylic acid, and secoisolariciresinol. However, the ADME profiles and bioavailability of these phytochemicals generally fall within the moderate to poor range, except for AHDPH, brevifolincarboxylic acid and secoisolariciresinol, which exhibit more favorable attributes. For forthcoming investigations, it would be interesting to explore the impact of these molecules on M^{Pro} inhibition and SARS-CoV-2 replication. This exploration could uncover their significance in combating the virus and could potentially yield novel insights for future therapeutic approaches. Consistent with previous computational and enzyme-based studies [44,45,46,47,48,49,50,51], three molecules—demethoxycurcumin, shikonin, and withaferin A—demonstrated high M^{Pro} inhibition activity in our docking analysis and were further experimentally validated for their inhibitory bioactivity against SARS-CoV-2 replication. Despite the observed cytotoxicity of two of these compounds, they hold translational potential by serving as a promising starting point for the development of more potent and selective SARS-CoV-2 inhibitors.

Methods

Molecular dynamics simulation

The protein structure of M^{Pro} protein was obtained from the Protein Data Bank[59] (PDB code: 5R80). Two independent molecular dynamics simulations of the ligand unbound protein were carried out utilizing the GROMACS 5.0.7 software[60]. The force field used for the protein was CHARMM36[61]. The models were immersed in a dodecahedron box with dimensions 5.0777 x 5.069 x 3.672 nm, containing TIP3P[62] water molecules. Subsequently, two chloride ions, twelve, and four sodium ions were added to model 1, 2, and 3, respectively to achieve neutrality of the system. Periodic boundary conditions and a 2fs time step were employed for each simulation. Particle Mesh Ewald[63] (PME) was employed to treat long-range electrostatic interactions and a cutoff of 0.9 nm was used for non-bonded interactions. Combination of steepest descent and

conjugate gradient methods of energy minimization (EM) were used to remove steric clashes and to minimize the interaction forces. EM was achieved in four steps[64]. In Step 1 the protein and ligand were restrained, in Step 2 the protein heavy atoms and ligand were restrained, in Step 3 the protein main chain was restrained and in Step 4, unrestrained EM was performed. Unrestrained minimization was carried out until convergence where the maximum atomic force was < 100 kJ/mol–nm. The minimized structures were first gradually heated to 300K in the NVT ensemble for 20ps using a position restraint of 1000 kJ/mol-nm constant force on the proteins. The systems were equilibrated under NPT (isobaric-isothermal) ensemble by 200ps by gradually decreasing the position restraint force to 700, 400, and then to 100 and finally to 10 kJ/mol–nm and then equilibrated for an additional 2 ns without any restraints. A production run for 300 ns were carried out under the NPT ensemble. The temperature was maintained at 300K with the Berendsen weak coupling method[65]. Bond lengths were constrained using the linear constraint solver (LINCS) algorithm[66] and the van der Waals cutoff was at 1.4 nm. The trajectories were sampled every 10ps for analysis in production dynamics. Molecular visualization and analysis were performed using VMD[67] and UCSF Chimera package[68].

Pocket Volume Analysis

POVME 3 [31,32] software was utilized for characterizing pocket volumes and shapes of ensembles of binding pockets from molecular dynamics simulation.

The volume of the active site is calculated by first defining the binding pocket region. The coordinates of Cβ atom of residue 165 located at the center of the active site was considered as the point of inclusion with 8 Å radius. Subsequently, the user-specified inclusion region encompassing all the binding-pocket conformations of the trajectory is filled with equidistant points spaced at 1.0 Å. The program then removes all the points that are close to the receptor alpha carbon atoms and leaves the points that are positioned within the pockets. Two algorithms, gift wrapping, and Akl-Toussaint heuristic are used to define the convex hull of receptor atoms near the inclusion sphere; any points outside the convex hull are then removed. Lastly, to remove the isolated patches of points lying outside the primary pockets or the pockets not contiguous with the primary pockets are removed. POVME scripts (binding_site_overlap.py and cluster.py) were used to calculate the similarity of the pockets from all protein conformations, to measure the binding site overlap, and to perform hierarchical clustering.

Structure-Based Virtual screening

Protein and Ligand Preparation

The representative protein structures derived post clustering were prepared for docking utilizing the “Protein preparation wizard” tool in Schrödinger suite[69]. The protocol included, the removal of water molecules and cofactors, rectifying misidentified elements, introducing hydrogen atoms, determining bond orders, and optimizing hydrogen bonds. Hydrogen bond assignment was done using PROPK at pH 7.0, the orientation of hydroxyl groups, side chain amide groups of Asn and Gln, and the charge state of His residues were optimized. The protein structures were minimized to an RMSD limit from the starting structure of 0.3 Å using the OPLS3 force field[70]. The prepared proteins were subsequently utilized for grid creation using the "Receptor Grid Generation" panel within the Glide module of the Schrödinger suite[71]. Binding site was defined by selecting key residues located at the subsites.

The library of phytochemicals was prepared using Schrödinger's Maestro LigPrep tool[72], which involved applying the OPLS4 force field, optimizing the structures, and incorporating hydrogen atoms. Additionally, Epik, integrated into the process, was utilized to assign probable protonation states within a pH range of 7 ± 2 and determine tautomeric forms for each compound.

Molecular Docking

Virtual screening was performed using Glide program[71]. A ligand-flexible docking of prepared ligands to the binding sites of five protein conformations was performed at two different levels, standard precision (SP) and extra precision (XP) using Virtual Screening Workflow protocol of GLIDE.

MM-GBSA Prediction

The examination of the protein and ligand complexes' free binding energies was carried out by utilizing the MM-GBSA (Molecular Mechanics, the Generalized Born model, and Solvent Accessibility) tool of Schrödinger. To determine the optimal binding energy the Prime module within the Schrödinger software was employed. Solvation model VSGB 2.0[73] and OPLS-AA force field[70] was used for the calculations. Additionally, the protocol incorporates physics-based modifications catering to π-π interactions, hydrophobic interactions, and self-contact interactions involving hydrogen bonding.

ADMET and Bioavailability Prediction

The SMILES (Simplified Molecular-Input Line-Entry System) strings for phytochemicals were imported into ADMET Predictor v11.0[43]. Physicochemical and pharmacokinetic properties were predicted using quantitative structure-activity relationship (QSAR) models within AP11.0.

For a comprehensive understanding of the model specifications, validation, and performance, please refer to the AP11.0 user manual and relevant publications[74,75].

Cytoxicity Assay

Vero cells were seeded using a multiDrop combi liquid dispenser (Thermo) into 384-well plates at a density of 500 cells/well suspended in 50 µL of media. Cells were allowed to recover and fully attach overnight (approximately 16 hours), at which point library compounds were dispensed into wells using an Echo 550 acoustic dispenser (Labcyte). A total of six final concentrations where tested (50 µM, 25 µM, 12.5 µM, 6.25 µM, 3.125 µM, and 1.5625 µM) and wells were back filled with DMSO such that all wells contained a fixed ratio of DMSO. Compounds were incubated with cells for 1 hour prior to addition of virus and then for an additional 24 hours, then fixed with 10% formalin, permeabilized 0.1% Triton X-100, washed, and stained for SARS-CoV-2 N protein using a specific antibody (Sino Biological # MM05) and fluorescently labelled secondary antibody. Cells were counter stained with Hoechst 33342 to detect cell nuclei, washed, and imaged with a Cytation 1 (Biotek) automated. Each image was then analyzed with a custom workflow in Cell Profiler (Broad Inst., Boston, MA) which involved counting of cell nuclei and infected cells. At least 4 replicates were used to construct dose response curves.

Statistics and data normalization

The growth rate index is calculated from cell counts using the following formula:

$$GR = 2^{\frac{\left(\frac{X_c}{X_0}\right)}{\left(\frac{X_{DMSO}}{X_0}\right)}} - 1$$

Where X_c is the observed cell count after drug exposure, X_0 is the median cell count from a plate that was fixed at the time when the drug was added, and X_{DMSO} is the median cell count of the on-plate DMSO control. The growth rate is then fit against the $\log_{10}$ molar concentration using a cascade of models (4 parameters logistic, 2 parameter logistic, and linear model). The model with the best overall fit, determined by minimizing the mean-squared-error, is then selected. Finally, compound effects are ranked using either potency (GR_{50}, LD_{50}) or the area under/over the curve (AUC_GRI, AOC_GR, and AOC_LD) calculated from the optimized fit.

Viral Assay

As for the cytotoxicity assay, Vero cells were seeded into 384-well plates to a density of 70% per well. Cells were allowed to recover and fully attach overnight (approximately 16 hours), at which point library compounds were added to cells. A total of ten final concentrations where tested ranging from 10 µM down to 20 nM. Compounds were incubated with cells for 1 hour prior to addition of SARS-CoV-2 virus, Washington strain, and then for an additional 36 hours. The cells were then fixed with 10% formalin, permeabilized 0.1% Triton X-100, washed, and stained for SARS-CoV-2 N protein using a specific antibody (Sino Biological # MM05) and fluorescently labelled secondary antibody. Cells were counter stained with Hoechst 33342 to detect cell nuclei, washed, and imaged with a Cytation 1 (Biotek) automated. Each image was then analyzed with a custom workflow in Cell Profiler (Broad Inst., Boston, MA) which involved counting of cell nuclei and infected cells. At least 4 replicates were used to construct dose response curves.

Infected cells/total cell count ratios were calculated for each well. Then, each well was normalized to the average of the 14 negative control wells on each plate. These normalized ratios were input into GraphPad Prism (7.05), and curves were fitted to the data. A dose response curve ([Inhibitor] vs. response - Variable slope (four parameters); constrained to 0) was used. $R^2 > 0.80$ was used as a threshold for a good fit of the dose-response curve. Total cell count was added to the right y-axis as a surrogate for potential cytotoxicity and dose-response curves were also run for total cell count following the same formula and constraint.

In Supplementary Table 2, the results are labeled accordingly: red, did not converge on a line of best fit; yellow, converged on a line of best fit but $R2 < 0.80$; green, converged on a line of best fit and $R^2 > 0.80$.

Declarations

Acknowledgments

We express our sincere gratitude to Charlie Burns and Samantha Roloff for their dedicated efforts in procuring the phytochemicals for testing. Additionally, we want to acknowledge Kong Vo for providing exceptional technical support throughout the duration of this project.

Authors Contribution

K.S. carried out the in silico portion of the research, including MD simulation, docking, MM-GBSA. and ADMET prediction. She analyzed the results and wrote the manuscript. J.P. conducted and analyzed the results of the viral replication assay. A.D. performed and analyzed the

cytotoxicity assay data. All authors reviewed the manuscript.

Data availability

The data that support the findings of this study are available from the corresponding author, Khushboo Singh upon reasonable request.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wu, F., *et al.* Author Correction: A New Coronavirus Associated with Human Respiratory Disease in China. Nature. **580,** E7 (2020).
2. Gorbalenya, A.E. *et al.* The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-ncov and naming it SARS-COV-2. *Nature Microbiol.* **5,** 536–544 (2020).
3. Galan, L.E. *et al.* Phase 2 randomized study on chloroquine, hydroxychloroquine or ivermectin in hospitalized patients with severe manifestations of SARS-COV-2 infection. *Pathog. Glob. Health*. **115,** 235–242 (2021).
4. Vallejos, J. *et al.* Ivermectin to prevent hospitalizations in patients with covid-19 (IVERCOR-covid19) a randomized, double-blind, placebo-controlled trial. *BMC Infect. Dis*. **21,** 635 (2021)
5. Dhawan, B.N. Anti-viral activity of Indian plants. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* **82,** 209–224 (2012).
6. H.C.C, M. *et al.* Phytochemical, antioxidant and microbial inhibitory effects of spondias mombin leaf and stem bark extracts. *IOSR J. Pharm. Biol. Sci*. **9,** pp. 14–17 (2014).
7. Khanna, K. *et al.* Herbal Immune-boosters: Substantial warriors of pandemiccovid-19 battle. *Phytomedicine*. **85,** 153361 (2021).
8. Ling, C. Traditional Chinese medicine is a resource for drug discovery against 2019 novel coronavirus (SARS-COV-2). *J. Integr. Med*. **18,** 8788 (2020).
9. Remali, J. & Aizat, W.M. A review on plant bioactive compounds and their modes of action against coronavirus infection. *Front. Pharmacol.***11,** 589044 (2021).
10. Fei, J. *et al.* Contribution of traditional Chinese medicine combined with conventional western medicine treatment for the novel coronavirus disease (COVID-19), current evidence with systematic review and meta-analysis. *Phytother. Res.***35, (**2021).
11. Lawson, A.D., MacCoss, M. & Heer, J.P. Importance of rigidity in designing small molecule drugs to tackle protein–protein interactions (ppis) through stabilization of desired conformers. *J. Med. Chem*. **61,** 4283–4289 (2017).
12. Henrich, C.J. & Beutler, J.A. Matching the power of high throughput screening to the chemical diversity of natural products. *Nat. Prod. Rep.***30,** 1284 (2013).
13. Romano, J.D. & Tatonetti, N.P. Informatics and computational methods in natural product drug discovery: A review and Perspectives. *Front. Genet*. **10,** 368 (2019).
14. Zeng, X. *et al.* CMAUP: A database of collective molecular activities of useful plants. *Nucleic Acids Res*. **47,** D1118-D1127 (2018).
15. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. **579,** 270–273 (2020).
16. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature*. **579,** 265–269 (2020).
17. Anand, K. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. *EMBO J.* ***21,*** 3213–3224 (2002).
18. Yang, H. *et al.* The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci.* ***100,*** 13190–13195 (2003).
19. Xue, X. *et al.* Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* ***82,*** 2515–2527 (2008).
20. Ren, Z. *et al.* The newly emerged SARS-like coronavirus HCoV-EMC also has an "Achilles' heel": current effective inhibitor targeting a 3C-like protease. *Protein Cell.* ***4,*** 248–250 (2013).
21. Wang, F. *et al.* Structure of main protease from human coronavirus NL63: insights for wide spectrum anti-coronavirus drug design. *Sci. Rep.***6,** 22677 (2016).
22. Jin, Z. *et al.* Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature*. **582,** 289-293 (2020).
23. Kneller, D.W., Kovalevsky, A. & Coates, L. Structural plasticity of the SARS-COV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nature Commun*. **11,** 3202 (2020).

24. Bzówka, M., Mitusińska, K., Raczyńska, A., Samol, A., Tuszyński, J.A., & Góra, A. Structural and Evolutionary Analysis Indicate That the SARS-CoV-2 Mpro Is a Challenging Target for Small-Molecule Inhibitor Design. *Int. J. Mol. Sci.***21,** 3099 (2020).
25. Gossen, J. *et al.* A Blueprint for High Affinity SARS-CoV-2 Mpro Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics. *ACS Pharmacol. Transl. Sci.* **4,** 1079-1095 (2021).
26. Flynn, J.M. *et al.* Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms. *Elife.***11,** e77433 (2022).
27. Estrada, E. Topological analysis of SARS CoV-2 main protease. *Chaos*. **30,** 061102 (2020).
28. Bharadwaj, K.K. *et al.* Macrolactin A as a Novel Inhibitory Agent for SARS-CoV-2 M^{pro}: Bioinformatics Approach. *Appl. Biochem. Biotechnol.* **193,** 3371-3394 (2021).
29. Abdusalam, A.A.A., & Murugaiyah, V. Identification of Potential Inhibitors of 3CL Protease of SARS-CoV-2 From ZINC Database by Molecular Docking-Based Virtual Screening. *Front. Mol. Biosci*. **7,** 603037 (2020).
30. Gupta, A. *et al.* Structure-Based Virtual Screening and Biochemical Validation to Discover a Potential Inhibitor of the SARS-CoV-2 Main Protease. *ACS Omega*. **5,** 33151-33161 (2020).
31. Durrant, J.D. *et al.* POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput*. **10,** 5047-5056 (2014).
32. Wagner, J.R. *et al.* POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.***13,** 4584-4592 (2017).
33. Durdagi, S. *et al.* Near-physiological-temperature serial crystallography reveals conformations of SARS-CoV-2 main protease active site for improved drug repurposing. *Structure*. **29,** 1382-1396 (2021).
34. Sztain, T., Amaro, R., & McCammon, J.A. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 protease. *J. Chem. Inf. Model.* **61,** 3495–3501 (2021).
35. Yang, X. *et al.* Screening of potential inhibitors targeting the main protease structure of SARS-CoV-2 via molecular docking. *Front. Pharmacol.***13,** 962863 (2022).
36. da Fonseca, A.M. *et al.* Screening of Potential Inhibitors Targeting the Main Protease Structure of SARS-CoV-2 via Molecular Docking, and Approach with Molecular Dynamics, RMSD, RMSF, H-Bond, SASA and MMGBSA. *Mol. Biotechnol.* Preprint at 10.1007/s12033-023-00831-x (2023).
37. Teli, D.M. et al. Fragment-based design of SARS-CoV-2 Mpro inhibitors. *Struct Chem*. **33,** 2155-2168 (2022).
38. Jamhour, R.M.A.Q. *et al.* Phytochemicals As a Potential Inhibitor of COVID-19: An In-Silico Perspective. *Russ. J. Phys. Chem.***96,** 1589-97 (2022).
39. Agrawal, P.K., & Blunden, G. Phytochemicals Against SARS-COV-2 Infection. *Nat. Prod. Commun.***18,** (2023).
40. Biancatelli, R.M.L.C., Berrill, M., Catravas, J., & Marik, P.E. Quercetin and vitamin C: an experimental, synergistic therapy for the prevention and treatment of SARS-CoV-2 related disease (COVID-19). *Front. Immunol.***11,** 1451 (2020).
41. Mani J.S *et al.* Natural product-derived phytochemicals as potential agents against coronaviruses: a review. *Virus Res.***284,** 197989. (2020).
42. Issa S.S. *et al.* The Main Protease of SARS-CoV-2 as a Target for Phytochemicals against Coronavirus. *Plants*. **11,** 1862 (2022).
43. AP11.0, SimulationsPlus LLC, Lancaster, CA.
44. Alici, H., Tahtaci, H., & Demir, K. Design and various in silico studies of the novel curcumin derivatives as potential candidates against COVID-19 -associated main enzymes. *Comput. Biol. Chem.* (2022).
45. Mulu, A. *et al.* The impact of curcumin-derived polyphenols on the structure and flexibility COVID-19 main protease binding pocket: a molecular dynamics simulation study. *PeerJ*. **9,** e11590 (2021).
46. Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., & Soetjipto, S. Potential Inhibitor of COVID-19 Main Protease (M^{pro}) From Several Medicinal Plant Compounds by Molecular Docking Study. Preprint at https://doi.org/10.20944/preprints202003.0226.v1 (2020).
47. Ghosh, A. *et al.* Structure-activity relationship (SAR) and molecular dynamics study of withaferin-A fragment derivatives as a potential therapeutic lead against the main protease (Mpro) of SARS-CoV-2. *J. Mol. Model*. **27,** 97 (2021).
48. Chakraborty, S. *et al.* The Natural Products Withaferin A and Withanone from the Medicinal Herb Withania somnifera Are Covalent Inhibitors of the SARS-CoV-2 Main Protease. *J. Nat. Prod*. **85,** 2340-2350 (2022).
49. Zhang, Y. *et al.* Structure-Based Discovery and Structural Basis of a Novel Broad-Spectrum Natural Product against the Main Protease of Coronavirus. *J. Virol.***96,** e0125321 (2022).
50. Ma, C. *et al.* Ebselen, Disulfiram, Carmofur, PX-12, Tideglusib, and Shikonin Are Nonspecific Promiscuous SARS-CoV-2 Main Protease Inhibitors. *ACS Pharmacol. Transl. Sci*. **3,** 1265-1277 (2020).
51. Gupta, S. *et al.* Identification of potential natural inhibitors of SARS-CoV2 main protease by molecular docking and simulation studies. *J. Biomol. Struct. Dyn*. **39,** 4334-4345 (2021).

52. Lachance, H. *et al.* Charting, navigating, and populating natural product chemical space for drug discovery. *J. Med. Chem.* **55,** 5989-6001 (2012).
53. Kaur, S. *et al.* How do plants defend themselves against pathogens-Biochemical mechanisms and genetic interventions. *Physiol. Mol. Biol. Plants*. **28,** 485-504 (2022).
54. Protein Data Bank. Retrieved August 29, 2023, from https://www.rcsb.org
55. Zhang, L., & Hilgenfeld, R. Crystal structure of SARS-CoV-2 Mpro in complex with the activity-based probe, biotin-PEG(4)-Abu-Tle-Leu-Gln-vinylsulfone. https://doi.org/10.2210/pdb6Z2E/pdb (2020).
56. Dai, W., Zhang, B., Jiang, X.M., Su, H., & Li, J. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*.**368,** 1331-1335 (2020).
57. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for the design of anti-SARS drugs. *Science*. **300,** 1763-1767 (2020).
58. Cherrak, S.A., Merzouk, H., & Mokhtari-Soulimane, N. Potential bioactive glycosylated flavonoids as SARS-CoV-2 main protease inhibitors: A molecular docking and simulation studies. *PLoS One*. **15,** e0240653 (2020).
59. Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., & Lukacik, P. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun*. **11,** 5047 (2020).
60. Abraham, M.J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1,** 19-25 (2015).
61. Huang, J., & MacKerell, A.D. Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem*. **34,** 2135-2145 (2013).
62. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926 (1983).
63. Darden, T., York, D., & Pedersen, L. Particle mesh Ewald: An *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys. **98**, 10089-10092 (1993).*
64. Singh,K. & Briggs, J.M. Impact of lymphoma-linked Asn11Tyr point mutation on the interaction between Bcl-2 and a BH3 mimetic: Insights from molecular dynamics simulation. *Chem. Biol. Drug Design.***95,** 435-450 (2020).
65. Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A., & Haak, J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.***81,** 3684-3690 (1984).
66. Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.***18,** (1997).
67. Humphrey, W., Dalke, A., & Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics*. **14,** 33-38 (1996).
68. Pettersen, E. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comp. Chem*. **25,** 1605-1612 (2004).
69. Schrödinger Release 2022-4: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2022.
70. Banks, J.L. *et al.* Integrated modeling program applied chemical theory (IMPACT). *J. Comp. Chem.***26,** 1752-1780 (2005).
71. Schrödinger Release 2022-4: Glide, Schrödinger, LLC, New York, NY, 2022
72. Schrödinger Release 2022-4: LigPrep, Schrödinger, LLC, New York, NY, 2022
73. Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner RA. The VSGB 2.0 model: a next-generation energy model for high-resolution protein structure modeling. *Proteins*. **79,** 2794-2812 (2011).
74. Cappelli, C.I., Manganelli, S., Lombardo, A., Gissi, A., & Benfenati, E. Validation of quantitative structure-activity relationship models to predict water-solubility of organic compounds. *Sci. Total Environ*. **463-464,** 781-789 (2013).
75. Dearden, J.C. In silico prediction of aqueous solubility. *Expet Opin. Drug Discov*. **1,** 31-52 (2006).

Figures

Figure 1

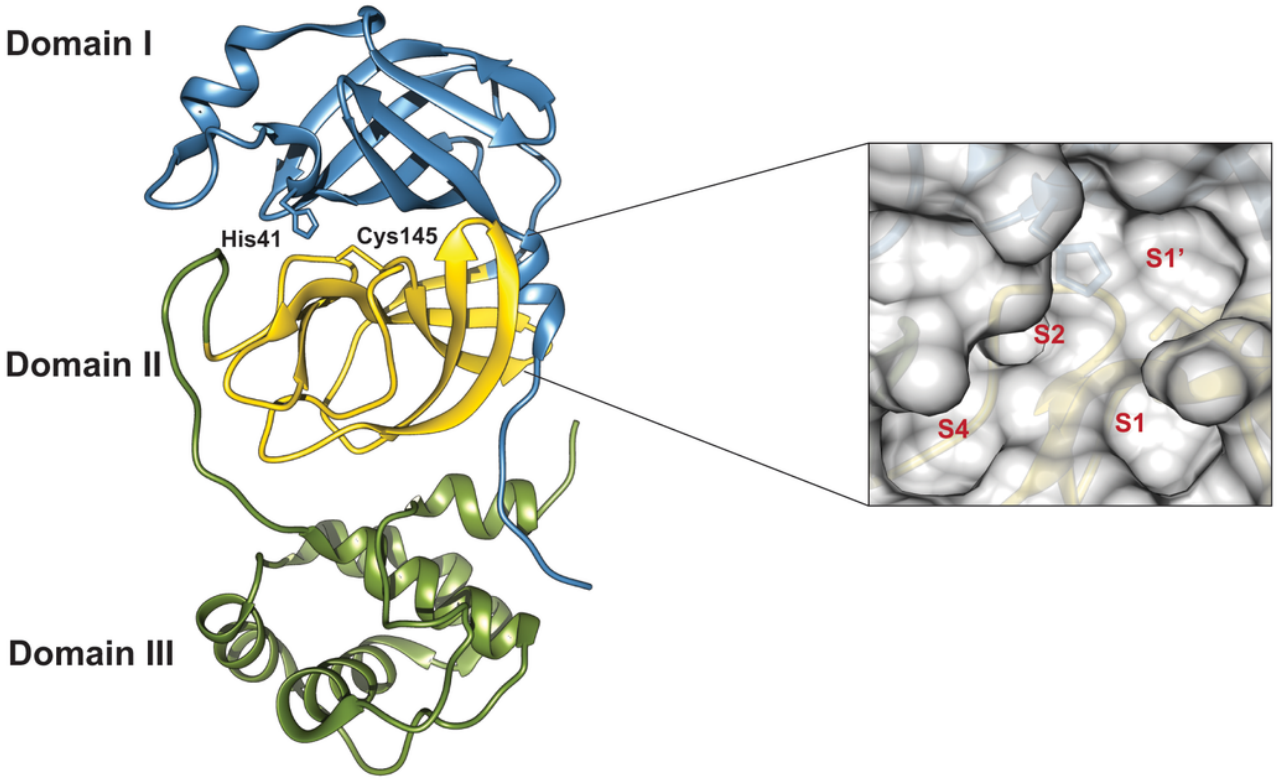

Figure 1

[Left] Cartoon representation of one protomer of the dimeric M^{Pro} protein. The three distinct domains of the protein are indicated in blue (Domain I), yellow (Domain II), and green (Domain III). The catalytic residues His41 (blue) and Cys145 (yellow) are represented in sticks. [Right] Surface representation of the substrate-binding pocket, including S1, S1', S2, and S4 subsites.

Figure 2

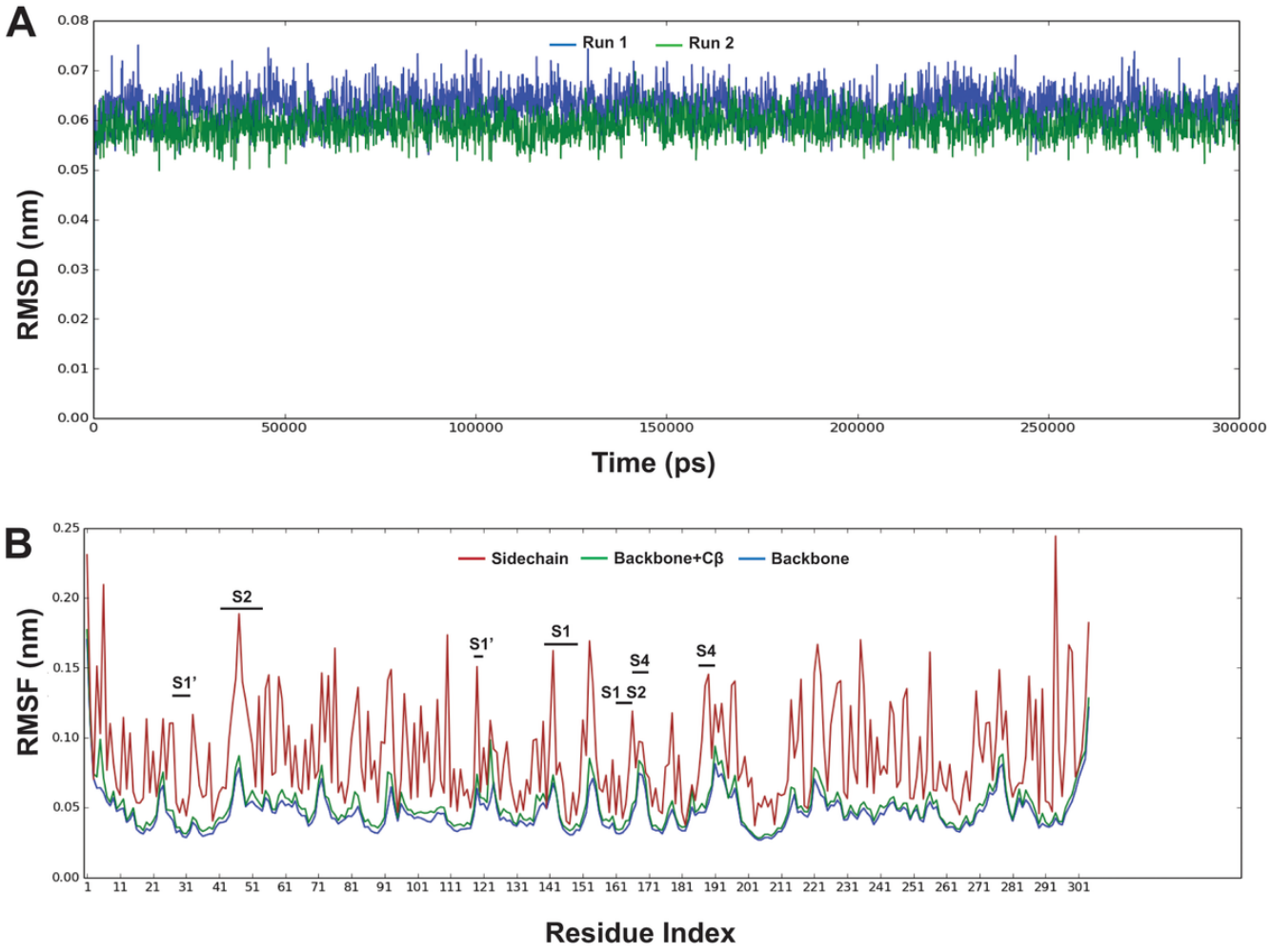

Figure 2

RMSD and RMSF plots of the MD simulations of M^{Pro} generated over 300 ns using GROMACS. (A) RMSD of Cα atoms versus time graph of the apo M^{Pro} protein. The data is collected from two independent runs. (B) Analysis of RMSF trajectories of sidechain, backbone + Cβ, and the backbone atoms of the residues.

Figure 3

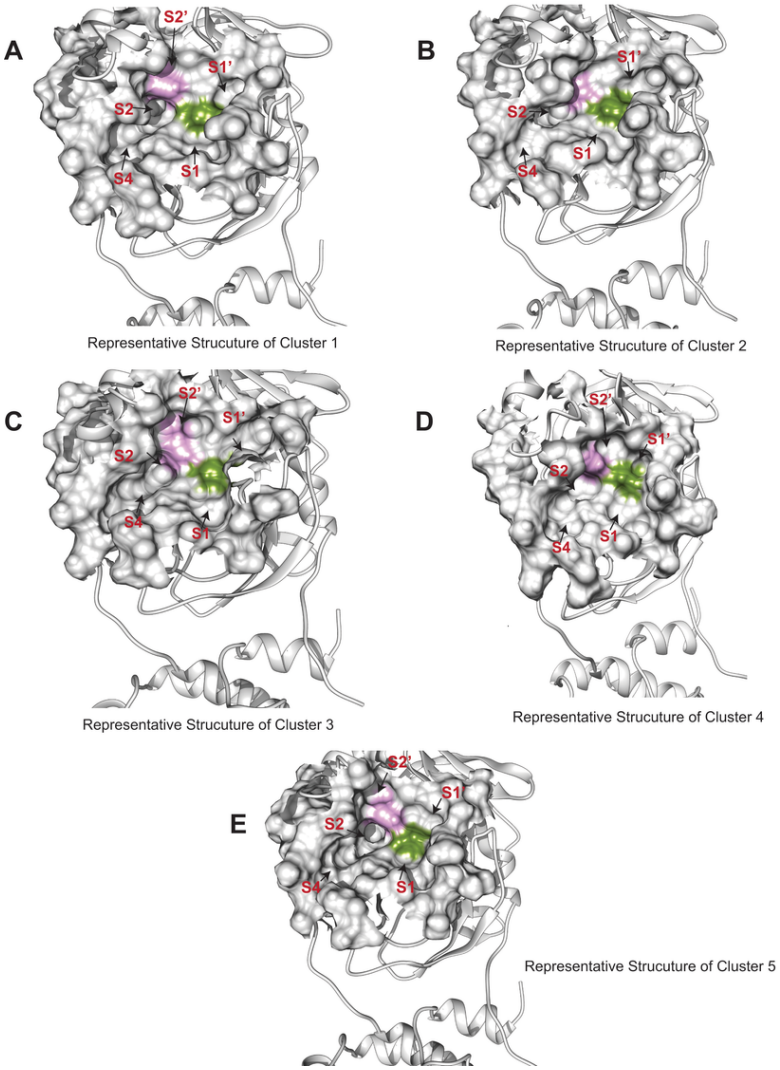

Figure 3

Surface depictions showcasing the binding site of the representative structures from each of the five clusters. The catalytic residues, His41 and Cys145 are colored in pink and green, respectively.

Figure 4

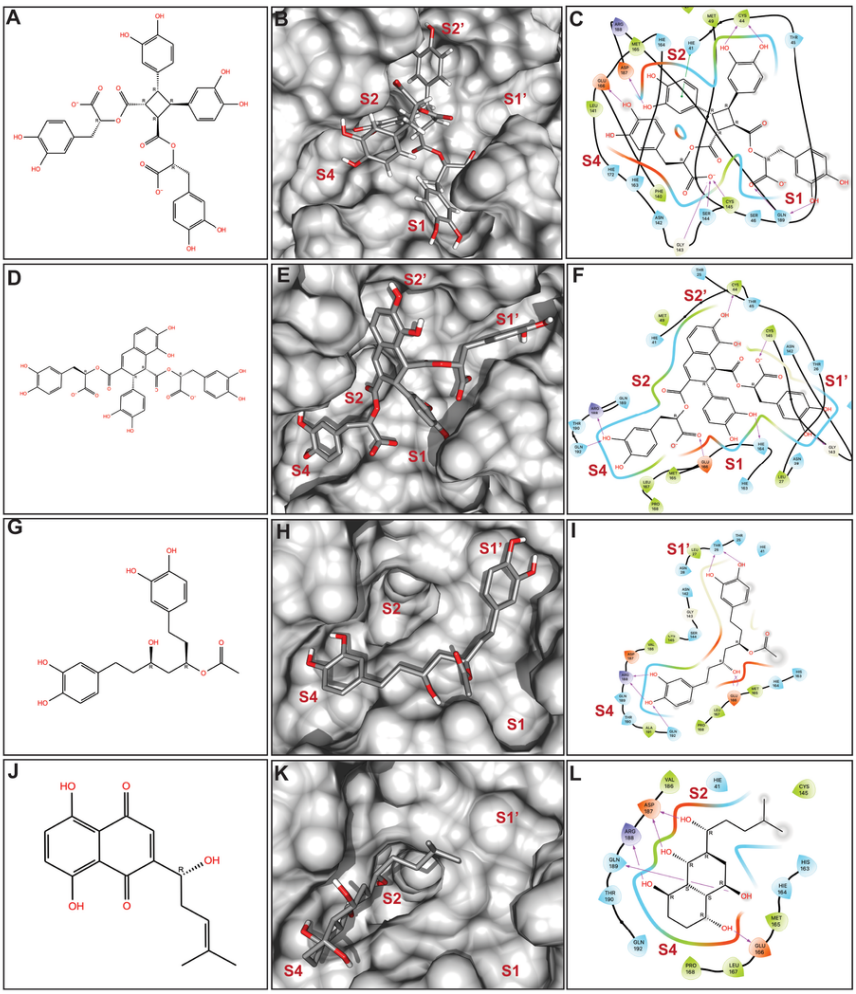

Figure 4

Representation of the phytochemical structure, its binding conformation, and residue interaction diagram of the non-glycoside phytochemicals with high docking scores for all five protein conformations. (A, B, and C) Shimobashiric acid C (SAC), (D, E, and F) Salvianolic acid L (SAL), (G, H, and I) AHDPH, and (J, K, and L) Shikonin.

Figure 5

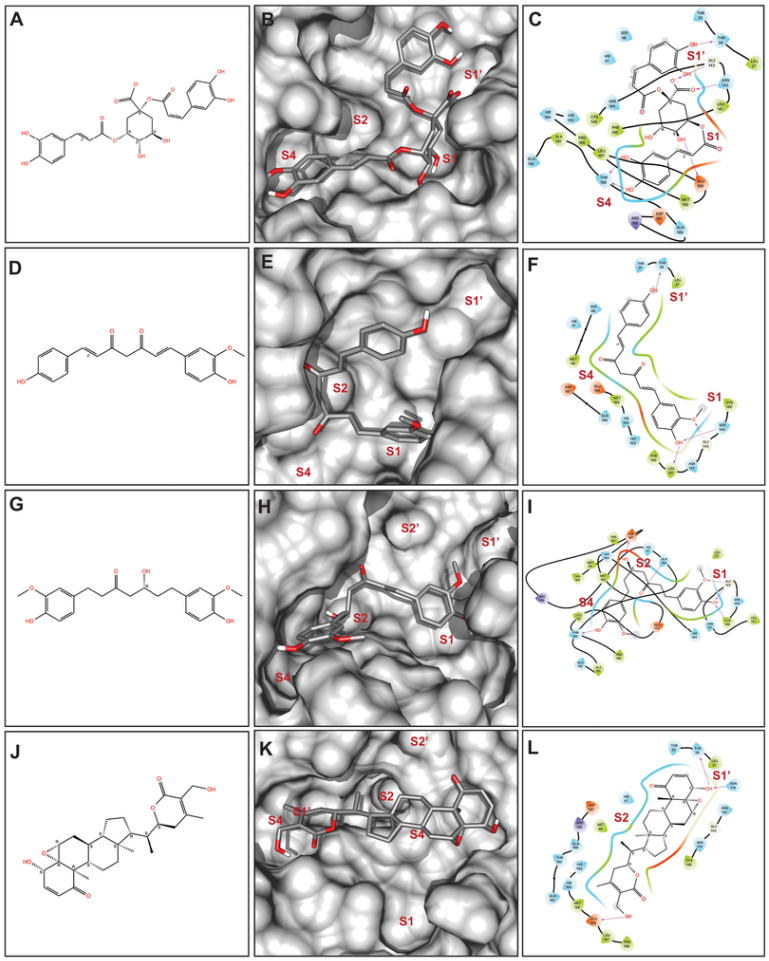

Figure 5

Representation of the phytochemical structure, its binding conformation, and residue interaction diagram of phytochemicals with high docking scores for at least one protein conformation. (A, B, and C) Cynarin, (D, E, and F) Demethoxycurcumin, (G, H, and I) Hexahydrocurcumin, and (J, K, and L) Withaferin A.

Figure 6

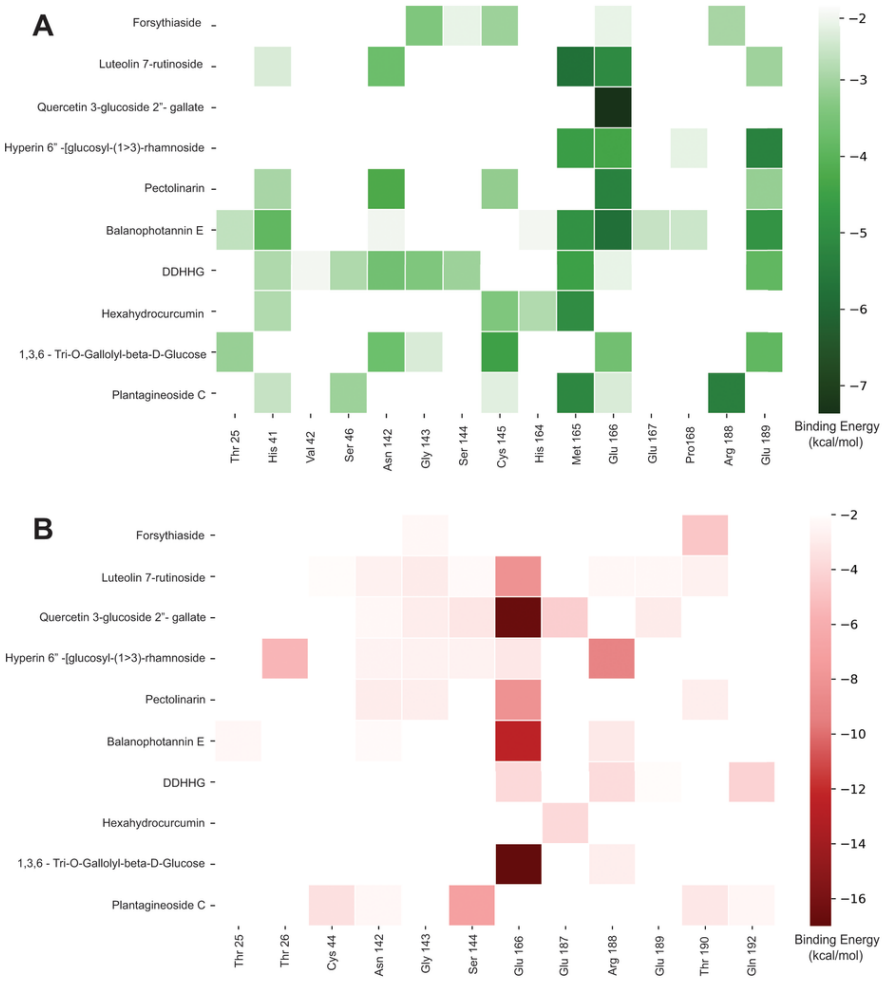

Figure 6

Heatmap illustrating key residues of the M^{Pro} binding site interacting with the top ten phytochemicals based on MM-GBSA binding energy. (A) Residues exhibiting van der Waal interaction energy above -2 kcal/mol. (B) Residues exhibiting columbic interaction energy above -2 kcal/mol.

Figure 7

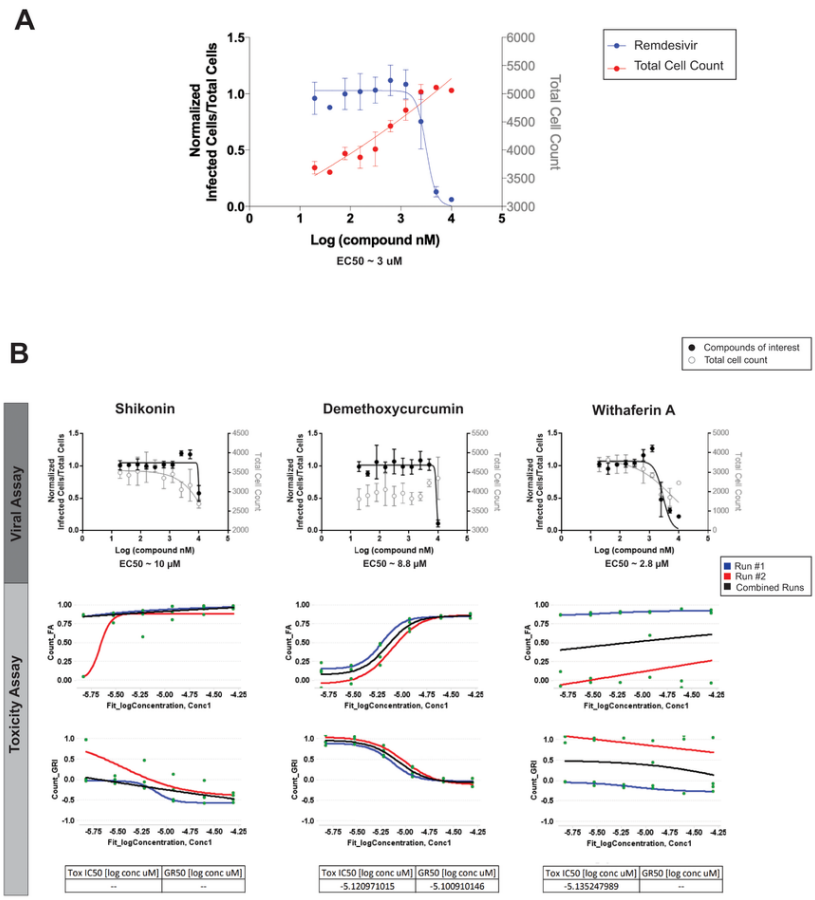

Figure 7

SARS-Cov-2 replication inhibition and Cytoxicity of A) Shikonin,

B) Demethoxycurcumin, C) Withaferin A. FA denotes the fraction of cells affected and GRI is the effect of the compound at an infinite concentration.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.pdf
- SupplementaryTable2.pdf