

Article

# In Silico Identification of Potential Natural Product Inhibitors of Human Proteases Key to SARS-CoV-2 Infection

R.P. Vivek-Ananth <sup>1,2</sup>, Abhijit Rana <sup>2,3</sup>, Nithin Rajan <sup>1</sup>, Himansu S. Biswal <sup>2,3,\*</sup> and Areejit Samal <sup>1,2,\*,†</sup>

<sup>1</sup> The Institute of Mathematical Sciences (IMSc), Chennai 600113, India; vivekananth@imsc.res.in (R.P.V.-A.); nithinofficial2912@gmail.com (N.R.)

<sup>2</sup> Homi Bhabha National Institute (HBNI), Mumbai 400094, India; abhijit.rana@niser.ac.in

<sup>3</sup> School of Chemical Sciences, National Institute of Science Education and Research (NISER), Bhubaneswar 752050, India

\* Correspondence: himansu@niser.ac.in (H.S.B.); asamal@imsc.res.in (A.S.)

† Lead Contact.

Academic Editor: Antonio Carta

Received: 27 July 2020; Accepted: 16 August 2020; Published: 22 August 2020



**Abstract:** Presently, there are no approved drugs or vaccines to treat COVID-19, which has spread to over 200 countries and at the time of writing was responsible for over 650,000 deaths worldwide. Recent studies have shown that two human proteases, TMPRSS2 and cathepsin L, play a key role in host cell entry of SARS-CoV-2. Importantly, inhibitors of these proteases were shown to block SARS-CoV-2 infection. Here, we perform virtual screening of 14,011 phytochemicals produced by Indian medicinal plants to identify natural product inhibitors of TMPRSS2 and cathepsin L. AutoDock Vina was used to perform molecular docking of phytochemicals against TMPRSS2 and cathepsin L. Potential phytochemical inhibitors were filtered by comparing their docked binding energies with those of known inhibitors of TMPRSS2 and cathepsin L. Further, the ligand binding site residues and non-covalent interactions between protein and ligand were used as an additional filter to identify phytochemical inhibitors that either bind to or form interactions with residues important for the specificity of the target proteases. This led to the identification of 96 inhibitors of TMPRSS2 and 9 inhibitors of cathepsin L among phytochemicals of Indian medicinal plants. Further, we have performed molecular dynamics (MD) simulations to analyze the stability of the protein-ligand complexes for the three top inhibitors of TMPRSS2 namely, qingdainone, edgeworoside C and adlumidine, and of cathepsin L namely, ararobinol, (+)-oxoturkiyenine and 3 $\alpha$ ,17 $\alpha$ -cinchophylline. Interestingly, several herbal sources of identified phytochemical inhibitors have antiviral or anti-inflammatory use in traditional medicine. Further in vitro and in vivo testing is needed before clinical trials of the promising phytochemical inhibitors identified here.

**Keywords:** COVID-19; TMPRSS2; cathepsin L; molecular docking; molecular dynamics; non-covalent interactions; phytochemical inhibitors

## 1. Introduction

In December 2019, a new respiratory disease with unknown cause with clinical symptoms of fever, cough, shortness of breath, fatigue and pneumonia was first reported in Wuhan, (China) [1–3]. While most cases of this new disease show mild to moderate symptoms, a small fraction of cases, especially those with comorbid conditions like diabetes and hypertension, can develop fatal conditions such as acute respiratory distress syndrome (ARDS) due to severe lung damage [4]. In January 2020, a novel betacoronavirus, initially named 2019-nCoV, was discovered to be the etiological agent of this

new disease [1–3]. Subsequently, human-to-human transmission of this disease was confirmed [4]. By 30 January 2020, the 2019-nCoV had spread to more than 20 countries and the World Health Organization (WHO) declared a public health emergency of international concern. On 11 February 2020, the international committee on taxonomy of viruses permanently named 2019-nCoV as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the WHO named the associated disease as coronavirus disease 2019 (COVID-19). By 11 March 2020, COVID-19 had spread to more than 150 countries across six continents and the WHO upgraded the status of the epidemic to pandemic. As of 27 July 2020, the number of laboratory confirmed COVID-19 cases and deaths have already surpassed 16 million and 650,000, respectively, with the worst affected countries as USA, Brazil, UK, Mexico, Italy, India, France and Spain (<https://www.worldometers.info/coronavirus/>). In short, the COVID-19 pandemic is an unprecedented public health and economic emergency for humankind.

Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses with large viral genomes. The publication of SARS-CoV-2 genome in January 2020 led to its taxonomic classification into the family *Coronaviridae* and genus *Betacoronavirus* [1–3]. Bats are natural reservoirs of coronaviruses [5]. The SARS-CoV-2 genome shares 96% nucleotide identity with bat coronavirus RaTG13, which suggests a probable zoonotic transfer to humans via an intermediate animal host [6]. Prior to SARS-CoV-2 epidemic, there are two precedences of zoonotic transfer of *betacoronaviruses* to humans, namely the severe acute respiratory syndrome coronavirus (SARS-CoV) and the middle east respiratory syndrome coronavirus (MERS-CoV), which had also led to outbreaks of severe respiratory disease [7]. In 2002–2003, SARS-CoV emerged in China and led to 8098 infections and 774 deaths across the world [7]. Interestingly, the SARS-CoV-2 genome shares ~80% nucleotide identity with SARS-CoV [1–3]. In 2012, MERS-CoV emerged in Saudi Arabia and led to 2521 infections and 866 deaths that were largely limited to Middle Eastern countries [7]. Unlike SARS and MERS, the geographic spread of COVID-19 and the ensuing mortality is significantly higher. To date, there are no approved antiviral drugs or vaccines against *betacoronavirus* infections including COVID-19 [8]. Hence, the current measures to contain COVID-19 include social distancing, aggressive testing, patient isolation, contact tracing and travel restrictions. In present circumstances, an immediate goal of biomedical research is to develop antivirals or anti-COVID therapeutics for SARS-CoV-2 [8].

The SARS-CoV-2 genome comprises ~30,000 bases with 14 open reading frames (ORFs) coding for 27 proteins [9]. The genome organization of SARS-CoV-2 is similar to other *betacoronaviruses* with the 5'-region coding for non-structural proteins and the 3'-region coding for structural proteins. Important structural proteins of SARS-CoV-2 coded by the 3'-region include the spike (S) surface glycoprotein, the envelope (E) protein, the matrix (M) protein and the nucleocapsid (N) protein [9]. The 5'-region of the SARS-CoV-2 genome contains the replicase gene which codes for two overlapping polyproteins, pp1a and pp1ab, which are proteolytically cleaved by two important non-structural proteins, 3-chymotrypsin like protease (3CL<sup>pro</sup>) and papain-like protease (PL<sup>pro</sup>), to produce functional (non-structural) proteins. Other important non-structural proteins of SARS-CoV-2 for the viral life cycle include the RNA-dependent RNA polymerase (RdRp) and helicase. Accordingly, the 4 non-structural proteins, 3CL<sup>pro</sup>, PL<sup>pro</sup>, RdRp and helicase, along with the spike glycoprotein of SARS-CoV-2 are among the most attractive targets for anti-COVID drugs [8].

Rather than targeting important SARS-CoV-2 proteins for viral life cycle, an alternative approach to anti-COVID drugs involves targeting host factors key to SARS-CoV-2 infection [8]. For host cell entry, SARS-CoV-2 employs the spike (S) protein whose S1 subunit has a receptor binding domain (RBD) that specifically recognizes the cell surface receptor angiotensin converting enzyme 2 (ACE2) [10–14]. Notably, both SARS-CoV-2 and SARS-CoV employ ACE2 as the cell entry receptor [10–14]. After attachment of S protein to ACE2 receptor, the membrane fusion of virus and host cell depends on proteolytic activation of S protein by host proteases which involves cleavage of S1 subunit at S1/S2 and S2' sites resulting in dissociation of S1 subunit and structural change in S2 subunit of S protein [10–14]. Hoffmann et al. [10] showed that the host cell proteases, Transmembrane Protease Serine 2 (TMPRSS2) and cathepsin L or cathepsin B, can carry out S protein priming required for SARS-CoV-2 entry. Hoffmann et al. [10]

also showed that TMPRSS2 is more essential for S protein priming and SARS-CoV-2 entry. In parallel, Ou et al. [11] used specific inhibitors of cathepsin L and cathepsin B to show that cathepsin L rather than cathepsin B is essential for S protein priming of SARS-CoV-2 and membrane fusion in lysosomes. These studies highlight at least two alternate pathways for host cell entry of SARS-CoV-2. On the one hand, after SARS-CoV-2 attachment to ACE2, the membrane fusion and cytoplasmic entry can occur at the plasma membrane provided the cell surface protease TMPRSS2 is available to carry out S protein priming. On the other hand, after SARS-CoV-2 attachment to ACE2, the virus can be internalized as part of endosomes in the endocytic pathway, and later, the membrane fusion and cytoplasmic entry will occur in lysosomes provided the lysosomal protease cathepsin L is available to carry out S protein priming [10–12]. Depending on the target cell and associated expression of host cell proteases, SARS-CoV-2 may use one of the alternative pathways for host cell entry. Importantly, the above-mentioned studies also showed that known inhibitors camostat mesylate and nafamostat mesylate of TMPRSS2 and E-64d and PC-0626568 (SID26681509) of cathepsin L can block or significantly reduce the host cell entry of SARS-CoV-2 [10–12]. In conclusion, human proteases TMPRSS2 and cathepsin L are key factors for host cell entry and are important targets for anti-COVID drugs [10–12,15].

To expedite this search for anti-COVID drugs, several computational studies have used homology modeling or published crystal structures of SARS-CoV-2 proteins, molecular docking and diverse small molecule libraries, to predict potential inhibitors of SARS-CoV-2 proteins including among existing approved drugs for repurposing and natural compounds (see e.g., [16–24]). In comparison, fewer computational studies [16,19] have focussed on identification of potential inhibitors of host factors. In this work, we perform virtual screening of a large phytochemical library specific to Indian medicinal plants to identify potential natural product inhibitors of TMPRSS2 and cathepsin L.

Plant-based natural products have made immense contributions to drug discovery [25]. Specifically, ~40% of the small-molecule drugs approved to date by the US Food and Drug Administration (FDA) are either natural products or natural product derivatives. Recently, there are reports from China of successful use of traditional Chinese medicine and associated herbs in treatment of COVID-19 patients [26]. On similar lines, there have been suggestions to tap the rich legacy of traditional Indian medicine and information on phytochemicals of Indian herbs in the search for anti-COVID drugs [27]. Previously, some of us have built IMPPAT, the largest resource to date on phytochemicals of Indian herbs [28]. In this work, we perform molecular docking using a large library of 14,011 phytochemicals compiled mainly from IMPPAT to identify potential natural product inhibitors of TMPRSS2 and cathepsin L. Subsequently, we have also performed molecular dynamics (MD) simulations to investigate the stability of protein-ligand complexes for the top phytochemical inhibitors of TMPRSS2 and cathepsin L identified in this work.

## 2. Results

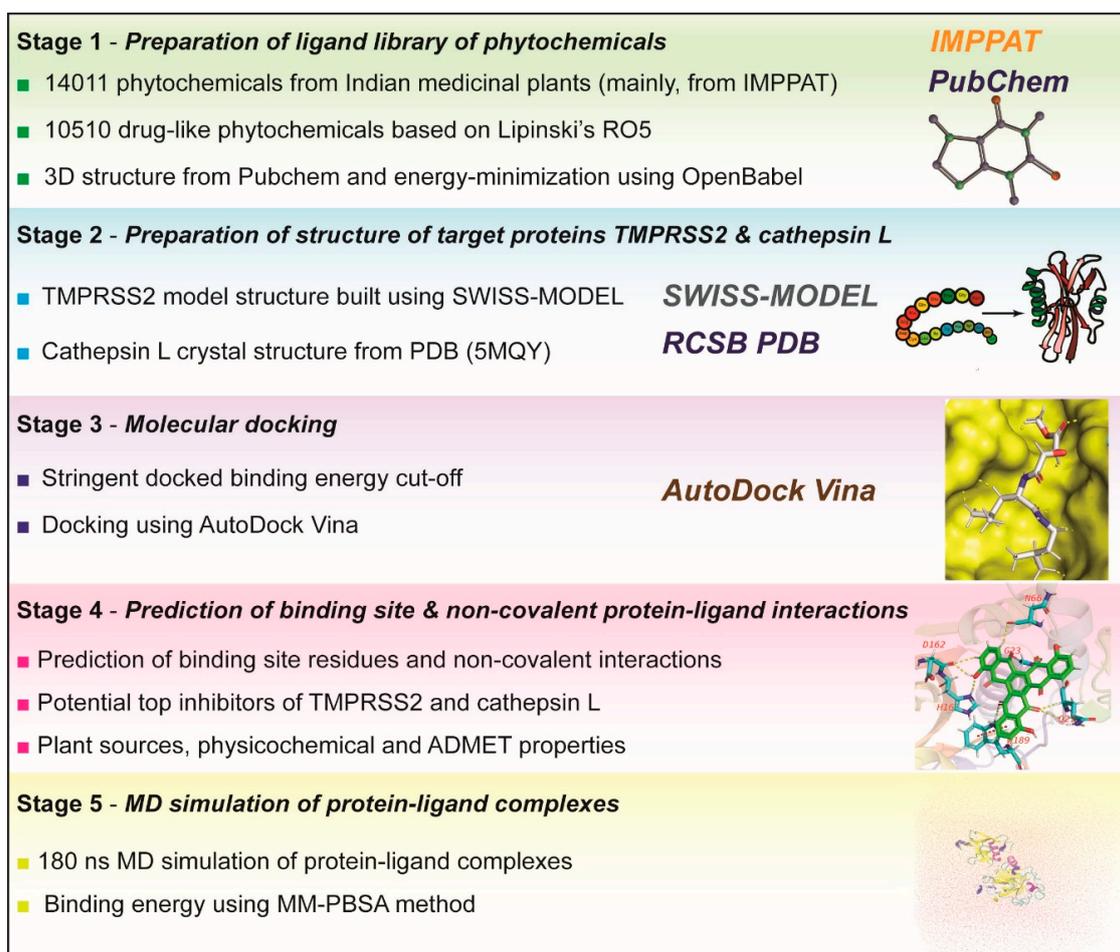
### 2.1. Workflow for Virtual Screening

This computational study aims to predict potential phytochemical inhibitors of human proteases, TMPRSS2 and cathepsin L, that are important for priming of S protein and cell entry of SARS-CoV-2 [10–12]. Briefly, the workflow for this virtual screening is as follows (Figure 1).

In the first stage, we prepared the ligands for molecular docking with target proteases. We compiled a library of 14,011 phytochemicals produced by medicinal plants used in traditional Indian medicine, and the main source of this compilation was IMPPAT [28], the largest resource on phytochemicals of Indian herbs to date (see Section 3). Next, the standard drug-likeness measure, Lipinski's rule of five (RO5) [29], was used to filter a subset of 10,510 drug-like phytochemicals. Next, the filtered phytochemicals were prepared for docking by retrieving their 3D structures from Pubchem [30] followed by energy-minimization using OpenBabel [31] (Section 3).

In the second stage, we prepared the target proteins for docking with prepared ligands. For TMPRSS2, the 3D structure is yet to be determined experimentally, and thus, we built a homology

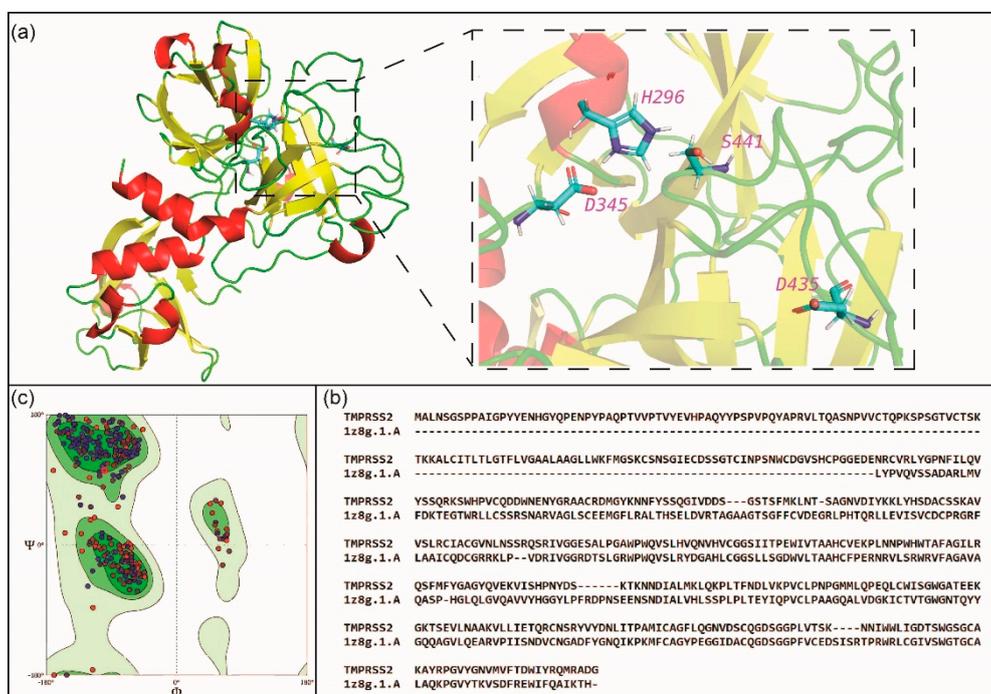
model of TMPRSS2 using SWISS-MODEL [32] which was used for docking after energy-minimization using UCSF Chimera [33] (Figure 2; Section 3). Figure 2 displays the TMPRSS2 model structure with the catalytic triad S441, H296 and D345 and the substrate binding residue D435 in the S1 subsite. For cathepsin L, the crystal structure (PDB 5MQY) with 1.13 Å resolution was used for docking after energy-minimization using UCSF Chimera (Figure 3; Section 3). Figure 3 displays the cathepsin L structure with the catalytic dyad C25 and H163 in S1 subsite, and other important residues in S2 and S1' subsites.



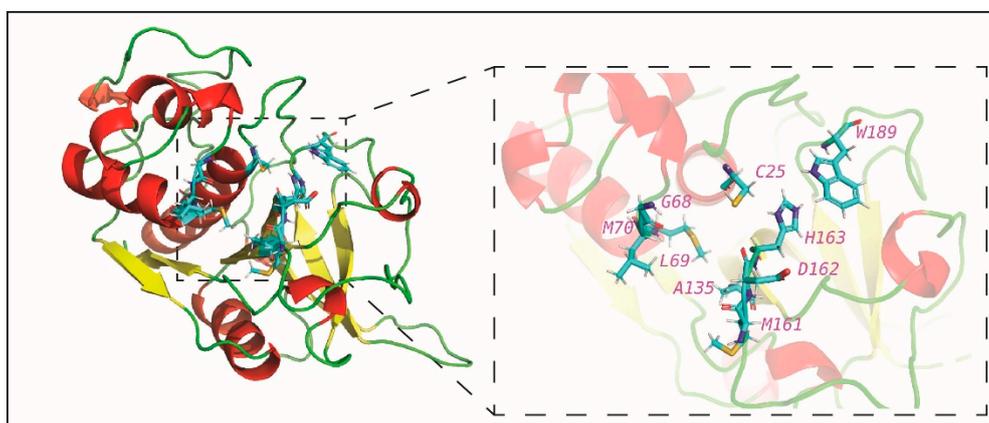
**Figure 1.** Virtual screening workflow to identify potential phytochemical inhibitors of human proteases TMPRSS2 and cathepsin L.

In the third stage, we performed protein-ligand docking using AutoDock Vina [34]. For protein-ligand docking, an appropriate grid box was manually determined for TMPRSS2 and cathepsin L (Section 3). To decide on a stringent binding energy cut off for the identification of potential inhibitors, docking was first performed for known inhibitors of target proteins (Section 3). Based on the docking binding energies of the known inhibitors, camostat and nafamostat, to TMPRSS2, we decided on a stringent criteria of binding energy  $\leq -8.5$  kcal/mol for the best docked pose of screened ligands to identify potential inhibitors of TMPRSS2 (Section 3). Similarly, based on the docking binding energies of the known inhibitors, E-64d and PC-0626568, to cathepsin L, we decided on a stringent criteria of binding energy  $\leq -8.0$  kcal/mol for the best docked pose of screened ligands to identify potential inhibitors of cathepsin L (Section 3). Thereafter, docking was performed for the prepared ligands in the phytochemical library against the prepared structures of TMPRSS2 and cathepsin L (Section 3). Lastly, we filtered the subset of phytochemicals whose binding energy in the best docked pose with TMPRSS2 (respectively, cathepsin L) is  $\leq -8.5$  kcal/mol (respectively,  $\leq -8.0$  kcal/mol). Moreover, the best docked

pose with TMPRSS2 or cathepsin L of each filtered phytochemical was separated from AutoDock Vina output file, and then, combined with the target protein structure to obtain the docked protein-ligand complex in .pdb format (Section 3). At the end of third stage, we obtained 101 phytochemicals whose binding energy in the best docked pose with TMPRSS2 is  $\leq -8.5$  kcal/mol and 16 phytochemicals whose binding energy in the best docked pose with cathepsin L is  $\leq -8.0$  kcal/mol.



**Figure 2.** (a) Cartoon representation of the homology model structure of TMPRSS2 which has been energy-minimized using UCSF Chimera. The figure zooms into the region containing the catalytic triad Ser441 (S441), His296 (H296) and Asp345 (D345), and the substrate binding residue Asp435 (D435) in the S1 subsite of the enzyme. (b) Alignment of protein sequences for TMPRSS2 and hepsin (PDB 1Z8G) which was used as a template to model the structure of TMPRSS2. (c) General Ramachandran plot of the energy-minimized model structure of TMPRSS2, which displays the torsional angles, phi ( $\phi$ ) and psi ( $\psi$ ), of the amino acid residues in the protein.



**Figure 3.** Cartoon representation of the crystal structure of human cathepsin L (PDB 5MQY). The figure zooms into the region containing the catalytic residues Cys (C25) and His163 (H163) in the S1 subsite, residues Asp162 (D162), Met161 (M161), Ala135 (A135), Met70 (M70) and Leu (L69) in the S2 subsite, Trp189 (W189) at the centre of S1' subsite and the conserved residue Gly68 (G68) in the S3 subsite of the enzyme.

In the fourth stage, the structure of docked protein-ligand complex in .pdb format for each filtered phytochemical from third stage was used to determine the ligand binding site residues in the target protein and different non-covalent interactions such as hydrogen bond, halogen bond, hydrophobic interactions, etc. between ligand and target protein (Section 3). In case of TMPRSS2, the specificity of this trypsin-like protease is determined by the conserved substrate binding residue D435 in the S1 pocket [35] (Section 3), and therefore, a potent inhibitor should either bind to or form non-covalent interactions with D435. In case of cathepsin L, the specificity of this cysteine protease is dependent on the catalytic residues, C25 and H163, in the S1 subsite (Section 3), and therefore, a potent inhibitor should either bind to or form non-covalent interactions with the catalytic residues. In this work, we consider a phytochemical to be a potential inhibitor of TMPRSS2 (respectively, cathepsin L) only if the ligand binding energy in the best docked pose is  $\leq -8.5$  kcal/mol (respectively,  $\leq -8.0$  kcal/mol) and the ligand binds to or forms non-covalent interactions with the residue D435 in TMPRSS2 (respectively, residues C25 and H163 in cathepsin L).

At the end of fourth stage, we obtained 96 phytochemicals (labelled T1–T96; Figure 5; Supplementary Table S1) as potential inhibitors of TMPRSS2 and 9 phytochemicals (labelled C1–C9; Figure 6; Supplementary Table S2) as potential inhibitors of cathepsin L. Using IMPPAT [28], we provide a list of Indian medicinal plants that can produce the identified phytochemical inhibitors of TMPRSS2 and cathepsin L (Tables 1 and 2; Supplementary Table S3). Furthermore, we have also compiled information on potential antiviral or anti-inflammatory use in traditional medicine of the herbal sources of the identified phytochemical inhibitors of TMPRSS2 and cathepsin L (Tables 1 and 2; Supplementary Table S3). In Supplementary Tables S4 and S5, we list the ligand binding site residues and non-covalent protein-ligand interactions for the identified phytochemical inhibitors of TMPRSS2 and cathepsin L. We have also predicted the physicochemical and ADMET properties of the identified phytochemical inhibitors of TMPRSS2 and cathepsin L (Section 3; Supplementary Tables S6 and S7).

Finally, in the fifth stage, for the top three inhibitors of TMPRSS2 namely, T1 (qingdainone), T2 (edgeworoside C) and T3 (adlumidine), and of cathepsin L namely, C1 (ararobinol), C2 ((+)-oxoturkiyenine) and C3 (3 $\alpha$ ,17 $\alpha$ -cinchophylline), their respective protein-ligand complexes were analyzed using 180 ns MD simulation (Section 3; Figures 8 and 9; Supplementary Figures S1 and S2) and their interaction binding energy was computed using MM-PBSA method (Section 3; Table 3).

## 2.2. Potential Phytochemical Inhibitors of TMPRSS2

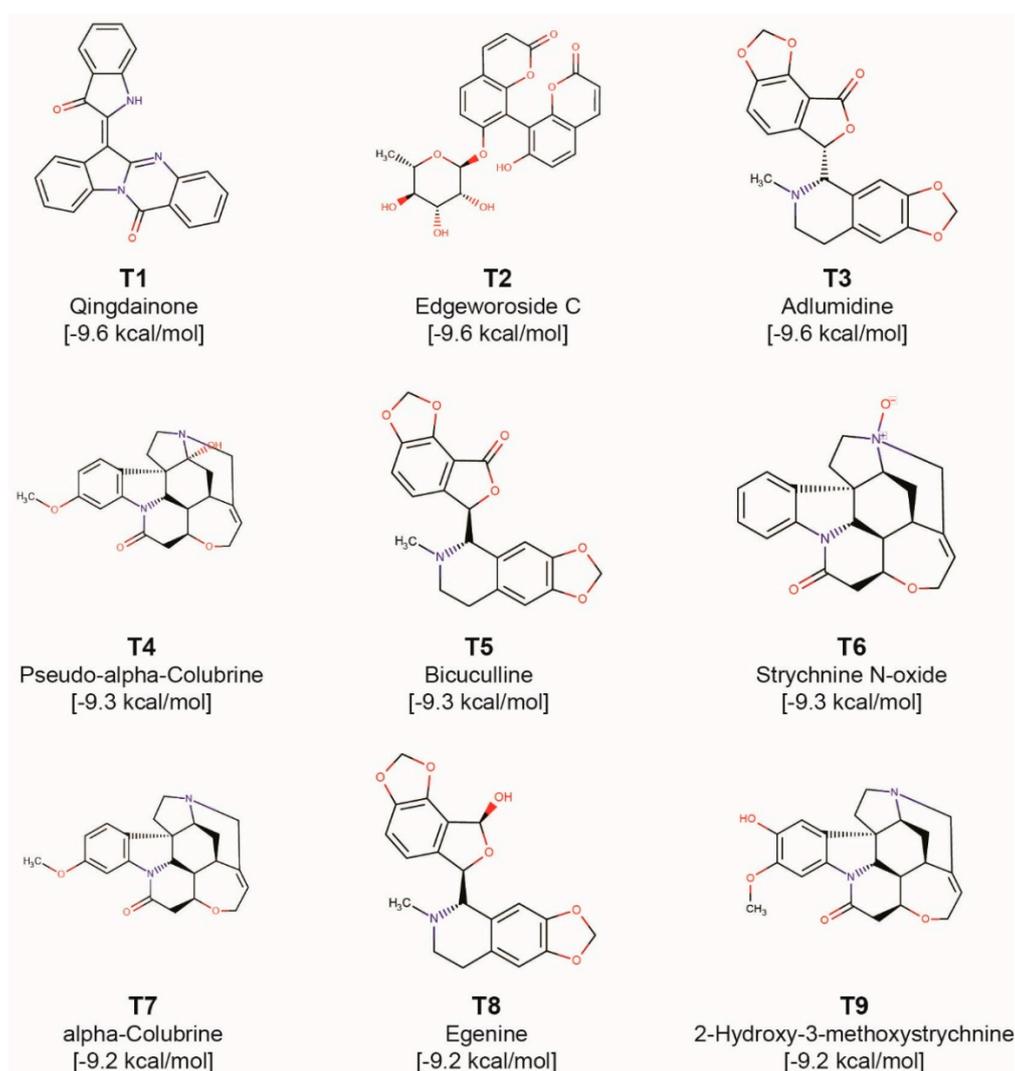
As mentioned above, we have identified 96 potential natural product inhibitors of TMPRSS2 by computational screening of 14,011 phytochemicals produced by Indian medicinal plants, and these 96 compounds labelled T1–T96 are listed in Supplementary Table S1 along with their PubChem identifier, common name, IUPAC name and structure in SMILES format. In this section, we provide a detailed discussion of the top nine phytochemical inhibitors (labelled as T1–T9) whose binding energies in the best docked poses with TMPRSS2 are  $\leq -9.2$  kcal/mol. Figure 4 displays the structure of these top 9 phytochemical inhibitors of TMPRSS2 and Table 1 provides a list of Indian medicinal plants that can produce them.

Phytochemical T1, qingdainone, has a binding energy of  $-9.6$  kcal/mol. T1 is a quinazoline alkaloid produced by *Strobilanthes cusia*, a herb with antiviral activity [36]. Figure 5a shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T1. TMPRSS2 residue D440 forms C-H $\cdots$ O type hydrogen bond with T1 whereas residue A399 forms C-H $\cdots$ N type hydrogen bond with T1. Further, T1 forms hydrophobic contacts with residues I381, S382, T387, E388, N398, A400, D440, C465 and A466.

Phytochemical T2, edgeworoside C, also has a binding energy of  $-9.6$  kcal/mol. T2 is a coumarin produced by *Edgeworthia gardneri*, a medicinal plant consumed as a herbal tea in Tibet [37]. In traditional medicine, *Edgeworthia gardneri* has been used to treat metabolic disorders including diabetes [38,39]. Figure 5b shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T2. T2 forms 12 hydrogen bonds with residues A386, N398, A399, V434, D435 and D440 of TMPRSS2.

The phenolic hydroxyl group of T2 acts as both acceptor and donor forming O-H...O and N-H...O type hydrogen bonds with the substrate binding residue D435 and C-H...O type hydrogen bond with residue V434. The hydroxyl groups attached to the pyran ring of T2 form hydrogen bonds with residues A386, N398 and D440. Further, T2 forms hydrophobic contacts with residues E260, I381, A400, N433, and A466.

Phytochemical T3, adlumidine, also has a binding energy of  $-9.6$  kcal/mol. T3 is produced by *Fumaria indica*, a herb used in traditional medicine to treat fever, cough, skin ailments and urinary diseases [40]. Adlumidine has also been reported to be an inhibitor of GABA receptor [41]. Figure 5c shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T3. The two 1,3-dioxole groups present in T3 facilitate the formation of an extensive hydrogen bond network with E388, E389, S436, C465 and A466. T3 also forms C-H...S type hydrogen bond [42] with C437. Further, T3 forms hydrophobic contacts with residues E260, I381, S382, T387, N398, A399 and A400.



**Figure 4.** Molecular structures of the top 9 phytochemical inhibitors (compounds T1–T9) of TMPRSS2. For each inhibitor, the figure shows the 2D structure, common name and docked binding energy of the ligand with TMPRSS2.

Phytochemicals T4 (pseudo- $\alpha$ -colubrine), T6 (strychnine N-oxide), T7 ( $\alpha$ -colubrine) and T9 (2-hydroxy-3-methoxystrychnine) have binding energies of  $-9.3$  kcal/mol,  $-9.3$  kcal/mol,  $-9.2$  kcal/mol and  $-9.2$  kcal/mol, respectively. These four phytochemicals are monoterpenoid indole alkaloids produced by *Strychnos nux-vomica*. The herb *Strychnos nux-vomica* is used in traditional Indian

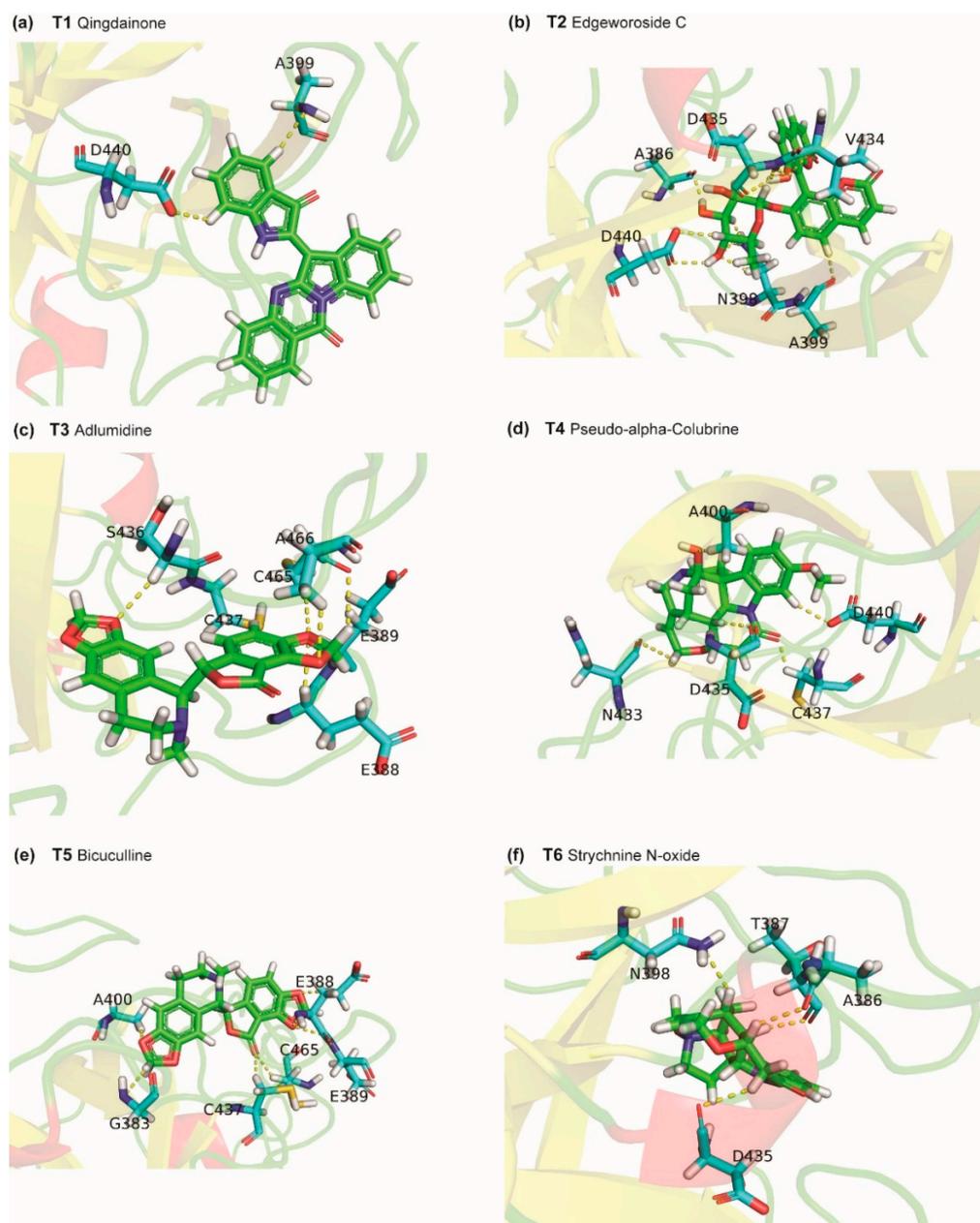
medicine and its alkaloids have been shown to exhibit anti-inflammatory, anti-oxidant, antitumor and hepatoprotective activities [43]. Note that *Strychnos nux-vomica* is a poisonous plant whose seeds are extensively used in Ayurveda only after proper detoxification procedure called *Shodhana* described in Ayurvedic texts [43]. Figure 5d shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T4. T4 forms C-H $\cdots$ O type hydrogen bonds with residues A400, N433, D435 (substrate binding residue), C437 and D440. Further, T4 forms hydrophobic contacts with residues E260, I381, S382, T387, N398, A399, V434, D440 and A466. Figure 5f shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T6. T6 forms a C-H $\cdots$ N type hydrogen bond with residue N398. The substrate binding residue D435 also forms a C-H $\cdots$ O type hydrogen bond with T6. Further, T6 forms hydrophobic contacts with residues N398, A400, V434 and A466. Supplementary Figure S3a shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T7. T7 forms five C-H $\cdots$ O type hydrogen bonds with residues N433, D435 (substrate binding residue), C437 and D440. Further, T7 forms hydrophobic contacts with residues E260, T387, N398, A399, A400, V434 and A466. Supplementary Figure S3c shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T9. The phenolic hydroxyl group of T9 forms hydrogen bonds with residues S382 and A399. The substrate binding site D435 also forms a C-H $\cdots$ O type hydrogen bond with T9. Further, T9 forms hydrophobic contacts with residues E260, N398, A399, A400 and V434.

Phytochemical T5, bicuculline, has a binding energy of  $-9.3$  kcal/mol, and it is a stereoisomer of T3. T5 is an isoquinoline alkaloid and is produced by herbs *Corydalis govaniana*, *Nerium oleander* and *Fumaria indica*. Bicuculline has also been reported to be a GABA receptor inhibitor [44]. Figure 5e shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T5. The two 1,3-dioxole groups present in T5 facilitate the formation of an extensive hydrogen bond network with residues E388, E389 and A400. The Furan-2-one ring also forms a C-H $\cdots$ O type hydrogen bond with C437. Further, T5 forms hydrophobic contacts with residues E260, T387, E388, N398, A399 and A466.

Phytochemical T8, egenine, has a binding energy of  $-9.2$  kcal/mol. T8 is an isoquinoline alkaloid produced by *Fumaria vaillantii*. In traditional medicine, *Fumaria vaillantii* has been reported to exhibit antifungal, anti-inflammatory and anti-psychotic activities [45]. Supplementary Figure S3b shows the TMPRSS2 residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with T8. The two 1,3-dioxole groups present in T8 form hydrogen bonds with G383, E388, E389 and A400. One of the hydroxyl groups present in T8 forms C-H $\cdots$ O type hydrogen bond with residue C437. Further, T8 forms hydrophobic contacts with residues T387, A399, E388, N398, E260, and A466.

**Table 1.** Herbal sources of top 9 phytochemical inhibitors of TMPRSS2. For each phytochemical, the table gives the symbol, docked binding energy, common name and plant source. Plant sources which have been reported to have antiviral or anti-inflammatory use in traditional medicine literature are shown in bold and marked with an [\*] sign.

Phytochemical Symbol	Binding Energy (kcal/mol)	Common Name	Plant Source
T1	-9.6	Qingdainone	<i>Strobilanthes cusia</i> [*]
T2	-9.6	Edgeworoside C	<i>Edgeworthia gardneri</i>
T3	-9.6	Adlumidine	<i>Fumaria indica</i> [*]
T4	-9.3	Pseudo- $\alpha$ -Colubrine	<i>Strychnos nux-vomica</i> [*]
T5	-9.3	Bicuculline	<i>Fumaria indica</i> [*], <i>Corydalis govaniana</i> [*], <i>Nerium oleander</i> [*] <i>Strychnos nux-vomica</i> [*],
T6	-9.3	Strychnine N-oxide	<i>Strychnos ignatii</i> [*], <i>Strychnos colubrina</i> [*] <i>Strychnos nux-vomica</i> [*],
T7	-9.2	$\alpha$ -Colubrine	<i>Strychnos ignatii</i> [*], <i>Strychnos colubrina</i> [*]
T8	-9.2	Egenine	<i>Fumaria vaillantii</i> [*]
T9	-9.2	2-Hydroxy-3-methoxystrychnine	<i>Strychnos nux-vomica</i> [*]



**Figure 5.** Cartoon representation of the protein-ligand interactions of the phytochemical inhibitors of TMPRSS2. Interactions of TMPRSS2 residues with atoms of (a) T1, (b) T2, (c) T3, (d) T4, (e) T5, and (f) T6. The carbon atoms of the ligand are shown in green colour while the carbon atoms of the amino acid residues in TMPRSS2 are shown in cyan colour. TMPRSS2 residues interacting with the ligand atoms via hydrogen bonds or  $\pi$ - $\pi$  stacking are labelled with the corresponding single letter residue code along with their position in the protein sequence. The hydrogen bonds and  $\pi$ - $\pi$  stacking are displayed using yellow and red dotted lines, respectively.

### 2.3. Potential Phytochemical Inhibitors of Cathepsin L

As mentioned above, we have identified nine potential natural product inhibitors of cathepsin L by computational screening of 14,011 phytochemicals produced by Indian medicinal plants, and these compounds labelled C1–C9 are listed in Supplementary Table S2 along with their PubChem identifier, common name, IUPAC name and structure in SMILES format. Figure 6 displays the structure of these top nine phytochemical inhibitors of cathepsin L and Table 2 provides a list of Indian medicinal plants that produce them.

Phytochemical C1, ararobinol, has a binding energy of  $-8.9$  kcal/mol. C1 is a bianthraquinone produced by herb *Senna occidentalis* used in Ayurveda. In traditional medicine, *Senna occidentalis* has been reported for antibacterial, antifungal, anti-inflammatory, anti-diabetic and anti-cancer activities [46]. Interestingly, there is a Chinese patent application [47] on potential use of ararobinol to treat human influenza virus infections, however, this suggests only a potential antiviral activity of C1 not specific to SARS-CoV-2 which further needs to be verified through in vitro and in vivo experimental studies. Figure 7a shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C1. Residues Q19 and A138 form C-H $\cdots$ N and C-H $\cdots$ O type of hydrogen bonds, respectively, with C1. Also, the residue W189 forms both face-to-edge and face-to-face type of  $\pi$ - $\pi$  stacking interaction with C1. Further, C1 forms hydrophobic contacts with residues C25, G139, L144, H163 and W189.

Phytochemical C2, (+)-oxoturkiyenine, has a binding energy of  $-8.3$  kcal/mol. C2 is an isoquinoline-derived alkaloid produced by the herb *Hypecoum pendulum* [48]. Figure 7b shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C2. The 2,5-dihydro-furan ring present in C2 forms two N-H $\cdots$ O type hydrogen bonds with residue Q19 and W189. The catalytic residue H163 forms N-H $\cdots$ O type hydrogen bond with C2. Also, the residue W189 forms a face-to-edge type of  $\pi$ - $\pi$  stacking interaction with C2. Further, C2 forms hydrophobic contacts with residues G139, H140, H163 and W189.

Phytochemical C3,  $3\alpha,17\alpha$ -cinchophylline, has a binding energy of  $-8.3$  kcal/mol. C3 is a cinchophylline-type of alkaloid produced by the herb *Cinchona calisaya*. The *Cinchona* alkaloids have been reported for their antimicrobial, antiparasitic and anti-inflammatory activities [49]. Figure 7c shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C3. C3 forms 8 hydrogen bonds with residues of cathepsin L. One of the catalytic residue C25 forms C-H $\cdots$ S and N-H $\cdots$ S type hydrogen bonds with C3. The other catalytic residue H163 forms C-H $\cdots$ N and N-H $\cdots$ N type hydrogen bonds with C3. Further, one of the two pyrrole rings present in C3 forms hydrogen bond with residue G23. Lastly, M70 forms a C-H $\cdots$ S type hydrogen bond with C3. Further, C3 forms hydrophobic contacts with residues Q21, C22, L69, M70, A135 and W189.

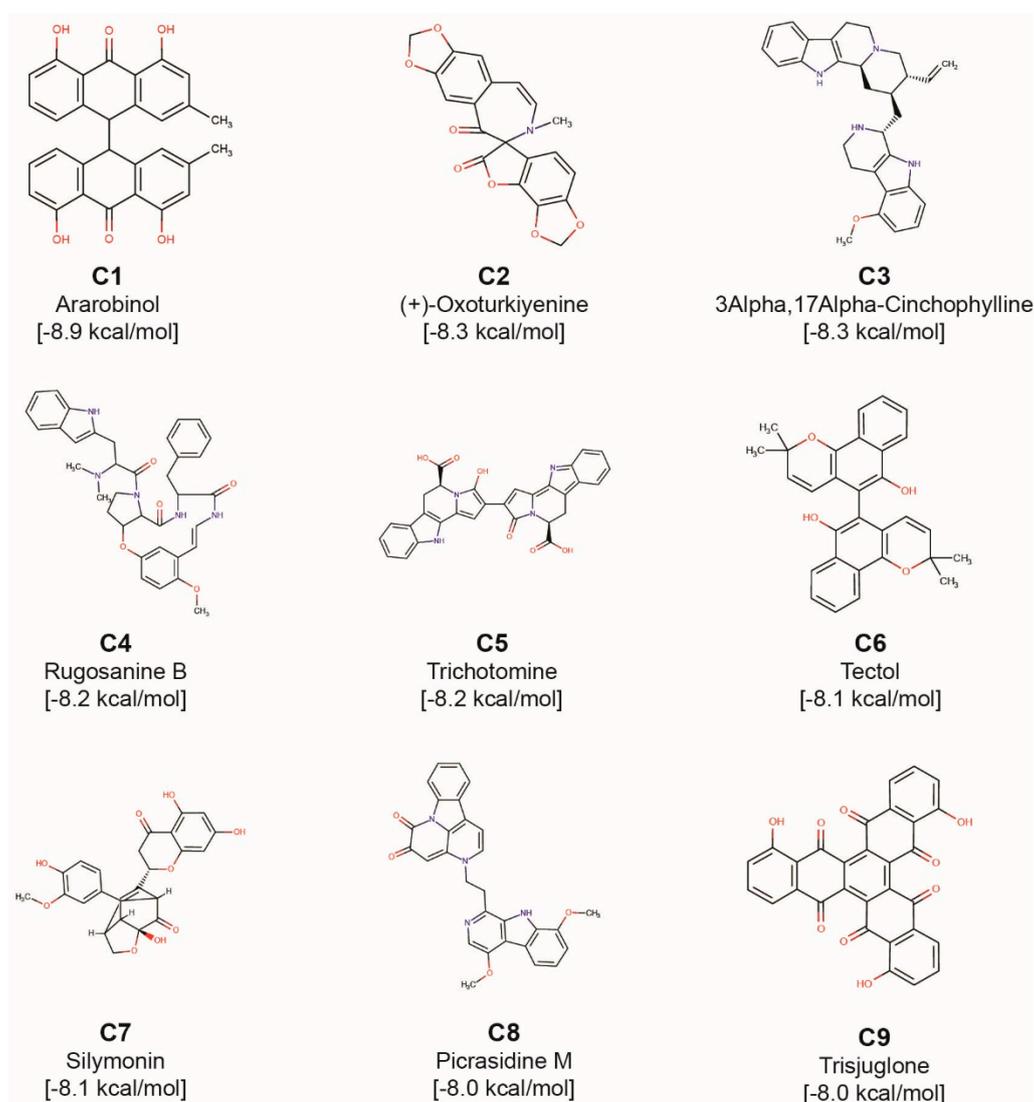
Phytochemical C4, rugosanine B, has a binding energy of  $-8.2$  kcal/mol. C4 is a cyclopeptide alkaloid produced by the bark of *Ziziphus rugosa* [50]. Various parts of *Ziziphus rugosa* have been reported for their antibacterial, antifungal, anti-inflammatory and analgesic activities [51]. Figure 7d shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C4. The pyrrole ring present in C4 forms a N-H $\cdots$ O type hydrogen bond with residue D162. Moreover, C4 forms C-H $\cdots$ O type hydrogen bonds with A138, D162 and L69. Also, the residue W189 forms a face-to-edge type of  $\pi$ - $\pi$  stacking interaction with C4. Further, C4 forms hydrophobic contacts with residues G23, C25, G67, M70, A135, A138, D162, H163, G164, W189 and A214.

Phytochemical C5, trichotomine, has a binding energy of  $-8.2$  kcal/mol. C5 is a bisindole alkaloid present in *Clerodendrum trichotomum*. *Clerodendrum trichotomum* has been reported for its use to treat cold, hypertension, rheumatism, dysentery and other inflammatory diseases [52]. Figure 7e shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C5. The carboxylic acid group present in C5 forms hydrogen bonds with residues Q19 and H163. The indole ring of C5 forms a N-H $\cdots$ N type hydrogen bond with residue G68. Further, C5 forms hydrophobic contacts with residues G23, C25, G67, G68, L69 and Y72.

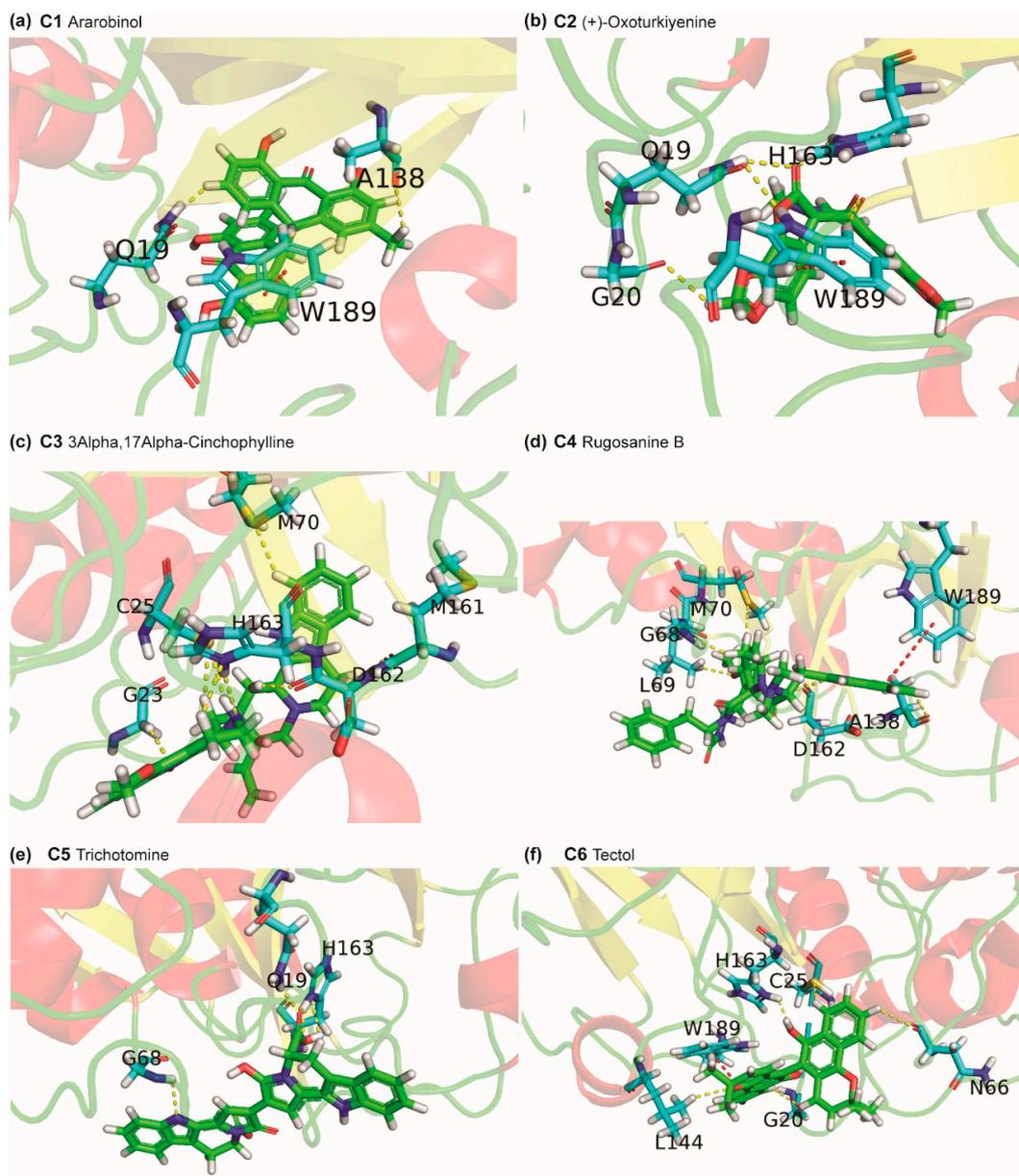
Phytochemical C6, tectol, has binding energy of  $-8.1$  kcal/mol. C6 is a naphthoquinone derivative [53] present in *Tectona grandis* and *Tecomella undulata*. *Tectona grandis* has been reported to have anti-inflammatory and antiparasitic activities [45]. *Tecomella undulata* has been used to treat syphilis and also reported to have anti-inflammatory and anti-HIV activities [54]. Additionally, Tectol has been reported to inhibit farnesyltransferase enzyme [55]. Figure 7f shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C6. The pyran group of C6 is involved in a C-H $\cdots$ O type hydrogen bond with residue L144. The catalytic residue C25 forms a C-H $\cdots$ S type hydrogen bond with C6. The other catalytic residue H163 forms a N-H $\cdots$ O type hydrogen bond with C6. Also, the residue W189 forms both face-to-face and face-to-edge type of  $\pi$ - $\pi$  stacking interaction with C6. Further, C6 forms hydrophobic contacts with G23, L144 and W189.

Phytochemical C7, silymonin, has a binding energy of  $-8.1$  kcal/mol. C7 is a flavanolignan [56] present in *Silybum marianum*. *Silybum marianum* has been used as a hepatoprotective agent and is reported to have anti-oxidant and anti-inflammatory activities [57]. Supplementary Figure S4a shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C7. C7 has four hydroxyl groups which help in the formation of an extensive network of hydrogen bonds with residues Q21, A138 and G139. C7 also forms two N-H $\cdots$ O type hydrogen bonds with H163 and W189. Also, the residue W189 forms a face-to-edge type of  $\pi$ - $\pi$  stacking interaction with C7. Further, C7 forms hydrophobic contacts with residues G23, A138, L144, H163 and W189.

Phytochemical C8, picrasidine M, has a binding energy of  $-8.0$  kcal/mol. C8 is a  $\beta$ -carboline alkaloid present in *Picrasma quassioides*. *Picrasma quassioides* has been reported to have antiviral and antifungal activities [28]. Supplementary Figure S4b shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C8. The carboxylic group of residue D162 forms two C-H $\cdots$ O type hydrogen bonds with C8. Also, residues M70 and G23 form hydrogen bonds of type C-H $\cdots$ S and C-H $\cdots$ O, respectively, with C8. Also, the residue W189 forms a face-to-edge type of  $\pi$ - $\pi$  stacking interaction with C8. Further, C8 forms hydrophobic contacts with residues G23, L69, D162 and W189.



**Figure 6.** Molecular structures of the top 9 phytochemical inhibitors (compounds C1–C9) of cathepsin L. For each inhibitor, the figure shows the 2D structure, common name and docked binding energy of the ligand with cathepsin L.



**Figure 7.** Cartoon representation of the protein-ligand interactions of the phytochemical inhibitors of cathepsin L. Interactions of cathepsin L residues with atoms of (a) C1, (b) C2, (c) C3, (d) C4, (e) C5, and (f) C6. The carbon atoms of the ligand are shown in green colour while the carbon atoms of the amino acid residues in cathepsin L are shown in cyan colour. Cathepsin L residues interacting with the ligand atoms via hydrogen bonds or  $\pi$ - $\pi$  stacking are labelled with the corresponding single letter residue code along with their position in the protein sequence. The hydrogen bonds and  $\pi$ - $\pi$  stacking are displayed using yellow and red dotted lines, respectively.

Phytochemical C9, trisjuglone, has a binding energy of  $-8.0$  kcal/mol. C9 is a naphthoquinone present in *Juglans regia* (i.e., walnut). In traditional medicine, *Juglans regia* has been reported to have anti-inflammatory, antifungal and antimicrobial activities [45]. Supplementary Figure S4c shows the cathepsin L residues that form hydrogen bonds or  $\pi$ - $\pi$  stacking interactions with C9. The benzoquinone moiety present in C9 forms two C-H $\cdots$ O type hydrogen bonds with residue Q21 and one N-H $\cdots$ O type hydrogen bond with W189. In contrast, the other catalytic residue H163 forms a N-H $\cdots$ O type hydrogen bond with C9. Also, the residue W189 forms a face-to-edge type of  $\pi$ - $\pi$  stacking interaction with C9. Further, C9 forms hydrophobic contacts with residues Q21, G23, C25, L144 and W189.

**Table 2.** Herbal sources of top 9 phytochemical inhibitors of Cathepsin L. For each phytochemical, the table gives the symbol, docked binding energy, common name and plant source. Plant sources which have been reported to have antiviral or anti-inflammatory use in traditional medicine literature are shown in bold and marked with an [\*] sign.

Phytochemical Symbol	Binding Energy (kcal/mol)	Common Name	Plant Source
C1	−8.9	Ararobinol	<i>Senna occidentalis</i> [*]
C2	−8.3	(+)-Oxoturkiyenine	<i>Hypocoum pendulum</i>
C3	−8.3	3Alpha,17Alpha-Cinchophylline	<i>Cinchona calisaya</i> [*]
C4	−8.2	Rugosanine B	<i>Ziziphus rugosa</i> [*]
C5	−8.2	Trichotomine	<i>Clerodendrum trichotomum</i> [*]
C6	−8.1	Tectol	<i>Tectona grandis</i> [*], <i>Tecomella undulata</i> [*]
C7	−8.1	Silymonin	<i>Silybum marianum</i> [*]
C8	−8	Picrasidine M	<i>Picrasma quassioides</i> [*]
C9	−8	Trisjuglone	<i>Juglans regia</i> [*]

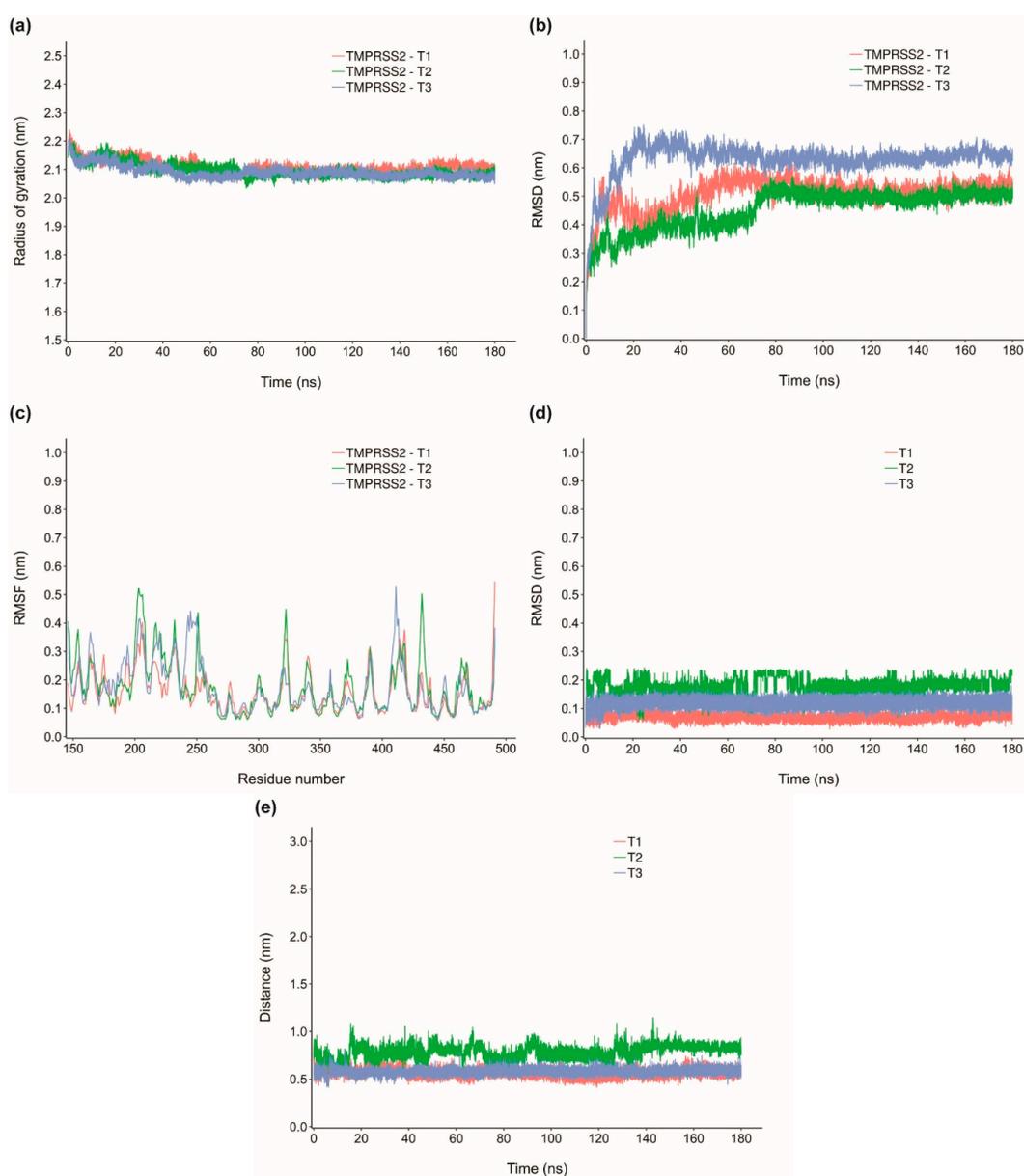
#### 2.4. Molecular Dynamics Simulation of Top Inhibitors

In order to investigate the stability of the protein-ligand complexes of the identified inhibitors, MD simulation of 180 ns was performed for top three inhibitors of TMPRSS2 namely, qingdainone (T1), edgeworoside C (T2) and adlumidine (T3), and of cathepsin L namely, ararobinol (C1), (+)-oxoturkiyenine (C2) and 3 $\alpha$ ,17 $\alpha$ -cinchophylline (C3). Specifically, we have performed six 180 ns MD runs for protein-ligand complexes (TMPRSS2-T1, TMPRSS2-T2, TMPRSS2-T3, cathepsin L-C1, cathepsin L-C2 and cathepsin L-C3) and two 180 ns MD runs for free TMPRSS2 and cathepsin L proteins (Section 3). To access the stability of the six protein-ligand complexes, we have computed radius of gyration ( $R_g$ ) of the protein, root mean square deviation (RMSD) of the C $\alpha$  atoms of the protein, root mean square fluctuations (RMSF) of the C $\alpha$  atoms of the protein, RMSD of the ligand and finally distance of the center of mass of the ligand from the center of mass of the catalytic residues or substrate binding residues of the protein in complex with the ligand (Figures 8 and 9).

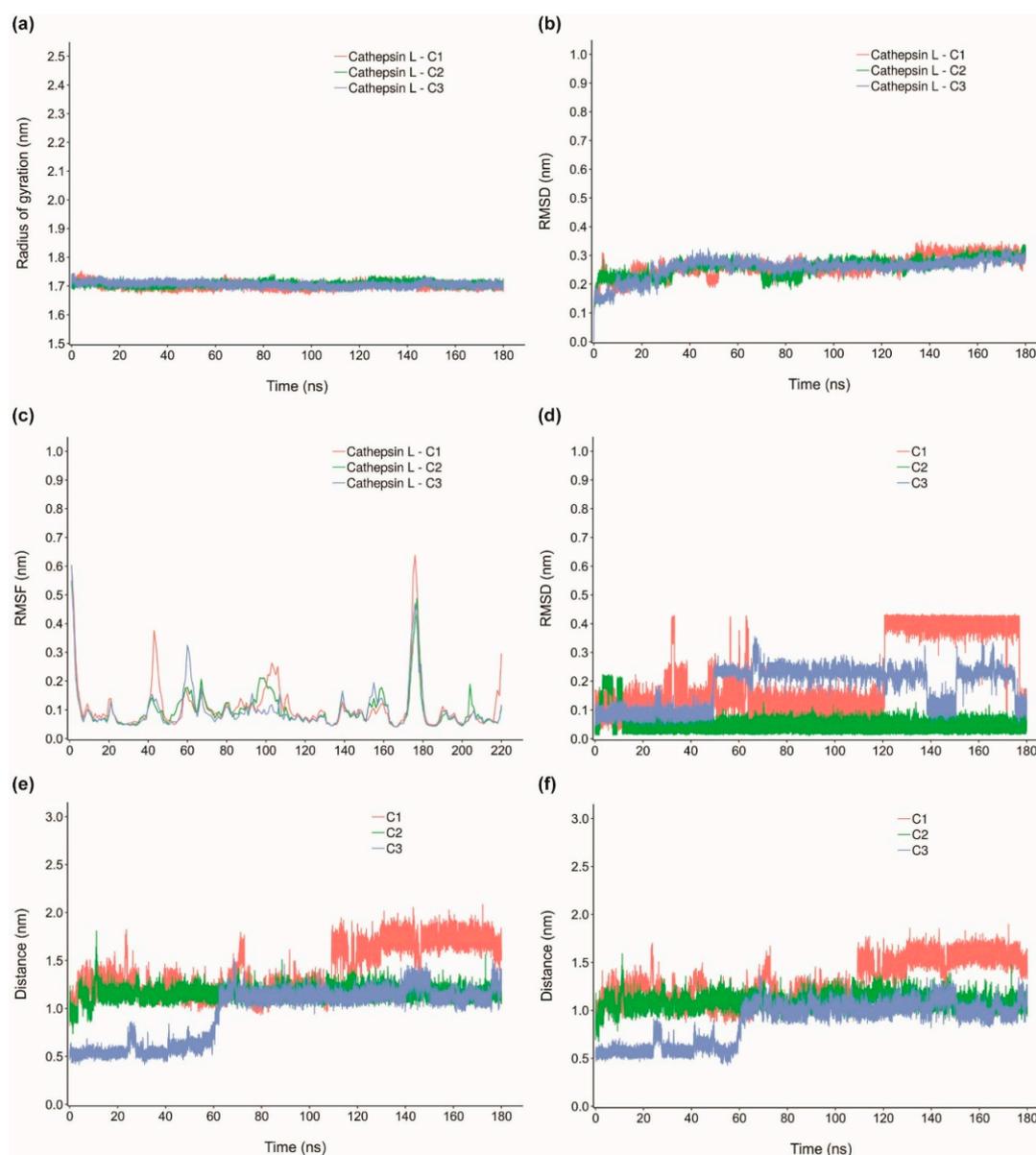
The  $R_g$  value of TMPRSS2 in complex with T1, T2 and T3 remains largely stable throughout the MD simulation (Figure 8a). This implies that the top inhibitors of TMPRSS2 namely, T1, T2 and T3 do not induce any major structural changes to TMPRSS2 and TMPRSS2 remains structurally stable in complex with these inhibitors. TMPRSS2 in complex with T1, T2 and T3 has an average  $R_g$  value of  $2.110 \pm 0.022$  nm,  $2.096 \pm 0.024$  nm and  $2.091 \pm 0.021$  nm, respectively. Further, RMSD value of the C $\alpha$  atoms of TMPRSS2 in complex with T1, T2 and T3 become stable after 80 ns (Figure 8b). Over the 80 ns to 180 ns time interval, TMPRSS2 in complex with T1, T2 and T3 has an average RMSD value of  $0.525 \pm 0.023$  nm,  $0.501 \pm 0.016$  nm and  $0.634 \pm 0.018$  nm, respectively. Lastly, Figure 8c shows the RMSF value per residue for TMPRSS2 in complex with T1, T2 and T3.  $R_g$ , RMSD and RMSF values of TMPRSS2 in complex with T1, T2 and T3 closely follow  $R_g$ , RMSD and RMSF values of TMPRSS2 free protein (Figure 8a–c; Supplementary Figure S1a–c). Supplementary Figure S2a–c show the superimposed snapshots at 120 ns, 140 ns and 160 ns of TMPRSS2-T1, TMPRSS2-T2 and TMPRSS2-T3 complexes, respectively. To further quantify the stability of the inhibitors T1, T2 and T3 bound to TMPRSS2, we have computed the RMSD of T1, T2 and T3 (Figure 8d) and distance of the center of mass of T1, T2 and T3 from the center of mass of the substrate binding residue D435 in TMPRSS2 (Figure 8e). Both RMSD of T1, T2 and T3 bound with TMPRSS2 and distance of the center of mass of T1, T2 and T3 from the center of mass of the substrate binding residue D435 becomes largely stable after 100 ns of the MD simulation (Figure 8d,e).

Also,  $R_g$  value of cathepsin L in complex with C1, C2 and C3 is stable throughout the MD simulation implying C1, C2 and C3 do not induce any major structural changes to cathepsin L and cathepsin L remains structurally stable in complex with these inhibitors (Figure 9a). Cathepsin L in complex with C1, C2 and C3 has an average  $R_g$  value of  $1.700 \pm 0.010$  nm,  $1.706 \pm 0.007$  nm and  $1.705 \pm 0.008$  nm, respectively. Similarly, RMSD value of the C $\alpha$  atoms of cathepsin L in complex

with C1, C2 and C3 become largely stable after 80 ns (Figure 9b). Over the 80 ns to 180 ns time interval, cathepsin L in complex with C1, C2 and C3 has an average RMSD value of  $0.281 \pm 0.028$  nm,  $0.276 \pm 0.022$  nm and  $0.270 \pm 0.014$  nm, respectively. Figure 9c shows the RMSF value per residue for cathepsin L in complex with C1, C2 and C3. As in the case of TMPRSS2,  $R_g$ , RMSD and RMSF values of cathepsin L in complex with C1, C2 and C3 closely follow  $R_g$ , RMSD and RMSF values of cathepsin L free protein (Figure 9a–c; Supplementary Figure S1d–f). Supplementary Figure S2d–f show the superimposed snapshots at 120 ns, 140 ns and 160 ns of cathepsin L–C1, cathepsin L–C2 and cathepsin L–C3 complexes, respectively. In order to quantify the stability of the inhibitors C1, C2 and C3 bound to cathepsin L we have also computed the RMSD of C1, C2 and C3 (Figure 9d) and distance of the center of mass of C1, C2 and C3 from the center of mass of the catalytic residues C25 (Figure 9e) and H163 (Figure 9f) in cathepsin L.



**Figure 8.** (a) Radius of gyration for TMPRSS2 in complex with T1, T2 and T3, (b) RMSD for TMPRSS2 in complex with T1, T2 and T3, (c) RMSF for TMPRSS2 in complex with T1, T2 and T3, (d) RMSD of T1, T2 and T3, and (e) Distance of the center of mass of T1, T2 and T3 from the substrate binding residue D435 in TMPRSS2.



**Figure 9.** (a) Radius of gyration for cathepsin L in complex with C1, C2 and C3, (b) RMSD for cathepsin L in complex with C1, C2 and C3, (c) RMSF for cathepsin L in complex with C1, C2 and C3, (d) RMSD of C1, C2 and C3, (e) Distance of the center of mass of C1, C2 and C3 from the catalytic residue C25 in cathepsin L, and (f) Distance of the center of mass of C1, C2 and C3 from the catalytic residue H163 in cathepsin L.

C1 has a largely stable RMSD after 120 ns of the MD simulation, C2 has the lowest and most stable RMSD in comparison with C1 and C3, and C3 shows a largely stable RMSD from 50 ns to 130 ns and from 150 ns to 170 ns of the MD simulation (Figure 9d). Distance of the center of mass of C1, C2 and C3 from the center of mass of the catalytic residues C25 and H163 in cathepsin L also remains largely consistent after 120 ns of the MD simulation (Figure 9e,f).

### 2.5. MM-PBSA Binding Energy of Top Inhibitors

Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) is a widely used method to compute the binding energy of small molecules with biological macromolecules such as proteins [58]. Notably, the protein–ligand binding energy obtained using MM-PBSA method has been found to be

more accurate than that obtained from protein-ligand docking [58]. Therefore, we have computed the binding energy for the top three inhibitors of TMPRSS2 and cathepsin L using MM-PBSA method. From the 180 ns MD simulation of the six protein-ligand complexes (TMPRSS2-T1, TMPRSS2-T2, TMPRSS2-T3, cathepsin L-C1, cathepsin L-C2 and cathepsin L-C3), 80 snapshots were obtained between 100 ns to 180 ns of the simulation at an interval of 1 ns along the trajectory, and thereafter, the 80 snapshots were used to compute the binding energy using *g\_mmpbsa* (Section 3) [59,60]. The final binding energy is the sum of contributions from van der Waals, electrostatic, polar solvation, and solvent accessible surface area (SASA) energy components. The contribution of each of the above components to the binding energy of the top inhibitors is shown in Table 3.

**Table 3.** MM-PBSA based binding energy for top three inhibitors of TMPRSS2 and cathepsin L.

Protein-Ligand Complex	Binding Energy (kcal/mol)	Van Der Waals Energy (kcal/mol)	Electrostatic Energy (kcal/mol)	Polar Solvation Energy (kcal/mol)	SASA Energy (kcal/mol)
TMPRSS2-T1	$-39.15 \pm 2.799$	$-54.285 \pm 2.903$	$-3.031 \pm 1.439$	$22.844 \pm 2.44$	$-4.678 \pm 0.237$
TMPRSS2-T2	$-30.284 \pm 3.585$	$-49.048 \pm 3.838$	$-12.501 \pm 4.884$	$35.978 \pm 5.226$	$-4.712 \pm 0.320$
TMPRSS2-T3	$-27.386 \pm 2.077$	$-39.379 \pm 2.109$	$-8.846 \pm 1.423$	$24.359 \pm 2.157$	$-3.52 \pm 0.210$
cathepsin L-C1	$-22.384 \pm 3.420$	$-25.296 \pm 3.127$	$-2.214 \pm 1.661$	$7.988 \pm 4.103$	$-2.861 \pm 0.366$
cathepsin L-C2	$-20.577 \pm 3.600$	$-30.129 \pm 3.154$	$-4.572 \pm 2.138$	$16.891 \pm 3.533$	$-2.767 \pm 0.234$
cathepsin L-C3	$-26.156 \pm 3.433$	$-37.165 \pm 3.308$	$-2.093 \pm 1.379$	$16.958 \pm 4.513$	$-3.856 \pm 0.319$

Although T1, T2 and T3 have the same docked binding energy value of  $-9.6$  kcal/mol to TMPRSS2, there are differences in their binding energy computed using MM-PBSA method. While TMPRSS2-T1 complex has the lowest binding energy value of  $-39.15 \pm 2.799$  kcal/mol, TMPRSS2-T2 and TMPRSS2-T3 complexes have binding energy value of  $-30.284 \pm 3.585$  kcal/mol and  $-27.386 \pm 2.077$  kcal/mol, respectively (Table 3). In case of cathepsin L, cathepsin L-C1, cathepsin L-C2 and cathepsin L-C3 complexes have binding energy value of  $-22.384 \pm 3.420$  kcal/mol,  $-20.577 \pm 3.600$  kcal/mol and  $-26.156 \pm 3.433$  kcal/mol, respectively (Table 3). Moreover, in comparison to binding energy obtained from docking using AutoDock Vina, binding energy for the TMPRSS2 and cathepsin L complexes with their top inhibitors computed using MM-PBSA method were found to be 2-fold to 4-fold lower, signifying even stronger binding (Tables 1–3)

### 3. Materials and Methods

#### 3.1. Phytochemical Library and Drug-Likeness Evaluation

Previously, some of us have built the Indian Medicinal Plants, Phytochemistry And Therapeutics (IMPPAT) database [28] which is the largest resource on phytochemicals of Indian herbs to date. For this study, we compiled a ligand library of 14,011 phytochemicals by augmenting the information in IMPPAT [28] with additional information compiled from other literature sources [61–70]. Thereafter, the widely used drug-likeness measure, Lipinski's rule of five (RO5) [29], was employed to filter the potential drug-like molecules within the ligand library of 14,011 phytochemicals. Specifically, 10,510 phytochemicals passed the RO5 drug-likeness filter. We then retrieved the three-dimensional (3D) structures of these phytochemicals from Pubchem [30]. Next the 3D structures of the drug-like phytochemicals were energy-minimized using *obminimize* within the OpenBabel toolbox [31]. Finally, the energy-minimized 3D structures of ligands in .sdf format were converted to .pdb format using OpenBabel.

#### 3.2. Homology Modeling of TMPRSS2 Structure

TMPRSS2 is a trypsin-like serine protease whose catalytic site consists of the triad Ser441 (S441), His296 (H296) and Asp345 (D345) [35]. It is well established that trypsin-like serine proteases cleave peptide bonds following positively charged amino acid residues such as arginine or lysine, and this

specificity of the enzyme is determined by a negatively charged aspartate residue located at the bottom of its S1 pocket [71]. In TMPRSS2, this specificity is determined by the conserved negatively charged residue Asp435 (D435) at the bottom of the S1 pocket [35].

To date the 3D structure of TMPRSS2 has not been experimentally determined, and thus, we have used SWISS-MODEL [32] webserver (<https://swissmodel.expasy.org/interactive>) to perform homology modeling of TMPRSS2. We submitted the TMPRSS2 protein sequence (NCBI reference sequence NP\_005647.3) to SWISS-MODEL and selected the crystal structure of human protein hepsin (PDB 1Z8G) [72] as the template to build the model structure (Figure 2a). Note that hepsin (PDB 1Z8G) is also a Type II transmembrane trypsin-like serine protease, and it shares 38% sequence similarity with TMPRSS2 (NP\_005647.3) (Figure 2b). Subsequently, UCSF Chimera [33] was used to minimize the energy of the TMPRSS2 model structure obtained from SWISS-MODEL. Thereafter, the energy-minimized TMPRSS2 model structure was assessed using the structure assessment tool within SWISS-MODEL. In the TMPRSS2 model, 94.19% of the amino acid residues were found to be in the Ramachandran favoured regions in the Ramachandran plot (Figure 2c) and the model structure has a MolProbity [73] score of 1.50.

### 3.3. Protein Structure of Cathepsin L

We use the crystal structure (PDB 5MQY) [74] of human cathepsin L with 1.13Å resolution obtained from Protein Data Bank for virtual screening. UCSF Chimera was used to minimize the energy of the cathepsin L structure. Figure 3 displays the cathepsin L structure with important residues in S1, S2 and S1' subsites of the enzyme [75]. Previous research has also revealed that S1 and S2 subsites of cathepsin L are important for the specificity of the enzyme [75,76]. In cathepsin L, the catalytic site consists of Cys25 (C25) and His163 (H163) in the S1 subsite, and Trp189 (W189) is at the center of the S1' subsite [75] (Figure 3). In cathepsin L, the S2 subsite with important residues Asp162 (D162), Met161 (M161), Ala135 (A135), Met70 (M70) and Leu69 (L69) forms a deep hydrophobic pocket, and lastly, the conserved residue Gly68 (G68) is at the center of the S3 subsite [75,77] (Figure 3).

### 3.4. Molecular Docking

AutoDock Vina [34] was used to perform the molecular docking of energy-minimized 3D structures of ligands with energy-minimized structure of target proteins. Accordingly, the 3D structures of prepared ligands in .pdb format were converted to .pdbqt format using the python script *prepare\_ligand4.py* from AutoDockTools [78]. Similarly, the energy-minimized structure of TMPRSS2 and cathepsin L in .pdb format were converted to .pdbqt format using the python script *prepare\_receptor4.py* from AutoDockTools [78].

For protein-ligand docking, the appropriate grid box specified by the search space centre and search space size for TMPRSS2 and cathepsin L was manually determined by considering the key residues in target proteins such as the catalytic residues and substrate binding residues, which are important for the function and specificity of the considered proteases as reported in the literature. For TMPRSS2, the grid box was defined by the search space centre (40.50, -6.00, 25.70) and search space size of 25 Å × 25 Å × 25 Å. For cathepsin L, the grid box was defined by the search space centre (55.06, 48.65, 18.12) and search space size of 25 Å × 25 Å × 25 Å.

For both target proteins, the molecular docking of prepared ligands was performed using AutoDock Vina with the exhaustiveness set as 8. The top conformation of the docked ligand with lowest binding energy, i.e., the best docked pose, was obtained from the output of AutoDock Vina using the associated script *vina\_split*. Subsequently, a combined structure file in .pdb format of the docked protein-ligand complex (with ligand in the best docked pose) was prepared using a custom script and pdb-tools [79].

### 3.5. Identification of Protein-Ligand Interactions

We searched the combined structure file of the docked protein-ligand complex for ligand binding residues in protein and different non-covalent interactions that can facilitate the binding of the ligand with the protein. These non-covalent protein-ligand interactions were identified using different geometric criteria which are specific to different types of interactions:

**Binding site residue.** Ligand binding site residues are defined as amino acids in protein which have at least one non-hydrogen atom in the proximity of at least one non-hydrogen atom of the ligand. The distance cut off to determine this proximity between non-hydrogen atoms of protein and ligand is taken to be the sum of their van der Waals radius plus 0.5 Å [80].

**Hydrogen bonds.** The accepted geometric criteria for hydrogen bonds of type D-H...A are as follows. Firstly, the distance between the hydrogen (H) and acceptor (A) atom should be less than the sum of their van der Waals radii. Secondly, the angle formed by donor (D), H and A atoms should be  $>90^\circ$  (Supplementary Figure S5a). Moreover, carbon (C), nitrogen (N), oxygen (O) or sulfur (S) atoms can be donors while N, O or S atoms can be acceptors [42,81,82].

**Chalcogen bonds.** In contrast to hydrogen bonds, chalcogen bonds are of type C-Y...A, where Y can be a S or selenium (Se) atom and A can be a N, O or S atom. The accepted geometric criteria for chalcogen interactions are as follows. Firstly, the distance between Y and A should be less than the sum of their van der Waals radii. Secondly, the angle formed by the triad, that is  $\angle C-Y...A$ , should lie in the range  $150^\circ$  to  $180^\circ$  (Supplementary Figure S5b) [83].

**Halogen bonds.** Halogen bonds are of type C-Y...A-B, where halogen Y can be a Fluorine (F), Chlorine (Cl), Bromine (Br) or Iodine (I) atom and A can be a N, O or S atom. The formation of the halogen bond is favoured when the distance between Y and A is  $\leq 3.7$  Å and the angle  $\theta_1$  of the A atom relative to the C-Y bond, and the angle  $\theta_2$  of the halogen Y relative to the A-B bond should be  $\geq 90^\circ$  (Supplementary Figure S5c) [84].

**$\pi$ - $\pi$  stacking.** This interaction occurs between two aromatic rings and can be majorly classified into two types, namely, face-to-face and face-to-edge. In the case of face-to-face type of  $\pi$ - $\pi$  interaction, the distance between the centroids of the two participating aromatic rings should be  $<4.4$  Å and the angle between their ring planes should be  $<30^\circ$ . In the case of face-to-edge type of  $\pi$ - $\pi$  interaction, the distance between the centroids of the two participating aromatic rings should be  $<5.5$  Å and the angle formed by the ring planes should be in the range  $60^\circ$  to  $120^\circ$  (Supplementary Figure S5d,e).

**Hydrophobic interactions.** The geometric criteria for the formation of hydrophobic interactions between atoms in protein and ligand are as follows [85]. The distance between a carbon atom in protein or ligand and a carbon, halogen or sulfur atom in ligand or protein, respectively, should be  $\leq 4$  Å. Furthermore, we ensure that the involved atoms in a hydrophobic interaction between protein and ligand do not form hydrogen, chalcogen or halogen bonds between them [85].

In order to detect the above-mentioned protein-ligand interactions, an in-house Python program was written to enable batch processing of combined structure files containing docked protein-ligand complexes for our large phytochemical library.

### 3.6. Comparison with Reference Inhibitors of TMPRSS2 and Cathepsin L

In order to identify potent phytochemical inhibitors of target proteins, we decided to compare the binding energy of the best docked pose of ligands with binding energies of the best docked pose of known inhibitors of TMPRSS2 and cathepsin L obtained from AutoDock Vina.

Recent experiments have shown that both camostat mesylate and nafamostat mesylate, which are approved for human use in Japan, can block the TMPRSS2-dependent cell entry of SARS-CoV-2 [10,15]. By docking these two inhibitors to TMPRSS2 using AutoDock Vina with exhaustiveness set at 50, the predicted binding energies of camostat and nafamostat was found to be  $-7.4$  kcal/mol and  $-8.5$  kcal/mol, respectively. Supplementary Figure S6a,b show the best docked poses of nafamostat and camostat with TMPRSS2, and it is seen that both molecules form hydrogen bonds with the substrate binding residue D435. Importantly, in comparison to camostat mesylate, nafamostat mesylate in

a recent experiment was shown to inhibit the TMPRSS2-dependent cell entry with 15-fold higher efficiency and an  $EC_{50}$  value in lower nanomolar range [15], and thus, the docked binding energies of these two known inhibitors are in line with experiments. Based on above observations, we decided on a stringent criteria of docked binding energy  $\leq -8.5$  kcal/mol for screened ligands to be identified as potential inhibitors of TMPRSS2.

Recent experiments have shown that the small molecules E-64d and PC-0626568 (SID26681509) can block the cathepsin L-dependent cell entry of SARS-CoV-2 [10–12]. Note that cathepsin L is one of 11 cysteine cathepsin proteases encoded by the human genome, and the cathepsins share a high sequence similarity to papain, a non-specific plant protease [86]. E64-d is a broad spectrum inhibitor which can inhibit proteases cathepsins B, H, L and calpain, while PC-0626568 is a specific inhibitor of cathepsin L [11]. Moreover, a recent study [11] used the specific inhibitor PC-0626568 of cathepsin L to conclude that cathepsin L rather than cathepsin B is important for cell entry of SARS-CoV-2. As both cathepsin L and cathepsin B are expressed in several mammalian tissues, it is important to design specific inhibitors of cathepsin L [75,76] to avoid any off-target toxicity. Along with the above-mentioned two inhibitors of cathepsin L, we have also considered the co-crystallized inhibitor GH4 present in the crystal structure of cathepsin L (PDB 5MQY) as another reference inhibitor of cathepsin L. We remark that while recent experiments [10–12] have shown that E-64d and PC-0626568 can block the cathepsin L-dependent cell entry of SARS-CoV-2, similar experimental data specific to SARS-CoV-2 infection is lacking for cathepsin L inhibitor GH4. By docking these—three known inhibitors to cathepsin L using AutoDock Vina with exhaustiveness set at 50, the predicted binding energies of E-64d, PC-0626568 and GH4 were found to be  $-5.0$  kcal/mol,  $-8.0$  kcal/mol and  $-6.3$  kcal/mol, respectively. Notably, the docked binding energies are in line with known specificity of E-64d and PC-0626568 to cathepsin L. Supplementary Figure S6c–e show the best docked poses of PC-0626568, E-64d and GH4 with cathepsin L. It is seen that both E-64d and PC-0626568 form hydrogen bonds with both catalytic residues C25 and H163. Based on above observations, we decided on a stringent criteria of docked binding energy  $\leq -8.0$  kcal/mol for screened ligands to be identified as potential inhibitors of cathepsin L. We have also compared the docked pose of GH4 obtained from AutoDock Vina with the pose of GH4 in the co-crystallized structure of cathepsin L, and the RMSD between the heavy atoms in the two poses was found to be  $0.786$  Å. Supplementary Figure S7 shows the superimposed structures of the docked pose of GH4 from AutoDock Vina and the pose of GH4 in the co-crystallized structure of cathepsin L. Apart from PC-0626568 and E-64d, seven other cathepsin-L inhibitors namely, dec-RVKR-CMK, K11777, small molecule 5705213, MDL28170, SSAA09E1, EST and oxocarbazate, have been reported in literature to exhibit ant-coronavirus activity [87].

### 3.7. Molecular Dynamics Simulations

Using GROMACS 5.1.5 [88] along with GROMOS96 54a7 force field [89], we have performed MD simulations for the top inhibitors of TMPRSS2 and cathepsin L to assess the stability of their protein-ligand complexes. The Automated Topology Builder (ATB) version 3.0 (<https://atb.uq.edu.au/>) was used to generate topology for the top inhibitors [90]. The protein-ligand complex was placed in the center of a cubic periodic box and was solvated by the addition of simple point charge (SPC) waters. Then net charge of the system was neutralized by addition of  $Na^+$  and  $Cl^-$  ions. Energy minimization was performed using the steepest descent algorithm. Then the system was heated to 310 K during a constant number of particles, volume, and temperature (NVT) simulation of 500 ps with 2 fs time step. Then the pressure was equilibrated to 1 bar during a constant number of particles, pressure, and temperature (NPT) simulation of 500 ps with 2 fs time step. In both the above simulations, protein and ligand were position restrained. Thereafter, the position restraint was removed and a production MD run was performed for a period of 180 ns with 2 fs time step. During the MD simulation, the structural coordinates were saved after every 2 ps. Temperature was maintained at 310 K using the v-rescale temperature [91]. Pressure was maintained at 1 bar using Parrinello-Rahman pressure coupling method [92]. Time constant for the temperature and pressure coupling were kept at 0.1 ps

and 2 ps, respectively. The short-range interactions were computed for the atom pairs within the cutoff of 1.4 nm, whereas the long-range electrostatic interactions were calculated using Particle mesh Ewald (PME) method with fourth order cubic interpolation and 0.16 nm grid spacing. All bonds were constrained using the LINCS method. The trajectories obtained from the MD simulations were then used to compute and analyze the radius of gyration of the protein ( $R_g$ ), root mean square deviation (RMSD) of the protein backbone  $C\alpha$  atoms and root mean square fluctuation (RMSF) of the protein backbone  $C\alpha$  atoms using GROMACS scripts.

### 3.8. MM-PBSA Calculation

Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method was used to compute the binding energy of the interactions between the protein-ligand complexes for the top inhibitors of TMPRSS2 and cathepsin L. From the 180 ns trajectory obtained from the MD simulation, 80 snapshots were obtained between 100 ns to 180 ns of the simulation at an interval of 1 ns along the trajectory, and thereafter, the binding energy was computed using `g_mmpbsa` [59,60]. `g_mmpbsa` computes the Gibbs free energy of binding ( $\Delta G_{\text{binding}}$ ) using MM-PBSA method as described by the following equation:

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$$

where  $G_{\text{complex}}$ ,  $G_{\text{protein}}$  and  $G_{\text{ligand}}$  represent the total binding energy of the protein-ligand complex, the free protein and the unbounded ligand, respectively.

### 3.9. Validation of the TMPRSS2 Model Structure

In order to validate the model structure of TMPRSS2 built using SWISS-MODEL, we have used two approaches. Firstly, we have run a 180 ns MD simulation of the TMPRSS2 free protein and have computed radius of gyration ( $R_g$ ), RMSD and RMSF for the TMPRSS2 free protein from its MD trajectory (Supplementary Figure S1a–c). The  $R_g$  of the TMPRSS2 free protein remains largely stable throughout the 180 ns MD simulation (Supplementary Figure S1a). This signifies that the TMPRSS2 model structure remains structurally intact and stable during the MD simulation. The RMSD of the TMPRSS2 free protein becomes stable after 60 ns (Supplementary Figure S1b).

Secondly, we have used the MD trajectory of TMPRSS2 in complex with the top three potential inhibitors (T1, T2 or T3) from 40 ns to 80 ns to validate the TMPRSS2 model structure. The MD trajectory of TMPRSS2 in complex with the inhibitors from 40 ns to 80 ns was used for validation, 80 ns to 100 ns was used for equilibration, and from 100 ns to 180 ns for finding the binding energy of T1, T2 or T3 with TMPRSS2 using MM-PBSA method. To validate the TMPRSS2 model structure, the binding energy computed using MM-PBSA method of T1, T2 or T3 with TMPRSS2 during 40 ns to 80 ns was compared with binding energy of T1, T2 or T3 with TMPRSS2 during 100 ns to 180 ns of the MD simulation. We find that T1, T2 and T3 have average binding energies of  $-37.507 \pm 2.537$  kcal/mol,  $-30.383 \pm 4.005$  kcal/mol and  $-26.586 \pm 2.506$  kcal/mol, respectively, with TMPRSS2 during 40 ns to 80 ns of the MD simulation. The above reported binding energies of T1, T2 and T3 with TMPRSS2 during 40 ns to 80 ns are very close to the binding energies computed for the same inhibitors in complex with TMPRSS2 during 100 ns to 180 ns of the MD simulation as reported in Table 3. These results further validate the TMPRSS2 model structure used for binding energy computations during 100 ns to 180 ns of the MD simulations.

### 3.10. Prediction of ADMET Properties

In order to assess the pharmacokinetic properties and potential toxicity of the inhibitors of TMPRSS2 and cathepsin L predicted from this study, we have also computed the Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties of the inhibitors using SwissADME [93] and vNN-ADMET [94].

#### 4. Conclusions

The current COVID-19 pandemic is a serious threat to humankind. As of 27 July 2020, COVID-19 has led to more than 650,000 deaths worldwide. Due to the absence of approved therapeutics or vaccines against SARS-CoV-2, several countries have been forced to implement partial or complete lockdown measures to restrict infection spread; however, such measures have in turn resulted in an economic catastrophe. Consequently, there is an urgent need to develop antivirals and vaccines against SARS-CoV-2 to protect humankind. In this direction, protein-ligand docking and MD simulation are powerful computational methods to expedite the search for anti-COVID drugs by rapid identification of promising lead molecules. Here, we have used molecular docking and MD simulations in the search of natural compound inhibitors of two human proteases, TMPRSS2 and cathepsin L, that are key host factors in SARS-CoV-2 infection [10–12].

Since early civilization, humans have used medicinal plants in different systems of traditional medicine to treat various ailments [95]. Specifically, traditional systems of Indian medicine including Ayurveda, Siddha and Unani have over centuries acquired invaluable knowledge on medicinal plants spanning the rich biodiversity of the subcontinent for treating various ailments including viral infections [28]. As plant-based natural products have been an indomitable source of lead molecules in the course of modern drug discovery [25], it is worthwhile to search for potential anti-COVID drugs among phytochemicals of Indian medicinal plants. In this direction, some of us have built IMPPAT [28], the largest resource on phytochemicals of Indian medicinal plants to facilitate natural product-based drug discovery. In this work, we have performed virtual screening of 14,011 phytochemicals that are produced by Indian medicinal plants to identify potential inhibitors of key host factors, TMPRSS2 and cathepsin L, for SARS-CoV-2 infection. For the identified top inhibitors of TMPRSS2 and cathepsin L, we have performed MD simulation, and thereafter, employed MM-PBSA method to compute binding energies of the protein-ligand complexes.

We have predicted 96 potential phytochemical inhibitors of TMPRSS2, of which the top three candidates, namely, qingdainone (T1), edgeworoside C (T2) and adlumidine (T3), have binding energy value of  $-39.15 \pm 2.799$  kcal/mol,  $-30.284 \pm 3.585$  kcal/mol and  $-27.386 \pm 2.077$  kcal/mol, respectively (Table 3). We have also predicted nine potential phytochemical inhibitors of cathepsin L, of which the top three candidates, namely, ararobinol (C1), (+)-oxoturkiyenine (C2) and  $3\alpha,17\alpha$ -cinchophylline (C3), have binding energy value of  $-22.384 \pm 3.420$  kcal/mol,  $-20.577 \pm 3.600$  kcal/mol and  $-26.156 \pm 3.433$  kcal/mol, respectively (Table 3). Furthermore, most of the herbal sources of the identified phytochemical inhibitors of TMPRSS2 and cathepsin L have been reported to have antiviral or anti-inflammatory use in traditional medicine (Tables 1 and 2). Additional *in vitro* and *in vivo* testing of the identified phytochemical inhibitors of TMPRSS2 and cathepsin L is needed before these molecules can enter clinical trials against COVID-19. In conclusion, we expect the natural product inhibitors identified in this computational study for TMPRSS2 and cathepsin L will likely inform future research toward natural product-based anti-COVID therapeutics.

**Supplementary Materials:** The following are available online: Figure S1: (a) Radius of gyration, (b) RMSD, and (c) RMSF for TMPRSS2 free protein. (d) Radius of gyration, (e) RMSD, and (f) RMSF for cathepsin L free protein, Figure S2: Superimposition of the snapshots at 120 ns, 140 ns and 160 ns of (a) TMPRSS2-T1 complex, (b) TMPRSS2-T2 complex, (c) TMPRSS2-T3 complex, (d) cathepsin L-C1 complex, (e) cathepsin L-C2 complex and (f) cathepsin L-C3 complex obtained from their respective MD simulation trajectories, Figure S3: Cartoon representation of the protein-ligand interactions of the phytochemical inhibitors of TMPRSS2. Interactions of TMPRSS2 residues with atoms of (a) T7, (b) T8, and (c) T9. The carbon atoms of the ligand are shown in green colour while the carbon atoms of the amino acid residues in TMPRSS2 are shown in cyan colour. TMPRSS2 residues interacting with the ligand atoms via hydrogen bonds or  $\pi$ - $\pi$  stacking are labelled with the corresponding single letter residue code along with their position in the protein sequence. The hydrogen bonds and  $\pi$ - $\pi$  stacking are displayed using yellow and red dotted lines, respectively, Figure S4: Cartoon representation of the protein-ligand interactions of the phytochemical inhibitors of cathepsin L. Interactions of cathepsin L residues with atoms of (a) C7, (b) C8, and (c) C9. The carbon atoms of the ligand are shown in green colour while the carbon atoms of the amino acid residues in cathepsin L are shown in cyan colour. Cathepsin L residues interacting with the ligand atoms via hydrogen bonds or  $\pi$ - $\pi$  stacking are labelled with the corresponding single letter residue code along with their position in the protein sequence. The hydrogen bonds and  $\pi$ - $\pi$  stacking are displayed using yellow and red dotted lines, respectively, Figure S5: Geometric criteria for the identification of protein-ligand



inhibitor predicted by vNNADMET, Cytochrome P450 2D6 inhibitor predicted by SwissADME, Cytochrome P450 2D6 inhibitor predicted by vNNADMET, Cytochrome P450 3A4 inhibitor predicted by SwissADME, Cytochrome P450 3A4 inhibitor predicted by vNNADMET, Human liver microsomal stability predicted by vNNADMET (HLM); several toxicity properties such as PAINS-Number of alerts predicted by SwissADME, Brenk-Number of alerts predicted by SwissADME, Drug-induced liver injury predicted by vNNADMET (DILI), Cytotoxicity predicted by vNNADMET, hERG Blocker predicted by vNNADMET, Mitochondrial toxicity predicted by vNNADMET (MMP) and Mutagenicity predicted by vNNADMET (AMES).

**Author Contributions:** Conceptualization, R.P.V.-A. and A.S.; Data curation, R.P.V.-A., N.R. and A.S.; Formal analysis, R.P.V.-A., A.R., H.S.B. and A.S.; Methodology, R.P.V.-A., A.R., H.S.B. and A.S.; Supervision, H.S.B. and A.S.; Writing—original draft, R.P.V.-A., A.R., H.S.B. and A.S.; Writing—review & editing, R.P.V.-A., H.S.B. and A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has received no specific external funding.

**Acknowledgments:** Authors thank S. Krishnaswamy for guidance and help with MD simulations. Authors also thank Dhiraj Kumar, Vinay Nandicoori and Pinaki Saha for discussions. A.S. and R.P.V.-A. Thank the computational staff of IIMSc for maintaining resources during the Covid-19 associated lockdown. H.S.B. and A.R. acknowledge the computational facilities of NISER. A.S. would like to acknowledge financial support from the Science and Engineering Research Board (SERB) India through the award of a Ramanujan fellowship (SB/S2/RJN-006/2014), Max Planck Society Germany through the award of a Max Planck Partner Group in Mathematical Biology, and the Abdus Salam International Centre for Theoretical Physics Italy for the award of a Simons Associateship. The funders have no role in study design, data collection, data analysis, manuscript preparation or decision to publish.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

MD	Molecular dynamic
COVID-19	Coronavirus disease 2019
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SARS-CoV	Severe acute respiratory syndrome coronavirus
MERS-CoV	Middle East respiratory syndrome coronavirus
TMPRSS2	Transmembrane Protease Serine 2
RBD	Receptor binding domain
ACE-2	Angiotensin converting enzyme 2
ORF	Open reading frame
FDA	Food and drug administration
ADMET	Absorption, distribution, metabolism, excretion and toxicity
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area

## References

- Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [[CrossRef](#)]
- Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395*, 470–473. [[CrossRef](#)]
- Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [[CrossRef](#)]
- Chan, J.F.-W.; Yuan, S.; Kok, K.-H.; To, K.K.-W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.-Y.; Poon, R.W.-S.; et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **2020**, *395*, 514–523. [[CrossRef](#)]
- Li, W. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* **2005**, *310*, 676–679. [[CrossRef](#)] [[PubMed](#)]
- Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [[CrossRef](#)]
- De Wit, E.; Van Doremalen, N.; Falzarano, D.; Munster, V. SARS and MERS: Recent insights into emerging coronaviruses. *Nat. Rev. Genet.* **2016**, *14*, 523–534. [[CrossRef](#)]

8. Li, G.; De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* **2020**, *19*, 149–150. [[CrossRef](#)]
9. Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J.; et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **2020**, *27*, 325–328. [[CrossRef](#)] [[PubMed](#)]
10. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T.S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181*, 271–280.e8. [[CrossRef](#)] [[PubMed](#)]
11. Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **2020**, *11*, 1620. [[CrossRef](#)] [[PubMed](#)]
12. Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11727–11734. [[CrossRef](#)] [[PubMed](#)]
13. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [[CrossRef](#)] [[PubMed](#)]
14. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444–1448. [[CrossRef](#)]
15. Hoffmann, M.; Schroeder, S.; Kleine-Weber, H.; Müller, M.A.; Drosten, C.; Pöhlmann, S. Nafamostat Mesylate Blocks Activation of SARS-CoV-2: New Treatment Option for COVID-19. *Antimicrob. Agents Chemother.* **2020**, *64*, 64. [[CrossRef](#)]
16. Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* **2020**, *10*, 766–788. [[CrossRef](#)]
17. ElFiky, A.A. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sci.* **2020**, *253*, 117592. [[CrossRef](#)]
18. Islam, R.; Parves, R.; Paul, A.S.; Uddin, N.; Rahman, S.; Al Mamun, A.; Hossain, N.; Ali, A.; Halim, M.A. A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, 1–12. [[CrossRef](#)]
19. Rahman, N.; Basharat, Z.; Yousuf, M.; Castaldo, G.; Rastrelli, L.; Khan, H. Virtual Screening of Natural Products against Type II Transmembrane Serine Protease (TMPRSS2), the Priming Agent of Coronavirus 2 (SARS-CoV-2). *Molecules* **2020**, *25*, 2271. [[CrossRef](#)]
20. Shah, B.; Modi, P.; Sagar, S.R. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. *Life Sci.* **2020**, *252*, 117652. [[CrossRef](#)]
21. Kumar, A.; Choudhir, G.; Shukla, S.K.; Sharma, M.; Tyagi, P.; Bhushan, A.; Rathore, M. Identification of phytochemical inhibitors against main protease of COVID-19 using molecular modeling approaches. *J. Biomol. Struct. Dyn.* **2020**, 1–11. [[CrossRef](#)]
22. Liu, S.; Zheng, Q.; Wang, Z. Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. *Bioinformatics* **2020**, *36*, 3295–3298. [[CrossRef](#)] [[PubMed](#)]
23. Olubiyi, O.O.; Olagunju, M.O.; Keutmann, M.; Loschwitz, J.; Strodel, B. High Throughput Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-CoV-2. *Molecules* **2020**, *25*, 3193. [[CrossRef](#)] [[PubMed](#)]
24. Ahmad, M.; Dwivedy, A.; Mariadasse, R.; Tiwari, S.; Kar, D.; Jeyakanthan, J.; Biswal, B.K. Prediction of Small Molecule Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus-2 RNA-dependent RNA Polymerase. *ACS Omega* **2020**. [[CrossRef](#)] [[PubMed](#)]
25. Newman, D.J.; Cragg, G.M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311–335. [[CrossRef](#)]
26. Ren, J.-L.; Zhang, A.-H.; Wang, X.-J. Traditional Chinese medicine for COVID-19 treatment. *Pharmacol. Res.* **2020**, *155*, 104743. [[CrossRef](#)]
27. Vellingiri, B.; Jayaramayya, K.; Iyer, M.; Narayanasamy, A.; Govindasamy, V.; Giridharan, B.; Ganesan, S.; Venugopal, A.; Venkatesan, D.; Ganesan, H.; et al. COVID-19: A promising cure for the global panic. *Sci. Total Environ.* **2020**, *725*, 138277. [[CrossRef](#)]

28. Mohanraj, K.; Karthikeyan, B.S.; Vivek-Ananth, R.P.; Chand, R.P.B.; Aparna, S.R.; Mangalapandi, P.; Samal, A. IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Sci. Rep.* **2018**, *8*, 4329. [[CrossRef](#)]
29. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [[CrossRef](#)]
30. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2018**, *47*, D1102–D1109. [[CrossRef](#)] [[PubMed](#)]
31. O’Boyle, N.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, 33. [[CrossRef](#)] [[PubMed](#)]
32. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; Beer, T.A.P.D.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)] [[PubMed](#)]
33. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [[CrossRef](#)] [[PubMed](#)]
34. 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.* **2009**, *31*, 455–461. [[CrossRef](#)]
35. Paoloni-Giacobino, A.; Chen, H.; Peitsch, M.C.; Rossier, C.; Antonarakis, S.E. Cloning of the TMPRSS2 Gene, Which Encodes a Novel Serine Protease with Transmembrane, LDLRA, and SRCR Domains and Maps to 21q22.3. *Genomics* **1997**, *44*, 309–320. [[CrossRef](#)]
36. Tsai, Y.-C.; Lee, C.-L.; Yen, H.-R.; Chang, Y.-S.; Lin, Y.-P.; Huang, S.-H.; Lin, C.-W. Antiviral Action of Tryptanthrin Isolated from *Strobilanthes cusia* Leaf against Human Coronavirus NL63. *Biomolecules* **2020**, *10*, 366. [[CrossRef](#)]
37. Zhao, D.-G.; Zhou, A.-Y.; Du, Z.; Zhang, Y.; Zhang, K.; Ma, Y.-Y. Coumarins with  $\alpha$ -glucosidase and  $\alpha$ -amylase inhibitory activities from the flower of *Edgeworthia gardneri*. *Fitoterapia* **2015**, *107*, 122–127. [[CrossRef](#)]
38. Gao, D.; Zhang, Y.-L.; Xu, P.; Lin, Y.-X.; Yang, F.; Liu, J.-H.; Zhu, H.-W.; Xia, Z. In vitro evaluation of dual agonists for PPAR $\gamma$ / $\beta$  from the flower of *Edgeworthia gardneri* (wall.) Meisn. *J. Ethnopharmacol.* **2015**, *162*, 14–19. [[CrossRef](#)]
39. Gao, D.; Zhang, Y.-L.; Yang, F.; Li, F.; Zhang, Q.-H.; Xia, Z.-N. The flower of *Edgeworthia gardneri* (wall.) Meisn. suppresses adipogenesis through modulation of the AMPK pathway in 3T3-L1 adipocytes. *J. Ethnopharmacol.* **2016**, *191*, 379–386. [[CrossRef](#)]
40. Gupta, P.C.; Sharma, N.; Rao, C.V. A review on ethnobotany, phytochemistry and pharmacology of *Fumaria indica* (Fumitory). *Asian Pac. J. Trop. Biomed.* **2012**, *2*, 665–669. [[CrossRef](#)]
41. Kardos, J.; Blaskó, G.; Kerekes, P.; Kovács, I.; Simonyi, M. Inhibition of [3H]GABA binding to rat brain synaptic membranes by bicuculline related alkaloids. *Biochem. Pharmacol.* **1984**, *33*, 3537–3545. [[CrossRef](#)]
42. Chand, A.; Sahoo, D.K.; Rana, A.; Jena, S.; Biswal, H.S. The Prodigious Hydrogen Bonds with Sulfur and Selenium in Molecular Assemblies, Structural Biology, and Functional Materials. *Acc. Chem. Res.* **2020**. [[CrossRef](#)]
43. Acharya, R.; Mitra, S.; Shukla, V. Effect of Shodhana (processing) on Kupeelu (*Strychnos nux-vomica* Linn.) with special reference to strychnine and brucine content. *AYU (An. Int. Q. J. Res. Ayurveda)* **2011**, *32*, 402–407. [[CrossRef](#)] [[PubMed](#)]
44. Ueno, S.; Bracamontes, J.; Zorumski, C.; Weiss, D.S.; Steinbach, J.H. Bicuculline and Gabazine Are Allosteric Inhibitors of Channel Opening of the GABAA Receptor. *J. Neurosci.* **1997**, *17*, 625–634. [[CrossRef](#)] [[PubMed](#)]
45. Khare, C.P. *Indian Medicinal Plants: An Illustrated Dictionary*; Springer: New York, NY, USA, 2007.
46. Yadav, J.P.; Arya, V.; Yadav, S.; Panghal, M.; Kumar, S.; Dhankhar, S. *Cassia occidentalis* L.: A review on its ethnobotany, phytochemical and pharmacological profile. *Fitoterapia* **2010**, *81*, 223–230. [[CrossRef](#)]
47. Zhu, S.H.; Yang, Z.Q.; Peng, X.L. Application of Anthraquinone Derivative in Resisting Influenza Virus and Bird Flu Virus H5N1. China Patent CN20051134677, 21 December 2005.
48. Mete, I.E.; Gözler, T. (+)-Oxoturkiyenine: An Isoquinoline-Derived Alkaloid from *Hypecoum pendulum*. *J. Nat. Prod.* **1988**, *51*, 272–274. [[CrossRef](#)]

49. Gurung, P.; De, P. Spectrum of biological properties of cinchona alkaloids: A brief review. *J. Pharmacogn. Phytochem.* **2017**, *6*, 162–166.
50. Tripathi, Y.C.; Maurya, S.; Singh, V.; Pandey, V. Cyclopeptide alkaloids from *Zizyphus rugosa* bark. *Phytochemistry* **1989**, *28*, 1563–1565. [[CrossRef](#)]
51. Yadav, A.; Singh, P. Analgesic and anti-inflammatory activities of *Zizyphus rugosa* root barks. *J. Chem. Pharm. Res.* **2010**, *2*, 255–259.
52. Wang, J.-H.; Luan, F.; He, X.-D.; Wang, Y.; Li, M.-X. Traditional uses and pharmacological properties of *Clerodendrum* phytochemicals. *J. Tradit. Complement. Med.* **2018**, *8*, 24–38. [[CrossRef](#)]
53. Sumthong, P.; Romero-González, R.R.; Verpoorte, R. Identification of Anti-Wood Rot Compounds in Teak (*Tectona grandis*L.f.) Sawdust Extract. *J. Wood Chem. Technol.* **2008**, *28*, 247–260. [[CrossRef](#)]
54. Jain, M.; Kapadia, R.; Jadeja, R.N.; Thounaojam, M.C.; Devkar, R.V.; Mishra, S.H. Traditional uses, phytochemistry and pharmacology of *Tecomella undulata*—A review. *Asian Pac. J. Trop. Biomed.* **2012**, *2*, 1918. [[CrossRef](#)]
55. Cadelis, M.; Bourguet-Kondracki, M.-L.; Dubois, J.; Valentin, A.; Barker, D.; Copp, B.R. Discovery and preliminary structure–activity relationship studies on tecomaquinone I and tectol as novel farnesyltransferase and plasmodial inhibitors. *Bioorganic Med. Chem.* **2016**, *24*, 3102–3107. [[CrossRef](#)] [[PubMed](#)]
56. Bosisio, E. Effect of the flavanolignans of *Silybum marianum* L. On lipid peroxidation in rat liver microsomes and freshly isolated hepatocytes. *Pharmacol. Res.* **1992**, *25*, 147–165. [[CrossRef](#)]
57. Bahmani, M.; Shirzad, H.; Rafieian, S.; Rafieian-Kopaei, M. *Silybum marianum*: Beyond Hepatoprotection. *J. Evid. -Based Complementary Altern. Med.* **2015**, *20*, 292–301. [[CrossRef](#)] [[PubMed](#)]
58. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [[CrossRef](#)]
59. Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041. [[CrossRef](#)]
60. Kumari, R.; Kumar, R.; Lynn, A. Open Source Drug Discovery Consortium g\_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962. [[CrossRef](#)]
61. *The Wealth of India the wealth of India: A Dictionary of Indian Raw Materials & Industrial Products. First Supplement Series (Raw Materials)*; Publications & Information Directorate, Council of Scientific & Industrial Research: New Delhi, India, 2000; Volume 1.
62. *The Wealth of India the Wealth of India: A Dictionary of Indian Raw Materials & Industrial Products. First Supplement Series (Raw Materials)*; Publications & Information Directorate, Council of Scientific & Industrial Research: New Delhi, India, 2000; Volume 2.
63. *The Wealth of India the wealth of India: A Dictionary of Indian Raw Materials & Industrial Products. First Supplement Series (Raw Materials)*; Publications & Information Directorate, Council of Scientific & Industrial Research: New Delhi, India, 2000; Volume 3.
64. *The Wealth of India the Wealth of India: A Dictionary of Indian Raw Materials & Industrial Products. First Supplement Series (Raw Materials)*; Publications & Information Directorate, Council of Scientific & Industrial Research: New Delhi, India, 2000; Volume 4.
65. *The Wealth of India the Wealth of India: A Dictionary of Indian Raw Materials & Industrial Products. First Supplement Series (Raw Materials)*; Publications & Information Directorate, Council of Scientific & Industrial Research: New Delhi, India, 2000; Volume 5.
66. Rastogi, R.P. (Ed.) Drug Research Perspectives. In *Compendium of Indian Medicinal Plants*; Central Drug Research Institute: Lucknow, India, 1990; Volume 1, pp. 1960–1969.
67. Rastogi, R.P. (Ed.) Drug Research Perspectives. In *Compendium of Indian Medicinal Plants*; Central Drug Research Institute: Lucknow, India, 1991; Volume 2, pp. 1970–1979.
68. Rastogi, R.P. (Ed.) Drug Research Perspectives. In *Compendium of Indian Medicinal Plants*; Central Drug Research Institute: Lucknow, India, 1993; Volume 3, pp. 1980–1984.
69. Rastogi, R.P. (Ed.) Drug Research Perspectives. In *Compendium of Indian Medicinal Plants*; Central Drug Research Institute: Lucknow, India, 1995; Volume 4, pp. 1985–1989.
70. Rastogi, R.P. (Ed.) Drug Research Perspectives. In *Compendium of Indian Medicinal Plants*; Central Drug Research Institute: Lucknow, India, 1998; Volume 5, pp. 1990–1994.
71. Evnin, L.B.; Vasquez, J.R.; Craik, C.S. Substrate specificity of trypsin investigated by using a genetic selection. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6659–6663. [[CrossRef](#)]

72. Herter, S.; Piper, D.E.; Aaron, W.; Gabriele, T.; Cutler, G.; Cao, P.; Bhatt, A.S.; Choe, Y.; Craik, C.S.; Walker, N.; et al. Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. *Biochem. J.* **2005**, *390*, 125–136. [[CrossRef](#)]
73. Chen, V.B.; Arendall, W.B.; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 12–21. [[CrossRef](#)] [[PubMed](#)]
74. Kuhn, B.; Tichý, M.; Wang, L.; Robinson, S.; Martin, R.E.; Kuglstatter, A.; Benz, J.; Giroud, M.; Schirmeister, T.; Abel, R.; et al. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. *J. Med. Chem.* **2017**, *60*, 2485–2497. [[CrossRef](#)] [[PubMed](#)]
75. Fujishima, A.; Imai, Y.; Nomura, T.; Fujisawa, Y.; Yamamoto, Y.; Sugawara, T. The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett.* **1997**, *407*, 47–50. [[CrossRef](#)]
76. Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D. Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta (BBA)-Proteins Proteom.* **2012**, *1824*, 68–88. [[CrossRef](#)] [[PubMed](#)]
77. Adams-Cioaba, M.A.; Krupa, J.C.; Xu, C.; Mort, J.S.; Min, J. Structural basis for the recognition and cleavage of histone H3 by cathepsin L. *Nat. Commun.* **2011**, *2*, 197. [[CrossRef](#)]
78. Morris, G.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)]
79. Rodrigues, J.P.; Mc Teixeira, J.; Trellet, M.; Bonvin, A.M.J.J. pdb-tools: A swiss army knife for molecular structures. *F1000Research* **2018**, *7*, 1961. [[CrossRef](#)]
80. Schmidt, T.; Haas, J.; Cassarino, T.G.; Schwede, T. Assessment of ligand-binding residue predictions in CASP9. *Proteins Struct. Funct. Bioinform.* **2011**, *79*, 126–136. [[CrossRef](#)]
81. Sarkhel, S.; Desiraju, G.R. N-H ... O, O-H ... O, and C-H ... O hydrogen bonds in protein-ligand complexes: Strong and weak interactions in molecular recognition. *Proteins Struct. Funct. Bioinform.* **2003**, *54*, 247–259. [[CrossRef](#)]
82. Zhou, P.; Tian, F.; Lv, F.; Shang, Z. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins Struct. Funct. Bioinform.* **2008**, *76*, 151–163. [[CrossRef](#)]
83. Kříž, K.; Fanfrlík, J.; Lepsik, M. Chalcogen Bonding in Protein–Ligand Complexes: PDB Survey and Quantum Mechanical Calculations. *Chem. Phys. Chem.* **2018**, *19*, 2540–2548. [[CrossRef](#)] [[PubMed](#)]
84. Borozan, S.; Stojanović, S.Đ. Halogen bonding in complexes of proteins and non-natural amino acids. *Comput. Biol. Chem.* **2013**, *47*, 231–239. [[CrossRef](#)] [[PubMed](#)]
85. De Freitas, R.F.; Schapira, M. A systematic analysis of atomic protein–ligand interactions in the PDB† Electronic supplementary information (ESI) available. See doi:10.1039/c7md00381a. *Med. Chem. Comm.* **2017**, *8*, 1970–1981. [[CrossRef](#)]
86. Turk, V.; Turk, B.; Turk, D. NEW EMBO MEMBERS’ REVIEW: Lysosomal cysteine proteases: Facts and opportunities. *EMBO J.* **2001**, *20*, 4629–4633. [[CrossRef](#)] [[PubMed](#)]
87. Liu, T.; Luo, S.; Libby, P.; Shi, G. Cathepsin L-selective inhibitors: A potentially promising treatment for COVID-19 patients. *Pharmacol. Ther.* **2020**, *213*, 107587. [[CrossRef](#)]
88. Abraham, M.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *19*–25. [[CrossRef](#)]
89. Schmid, N.; Eichenberger, A.P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.E.; Van Gunsteren, W.F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856. [[CrossRef](#)]
90. Stroet, M.; Caron, B.; Visscher, K.M.; Geerke, D.P.; Malde, A.K.; Mark, A.E. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *J. Chem. Theory Comput.* **2018**, *14*, 5834–5845. [[CrossRef](#)]
91. Berendsen, H.J.C.; Postma, J.P.M.; Van Gunsteren, W.F.; DiNola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690. [[CrossRef](#)]
92. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182. [[CrossRef](#)]

93. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [[CrossRef](#)] [[PubMed](#)]
94. Schyman, P.; Liu, R.; Desai, V.; Wallqvist, A. vNN Web Server for ADMET Predictions. *Front. Pharmacol.* **2017**, *8*, 889. [[CrossRef](#)] [[PubMed](#)]
95. Yuan, H.; Ma, Q.; Ye, L.; Piao, G. The Traditional Medicine and Modern Medicine from Natural Products. *Molecules* **2016**, *21*, 559. [[CrossRef](#)] [[PubMed](#)]

**Sample Availability:** Samples of the compounds are not available from the authors.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).