VirusDis. (January–March 2021) 32(1):98–107
https://doi.org/10.1007/s13337-021-00654-x

ORIGINAL ARTICLE

Phytochemicals as potential drug candidates for targeting SARS CoV 2 proteins, an in silico study

Anish Nag[1] · Ritesh Banerjee[2] · Rajshree Roy Chowdhury[1] · Chandana Krishnapura Venkatesh[1]

Received: 26 August 2020 / Accepted: 18 January 2021 / Published online: 5 April 2021

Abstract Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a member of the family Coronaviridae, and the world is currently witnessing a global pandemic outbreak of this viral disease called COVID-19. With no specific treatment regime, this disease is now a serious threat to humanity and claiming several lives daily. In this work, we selected 24 phytochemicals for an in silico docking study as candidate drugs, targeting four essential proteins of SARS-CoV-2 namely Spike glycoprotein (PDB id 5WRG), Nsp9 RNA binding protein (PDB id 6W4B), Main Protease (PDB id 6Y84), and RNA dependent RNA Polymerase (PDB id 6M71). After statistical validation, the results indicated that a total of 11 phytochemicals divided into two clusters might be used as potential drug candidates against SARS-CoV-2.

Keywords SARS-CoV-2 · In silico · Docking · Phytochemicals · Remdesivir

Introduction

Coronaviruses are single-stranded RNA viruses belonging to the family Coronaviridae and were known for causing mild respiratory infections in birds and mammals. These viruses were considered as minor pathogens for human until the emergence of two infamous zoonotic members of this family, the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) [20] and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) [51]. In the last two decades, it caused severe and life-threatening respiratory infections in humans across the globe. However, the world is recently witnessing the new and deadlier outbreak of acute pneumonia disease called as 'Coronavirus diseases 2019' (COVID-19) caused by the same viral family member Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [45, 53]. In late 2019, the novel coronavirus SARS-CoV-2 was first reported and identified in Wuhan of Hubei province of China and spread rapidly afterwards throughout the globe, causing severe to fatal respiratory illnesses [3]. SARS-CoV-2 is a highly transmittable pathogenic virus with an estimated reproductive number (Ro) of 2.2. World Health Organization (WHO) declared COVID-19 as a public health emergency of international concern with 23,752,965 confirmed infections and 815,038 death reports worldwide so far (26th August 2020) [47]. Some common symptoms of COVID-19 are dry cough, high fever, shortness of breath, muscle aches, fatigue, etc. that may arise within 3 to 14 days after pathogen exposure. In some severe cases, it may also cause Acute Respiratory Distress Syndrome (ARDS) leading to septic shock, and multi-organ failure due to fluid builds up within and around the lungs, drastically reduced blood pressure and oxygen starvation [29].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13337-021-00654-x.

✉ Anish Nag
anish.nag@christuniversity.in

[1] Department of Life Sciences, CHRIST (Deemed to be University), Central Campus, Bangalore 560029, India

[2] Department of Botany, University of Calcutta, Kolkata 700019, India

SARS-CoV-2 is a Beta Coronavirus containing a lipid membrane with envelope protein, hemagglutinin esterase dimer protein, membrane glycoprotein, spike protein, and positive-sense single-stranded RNA (~ 30 kb) with nucleocapsid protein [2]. CoVs invade the pulmonary epithelial cells of the lower respiratory system of the patient, deliver their nucleocapsid in the cell, and replicate in the cytoplasm by hijacking the cellular machinery [37]. Homotrimer transmembrane spike (S) glycoproteins of SARS-CoV-2 promote host attachment with the help of its S1 subunit and help the virus to enter into the host cell by virus-cell membrane fusion through S2 subunit [42, 45]. Cellular proteases cleave the S protein at the S1/S2 and S2' sites to allow the entry of the viral particles, followed by fusion of viral capsid with the cellular membrane [13]. After the entry, the virion releases the viral RNA inside the cell and translates several polyproteins using the machinery of the host cell that are cleaved subsequently into 27 viral proteins by internally encoded proteases.

Further processing leads to the production of several non-structural proteins (Nsps) and structural proteins that play an essential role in the synthesis of viral RNA and assembly of the virions, respectively [22, 25]. On the other hand, the virion uses RNA dependent RNA polymerase (RdRp) to replicate its daughter RNA genome [14]. These viral proteins can be the primary targets of effective drugs to suppress viral entry and replication.

Currently, there is no vaccine or specific drugs available for COVID-19 except for symptomatic supportive therapy. The treatment of the infected patients is limited to isolation and application of some broad-spectrum antiviral drugs [49]. Some antiviral medications like Remdesivir, Ganciclovir, Lopinavir, and Ritonavir are being tested clinically against COVID-19. Recently, antimalarial drug hydroxychloroquine and chloroquine had been used to treat COVID-19 infected individuals [6, 23]. Among these antiviral drugs, Remdesivir specifically was found to be effective against SARS-CoV as revealed by experimental and computational biology based evidence. Experimental studies indicated that the principal mechanism of this drug to block the viral RNA transcription. CoV is susceptible to the Remdesivir, targeting RNA dependent RNA polymerase and Non Structural proteins (NSps) [1, 19, 38, 46, 50]. Furthermore, Hall et al. [11] showed that Remdesivir, along with other drugs, could inhibit the main protease of SARS CoV in an in silico study. Considering all these studies, Remdesivir was selected as a control drug in this study. However, the search for an effective and specific cure for SARS-CoV-2 is still on.

Plant-derived natural products and metabolites have been used as traditional medicines to treat different diseases around the world for ages. These plant metabolites comprise several functional bioactive compounds that gained massive interest in the pharmacological and clinical industries to prevent and cure several diseases and disorders. Common phytochemicals like flavonoids, terpenoids, phenols, xanthophylls, carotenoids, and essential oils are used as potent sources of immunomodulatory, antitumor, antimicrobial, and antioxidative drugs for the treatment of several diseases [8, 15]. Several researchers demonstrated and strongly suggested the antiviral activity of several phytochemicals using various biological systems [24, 28]. The success of the quest for an appropriate antiviral drug entirely depends on the comprehensive pharmacodynamic screening and identification of potential broad-spectrum antiviral Phyto-compounds keeping in mind the bioavailability and stability. The interaction study between target proteins and drug compounds by experimental approaches are time-consuming and costly. The application of the latest biomedical tools and in-silico techniques are inexpensive techniques that help to find the efficacy of phytochemicals as the source of drugs within a short period thereby drastically reducing the time and cost of research and drug development. Hence for effective drug development against COVID-19, preliminary bioinformatics analysis of SARS-CoV-2 proteins and exploration of potential bioactive phytochemicals by in silico prediction of their interaction with the target proteins are of high importance for the best and appropriate use of limited resources. In this study, we performed a computational analysis to identify potent phytochemical compounds against different SARS-CoV-2 proteins. The drug-like properties of the selected phytochemicals were evaluated, followed by structural optimisation of the ligands. The molecular docking experiment was performed to assess the binding affinity of these phytochemicals to the SARS-CoV-2 protein receptors and predict the new potentially active bioactive compounds with antiviral properties.

Materials and methods

Selection of receptor and ligands

Twenty-Four phytochemicals were downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and considered as ligands for this study.

Three (3) proteins of novel Coronavirus (SARS-CoV-2) namely SARS-CoV spike glycoprotein (PDB id 5WRG 4.30 Å, Electron Microscopy), Nsp9 RNA binding protein (PDB id 6W4B, 2.95 Å, X-Ray diffraction), main protease (PDB id 6Y84, 1.39 Å, X-Ray diffraction), and SARS-Cov-2 RNA-dependent RNA polymerase (PDB id 6M71, 2.90 Å, Electron Microscopy)were considered as receptors. Remdesivir (Compound CID: 121,304,016), an antiviral drug, was taken as control.

Evaluation of drug-like properties

The canonical smile formats of the phytochemicals were uploaded in the SwissADME (http://www.swissadme.ch/) site to evaluate its pharmacological and drug-likeness.

Preparation of protein receptors

The proteins (PDB id 5WRG, 6W4B,6Y84, and 6M71) were prepared by retrieving the three-dimensional crystal structure of each from the RCSB protein bank (https://www.rcsb.org/), structure optimisation was performed by UCSF Chimera software [32].

Preparation of Ligands

The three-dimensional structures of the phytochemicals, as well as the control drug Remdesivir (Compound CID: 121,304,016), were downloaded from PubChem in '. SDF' format. Structural optimisation and conversion in the PDB format were done by Avogadro software [12] before conducting the molecular docking analysis.

Molecular docking

Following receptor and ligand preparation, molecular docking analysis was performed by DockThor web server (https://www.dockthor.lncc.br/v2/) [36] to evaluate the binding affinities. After the minimisation process, the grid box resolution was set along the x, y, and z points (size and center), respectively in a partially blind docking mode (dimension: x, y and z: 188.446, 193.4115 and 169.673 for 5WRG; 40.4825, − 12.5045 and 13.711 for 6W4B 11.6405, − 0.022 and 6.329 for 6Y84 and 121.009, 121.761 and124.981 for 6M71; Grid Size was set as 40) grid was centred onto the proteins. The control drugs (Remdesivir), as well as all 24 phytochemicals, were docked with all three protein receptors and the resulting interactions were compared with those calculated docking results of the Remdesivir with the same receptors. The visualisation and analysis of the docking sites were done by Discovery Studio 2020 (BIOVIA, San Diego, USA).

Statistical analysis (PCA and hierarchical clustering)

Multivariate data analysis based on Principal Component Analysis (PCA) tool was performed by Minitab software (Minitab 18). Generation of the single coloured heat map and hierarchical clustering analysis based on Pearson correlation was performed by Molecular Experiment Viewer 4.9.0 (MEV 4.9.0). We used the four docking scores (affinity kcal/mol) to four different targets as input to PCA to extract PC1 and PC2 coordinates as well as to construct the hierarchical clusters.

Chemical characterisation

ClassyFire (http://classyfire.wishartlab.com/) relies on a comprehensible, comprehensive, and computable chemical taxonomy. It is a free accessible web-based application for automated structural classification of chemical entities. Based on the PCA results, we performed ClassyFire based structural analysis of compounds to understand their intra and inter-cluster relationships.

Results and discussion

Evaluation of drug-like properties

The process of drug discovery is evolving since its inception. To increase the accessibility and effectiveness of the drug discovery process, researchers have been continuously striving to develop new tools such as SwissADME. SwissADME is an open-source web server, and it predicts ADME (Absorption, Distribution, Metabolism, and Excretion) parameters and computes physicochemical descriptors, pharmacokinetic properties, drug-like nature and medicinal chemistry friendliness of one or multiple small molecules to support drug discovery. In this study, 24 phytochemicals (and Remdesivir) were screened and evaluated for drug-like properties. The result was represented in the BOILED-Egg graphical classification model (Supplementary material). BOILED-Egg graphical interface can predict passive diffusion through passive human gastrointestinal absorption (HIA) and blood-brain barrier (BBB) permeation by position in a WLOGP-versus-TPSA physicochemical space [5]. In our study, 4 (four) phytochemicals were found to show passive human gastrointestinal absorption (HIA), and 17 (seventeen) compounds showed a blood-brain barrier (BBB) permeation property. Four (4) compounds were found to be out of range. A membrane-bound transporter PGP (P-glycoprotein)

Table 1 SwissADME results

Molecules	TPSA	iLOGP	ESOL Log S	ESOL class	Lipinski #violations	Leadlikeness #violations	Bioavailability score
5-Methylundecane	0	3.72	− 4.42	Moderately soluble	1	3	0.55
7-Hydroxyfavone	50.44	2.22	− 4.19	Moderately soluble	0	2	0.55
Alpha Pinene	0	2.63	− 3.51	Soluble	1	2	0.55
Arachidonic acid	37.3	4.64	− 5.20	Moderately soluble	1	2	0.56
Benzenemethanol	20.23	1.66	− 1.69	Very soluble	0	1	0.55
Beta cubebene	0	3.39	− 4.01	Moderately soluble	1	2	0.55
Campesterol	20.23	4.92	− 7.54	Poorly soluble	1	2	0.55
Camphene	0	2.58	− 3.34	Soluble	1	2	0.55
Capric acid	37.3	2.5	− 2.96	Soluble	0	3	0.56
Cartilagineol	20.23	3.27	− 5.37	Moderately soluble	1	2	0.55
Corynan-17-Ol	39.26	2.71	− 3.81	Soluble	0	0	0.55
Demecolcine	66.02	3.47	− 3.03	Soluble	0	1	0.55
Elatol	20.23	3.22	− 4.52	Moderately soluble	1	1	0.55
Ethylbenzene	0	2.06	− 2.97	Soluble	0	1	0.55
Flavone	30.21	2.55	− 4.09	Moderately soluble	0	2	0.55
Loliolide	46.53	1.88	− 1.69	Very soluble	0	1	0.55
Neophytadoiene	0	5.05	− 6.77	Poorly soluble	1	2	0.55
Octanedioic acid	74.6	1.15	− 1.11	Very soluble	0	1	0.56
Octanoic acid	37.3	1.95	− 2.26	Soluble	0	1	0.56
Phytol	20.23	4.71	− 5.98	Moderately soluble	1	2	0.55
Remdesivir	213.36	3.24	− 4.12	Moderately soluble	2	2	0.17
Rutin	269.43	2.43	− 3.30	Soluble	3	1	0.17
Squalene	0	6.37	− 8.69	Poorly soluble	1	3	0.55
Stigmastanol	20.23	5	− 7.27	Poorly soluble	1	2	0.55
Withaferin A	96.36	3.39	− 4.97	Moderately soluble	0	2	0.55

mediates efflux (active transport) of a wide range of structurally unrelated drugs and other xenobiotics out of the cells. Against a concentration gradient, PGP induces the efflux of various substrates leading to the reduction of their intracellular concentration, thereby affecting the oral bioavailability of drugs [4]. BOILED-Egg graphical presentation further showed that among 25, 8 (eight) phytochemicals could function as PGP substrates (PGP+). Additionally, drug likeliness parameters, as shown in Table 1, showed that with few parameter exceptions, these phytochemicals could be used as potential drug candidates.

Molecular docking analysis

The current outbreak of the CoV has caused significant concern in the field of drug and vaccine development. Researchers have started investigating all possible compounds that could work against it. The target proteins used for molecular docking in this study were based on the significant role that they play for the survival of the virus. The use of phytochemicals over the chemically synthesised drugs has increased in demand in the last few decades, mainly due to their effectiveness and lesser-known side effects. A variety of active phytochemicals have been studied regularly for drug development [27].

Molecular docking is an efficient technique to study ligand-protein interaction probability. The results, in general, are expressed with free binding energy (kcal/mol; binding affinity) which is expected to be lower in case of optimal docking poses. DockThor is considered as one of the most useful docking servers. In a study, by Santos et al., 2020 showed that it could dock 40 % of the cases with an overall backbone RMSD below 2.5 Å when the top-scored docking pose was considered in other software. Further, DockThor was capable of assessing the docking poses closest to the crystal structure (i.e., best-RMSD pose), with a success rate of 60% [36]. In our study, DockThor aided molecular docking study revealed that phytochemicals could effectively bind with viral proteins when compared with the commercial drug Remdesivir (CID 121,304,016) (Table 2; Fig. 1).

SARS-CoV spike glycoprotein (PDB id 5WRG) is the surface protein that is mainly responsible for the initial attachment with the host cells receptor, Angiotensin-converting enzyme (ACE2). It also has a critical role in penetrating the cell by fusing with the membranes. The precursor chain of his protein is initially synthesised as a single polypeptide chain of 300 amino acids and further cleaved by host furin-like proteases into an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit [10]. Hence, inhibiting spike protein of SAR CoV shall prevent the viral entry, fusogenicity and propagation. Pandey et al. 2020 [30] in their work targeted the SARS-CoV spike glycoprotein with ten naturally occurring phytochemicals in an in-silico docking study and compounds like Fisetin, Quercetin, and Kaempferol showed significant binding affinities for viral spike glycoprotein. The authors observed the binding energies of the phytochemicals within the range of − 6.7 to − 8.5 kcal/mol, which was significantly lower than the control drug hydroxychloroquine (− 5.6 kcal/mol). Lower docking binding energy reflects better binding affinity of the ligands towards target proteins. The performance of these compounds in terms of prevention of infectivity and virulence of the viral pathogens was correlated with low binding free energy in the docking study. In our research, while docked with SARS-CoV spike glycoprotein (PDB id 5WRG), nine (09) compounds namely Cartilagineol (CID 101,934,341), Flavone (CID 10,680), Stigmastanol (CID 241,572), Campesterol (CID

Table 2 Docking results (Affinity-Kcal/mol) as generated by DockThor analysis between phytochemicals and proteins

Molecules	PubChem ID	Compound	Affinity (Kcal/mol)			
			5WRG	6W4B	6Y84	6M71
1	CID 94,213	5-Methylundecane	− 6.727	− 7.483	− 6.185	− 6.835
2	CID 5,281,894	7-Hydroxyfavone	− 6.509	− 7.281	− 7.504	− 6.982
3	CID 6654	Alpha Pinene	− 7.043	− 7.275	− 6.957	− 7.200
4	CID 444,899	Arachidonic acid	− 6.938	− 7.21	− 6.694	− 6.183
5	CID 244	Benzenemethanol	− 6.054	− 6.16	− 6.395	− 6.521
6	CID 93,081	Beta cubebene	− 7.124	− 7.793	− 7.497	− 7.444
7	CID 173,183	Campesterol	− 7.491	− 7.627	− 7.78	− 7.287
8	CID 6616	Camphene	− 6.892	− 6.73	− 7.215	− 7.222
9	CID 2969	Capric acid	− 6.348	− 6.727	− 6.497	− 6.333
10	CID 101,934,341	Cartilagineol	− 7.780	− 7.586	− 7.058	− 7.081
11	CID 164,952	Corynan-17-Ol	− 7.487	− 6.676	− 6.733	− 7.770
12	CID 220,401	Demecolcine	− 7.294	− 7.652	− 6.922	− 7.440
13	CID 479,931	Elatol	− 7.405	− 7.58	− 7.124	− 6.921
14	CID 7500	Ethylbenzene	− 7.229	− 7.327	− 6.917	− 7.265
15	CID 10,680	Flavone	− 7.683	− 8.405	− 7.431	− 7.139
16	CID 100,332	Loliolide	− 6.620	− 6.604	− 6.492	− 6.614
17	CID 10,446	Neophytadoiene	− 7.076	− 6.604	− 6.743	− 7.319
18	CID 10,457	Octanedioic acid	− 6.309	− 6.057	− 6.081	− 5.956
19	CID 379	Octanoic acid	− 6.639	− 6.085	− 6.503	− 5.820
20	CID_5280435	Phytol	− 6.769	− 7.104	− 6.674	− 7.367
21	CID 121,304,016	Remdesivir	− 7.222	− 7.83	− 7.794	− 7.783
22	CID 5,280,805	Rutin	− 7.045	− 7.068	− 7.147	− 7.354
23	CID 638,072	Squalene	− 7.444	− 7.548	− 7.027	− 8.058
24	CID 241,572	Stigmastanol	− 7.634	− 7.761	− 7.44	− 7.735
25	CID_265237	Withaferin A	− 7.278	− 7.416	− 7.252	− 7.546

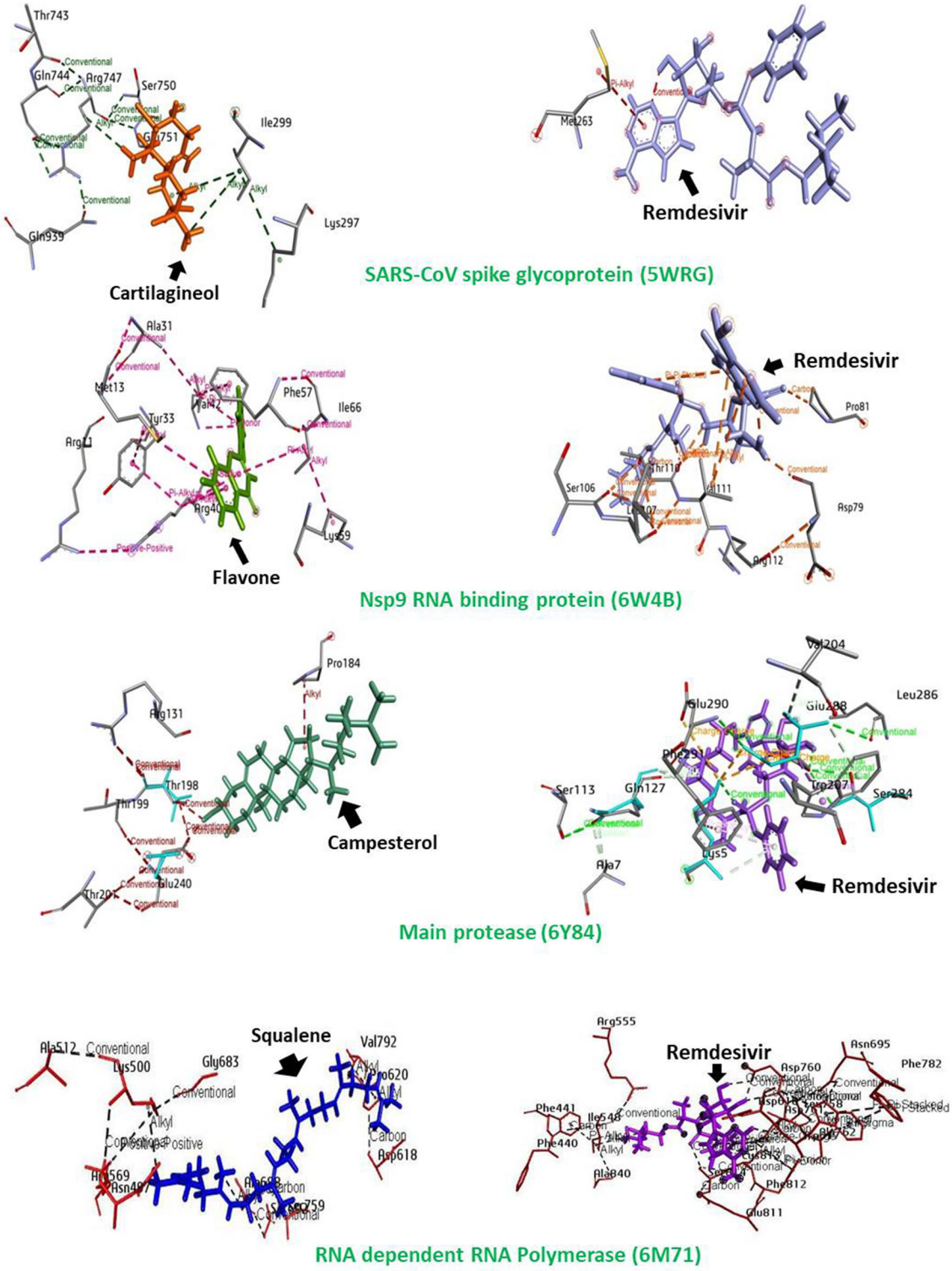

Fig. 1 Docking interaction of top ranking ligands with their respective proteins along with the drug Remdesivir

173,183), Corynan-17-Ol (CID 164,952), Squalene (CID 638,072), Elatol (CID 479,931), Demecolcine (CID 220,401) and Withaferin A (CID 265,237) showed better binding affinities (Kcal/mol) than the commercial drug Remdesivir (CID 121,304,016) within the range of − 7.278 to − 7.780 Kcal/mol. Cartilagineol interacted with ILE299 (chain A) and ARG747 (chain B) through Alkyl interactions of 5.37–5.40 and 5.48 Å distances respectively.

VanderWaal interactions were seen for the amino acids SER750 (chain B), GLY751 (chain B) and GLN939 (chain A). Flavone showed interaction with ARG977 (chain C) through H (3.35 Å) and Pi-alkyl (5.08 Å) bonds. Further, ASP976 (chain A) was observed as the ligand pocket amino acid. Corynan-17-Ol interacted with the amino acids ILE299 (chain A) and ALA754 (chain B) through two alkyls (5.17 Å)/pi-alkyl (5.21 Å) and three pi alkyl (4.56 Å)/pi-alkyl (4.70 Å) /alkyl (4.37 Å) bonds. VanderWaal interactions were observed for SER750 (chain A) and LYS297 (chain A). Amino acids, ALA754 (chain B) and SER750 (chain B) interacted with the ligand Elatol through alkyl (4.30 Å) and carbon-hydrogen (3.26 Å) bonds, respectively. ARG747 (chain B) showed Vanderwaal interaction. ARG977 (chain B) showed interaction with the phytochemical Ethylbenzene through Alkyl linkage (5 Å). ASP976 (chain C) showed Vanderwaal interaction. Squalene was found to interact with the amino acid namely, LEU736 (chain B), MET263 (chain A) and GLN737 (chain B) through Alkyl (5.21 and 4.89 Å) and Vanderwaal interactions respectively. The phytochemical Withaferin A contacted with the amino acid LYS946 (chain A) through conventional H bond (3.19 Å) and with SER289 (chain A) through Carbon-Hydrogen bond (3.34 Å). GLU285 (chain A) was found in the pocket. Interaction with Stigmastanol was noted by the amino acids ILE955 (chain A) and SER956 (chain A) through conventional H (3.06 Å) and Carbon Hydrogen (3.09 Å) bonds, respectively. For Demecolcine and Campesterol, only Vanderwaal interactions were observed. Control drug Remedesivir showed Pi-alkyl interaction (5.04 Å) with MET263 (chain A).

RNA dependent RNA polymerase (RdRp) is crucial for RNA viruses for synthesising daughter genome. Coronavirus expresses RdRp to synthesise daughter RNA genome [14]. Therefore, inhibition of RdRp can be an effective strategy to stop the growth of the viral population. In our study, Squalene (CID 638,072) showed better docking score in case of viral RNA dependent RNA polymerase protein (PDB Id 6M71) when compared with Remdesivir. Squalene, interacted with the protein through ALA688 (chain A, two Pi-alkyl bonds, 4.12 and 4.55 Å), LYS500 (chain A, Pi-alkyl bond, 4.49 Å) and PRO620 (chain A, Pi-alkyl bond, 4.34 Å). A few amino acids ASN497, ARG569, ASP618 and SER759 of chain A showed Vanderwaal interactions. Corynan-17-ol (CID 164,952) showed close association (− 7.770 Kcal/mol) with the Remdesivir score (− 7.783 Kcal/mol). This phytochemical showed multiple interaction sites with the protein. Amino acids namely ARG553 (chain A, two Pi-cation bonds, 4.34 and 4.82 Å), ARG624 (chain A, Pi-alkyl bond, 5.43 Å), TYR455 (chain A, Pi-Pi T shaped, 5.22 Å), LYS621 (chain A, three Pi-alkyl bonds, 4.64, 4.70 and 5.16 Å), ASP623 (chain A, conventional H bond, 3.10 Å) and ASP760 (Chain A, Conventional H bond, 1.66 Å). Remdesivir, interacted with the protein through the amino acids ILE548 (chain A, Pi alkyl bond, 4.94 Å), ASP760 (chain A, two conventional H bonds, 2.08 and 2.33 Å), ASP761 (chain A, three Pi-anion attractive charges, 3.61, 4.04 and 4.24 Å), SER814 (chain A, conventional H bond, 2.73 Å) and CYS813 (chain A, Pi-alkyl bond, 4.46 Å).

SARS-CoV-2 main protease (PDB id 6Y84) being reported to be involved in the viral translation process through the processing of the polyproteins [52] was considered as another target for this study. In a study, the drug likeliness of various phytochemicals from *Ocimum sanctum* was analysed against the CoV-2 main protease using docking protocols, and Tulsinol and Dihydrodieuginol B were identified to have a potent inhibitory effect on the viral protein [44]. Another study conducted by Jagdale et al. 2020 showed that phytochemicals namely taiwanhomflavone A from the tree *Cephalotaxus wilsoniana* and lactucopicrin15-oxolate from *Lactuca virosa* could inhibit SARS-CoV-2 main protease. Control drug Remdesivir showed interaction with the 6W4B through the amino acids THR 110 (chain A, conventional H bond, 3 Å), ASP79 (chain A, conventional H bond, 2.12 conventional H bond), VAL111 (chain A, two Pi-alkyl interactions, 4.81 and 5.28 Å) and PRO81 (chain A, Carbon Hydrogen bond, 3.52 Å). Although, Remdesivir was found to optimally bind with SARS-CoV-2 main protease (PDB id 6Y84) with the score − 7.794 Kcal/mol, however, Campesterol (CID 173,183) also showed close association (− 7.78 Kcal/mol) with the protein. Campesterol interacted with the amino acids GLU240 (chain A, conventional H bond, 2.92 Å), THR198 (chain A, conventional H bond, and PRO184 (chain A, Pi-alkyl bond, 4.29 Å). Remdesivir, on the other hand, showed interaction with the protein through LYS5 (chain A, conventional H bond, 3.91 Å), GLU288 (chain A, conventional H bond, 2.73 Å) and GLN127 (chain A, two Carbon Hydrogen bonds, 2.09 and 3.02 Å).

SARS CoV-2 Nsp9 RNA binding protein (PDB id 6W4B) is a non-structural protein that is presumed to have an essential role in binding with the RNA/DNA during replication; however, its direct involvement is still unclear [25]. Silva et al., 2020 [39] reported the best docking ligands for SARS—CoV Nsp15/NendoU as (E, E)-α-Farnesene,(E)-β-Farnesene, and (E, E)—Farnesol. (E, E)—Farnesol. Flavone (CID 10,680) was found to effectively bind with SARS CoV-2 Nsp9 RNA binding protein (PDB id 6W4B) with affinity − 8.405 Kcal/mol. Pi-sulfur bond (5.6 Å) interaction was observed between MET13 (chain A) of the protein and the phytochemical. Further, Flavone showed specific interactions of Pi-alkyl bonds between ILE66 (chain A) and ARG40 (chain A) and Flavone (5.44

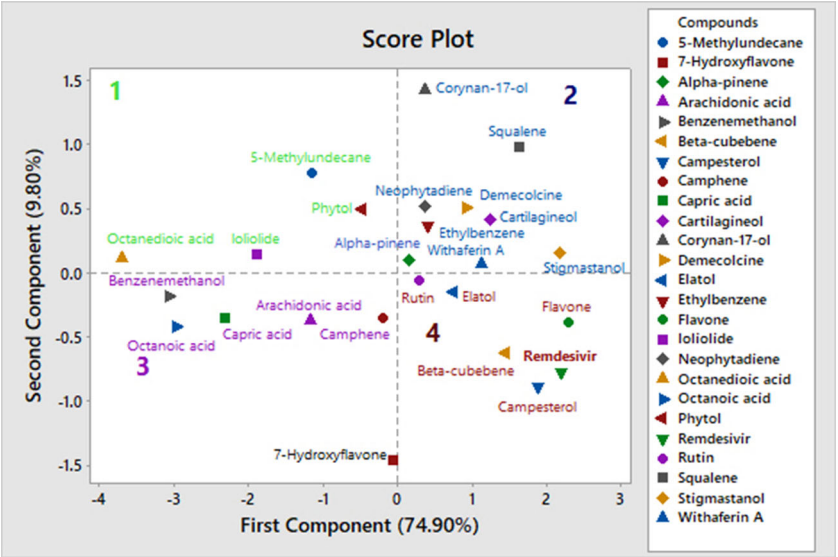

Fig. 2 PCA results of docking outputs (Affinity:Kcal/mol). Four clusters are shown in different colours

and 4.68 Å). MET13 (chain A) showed Pi-sulfur interaction (5.6 Å).

Finally, we identified compounds like Cartilagineol, Flavone, Campesterol, Corynan-17-Ol, Elatol, Ethylbenzene, Demecolcine, Beta cubebene and Squalene were found to be potential drug candidates for their respective targets. The extensive study on these natural phytochemicals also showed great results against other viruses like dengue, HIV, malaria, etc. [18, 43]. Many of these compounds are indicated in the literature to have pharmacological properties [48]. Cartilagineol has mainly derived from red algae *Laurencia* sp. Studies have been conducted to evaluate their properties, and it is well known for their antimicrobial and anti-inflammatory activity. Flavones are a large group of compounds that are naturally found in various plants such as Artemisia, Gnaphalium and Achyroclines and are medically acclaimed to have antioxidant, antimicrobial and anti-cancerous activity. Flavones are used for docking as a potential drug against viruses like Picorna virus [15], Dengue virus [26, 41] etc. On the other hand, phytochemicals like Stigastanol and Campesterol plant sterol derivatives found in plants like algae and aerial plants like *Caltopis gigantea* and *Carissa carandas* were studied in various molecular docking experiments of cancer [40] and Human Rhinovirus [17]. Squalene is obtained from plants like *Alliaria petiolata*. *Anacardium occidentale*, and *Carica papaya* and showed its efficacy against viral pathogens such as HIV, Dengue and Ebola in silico [7]. Praveena et al. [33] reported Corynan-17-ol and 18, 19-didehydro-10-methoxy from *Morinda tinctoria* fruit extract as lead molecules against breast cancer protein ErBb2 in an in silico docking study. Elatol a key phytochemical isolated from the marine algae red Seaweed *Laurencia dendroidea* showed anti-leishmanial activity *against Leishmania amazonensis* [34, 35]. Beta cubebene, as one of the constituents of essential oil of *Ocimum basilicum*, showed antioxidant and antiviral activities [34].

Phytochemicals with positive docking results in our study hence can be explored further as a potential SARS CoV 2 candidate drugs in agreement with the literature.

Statistical analysis: principal component and hierarchical clustering analysis

Principal Component Analysis (PCA) is a multidimensional data analysis tool which mainly deals with a large dataset and interprets them by reducing their dimensionality thereby making it easy to deduce with minimum loss of statistical information or "variability" [16]. It transforms measured variables into uncorrelated variables, i.e. principal components. Each of the principal components covers a separate dimension of variations of the measured dataset. While the first component shows the maximum variations of the dataset, the second component is orthogonal to the first one and covers remaining variations and so on [31].

Its working domain is vast, starting from biology, physiology, chemistry, engineering, physics and meteorology. The application of PCA ranges from data mining, quantitative structure-activity to 'omics' approaches [9]. To discern the overall quantitative relations among various phytochemicals, a PCA using the affinity values (Kcal/mol) of docked results were performed. First principal component (PC1) and the second component (PC2), as shown in Fig. 2 explained approximately 74.9 and 9.8% of the variance (total explained variations 84.7%), respectively. Considering both the components (PC1 and PC2), we observed four distinct groups in PCA analysis.

Phytochemicals namely Elatol, Flavone, Rutin, Beta-cubebene and Campesterol were placed in the same group as Remdesivir indicating statistically similar potentials.

Hierarchical cluster analysis is the iterative statistical method which involves multiple steps leading to the formation of small classes based the similar observations. To overview and interpret a large set of data, often those are grouped into smaller categories. By this, researchers can conveniently conclude [21]. Although we observed quite a few variations with PCA analysis, possibly due to different statistical algorithm, however, Remdesivir was placed with Beta-cubebene and Campesterol (Supplementary material).

Further, after performing chemotaxonomic analysis in ClassyFire, we observed that median values of the molecular weights for cluster 1, 2, 3 and 4 in PCA, were 185.22, 410.73, 158.24 and 367.19 g/mol respectively. All the PCA clusters were predominantly rich in lipid variants although with variations in their percentage content ($\sim$ 50, 75, 67 and 50% in Cluster 1, 2, 3 and 4 respectively) and chemical subclasses (fatty acyls-prenols, steroids-prenols, fatty acyls-prenols and steroid-prenols in cluster 1, 2, 3 and 4 respectively). Apart from lipid molecules, we also observed a significant presence of flavonoids in cluster 4. Cluster 1, however, showed the almost equal presence of saturated hydrocarbon-benzofuran along with lipids.

It can be concluded from the study that 11 (eleven) phytochemicals, as mentioned in the study, are capable of inhibiting specific target protein of SARS-CoV 2 and can be further explored as potential drug candidates.

Funding The authors declare no funding source.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Agostini ML, Erica LA, Amy C, Rachel LG, Timothy PS, Xiaotao L, et al. Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. MBio. 2018;9(2):e00221.
2. Beck BR, Bonggun S, Yoonjung C, Sungsoo P, Keunsoo K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol. 2020;18:784–90.
3. Bogoch II, Alexander W, Andrea TB, Carmen H, Moritz UGK, Kamran K. Potential for global spread of a novel coronavirus from China. J Travel Med. 2020;27(2):011.
4. Constantinides PP, Kishor MW. Lipid formulation strategies for enhancing intestinal transport and absorption of P-glycoprotein (P-gp) substrate drugs: in vitro/in vivo case studies. J Pharm Sci. 2007;96(2):235–48.
5. Daina A, Olivier M, Vincent Z. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7:42717.
6. Devaux CA, Jean-Marc R, Philippe C, Didier R. New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? Int J Antimicrob. 2020;55:105938.
7. Duke JA. Handbook of phytochemical constituent grass, herbs and other economic plants. Boca Raton: CRC Press; 1992.
8. Ezzat A, Min W, Xiao-Li L, Chee-Keong K. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. Brief Bioinform. 2019;20(4):1337–57.
9. Giuliani Al. The application of principal component analysis to drug discovery and biomedical data. Drug Discov Today. 2017;22(7):1069–76.
10. Gui M, Wenfei S, Haixia Z, Jingwei X, Silian C, Ye X, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res. 2017;27(1):119–29.
11. Hall J, Donald C, Hai-Feng J. A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Travel Med Infect Dis. 2020;35:101646.
12. Hanwell MD, Donald EC, David CL, Tim V, Eva Z, Geoffrey RH. Avogadro: an advanced semantic chemical editor, visualisation, and analysis platform. J Cheminform. 2012;4(1):17.
13. Hoffmann M, Hannah KW, Nadine K, Marcel AM, Christian D, Stefan P. The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. BioRxiv. 2020.
14. Huang J, Wenliang S, Hui H, Quancai S. Pharmacological therapeutics targeting RNA-dependent RNA polymerase, proteinase and spike protein: from mechanistic studies to clinical trials for COVID-19. J Clin Med. 2020;9(4):1131.
15. Ishitsuka H, Ohsawa C, Ohiwa T, Umeda I, Suhara Y. Antipicornavirus flavone Ro 09-0179. Antimicrob Agents Chemother. 1982;22(4):611–6.
16. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans R Soc A. 2016;374:20150202.
17. Kant K, Lal UR, Ghosh M In-silico discovery of natural lead hits from the genus of arisaema against Human Rhino Virus. In: Proceedings of the 20th international electronic conference on synthetic organic chemistry. 2016; 1–31.
18. Kehinde I, Pritika R, Manimbulu N, Michelle G. The pharmacokinetic properties of HIV-1 protease inhibitors: a computational perspective on herbal phytochemicals. Heliyon. 2019;5(10):e02565.
19. Khan RJ, Jha RK, Gizachew MA, Jain M, Singh E, Pathak A, et al. Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3 C-like proteinase and 2′-O-ribose methyltransferase. J Biomol Struct Dyn. 2020;2020:1–14.
20. Ksiazek TG, Dean E, Cynthia SG, Sherif RZ, Teresa P, Shannon E, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med. 2003;348(20):1953–66.
21. Levenstien MA, Yaning Y, Jürg O. Statistical significance for hierarchical clustering in genetic association and microarray expression studies. BMC Bioinform. 2003;4(1):62.
22. Littler D, Benjamin G, Rhys NC, Jamie R. Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. BioRxiv. 2020.
23. Liu C, Qiongqiong Z, Yingzhu L, Linda VG, Steve PW, Linda JC, et al. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. ACS Cent Sci. 2020;6:315–31.

24. Martin KW, Edzard E. Antiviral agents from plants and herbs: a systematic review. Antiviral Ther. 2003;8(2):77–90.
25. Miknis ZJ, Eric FD, Timothy CU, Ryan AR, Ralph SB, Wayne S. Severe acute respiratory syndrome coronavirus nsp9 dimerisation is essential for efficient viral growth. J Virol. 2009;83(7):3007–18.
26. Mir A, Humaira I, Sobiah R, Umar HKN. Identification of bio-flavonoid as fusion inhibitor of dengue virus using molecular docking approach. Inform Med. 2016;3:1–6.
27. Naithani R, Loredana CH, Louis EH, Deepak S, David LM, Rajendra GM, et al. Antiviral activity of phytochemicals: a comprehensive review. Mini-Rev Med Chem. 2008;8(11):1106–33.
28. Naithani R, Mehta RG, Shukla D, Chandersekera SN, Moriarty RM. Antiviral activity of phytochemicals: a current perspective. In: Watson RR, Zibadi S, Preedy VR, editors. Dietary components and immune function. Totowa: Humana Press; 2010. p. 421–68.
29. Pan C, Lu C, Cong L, Wei Z, Jia-An X, Michael CS, et al. Lung recruitability in COVID-19–associated acute respiratory distress syndrome: a single-center observational study. Am J Respir Crit Care Med. 2020;201:1294–7.
30. Pandey P, Rane JS, Chatterjee A, Kumar A, Khan R, Prakash A, et al. Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. J Biomol Struct Dyn. 2020;2020:1–11.
31. Patras A, Nigel PB, Gerard D, Ashish R, Keith W, Gwenole G. Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity. J Food Compos Anal. 2011;24:250–6.
32. Pettersen EF, Thomas DG, Conrad CH, Gregory SC, Daniel MG, Elaine CM, et al. UCSF Chimera—a visualisation system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.
33. Praveena A, Arthi S, Sudarmathi B. In vitro and in silico analysis to identify novel lead compound from Morinda tinctoria fruit against breast cancer. Indian J Pharm Sci. 2019;81(5):970–5.
34. Romeilah RM, Sayed AF, Ghada IM. Chemical compositions, antiviral and antioxidant activities of seven essential oils. Res J Appl Sci. 2010;6(1):50–62.
35. Santos AO, Phercyles VS, Tânia UN, Daniela BS, Éverson MB, Renato CP, et al. Effect of elatol, isolated from red seaweed Laurencia dendroidea, on Leishmania amazonensis. Mar Drugs. 2010;8(11):2733–43.
36. Santos KB, Isabella AG, Ana LMK, Laurent D. Highly flexible ligand docking: benchmarking of the DockThor program on the LEADS-PEP protein-peptide dataset. J Chem Inf Model. 2020;60:667–83.
37. Shah B, Palmi M, Sneha RS. In silico studies on therapeutic agents for COVID-19: drug repurposing approach. Life Sci. 2020;252:117652.
38. Siegel D, Hon CH, Edward D, Michael OC, Kwon C, Lijun Z, et al. Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo [2, 1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. J Med Chem. 2017;60(5):1648–61.
39. Silva JKR, Pablo LBF, Kendall GB, William NS. Essential oils as antiviral agents. Potential of essential oils to treat SARS–CoV–2 infection: an in -silico investigation. Int J Mol Sci. 2020;21(10):3426.
40. Sureshkumar P, Senthilraja P, Kalavathy S. In-silico docking analysis of *Calotropis gigantea* (L.) R. Br derived compound against anti-cervical cancer activity. World Res J Comput Aided Drug Des. 2012;1(1):9–12.
41. Thomas CM, Wood RC, Wyatt JE, Pendleton MH, Torrenegra RD. Anti-neoplastic activity of two flavone isomers derived from. PLos ONE. 2012;7(6):e39806.
42. Tortorici MA, David V. Structural insights into coronavirus entry. Adv Virus Res. 2019;105:93–116.
43. Ul Qamar T, Arooj M, Usman AA, Samia A, Tabeer F, Muhammad H, et al. Computer aided screening of phytochemicals from Garcinia against the dengue NS2B/NS3 protease. Bioinformation. 2104;10(3):115.
44. Varshney KK, Megha V, Bishamber N. Molecular modeling of isolated phytochemicals from Ocimum Sanctum towards exploring potential inhibitors of SARS coronavirus main protease and papain-like protease to treat COVID-19. 2020; SSRN 3554371.
45. Walls AC, Xiaoli X, Young-Jun P, Alejandra T, Joost S, Joel Q, et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. Cell. 2019;176(5):1026–39.
46. Wang M, Ruiyuan C, Leike Z, Xinglou Y, Jia L, Mingyue X, Zhengli S, Zhihong H, Wu Z, Gengfu X. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020;30(3):269–71.
47. World Health Organisation Coronavirus diseases (COVID 19) situation report-116. (2020) https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200515-covid-19-sitrep-116.pdf?sfvrsn=8dd60956_2. Accessed 26 Aug 2020
48. Xiao J. Phytochemicals in medicine and food. Berlin: Springer; 2015. p. 317–20.
49. Xu J, Shi PY, Li H, Zhou J. Broad spectrum antiviral agent niclosamide and its therapeutic potential. ACS Infect Dis. 2020;6(5):909–15.
50. Yin W, Mao C, Luan X, Shen DD, Shen Q, Su H, et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science. 2020;368(6498):1499–504.
51. Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367(19):1814–20.
52. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. 2020;368(6489):409–12.
53. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–33.