Discovery of Multi-Target-Directed Ligands by Targeting Host-specific SARS-CoV-2's Structurally Conserved Main Protease

Rakesh Joshi[1,2*#], Shounak Jagdale[1#], Sneha Bansode[1#], S. Shiva Shankar[1,2#], Meenakshi Tellis[1,3#], Vaibhav Kumar Pandya[1#], Ashok Giri[1,2*], Mahesh Kulkarni[1,2*]

[1]Biochemical Sciences Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India

[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, Uttar Pradesh, India

[3]Department of Botany, Savitribai Phule Pune University, Ganeshkhind, Pune 411007, Maharashtra, India

*Corresponding authors

Dr. Rakesh Joshi, Dr. Mahesh Kulkarni, Dr. Ashok Giri

Biochemical Sciences Division,

CSIR – National Chemical Laboratory

Dr. Homi Bhabha Road, Pune – 411 008, INDIA

Email: <u>rs.joshi@ncl.res.in</u>, <u>mj.kulkarni@ncl.res.in</u>, <u>ap.giri@ncl.res.in</u>

Tele (office): 020-2590-2541

[#]Authors contributed equally

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection has resulted in the current COVID-19 pandemic. Worldwide this disease has infected around 1.5 million individuals with a mortality rate ranging from 5 to 10%. It has also imposed extreme challenges on global health, economy, and social behavior. Due to the unavailability of therapeutics, several efforts are going on in the drug discovery to control the SARS-CoV-2 viral infection. The main protease (M^{Pro}) plays a critical role in viral replication and maturation, thus can serve as the primary drug target. To understand the structural evolution of M^{Pro}, we have performed phylogenetic and SSN analysis, that depicted divergence of Coronaviridae M^{Pro} in five clusters specific to viral hosts. This clustering was also corroborated with the comparison of M^{Pro} structures. Furthermore, it has been observed that backbone and binding site conformations are conserved despite variation in some of the residues. This conservation can be exploited to repurpose available viral protease inhibitors against SARS-CoV-2 M^{Pro}. In agreement with this, we performed screening of the custom-made library of ~7100 molecules including active ingredients present in the Ayurvedic anti-tussive medicines, anti-viral phytochemicals and synthetic anti-virals against SARS-CoV-2 M^{Pro} as the primary target. We identified several natural molecules that strongly binds to SARS-CoV-2 M^{Pro} among which top seven molecules are δ-Viniferin, Myricitrin, Taiwanhomoflavone A, Lactucopicrin 15-oxalate, Nympholide A, Biorobin and Phyllaemblicin B. Most of the predicted lead molecules are from *Vitis vinifera*, also reported for anti-tussive and/or antiviral activities. These molecules also showed strong binding with other main targets RdRp and hACE-2. We anticipate that our approach for identification of multi-target-directed ligand will provide new avenues for drug discovery against SARS-CoV-2 infection.

Keyword: Coronavirus; COVID-19; hACE-2; M^{Pro}; Multi-target-directed ligand; Protease inhibito; RdRp; SARS-CoV-2 virus

1. Introduction

Coronaviruses, a large family of viruses, causing upper-respiratory-tract infections in humans and other higher mammals[1]. Coronavirus outbreak has been reported three times in the 21st century namely SARS in 2002, MERS in 2012, and COVID-19 in November 2019. The recent outbreak of COVID-19 caused by SARS-CoV-2 has emerged from China. Human-human transmission of this highly infectious zoonotic virus has led to exponential growth in the number of infected cases, which resulted in its pandemic outbreak worldwide[2]. Currently, more than 200 countries are affected due to COVID-19 with more than 1 million people around the world and approximately 5% mortality rate[2]. Additionally, the lives of millions of people have been impacted due to mandatory lockdowns, isolations, and quarantines. Thus the severe effect of the COVID-19 outbreak has imposed major challenges for global health, society and economy[3]. Currently, there are no specific antiviral drugs or vaccines available for the treatment and management of COVID-19, which is further making the situation difficult to handle. At present preventive and supportive therapies are being implemented to prevent complications[4].

For drug discovery purposes, the main protease (M[Pro]) of coronaviruses has been studied extensively. They are papain-like proteases involved in the self-maturation and processing of viral replicase enzymes[5]. Due to their key role in virus replication, inhibition of these proteases poses to be a drug target. Furthermore, with their very low similarity with human proteases, inhibitors of M[Pro] are found to be very less cytotoxic[5,6]. Preliminary studies have suggested the potential use of protease inhibitor lopinavir/ritonavir, commonly used drugs for human immunodeficiency virus (HIV), for the treatment of COVID-19 patients[7]. Unfortunately, in the open-label randomized clinical study, these drugs are not found to be that impressive for COVID-19 treatment. Additionally, several other viral protease inhibitors like HCV Protease Inhibitor Danoprevir, HIV protease inhibitor Darunavir are under *in vivo* and clinical studies for the treatment of SARS-CoV-2 infection[8]. To find new or repositioning existing drug molecules, the understanding of sequence and structure of SARS-CoV- M[Pro] in regards to various other coronavirus strains could be important. Their comparative

understanding is enigmatic, which could be important to repurpose available antiviral protease inhibitors against SARS-CoV-2 M^{Pro}.

Along with M^{Pro}, RNA-dependent RNA polymerase (RdRp), which carries out the synthesis of viral RNA from RNA templates and is involved in the replication and transcription of viral genome can serve as a lucrative target[9]. Remdesivir, a nucleotide analog, has shown broad-spectrum antiviral activities, and preclinical studies have also shown promising human safety data[8]. Recently, a report suggested that the significant inhibition of RdRp of SARS-CoV-2 by Remdesivir, and clinical studies for the same are ongoing to evaluate the efficacy of this molecule in the COVID-19 patients (www.clinicaltrials.gov; Study No: NCT04280705). Apart from the maturation and replication, the entry of this virus by binding its surface spike protein to the human Angiotensin-converting-enzyme 2 receptor (hACE-2) is also a crucial process that can be targeted[10,11]. Since this host cell receptor is essential for the virus entry; targeting hACE-2 has a promise for preventing SARS-Cov-2 infection. Hoffman et al. have shown that clinically approved serine protease (TMPRSS2) inhibitor can also bind to the SARS-CoV-2 receptor hACE-2 and inhibit viral entry[12]. To date, the Chinese Clinical Trial Registry has recorded around 550 trials against SARS-CoV-2, mostly evaluating existing drug molecules. These trials include the application of antiviral (favipiravir, adalimumab, dihydroartemisininpiperaquine, leflunomide, lopinavir), antimalarial drugs (chloroquine or hydroxychloroquine), high-dose Vitamin C, etc (www.chictr.org.cn/index.aspx). From the available data of experimental and clinical studies, it has been found that molecules targeting single protein will be ineffective antiviral lead. Thus, to devise an effective strategy, there is a need for molecules that can target multiple key proteins such as RdRp and hACE-2 along with the M^{Pro} as a major drug target[13].

In silico screening of the drugs is a very vital and useful tool for rapid screening to meet the urgent demand for repurposing drugs for the treatment of SARS-CoV-2 infection. Therefore, in this study, we have performed a comparative analysis of the major target of SARS-CoV-2, M^{Pro} for its structural characteristics as compared to other viruses from the Coronoviridea family. This information was further used for the virtual

screening of the custom-made library of phytochemicals, active ingredients present in the commonly used ayurvedic anti-tussive medicines in India, and the synthetic anti-viral drugs against M^{Pro}. Top hit molecules from this screen were then docked against SARS-CoV-2 RdRP and hACE2 to find the multi-target-directed molecules. We believe that this study will provide a new approach and a base for the discovery of multi-targeted antiviral molecules against SARS-CoV-2 pathogenesis. Thus these compounds could be potential candidates for *in vitro* and *in vivo* anti-viral studies followed by clinical treatment of SARS-CoV-2.

2. Results and discussion

2.1 Main protease shows divergence due to sequence variation

Phylogenetic analysis of proteases (M^{Pro}) helped to understand the structural-functional variation of SARS-CoV-2 concerning other viruses from the Coronaviridae family (Figure 1A). This analysis suggested that the M^{Pro}'s from the Coronaviridae family are distributed in five clusters. It has been observed that these proteases are separated in different clades depending on their zoonotic sources. In the clade 1 (orange colour), which appears to be the ancestral clade in the phylogenetic tree have sequences from Porcine transmissible gastroenteritis and pandemic diarrhea coronavirus. This clade appears to be evolutionary primitive to avian coronavirus but also showed coevolutionary traits. In the clade 2 (cyan), most of the sequences are from the strains of Avian infectious bronchitis virus. It is then followed by clade 3 (blue), which mainly consists of the sequences from different strains of Bat coronavirus. It is also found to be the parental clade for the clade 4 (red) and the clade 5 (green), which consist of predominantly mammalian host coronaviruses. It is interesting to see that the clade 4, which comprises viruses from bats and the majority of human infectious coronaviruses like SARS, MERS and also recent SARS-CoV-2. Our phylogenetic analysis revalidated the bat origin of the SARS-CoV-2[14]. While the clade 5 has sequences from Bovine coronaviruses, it also has subclades that are indicative of a distinct host range. Part of the Clade 5 is from the bovine coronavirus and the other part is from murine coronavirus. Between these two subclades, there is a set of sequences that forms clad

5.2 comprises of human coronaviruses like strain OC43 (P0C6U7) and HKU1 (P0C6X3) which have bovine or murine as primary or intermediate host **(Figure 1A)**. As evident in the multiple sequence alignment **(Supplementary Figure 1)**, the M[Pro] sequences from various coronaviruses showed very high conservation in the active site. However, the sequence variation around the active site can lead to the separation of these proteases in distinct clades depending on their hosts.

Furthermore, the sequence similarity network (SSN) analysis was carried out to understand the similarities in M[Pro] sequences from viruses of the Coronaviridae family. It was noted that till e-value 1E-80, all the sequences were clustered together suggesting an extremely high sequence similarity amongst all the Coronaviridae proteases **(Supplementary Figure 2).** At e-value 1E-140, the sequences were segregated in 5 clusters **(Figure 1B)**. The largest cluster (green) consisted of sequences from bovine coronavirus, murine coronavirus and some human coronavirus such as OC43 and HKU1. The second cluster (orange) exclusively consisted of the porcine coronaviruses and related strains. Bat coronaviruses were grouped in a single cluster (blue), while cyan cluster housed avian coronaviruses. The red cluster consisted of human coronaviruses including SARS, MERS, SARS-CoV2 along with various other bat coronaviruses. The small cluster of 2 sequences that separated from the red cluster at e-value 1E-125 also belongs to the bat coronaviruses. The clusters of SSN are showing a high correlation with the clades of the phylogenetic tree (Figure 1A) suggesting high evolutionary divergence amongst the coronaviruses depending on their host reservoir. This high specificity of the variations related to the host might have resulted from the natural selection in the animal hosts before human transmission[14].

2.2 Coronavirus main protease structural backbone and active site conformation are conserved despite sequence variations

A comparison of 49 structures of M[Pro] from different coronavirus strains using the DALI server showed five clusters similar to the phylogenetic tree **(Figure 2A; Supplementary Data 1)**. Clusters 1 to 3 contains protease structures from porcine transmissible gastroenteritis and pandemic diarrhea coronavirus. Also, they include the

M^{Pro} structure from human coronavirus strains like NL63 and HUK. Furthermore, clusters 4 and 5 mostly consist of protease structures from the coronavirus strains causing respiratory infections like SARS, MERS and SARS-CoV-2. The structural similarity matrix with hierarchical clustering also corroborates with the phylogenetic tree and structural cladogram showing the classification of M^{Pro} in five major clusters **(Figure 2B)**. These results indicate that sequence variations in the M^{Pro} are also reflected in the structural variations resulting in similar clustering at the sequence and structure level.

Structural superimposition of representative proteases from each cluster of the structure cladogram and the similarity matrix depicted that the backbone of these proteases is conserved with RMSD less than 1Å in the majority of cases (**Figure 3A**). It suggests that the sequence variation leads to minor structural changes without major changes in the conformation. Furthermore, a comparison of the substrate-binding pocket of these selected proteases shows that active site residue conformation is highly conserved **(Figure 3B).** There are some variations in the binding pocket residues as shown in the multiple sequence alignment, but these variations are subtle, and overall binding pocket conformation is conserved. Structural conservation of backbone and binding pocket of M^{Pro} from various coronavirus strains suggest that substrate recognition and inhibitor binding would more or less likely to be similar in them. Hegyi et al. have shown the conservation in the substrate specificities and proteolytic processing of replicase polyproteins in main proteases across coronaviruses[15].

This similarity highlights the development of universal M^{Pro} assays and the design of broad-spectrum inhibitors of M^{Pro} from different coronavirus strains. This also suggests that the M^{Pro} inhibitor-based drug repositioning would be the rapid and near to accurate approach to screen the drugs against SARS-CoV-2 infection.

2.3 Inter-residue contacts and the structural flexibility is located dominantly in the N-terminal catalytic site harboring beta-barrels

Besides structural conservation, it is important to understand inter-secondary elements and inter-residue contacts within the protein to unravel the mechanism of these proteases. Secondary structure topology indicates that proteins are

asymmetrically populated by the beta-sheets at the N- terminus and helix at C- terminus **(Figure 4A)**. The chord plot between pairs of secondary structural elements (arc of a circle) displays the number of interactions (chord thickness) between the corresponding elements. In M^{Pro}, the contacts between the secondary structural elements indicated that the majority of contacts are present in the N-terminus beta-sheets, which makes the juxtaposed beta barrels connected with the loop that harbors substrate binding site.

Whereas, C-terminus globular cluster of helix showed comparatively less number of contacts indicating the rigid nature of these helix cluster involved in the protease dimerization **(Figure 4B)**. Inter-residue contact map showed that the population of inter-residue contact is far denser in the first 200 amino acids, which consist of beta barrels having the catalytic role. While C-terminus residues show less number of contacts suggesting the involvement of this part of the protein (globular cluster of helices) in a structural role **(Figure 4C).** To validate the flexibility of the beta-barrels, the normal mode simulation was performed to map the Root Means Square Fluctuations of the residues from its mean position. Structural trajectory and the fluctuation map indicate that most of the residues present in the beta barrels have fluctuation above 1Å, while the C-terminus is mostly rigid with very few residues having fluctuation above 1Å **(Figure 4D and 4E)**. Though the N-terminus show high fluctuations, the active sites (H41 and C145) showed very less variation, which is important to maintain the required conformation of these residues in the substrate-binding pocket. The high frequency of inter secondary elements and residue contacts in the N terminus might provide structural flexibility for the proteolytic processing of large replicase polypeptide and they are potential sites to be targeted while designing or screening the inhibitors[16].

2.4. Anti-tussive molecules like δ-Viniferin and myricitrin showed higher affinity for the active site of SARS-CoV-2 M^{Pro} and other targets

Screening of the custom made the library of ~7100 molecules comprising of different classes such as flavonoids, glucosinolates, anti-tussive, anti-influenza, anti-viral, terpenes, terpenoids, alkaloids and other predicated anti-COVID-19 molecules was done against SARS-CoV-2 M^{Pro} as the main target **(Supplementary Data 2)**. Sequence

similarity search of M^{Pro} against human protein sequences was performed, which suggests that the viral proteases do not have any homolog in the human, thus targeting M^{Pro} will be specific to viral proteases only. Docking analysis results demonstrated that several phytochemicals and anti-viral have strong binding in the active site of SARS-CoV-2 M^{Pro}. The selected molecules also displayed a high binding affinity to SARS-CoV-2 RdRp and hACE-2 **(Supplementary Data 3)**. Likewise M^{Pro}, other viral targets considered in this study, RdRp, also showed that there is no human homolog present, thus could be a potential viral target with M^{Pro}. Top molecules with the highest binding energy and with acceptable ADME criteria were further selected for interaction analysis **(Figure 5)**.

As one of the symptoms of COVID-19 is cough, a library of ayurvedic anti-tussive molecules in Indian medicine (82 molecules) are screened against SARS-CoV-2 M^{Pro}. The resulting top 20 molecules were further screened against other targets- RdRp and hACE 2 receptors. Amongst these molecules, δ-Viniferin, chrysanthemin, myritilin and myricitrin showed strong binding with SARS-CoV-2 M^{Pro}. The δ-Viniferin and myricitrin showed high solubility and bioavailability in computational ADME analysis; thus, analyzed for intermolecular interaction. With SARS-CoV-2 M^{Pro}, δ-Viniferin exhibited binding energy of -8.4 Kcal/mol with several non-covalent interactions in the active site and substrate-binding site. THR24 and HIS163 form polar contact with hydroxyl groups of δ-Viniferin, while active site residue HIS41 shows Pi-Pi interaction with the central aromatic scaffold of a ligand. Apart from this, MET165 is involved in Pi-sulfur interaction with ligand and other residues from the substrate-bindingg pocket, which are involved in either attractive van der Waal interaction or carbon-hydrogen bond **(Figure 6A)**. In addition to M^{Pro}, δ-Viniferin appeared to have a strong binding with the SARS-CoV-2 RdRp and hACE2 receptor with a score of -8.3 and -7.4 Kcal/mol, respectively. In the case of RdRp, binding pocket residues ALA762, TRP800, SER814 showed hydrogen bonds with the hydroxyl group of the ligand. While, active site and other key residues namely ARG553, ASP618, ASP623 and ASP761 display Pi interactions with the aromatic scaffolds of δ-Viniferin. The high density of these intermolecular non-covalent interactions is an indicator of strong and specific binding of δ-Viniferin with RdRp active

pocket. Furthermore, hydroxyl and aromatic groups of δ-Viniferin also interacts through polar contacts and Pi-Alkyl interaction with the GLU35, LEU39, LYS68 and ALA71 residues of hACE-2 **(Figure 6B and 6C)**. The interaction of δ-Viniferin with M[Pro], RdRp and hACE-2 suggests its high potential as a multi-target directed ligand against SARS-CoV-2. δ-Viniferin is a resveratrol dimer, a major stilbene produced by stressed grapevine leaves[17]. It is also one of the major stilbenes present along with resveratrol in the red wine[18]. Previous studies have also suggested that δ-Viniferin shows potent antiviral activity against a variety of viruses viz. rotavirus[19], Human Immunodeficiency Virus (HIV)[20]; hepatitis C virus (HCV)[21].

Another anti-tussive molecule, myricitrin, found to have a strong binding with the active site residues of the M[Pro] with a binding score of -8.9 Kcal/mol. Active site residues HIS41 and CYS145 form Pi-Alkyl and Pi-Sulfur interactions with the aromatic scaffold of ligand, respectively. TYR54, PHE140, GLY143, HIS163 and GLU166 are involved in the formation of 6 hydrogen bonds with hydroxyl and carboxyl oxygen of ligand. Apart from this, there are other additional interactions like van der Waals forces that stabilize the binding of the ligand to the enzyme **(Figure 6D)**. Myricitrin exhibited equally strong interaction with RdRp and hACE-2 with a binding score of -7.9 and -7.1 Kcal/mol, respectively. For RdRp, the carboxyl group of myricitrin forms two hydrogen bonds with ARG553, ARG555 and THR556. While the active site residue ASP623 is having Pi-Anion interaction with the aromatic rings of the ligand. Similarly, this ligand binds to hACE-2 by forming polar interaction with ALA348, PHE390, ARG393 and ASN394. This interaction is then further stabilized by Pi-stacking of PHE40 and Pi-anion interactions of ASP350, ASP382 with aromatic rings of the ligand **(Figure 6E and 6F)**.

Myricitrin is a glycosylated analog of myricetin, present in the *Myrica esculenta*. Although in our docking study, myricetin has shown relatively lesser binding affinity when compared to myricitrin, previous studies have shown that it possesses potent antiviral activity. Myricetin inhibited VP35 protein with double-Stranded RNA Interaction in Ebola virus[22]. The glycosylated myricetin, i.e. myricitrin has been found to inhibit HIV[23], herpes simplex virus[24], SARS coronavirus[25,26].

Together, our results suggest that δ-Viniferin and myricitrin could be potent molecules for the mitigation of SARS-CoV-2. Interestingly, top anti-tussive molecules in the present study mainly δ-Viniferin, chrysanthemin, myritilin and myricitrin are present in the extract of black grapes, which is one of the key ingredients of ayurvedic antitussive medicines, and energy and immune booster, Chyawanprash[27]. Thus, these medicines could be helpful in the management and mitigation of COVID-19. Particularly, the herbal formulation and beverages containing *Vitis vinifera* might be helpful in the management of COVID-19.

2.5. Potential of bioactive molecules from the medicinal plant as SARS-Cov-2 M^{Pro} inhibitor and multi-targeted drug ligand

The tapping of natural molecule libraries to predict inhibitors of M^{Pro} showed that several active biomolecules from different medicinal plants have a strong affinity against SARS-CoV-2 M^{Pro}. The flavonoid group of compounds showed a high binding score with all three targets. One of the molecules from this class, Taiwanhomoflavone A, isolated from *Cephalotaxus wilsoniana* showed a strong binding affinity with M^{Pro} (-9.6 Kcal/mol). *In silico* ADME study indicated that this molecule possesses high bioavailability and likeliness to be the lead like compound. Interaction analysis depicted that reactive groups of Taiwanhomoflavone A form hydrogen bonds with THR199 and stabilizing forces are provided by electrostatic interaction with LEU189 residue of M^{Pro}. The bulkiness of this ligand leads to a large area of interaction that may lead to a higher contribution of attractive van der Waal's force between ligand and binding site residues **(Figure 7A)**. Furthermore, taiwanhomoflavone A also showed extremely strong binding with the RdRp and hACE2 receptor, with a binding score of – 10.6 Kcal/mol and -7.6 Kcal/mol, respectively. The RdRp binding pocket residues LYS621 and TRP800 showed hydrogen bonds with the carboxyl group of the ligand. While, active site residues namely ASP760 and ASP761 form Pi interactions with the aromatic groups of Taiwanhomoflavone A. In the case of hACE-2, taiwanhomoflavone A makes polar contact with GLU75 and Pi interaction with GLU35 **(Figure 7B and 7C)**. Taiwanhomoflavone A has been explored for its anticarcinogenic, anti-inflammatory and

anti-neoplastic properties[28]. It would be interesting to consider this molecule further for anti-viral evaluation *in vitro* and *in vivo*.

In the terpenoids group, lactucopicrin 15-oxolate, a derivative of lactucin and oxalic acid showed higher interaction. The lactucarium compounds are found in *Lactuca virosa* (wild lettuce) and also in the latex of *Lactuca* species. The Lactucopirin 15-oxolate has a binding energy of -8.2 Kcal/mol with M^{Pro} by the virtue of three hydrogen bonds to SER144, GLY143, and CYS145 of the active site. It also showed a high binding affinity with RdRp and hACE2 receptor with the binding score of -8.7 Kcal/mol and -8.2 kcal/mol, respectively **(Figure 7D)**. With RdRp lactucopirin 15-oxolate makes hydrogen bonds at ARG553, ARG555, and THR556 residues, and it is stabilized by hydrophobic interactions of ASP623 and THR687 **(Figure 7E)**. In the case of the hACE2 receptor, it forms three hydrogen bonds with residues ALA99, TYR385, ASN394 and Pi-stacking interaction with PHE40 **(Figure 7F)**. The lactucopicrin and lactucin are known for their anti-malarial activity[29]. Based on the *in silico* ADME studies, the compound is found to be highly soluble, bioavailable, and high GI absorption criteria which can be utilized for further drug development against novel-SARS-CoV-2. Apart from this, Biorobin, Pedunculagin, Afzelin, Hyperin, Baicalin and other several molecules from alkaloid, flavonoids, and terpenoid that showed strong interaction with SARS-CoV-2 targets **(Table 1; Supplementary Data 3)**. These findings highlight that the traditional medicines and/or nutraceuticals along with synthetic drug molecules could be effective ways to tackle COVID-19.

2.6 Binding study repurposes application of anti-influenza and anti-HIV molecule against SARS-CoV-2 infection

Drug re-purposing is a strategy to investigate the already existing drugs for new uses. The advantage of repurposing is that the drugs are preclinically approved, and it helps in the formulation and development of the drug in ease. In this case, we have collected the existing anti-influenza and anti-viral drugs to study the treatment strategy for the SARS-CoV-2. We screened 17 compounds and found a macrocyclic derivative Nympholide A, a phytochemical present in the aquatic plant of Nymphaeaceae family[30]. It showed the highest binding energy with M^{pro} and RdRp with a score of -7.8 and -8.8

Kcal/mol, respectively. M^{pro} residues Met165, GLU166, LEU 167 have hydrogen bond interaction with the ligand. In the case of RdRp, residues LYS551, ARG553, LYS798, TYR 619 side chains have strong hydrogen bonding with the ligand and GLU811 makes the Pi-anion interaction which improves the binding in RdRp compared to protease. The hACE2 receptor has fewer interacting residues, as a result, shows a weaker affinity to Nympholide A **(Figure 8A-C)**.

In the case of antiviral compounds, Invirase (saquinavir) has the highest binding energy with M^{pro} (-9 Kcal/mol). RdRp and hACE2 targets have a binding energy of -8.6 and -7.5 Kcal/mol, respectively. Invirase aromatic sidechain makes Pi-sulfur and Pi-Pi stacking interaction with the Met49 and His41 residues of M^{pro}. These interactions help in strong binding to protease active site and further the interaction is stabilized by hydrogen bonding of GLY143, SER144, CYS145, GLU166 of M^{pro} residues **(Figure 8D)**. Similar to protease, the viral RdRp also have aromatic Pi-alkyl and Pi-anion interaction with ARG624 and ARG553 residues. The hydrogen bonding of ASP618, LYS551, LYS 621 residues of RdRp with the hydroxyl group of ligand give similar stability to protease. The hACE2 showed Pi-Pi stacking interaction of PHE40, PHE390 and one hydrogen bond of ASN394 with the ligand **(Figure 8E and 8F)**. Invirase has been known to inhibit the HIV-1 protease. In our study, the invirase has higher interaction with protease compared to RdRp and hACE-2[31]. As there are other viral protease inhibitors like danoprevir, ritonavir and lopinavir are being tested in clinical trials for potential treatment against SARS-CoV-2, it would be interesting to consider invirase for further studies[7].

3. Conclusion

This study provides a comprehensive overview of the structural characteristics of the Coronaviridea proteases and their potential inhibitors. Phylogenetic and SSN network analysis has illustrated the divergence in the proteases depending on their host reservoir suggesting the role of natural selection of host in conditioning the virus. Structural comparison depicted that the high structural similarities in the backbone and substrate binding pocket indicating conservation in the substrate specificities over different strains of Coronavirus. Asymmetrical structural flexibility in the M^{Pro} might be crucial for its activity, thus it can be targeted for inhibition. The structural similarity in the

binding pocket assisted the prediction of repositioning of known protease inhibitors as SARS-CoV-2 M^{Pro} inhibitors. Several molecules were identified as of the potent SARS-CoV-2 M^{Pro} inhibitors like delta δ-Viniferin, Myricitrin, chrysanthemin, myritilin, Taiwanhomoflavone A, Lactucopicrin 15-oxalate, Nympholide A, Biorobin, and Phyllaemblicin B. These molecules also equally showed strong binding with other targets of SARS-CoV-2, like RdRp and hACE-2. Interestingly, most of the potential lead molecules of our study are anti-tussive and are present in high amounts in the extract of black grapes, thus we speculate that the herbal formulation and beverages containing *Vitis vinifera* might be applicable in the management of COVID-19. These molecules could be potential lead molecules for multi-targeted drug development against SARS-CoV-2.

4. Materials and methods

4.1. Sequences retrieval and phylogenetic analysis

The amino acid sequences of main proteases from the Coronaviridae family were retrieved from the UniProtKB database (www.uniprot.org) depending on their similarity with COVID-19 M^{Pro}. The criteria for the selection of the sequences were 90% < coverage and similarity 50% <. A total of 57 sequences (**Supplementary Data 4)** were taken forward for the phylogenetic analysis by maximum likelihood (ML) method using MEGA6 software[32]. Jones-Taylor-Thornton (JTT) model was used for amino acid substitution with a bootstrap value of 1000. The ML tree was developed with a heuristic method with Nearest-Neighbor-Interchange (NNI) as a model. Tree representation was done using FigTree v1.4.4 software (http://tree.bio.ed.ac.uk/software/figtree/). SSN was carried out using SSNpipe (https://github.com/ahvdk/SSNpipe). The protocol was followed as per the online manual (https://github.com/ahvdk/SSNpipe/wiki/SSNpipe-Usage-Examples)[33]. The networks generated were visualized using Cytoscape 3.7.2[34].

4.2 Structural alignment and superimposition of M^{Pro}

The crystal structure of SARS-Cov-2 M^{Pro} (PDB ID: 6Y2E) is compared with the 49 crystal structure of the proteases from various coronaviruses **(Supplementary data 1)**

using the DALI server with all against all option[35]. The dendrogram was built based on similarity Z-score and the similarity matrix was built based on structural conservation. Representation of the similarity matrix was obtained using Multiple Experimental Viewer (Mev) software. Furthermore, the hierarchical clustering of the similarity matrix was performed and compared with the dendrogram. Representative structure from each cluster was then superimposed and compared with COVID M[Pro] using MatchMaker in Chimera 1.10.2 software. The structural backbone and active site conservation representation was obtained by Chimera 1.10.2 software[36].

4.3 Secondary structure element and inter-residue contacts analysis

Secondary structure topology and the non-covalent contact in COVID M[Pro] was analyzed by the Protein Contact Atlas tool (https://www.mrc-lmb.cam.ac.uk/rajini/index.html)[37]. Topology and the chord plot of M[Pro] are obtained as server output. Inter-residue contacts were analyzed by the Fast Coarse-Grained protein dynamics simulation using the CABS-flex server (http://biocomp.chem.uw.edu.pl/CABSflex2/)[38]. The output trajectory of the CABS-flex server was used to plot the fluctuation map of M[Pro].

4.4. Library preparation

Phytochemicals belonging to alkaloids, flavonoids, glucosinolates, phenolics, terpenes and terpenoids were prepared using available structures from PubChem[39]. Similarly, a library of active ingredients from the plants used in ayurvedic antitussive medicine was prepared. This library was further enriched with antiviral molecules available in PubChem, other databases and literature. We have also included several molecules that are predicted as anti-SARS-CoV-2 molecules in recent reports. All the molecules were checked for stereochemical properties and then converted to *.pdbqt format using Autodock Tools[40]. This library was used for further docking studies.

4.5. Preparation of the target molecules

Crystal structure of liganded MPro (PDB ID: 6Y2F) and hACE-2 complexed with viral spike protein (PDB ID: 2AJF) was downloaded from the RCSB Protein DataBank[5,11].

Water and other heteroatoms were deleted from these structures. Grid for M^{Pro} was set around active site residues H41 and C145 with dimension on 36 x 56 x 40Å using the AutoGrid program of AutoDock Tools[40]. The protein is converted to *.pdbqt for further docking studies. A similar process was performed on hACE2, with a grid dimension of 20 x 38 x 24Å around the viral spike protein recognition residues (K31, E35, D38, M82, K353)[11]. These target molecules were then further used for virtual screening.

4.6. Homology Modelling of SARS-CoV2-RdRp and preparation for docking

The protein sequence of SARS-CoV-2 RNA directed RNA polymerase (RdRp) was retrieved from the GeneBank database using the complete genome sequence of SARS-CoV-2 isolate 2019-nCoV/USA-IL2/2020 with the accession number MT044257.1. A homology model was built using the crystal structure of SARS Human-CoV with NSP12 complexed to NSP7 and NSP8 co-factors (PDB ID: 6NUR, chain A) as a template in Swiss Model Webserver[41]. The model was validated using ProSA and MolProbity sever[42,43]. The model was energy-optimized using UCSF chimera and after the addition of hydrogen atoms, the model was used for the docking study as described earlier[36].

4.7. Virtual screening using Autodock vina and ADME studies

Prepared receptor molecules from the custom-made libraries were set for the virtual screening by AutoDock Vina based Lamarckian Genetic Algorithm (LGA) parameter for ligand tethering of the proteins using 10 runs criteria[44]. Top hits of ligands were selected based on their docking score against all targets namely, M^{Pro}, RdRp and hACE2. A comparison of the docking score for these targets is represented as a heatmap of the binding score using MeV software (http://www.tm4.org/). Top ligands were then submitted to the SwissADME server for *in silico* prediction of absorption, distribution, metabolism and excretion parameters of these ligands (http://www.swissadme.ch/)[45]. Ligands with high solubility and bioavailability were further taken for the interaction analysis. Ligand binding position and interaction analysis were performed using PyMOL visualization software (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) and two-dimensional ligand interaction images are made using

Biovia Discovery Studio 4.5 (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016).

Acknowledgment

Conflict of Interest

All the authors declared no conflict of interest

References

1.   Bedford, J. *et al.* COVID-19: towards controlling of a pandemic. *Lancet* **6736**, 1015–1018 (2020).

2.   Fahmi, I. #Covid19 Coronavirus Disease 2019. *DroneEmprit* **2019**, 1–19 (2020).

3.   Fry, H. The Coronavirus , by the Numbers. *Nature* **579**, 482–483 (2020).

4.   Salata, C., Calistri, A., Parolin, C. & Palù, G. Coronaviruses: A paradigm of new emerging zoonotic diseases. *Pathog. Dis.* **77**, 1–5 (2020).

5.   Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **3405**, 1–9 (2020).

6.   Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) Structure: Basis for design of anti-SARS drugs. *Science (80-.).* **300**, 1763–1767 (2003).

7.   Cao, B. *et al.* A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N. Engl. J. Med.* 1–13 (2020). doi:10.1056/NEJMoa2001282

8.   Dong, L., Hu, S. & Gao, J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discov. Ther.* **14**, 58–60 (2020).

9.   Imbert, I. *et al.* A second, non-canonical RNA-dependent RNA polymerase in SARS Coronavirus. *EMBO J.* **25**, 4933–4942 (2006).

10. Kuhn, J. H., Li, W., Choe, H. & Farzan, M. Angiotensin-converting enzyme 2: A functional receptor for SARS coronavirus. *Cell. Mol. Life Sci.* **61**, 2738–2743 (2004).

11. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* 1–8 (2020). doi:10.1038/s41586-020-2179-y

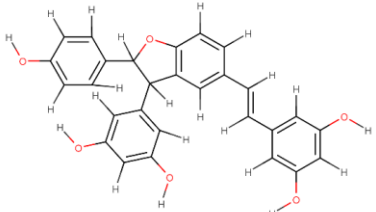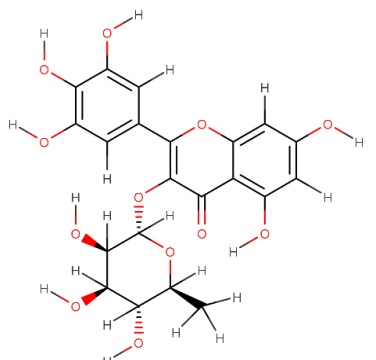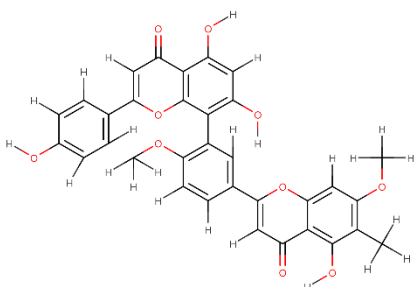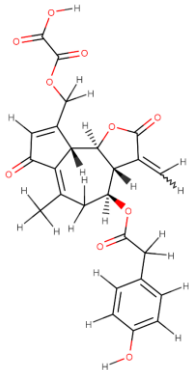12. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2

and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 1–10 (2020). doi:10.1016/j.cell.2020.02.052

13.   Wu, C. *et al.* Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* (2020). doi:10.1016/j.apsb.2020.02.008

14.   Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **89**, 44–48 (2020).

15.   Hegyi, A. & Ziebuhr, J. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**, 595–599 (2002).

16.   Li, C. *et al.* Conformational Flexibility of a Short Loop near the Active Site of the SARS-3CLpro is Essential to Maintain Catalytic Activity. *Sci. Rep.* **6**, 1–9 (2016).

17.   Aresta, A., Cioffi, N., Palmisano, F. & Zambonin, C. G. Simultaneous Determination of Ochratoxin A and Cyclopiazonic, Mycophenolic, and Tenuazonic Acids in Cornflakes by Solid-Phase Microextraction Coupled to High-Performance Liquid Chromatography. *J. Agric. Food Chem.* **51**, 5232–5237 (2003).

18.   Vitrac, X. *et al.* Determination of Stilbenes (δ-viniferin, trans-astringin, trans-piceid, cis- and trans-resveratrol, ε-viniferin) in Brazilian Wines. *J. Agric. Food Chem.* **53**, 5664–5669 (2005).

19.   Yu, B., Jiang, Y., Zhang, B., Yang, H. & Ma, T. Resveratrol dimer trans-ε-viniferin prevents rotaviral diarrhea in mice by inhibition of the intestinal calcium-activated chloride channel. *Pharmacol. Res.* **129**, 453–461 (2018).

20.   Pflieger, A. *et al.* Natural stilbenoids isolated from grapevine exhibiting inhibitory effects against HIV-1 integrase and eukaryote MOS1 transposase in vitro activities. *PLoS One* **8**, (2013).

21.   Lee, S. *et al.* Plant-derived purification, chemical synthesis, and in vitro/in vivo evaluation of a resveratrol dimer, viniferin, as an HCV Replication inhibitor.
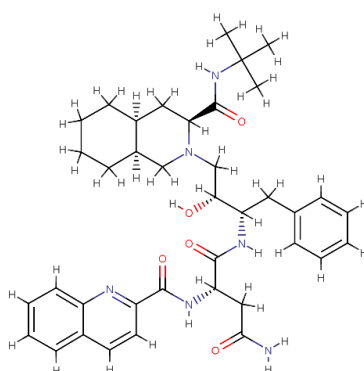
Viruses **11**, 1–18 (2019).

22.   Daino, G. L. *et al.* Identification of Myricetin as an Ebola Virus VP35–Double-Stranded RNA Interaction Inhibitor through a Novel Fluorescence-Based Assay. *Biochemistry* **57**, 6367–6378 (2018).

23.   Ortega, J. T. *et al.* The role of the glycosyl moiety of myricetin derivatives in anti-HIV-1 activity in vitro. *AIDS Res. Ther.* **14**, 1–6 (2017).

24.   Li, W. *et al.* Inhibition of herpes simplex virus by myricetin through targeting viral gD protein and cellular EGFR/PI3K/Akt pathway. *Antiviral Res.* **177**, 104714 (2020).

25.   Yu, M. S. *et al.* Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsP13. *Bioorganic Med. Chem. Lett.* **22**, 4049–4054 (2012).

26.   Keum, Y. S. & Jeong, Y. J. Development of chemical inhibitors of the SARS coronavirus: Viral helicase as a potential target. *Biochem. Pharmacol.* **84**, 1351–1358 (2012).

27.   Georgiev, V., Ananga, A. & Tsolova, V. Recent advances and uses of grape flavonoids as nutraceuticals. *Nutrients* **6**, 391–415 (2014).

28.   Kuo, Y.-H. *et al.* A Novel Cytotoxic C-Methylated Biflavone, Taiwanhomoflavone-B from the Twigs of *Cephalotaxus wilsoniana*. *Chem. Pharm. Bull.* **50**, 1607–1608 (2002).

29.   Bischoff, T. A. *et al.* Antimalarial activity of Lactucin and Lactucopicrin: sesquiterpene lactones isolated from Cichorium intybus L. *J. Ethnopharmacol.* **95**, 455–457 (2004).

30.   Elegami, A. A. *et al.* Two very unusual macrocyclic flavonoids from the water lily Nymphaea lotus. *Phytochemistry* **63**, 727–731 (2003).

31. Stefanidou, M., Herrera, C., Armanasco, N. & Shattock, R. J. Saquinavir inhibits early events associated with establishment of HIV-1 infection: Potential role for protease inhibitors in prevention. *Antimicrob. Agents Chemother.* **56**, 4381–4390 (2012).

32. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

33. Halary, S., Leigh, J. W., Cheaib, B., Lopez, P. & Bapteste, E. Network analyses structure genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 127–132 (2010).

34. Manual, C. U. Cytoscape User Manual.

35. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).

36. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).

37. The Protein Contacts Atlas. *Nat. Methods* **15**, 164 (2018).

38. Jamroz, M., Kolinski, A. & Kmiecik, S. CABS-flex: server for fast simulation of protein structure fluctuations. *Nucleic Acids Res.* **41**, W427–W431 (2013).

39. Kim, S. *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **47**, D1102–D1109 (2018).

40. Morris, G. & Huey, R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J.Compt. Chem.* **30**, 2785–2791 (2009).

41. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, (2014).

42. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-
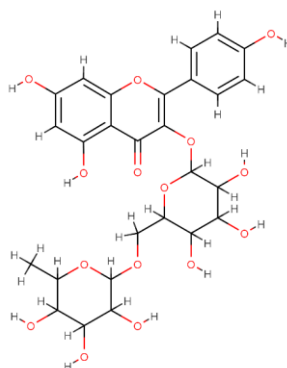
atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

43.   Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407--W410 (2007).

44.   Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–61 (2010).

45.   Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 42717 (2017).

Table 1: Potential multi-target-directed ligands against SARS-CoV-2 M^{Pro}, RdRp and hACE-2

Molecule Name	Structure	Pharmacological functions	Source
δ viniferin		Anti-tussive, Anti-oxidant, Anti-viral	*Vitis vinifera*
Myricitrin		Anti-tussive, Anti-oxidant Anti-inflammatory Anti-Viral	*Myristica fragrans*
Taiwanhomoflavone A		Anti-carcinogenic, Anti-inflammatory and Anti-neoplastic	*Cephalotaxus wilsoniana*
Lactucopicrin 15-oxalate		Anti-oxidant, Anti-malaria	*Lactuca virosa*
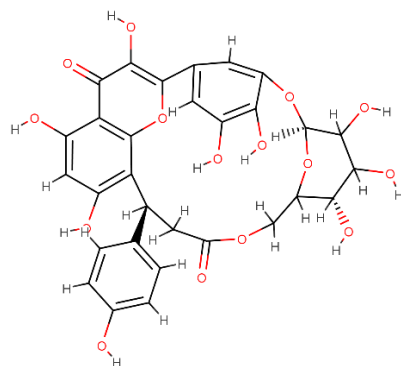
Nympholide-A		Anti-influenza, Analgesic, Sedative	*Nymphaea lotus*
Invirase		Anti-retroviral	*Synthetic*
Biorobin		Anti-virus	*Ficus benjamina*
Phyllaemblicin B		Anti-virus	*Phyllanthus emblica*
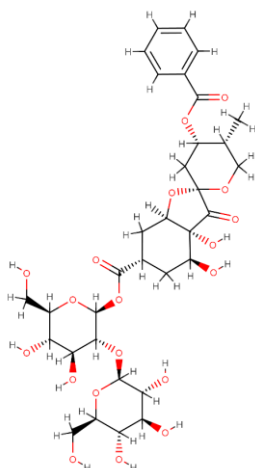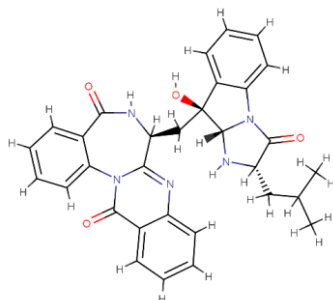
-(-)asperlicin		Antagonist: cholecystokinin receptor CCKA	*Aspergillus alliaceus*
Cassameridin		Antifungal activity	*Litsea kawakamii*
Chrysanthemin		Food additives	*Roselle plant*
Scalarane		Anti-viral Anti-bacterial Anti-inflammatory	*Hyrtios erectus*
Baicalin		Anti-cancerous Anti-inflammatory Anti bacterial	*Scutellaria baicalensis*
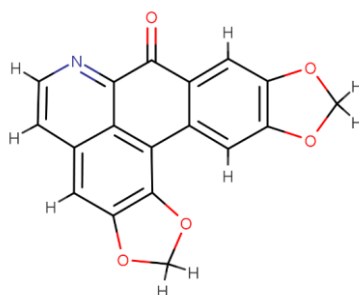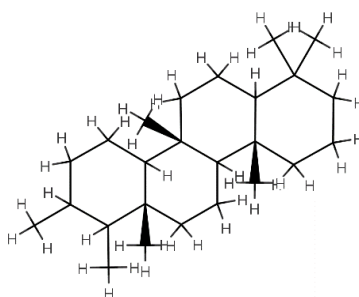
Figure legends

Fig. 1. Diversity of protease like proteins in the Coronaviridae family. **(A)** The phylogenetic tree depicts five clusters. Clade-1 (orange) main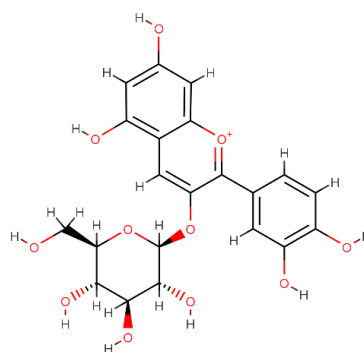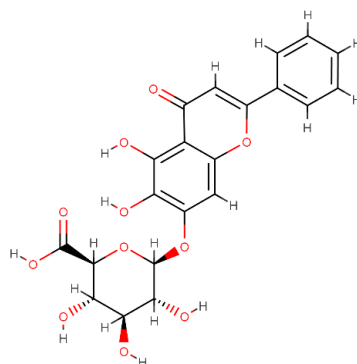ly consists of porcine transmissible coronavirus proteases. Clade-2 (blue) covers avian infectious bronchitis virus proteases. Clade-3 (purple) dominates with bat coronavirus sequences; followed by clade-4 (red) shared by the bat and human proteases. Clade-5 (green) has three distinct sub-clades that attribute to different hosts. The first sub-clade of clade 5 corresponds to bovine coronavirus and the third sub-clade exclusively involves murine coronavirus proteases. **(B)** The sequence similarity network (SSN) of the Coronaviridae proteases showed five clusters at the E-value of 1e-140. The cluster composition of sequences in the five clusters aligns with that of clades in the phylogeny (highlighted with colour code and symbols).

Fig. 2. (A) Dendrogram of the 49 structures of SARS-CoV-2 M^{Pro} like proteins from the Coronoviridea family **(B)** Structural similarity matrix of Coronaviridea proteases. Cluster 1,2,3 predominantly have porcine protease structures and cluster 4,5 have structures corresponding to human infectious coronavirus. The dendrogram and matrix both display five clusters (labeled in the Fig.), similar to that observed in phylogenetic analysis, indicating that the sequence variations also depict in the structural analysis of proteases.

Fig. 3. (A) Structural superimposition of selected proteases from Coronaviridae family, showed RMSD <1Å and the high conservation of the backbone structure **(B)** Superimposition of the active site and binding site residues displayed conservation of active site pocket shape despite few differences in the residues. This indicated that the substrate recognition and binding efficiency or function of the M^{Pro} are conserved across the virus family.

Fig. 4. (A) Secondary structure topology indicated that most of the beta-sheets towards the N-terminal end and helices to the C-terminal **(B)** The chord plot of secondary structure contacts in COVID-19 M^{Pro} wherein S indicate beta sheets and H indicate

helices. We observed that beta-sheets exhibit more contacts than helices **(C)** Inter-residue contact map within the M^{Pro}. It shows the presence of more interactions in the first 200 residues i.e towards N-terminus that mainly forms active sites **(D)** Superimposed topology with colours corresponding to secondary structure elements **(E)** Root Means Square Fluctuations of the residues from its mean position. High fluctuation (1Å<RMSF) is seen in the N-terminus as compared to C-terminal. Active site residues (H41, C145) are highlighted and they showed less fluctuation.

Fig. 5. Heatmap of binding energies of top hits for different types of molecules used for screening, namely flavonoids, glucosinolates, anti-tussive, anti-influenza, synthetic anti-viral, terpenes, terpenoids and alkaloids. In general, we noticed strong binding of ligands towards M^{Pro} and RdRp as compared to the hACE 2 receptor. Flavonoids showed promising results with better binding affinities than existing synthetic anti-viral drugs. For details of the docking score and names of the ligands check Supplementary Data 3.

Fig. 6. δ-Viniferin with active site residues of **(A)** M^{Pro} **(B)** RdRp and **(C)** hACE-2. It is followed by interaction map of myricitrin with **(D)** M^{Pro} **(E)** RdRp and **(F)** hACE-2. The interaction type is distinguished by coloured circles (residues). Dashed lines direct to the specific moiety in the ligand. Green residues symbolize van der Waals forces. Pink residues indicate those are involved in Pi-Pi stacking. Light pink indicates alkyl group interactions. Light orange colour indicates Pi-sulphur interactions. Dark orange show pi-anion interactions. Aromatic rings are involved in Pi-Pi, Pi-anion and Pi-sulphur interactions.

Fig. 7. Intermolecular interaction between Taiwanhomoflavone A with active site residues of **(A)** M^{Pro}, **(B)** RdRp and **(C)** hACE-2. Next is the interaction map of Lactucopicrin 15-oxalate with **(D)** M^{Pro}, **(E)** RdRp and **(F)** hACE-2. The representation of the interaction between different residues is similar to Fig. 6.

Fig. 8. Intermolecular interaction between natural compound Nympholide A with active site residues of **(A)** M^{Pro}, **(B)** RdRp and **(C)** hACE-2. Further is the interaction map of Invirase (synthetic anti-viral) with **(D)** M^{pro}, **(E)** RdRp and **(F)** hACE-2. The representation of the interaction between different residues is done similar to Fig. 6.

Supplementary information

Supplementary Data 1: Structures for the comparison using DALI server

Supplementary Data 2: Composition of custom-made library

Supplementary Data 3: Binding energy scores for all the top molecules from different classes of the molecules (Values used to prepare heatmaps).

Supplementary Data 4: Sequences of Coronaviridae main protease for phylogenetic analysis

Supplementary Figure 1: Multiple sequence alignment of SARS-CoV-2 M^Pro with other proteases. The conserved active site residues (H41, C145) are shown in red and the slightly variable binding site residues are shown in green colour (T25, T26, L27, M49, H140, L141, N142, G143, S144, C145, H163, H164, M165, E166, H172, R188, Q189).

Supplementary Figure 2: SSN Tree depicting the high sequence similarity among the proteases from the same host or reservoir. Up to e-value 1E-80 all the sequences were clustered together. This suggests an extremely high sequence similarity amongst all the Coronaviridae proteases. At e-value 1E-140 the sequences were clustered in mainly 5 clusters, an indicator of very specific clustering
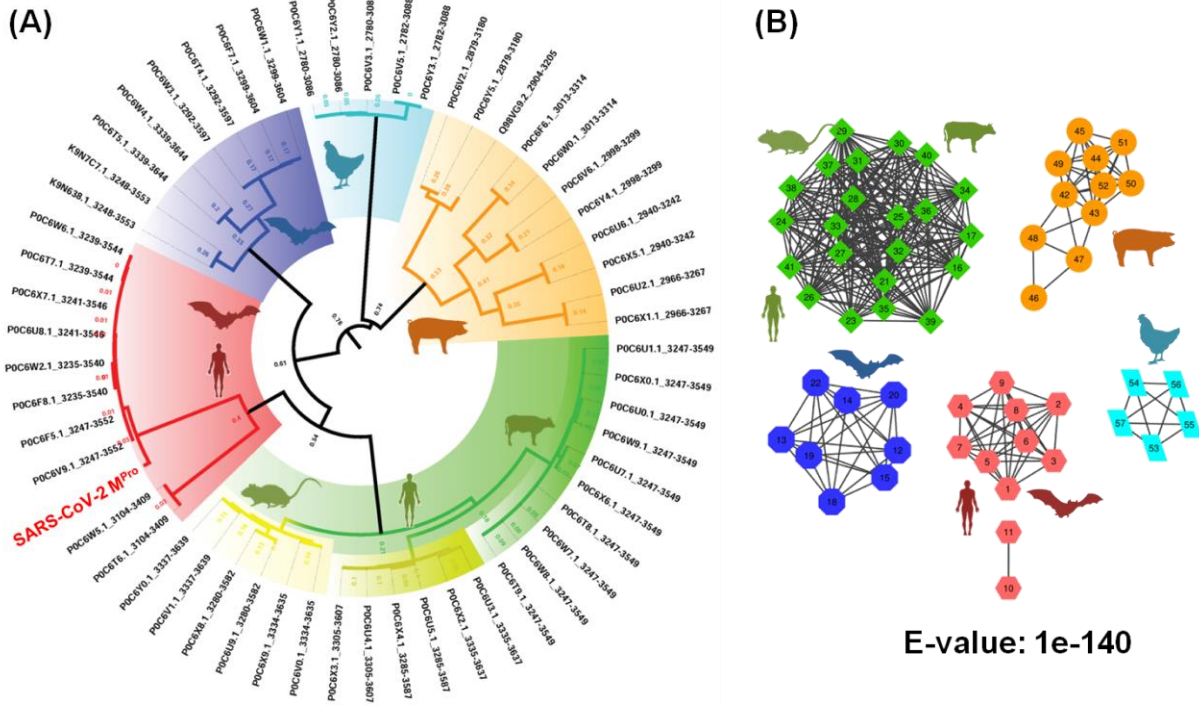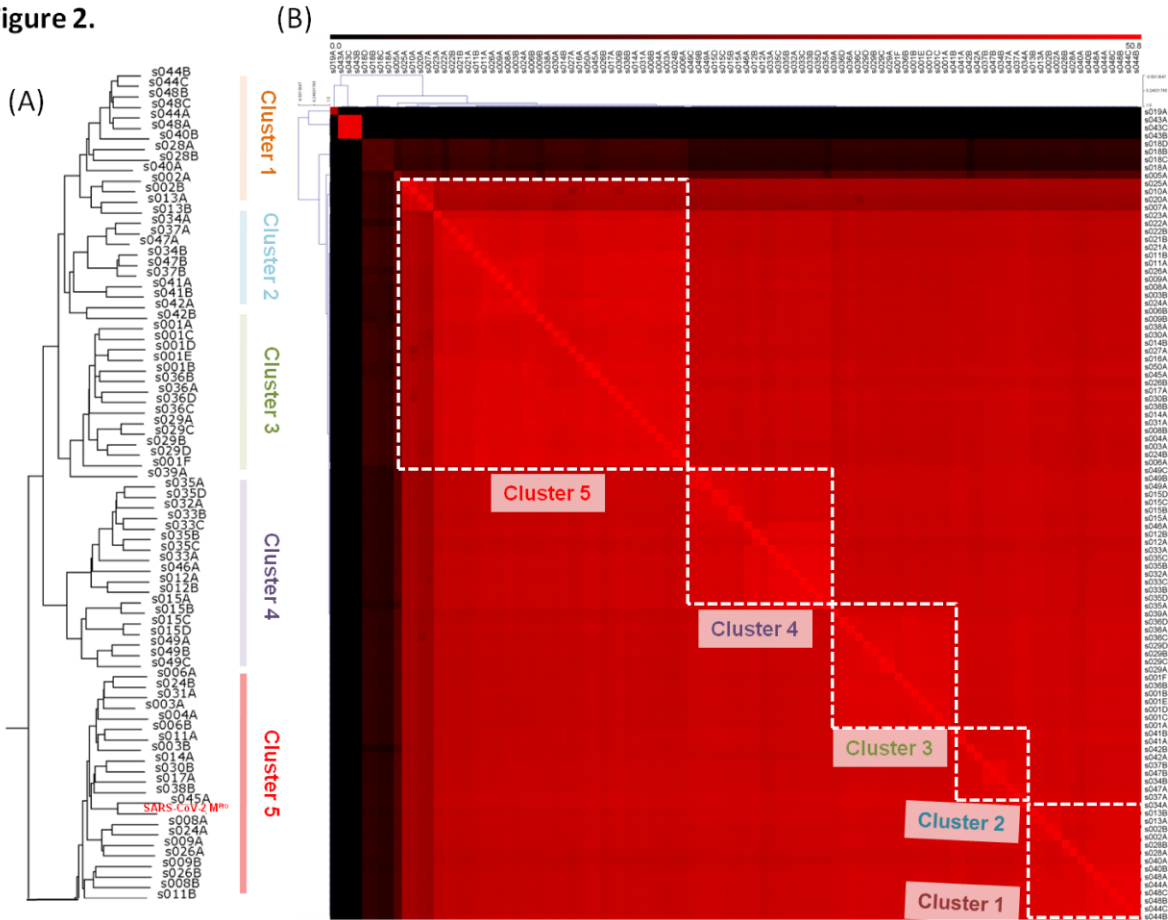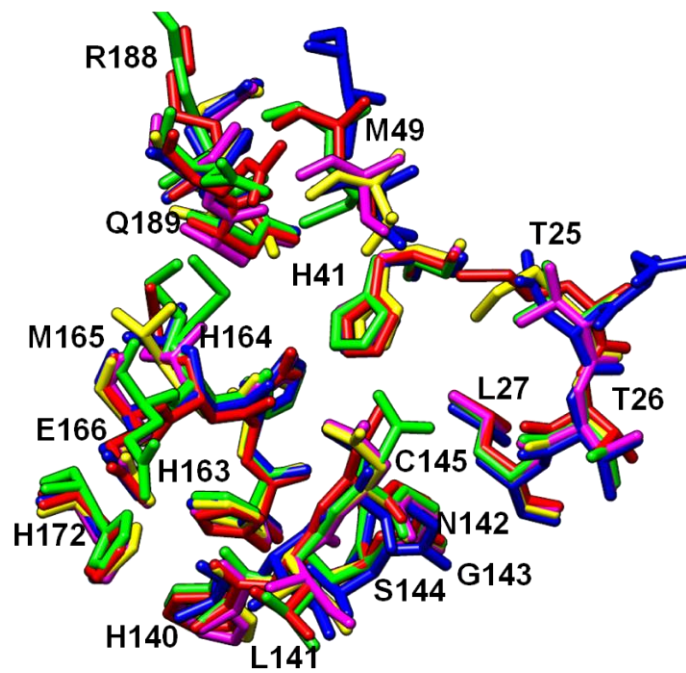
Figure 1

E-value: 1e-140

Figure 2.

Figure 3

(A)

(B)

Figure 4

Figure 5

Binding energy
−9.0 −6.0

Anti-tussive
Myricitrin
Chrysanthemin
Malvin
Delta viniferin
Myrtillin
Apigenin
Kaempferol
Oenin
Rosmarinic acid
Oleanolic acid

Anti-influenza
Nympholide A
Berberine
Deodarin
Matairesinol
Emidin
Nortrachelogenin
Ampelopsin
cedeodarin
Taxifolin
Cedrin

Anti-viral
Invirase
Isentress
Rescriptor
Prevymis
Prezista
Remdesivir
Selzentry
Aptivus
Kaletra
Daclatasvir

Terpenes
Scalarane
Taraxasterane
friedelane
"9-cis-beta,beta-carotene"
casbene
5?-cucurbitane
Tirucallane
miltiradiene
5?-cucurbitane
(Z)-biformene

Alkaloids
(-)-asperlicin C
cassameridin
oriciacridone F
(+)-asperlicin E
(+)-Atherospermoline
oxofangchirine
Dimethamine
terrequinone A
Dauricine

Terpenoids
Lactucopicrin 15-oxalate
Lactucopicrin
"(5S,9S,10S,13E)-labda-7,13-dien-
"11,13-dihydrolactucin 15-oxalate"
vitetrifolin D
metachromin S
metachromin T
integric acid
viteagnusin I
15-[(4-hydroxyphenyl)amino]retinal
cannabigerolic acid
camelliol C
cannabinerolic acid

Predicted anti-COVID-19
Biorobin
pedunculagin
Procyanidin
Gnidicin
Afzelin
Quercetin 3-alpha-L-arabinofuranos
Geraniin
Avicularin
Hirsutrin
Strictinin
Tellimagradin
Chrysin-7-O-?-glucuronide
Hyperin
Isoquercitrin
pelargonidin glycosides
Hesperidin
Isoquercitrin (CHEMBL1098724)
Houttuynoside A
Granatin A
Baicalin
Luteolin-7-glucoside
Neohesperidin
Hyperin (Quercetin-3-O-galactoside)
Isoquercitrin (Quercetin 3-.beta.-D-g
Carminic acid
Phyllaemblicin B
punicalin

Flavonoids
Cinnamtannin A1
epicatechin-(4beta->8)-epicatechin-(4beta->6)-catechin
Anemone purple anthocyanin 1
Anemone blue anthocyanin 3
6"-hydroxybenzoylhyperin
Cinnamtannin D2
Anemone blue anthocyanin 1
(-)-Epicatechin(4.beta.-->8)(-)-epigallocatechin
(+)-catechin-(4.beta.-->8)(-)-epigallocatechin
Gambiriin B2
Epicatechin-(4beta->8)-gallocatechin
Cinnamtannin A3
Eriodictioside
quercetin 3-glucosyl-(1->4)-rhamnoside
Gallocatechin-(4alpha->8)-epigallocatechin
"5,6,7,4'-tetrahydroxyflavonol-3-O-rutinoside"
Kaempferol 3-O-[6-(4-coumaroyl)-beta-D-glucosyl-(1->2)
taiwanhomoflavone A
(+)-gallocatechin-(4->8)-(+)-catechin
Gambiriin A3
kaempferol-3-O-alpha-L-rhamnopyranosyl-(1->4)-beta-D-
Cyanidin 3-O-[beta-D-xylosyl-(1->2)-beta-D-galactoside]

Glucosinolates
N-Sulfoindol-3-yl-methylglucosinolate
Glucobrassicin
2-Phenylethyl_glucosinolate
Benzyl_glucosinolate
Neoglucobrassicin
4-Methoxybenzyl_glucosinolate
p-Hydroxybenzyl_glucosinolate
2_R_-Hydroxy-2-phenylethyl_glucosinolate
3-Methylsufonylpropyl_glucosinolate
4-methylsulfinylbutyl_glucosinolate

(column headers throughout: Mpro, RdRp, hACE-2)

Figure 6

Figure 7

Figure 8