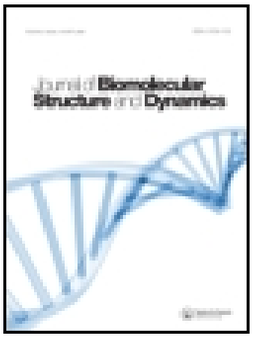


Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

The potential role of procyanidin as a therapeutic agent against SARS-CoV-2: a text mining, molecular docking and molecular dynamics simulation approach

Nikhil Maroli , Balu Bhasuran , Jeyakumar Natarajan & Ponmalai Kolandaivel

To cite this article: Nikhil Maroli , Balu Bhasuran , Jeyakumar Natarajan & Ponmalai Kolandaivel (2020): The potential role of procyanidin as a therapeutic agent against SARS-CoV-2: a text mining, molecular docking and molecular dynamics simulation approach, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2020.1823887

To link to this article: https://doi.org/10.1080/07391102.2020.1823887

View supplementary material

Published online: 22 Sep 2020.

Submit your article to this journal

Article views: 515

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tbsd20

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1823887

Taylor & Francis
Taylor & Francis Group

Check for updates

The potential role of procyanidin as a therapeutic agent against SARS-CoV-2: a text mining, molecular docking and molecular dynamics simulation approach

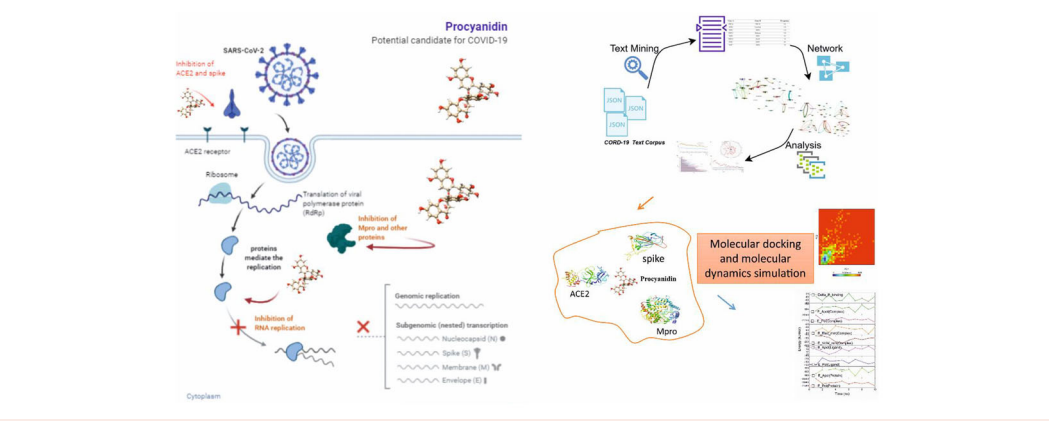
Nikhil Maroli[a,b], Balu Bhasuran[a], Jeyakumar Natarajan[c] and Ponmalai Kolandaivel[d]

[a]Computational Biology Division, DRDO Center for Life Sciences, Bharathiar University Campus, Coimbatore, Tamil Nadu, India; [b]Center for Condensed Matter Theory, Department of Physics, Indian Institute of Science, Bangalore, Karnataka, India; [c]Data Mining and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India; [d]Periyar University, Salem, Tamil Nadu, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

A novel coronavirus (SARS-CoV-2) has caused a major outbreak in human all over the world. There are several proteins interplay during the entry and replication of this virus in human. Here, we have used text mining and named entity recognition method to identify co-occurrence of the important COVID 19 genes/proteins in the interaction network based on the frequency of the interaction. Network analysis revealed a set of genes/proteins, highly dense genes/protein clusters and sub-networks of Angiotensin-converting enzyme 2 (ACE2), Helicase, spike (S) protein (trimeric), membrane (M) protein, envelop (E) protein, and the nucleocapsid (N) protein. The isolated proteins are screened against procyanidin-a flavonoid from plants using molecular docking. Further, molecular dynamics simulation of critical proteins such as ACE2, Mpro and spike proteins are performed to elucidate the inhibition mechanism. The strong network of hydrogen bonds and hydrophobic interactions along with van der Waals interactions inhibit receptors, which are essential to the entry and replication of the SARS-CoV-2. The binding energy which largely arises from van der Waals interactions is calculated (ACE2=-50.21 ± 6.3, Mpro=-89.50 ± 6.32 and spike=-23.06 ± 4.39) through molecular mechanics Poisson-Boltzmann surface area also confirm the affinity of procyanidin towards the critical receptors.



ARTICLE HISTORY
Received 19 July 2020
Accepted 10 September 2020

KEYWORDS
Text mining; molecular docking; molecular dynamics simulation; covid; procyanidin

Introduction

Recently a novel coronavirus emerged from the Wuhan province of China (Chen et al., 2020) that distress the world with over 9 million cases and 472,216 fatalities as of 23rd June 2020 according to COVID-19 Map - Johns Hopkins Coronavirus Resource Center. On March 11 of 2020, the World Health Organization (WHO) declared it as a pandemic, and currently no vaccines or drugs developed against the virus outbreak. This novel and highly pandemic virus named by WHO as "severe acute respiratory syndrome coronavirus 2" (SARS-CoV-2) and the disease caused by them is known as "coronavirus disease 2019" (COVID-19)(Lu et al., 2020; Paraskevis et al., 2020). SARS-CoV-2 is a virus that belongs to the coronaviridae family and it contains four structural proteins (Brian & Baric, 2005; C. Huang et al., 2020; Wu et al., 2020; Zhang et al., 1994). A spike protein (S), envelop protein (E), membrane (M) and nucleocapsids (N). The S protein binds to the receptors of the host organism and provides fusion and entry of the virus(Anand et al., 2002; Hilgenfeld,

CONTACT Nikhil Maroli scinikhil@gmail.com

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2020.1823887.

© 2020 Informa UK Limited, trading as Taylor & Francis Group

2014). The S protein initiates the receptor binding in the S1 subunit and S2 subunit that fuses and initiates the attachment to the host membrane. It is revealed that angiotensin-converting enzyme-2 (ACE2) is one of the major binding receptors of the SARS-CoV-2 virus in the lungs, where it was dipeptidyl peptidase-4 in case of MERS-CoV (Guo et al., 2008; F. Li, 2016; W. Li et al., 2003; McBride et al., 2014; Mubarak et al., 2019). The ACE2 receptor mainly expresses in the lung alveolar epithelial cells and its main function is to reduce the blood pressure through catalyzing the hydrolysis of angiotensin-II. Once the protein is bound with the receptor the nucleocapsid is released into the cytoplasm of the host cells, which contains ORF1a and ORF1b genes. ORF1a and ORF1b genes produce pp1a and pp1b that use host ribosomes for the translation process. The process of PPs leads to the production of 16 NSPs and all these NSPs having the indistinct function in the host cells such as suppression of host gene expression, formulation of multi-domain complexes and some of them act as primase(Stobart et al., 2013; Te Velthuis et al., 2012). M, E, and S proteins enter into the endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC) complex and produce viral envelope. At the same time, the ribonucleic-protein formation will occur once the replicated genomes bind to the N protein. Once the virus particles came out from ERGIC, the vesicle fuses with the plasma membrane and releases of the virus particles into the extracellular region occur. Transition between a metastable prefusion state to a stable post-fusion state is triggered by the binding of spike protein with ACE2. Several researchers studied and reviewed the binding mechanism of spike protein with ACE2 receptor and screened small molecule and other inhibitors towards different targets (Boopathi et al., 2020; Brielle et al., 2020; Freitas et al., 2020; Gupta et al., 2020; Jin et al., 2020; Muralidharan et al., 2020; Xue et al., 2020; Zhou et al., 2020). Also, the receptor-binding domain (RBD) dynamics and development of inhibitor targeting the RBD domain(Tai et al., 2020) and suggested the potential role of the RBD domain towards the therapeutic development against SARS CoV2.

Procyanidin is a member of proanthocyanidin that belongs to the class of flavonoids-the secondary metabolites of polyphenolic plant and fungus. Proanthocyanidins are the most abundant polyphenolic compounds after lignin (Souquet et al., 1996). This oligomeric compound formed from catechin and epicatechin molecules that mainly found in apples, maritime pine bark, aronia fruits, cinnamon, and grapes(Pacheco-Palencia et al., 2008; Vivas et al., 2006; Yang & Xiao, 2013). Though, the dietary content of procyanidin is not yet studied well. Some of the preliminary studies show the potential role of procyanidins a cardioprotective and anticancer agent(Bae et al., 2020) and the inhibition of ACE (Actis-Goretta et al., 2003). The large growing scientific literature collection of COVID 19 requires automated approaches for a better understanding of the disease. Text mining is one of the prominently adopted approaches in both academia and industry to extract information from the collection of literature(Bhasuran et al., 2016; Bhasuran & Natarajan, 2018). In this big data era, text mining is increasingly applied in various streams of science and technology for hypothesis generation and knowledge discovery. The text mining approach is implemented to identify the important COVID 19 genes/proteins and the biological connections between them. Particularly text mining methodology has been adapted to find the role of highly studied COVID-19 genes/proteins such as ACE2, NSP (1-10), Helicase, ORF, etc. To identify these genes/proteins and the biological interlink between them, a curated dictionary-based gene/protein entity mapping and co-occurrence-based text mining approach is performed on the COVID-19 scientific literature set.

In this study, we explored the potential role of procyanidin as a therapeutic agent against SARS-CoV-2 using molecular docking and molecular dynamics simulation methods. The selection of procyanidin is made through the screening of plant-based compounds collected from various literature along with the output from the literature mining. The proteins associated with SARS-CoV-2 were identified through named entity recognition method and 24 proteins are selected from the result which shows the higher frequency of the threshold value. The protein structures were collected from PDB databank and modeled using homology modeling. The molecular docking was performed to assess the binding affinity of procyanidin with the selected proteins. Further, crucial proteins such as ACE2, Mpro, and spike that directly involved in the entry of SARSCoV2 are selected for the MD simulation to reveal the inhibition mechanism of procyanidin. The MM-PBSA method was adopted to decipher the binding energy of the procyanidin.

Materials and methods

Text mining

The text mining pipeline employed in the current study follows as, COVID 19 literature data collection, named entity recognition of gene/proteins, gene-gene co-occurrence extraction, construction of a gene-gene interaction co-occurrence network and analysis.

COVID19 corpus

In the present study, we used the COVID-19 Open Research Dataset (CORD-19) (https://pages.semanticscholar.org/coronavirus-research) (Wang et al., 2020) downloaded on April 10th, 2020 as the literature data source. CORD-19 is a growing data source and is compiled with COVID-19 literature data and the past coronavirus research literature. The data set was created using a query-based search using "COVID-19" OR "Coronavirus" OR "Coronavirus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR "Severe Acute Respiratory Syndrome" OR "Middle East Respiratory Syndrome". Scientific literature matching this keyword query was collected from a total of 3200 journals. According to the developers the CORD-19 dataset consisting of research literature spanning various biological domains such as Virology (42.3%), Immunology (20.7%), Molecular biology (12.7%), Genetics

(8.0%), Intensive care medicine (6.7%) and Others (9.6%). The current study used a literature set consist of nearly 49,000 articles collected from CORD-19 literature corpus.

Identification and Co-occurrence extraction of COVID-19genes/proteins

Text mining is applied to the COVID-19 scientific literature set consist of nearly 49,000 articles collected from COVID-19 Open Research Dataset (CORD-19). Identification of which genes are playing an important role in disease from the scientific literature is a crucial task. In the current study, a dictionary-based named entity recognition (NER) approach is used to find the gene and protein entities mentioned in the COVID 19 scientific literature set. The dictionary was created using a manual screening of proteins mentioned as drug targets/biomarkers of COVID-19 in various scientific publications. The major proteins in the list are Angiotensin-converting enzyme 2 (ACE2), envelop (E) protein, E protein, Helicase, ExoN, M protein, main protease, membrane (M) protein, N protein, NendoU, Non-structural protein(1-16), ORF(1a, 1b,3a,6,7a,8,10), RNA-directed RNA polymerase(RdRp),S protein, spike (S) protein, nucleocapsid (N) protein, polyproteins, PP1a, PP1ab, Papain-like proteinase, and surface glycoprotein. A sentence level co-occurrence approach is based on the fact that two entities mentioned together in a sentence poses some form of relatedness. The current study employs a frequency-based co-occurrence approach and its aggregation to find the biological interlinks between important COVID 19 genes/proteins by text mining the CORD-19 corpus. To identify and extract the genes/proteins and interactions we have adapted our previous methodology(Bhasuran et al., 2018).

Network generation and analysis

Identified gene/protein entities and their relations were then subjected to create a gene-gene interaction network using the visualization tool Gephi 0.9.2(Bastian et al., 2009). Finally, various network analysis algorithms and parameters such as node degree, centrality measures, page rank, and sub-network analysis were performed on the COVID-19 gene-gene interaction network to identify the hub genes/proteins and highly dense genes/protein clusters. Centrality measures are used to identify the most important highly connected nodes within a graph. In the current study, three types of centrality measures such as closeness centrality, betweenness centrality, and eigenvector centrality were calculated from the generated gene-gene interaction network. In general, closeness centrality searches for the shortest path between the nodes, betweenness centrality calculates a node ability to connect the shortest path between any two nodes, and eigenvector measures the influence of a node based on its connectivity in the network. In the current study, two other important analyses performed on the network namely PageRank and Clustering. PageRank is one of the popular numerical weighting-based link analysis algorithms used for measuring the relative importance of a particular node in a given network based on its network connectivity. The schematic architecture of the text mining pipeline is depicted in Figure 1.

Molecular docking and MD simulation

The molecular docking of procyanidin with selected proteins was performed using Autodock VINA implemented in the SAMSON software package(NANO-D & I, 2016; Trott & Olson, 2010). The procyanidin molecule was obtained from the PubChem database and geometry optimization was performed using the Gaussian 09 software package(Frisch et al., 2009). The protein structure of ACE2, Mpro, and spike proteins are obtained from the respective RSCB PBD databank. The water molecules and other ligands in protein structures are removed before the docking. And the docking protocols are adapted from our previous studies(Maroli et al., 2019; 2019). Further, the receptor-procyanidin docked complexes were used to perform molecular dynamics simulation using the GROMACS 2020.1 software package(Abraham et al., 2015).The TIP3P water model with charmm36 forcefield (J. Huang & MacKerell, 2013; Mark & Nilsson, 2001)is used for the molecular dynamic's simulation with a 2-fs time step. The particle mesh Ewald method(Essmann et al., 1995) was used for the calculation of electrostatic and interactions. The Verlet cut-off distance of 1.4 mm was used for the short-range repulsive and attractive interactions. The temperature and pressure were maintained using the Nose-Hoover temperature coupling and Parrinello-Rahman algorithm at 310 K and 1 bar, respectively. The LINCS algorithm(Hess et al., 1997) was used to constrain all bond lengths and simulations were performed in the NPT ensemble for 100 ns. The equilibration of 10 ns in NVT and NPT ensembles were performed after the minimization of 5000-time steps. The trajectories are saved in every 100 ps and visualization was performed using UCSF Chimera and Ligplot software packages(Pettersen et al., 2004; Wallace et al., 1995).

Molecular mechanics Poisson – Boltzmann surface area

The binding energy of procyanidin with ACE2, Mpro, and spike proteins were calculated using the MM-PBSA method implemented in the g_mmpbsa tool(Kumari et al., 2014). The binding free energy of the protein-ligand complex in a solvent is defined as

$$\Delta E_{be} = E_{complex} - (E_{protein} + E_{procyanidin})$$

Where $E_{complex}$denote the total binding energy of protein-procyanidin complex, $E_{protein}$ and Eprocyanidin represent the total energy of the separated protein and procyanidin in the solvent, respectively.

$$<\Delta E_{be}> = <\Delta G_{mm}> + <\Delta E_{solv}> - <T\Delta S>$$
$$<\Delta E_{solv}> = <\Delta E_{pb}> + <\Delta E_{np}>$$
$$<\Delta E_{mm}> = <\Delta E_{vdw}> + <\Delta E_{ele}>$$

$\langle \Delta E_{MM} \rangle$ is the total molecular mechanic's energy in the gas phase, $\langle \Delta G_{solv} \rangle$ is salvation free energy, $T\Delta S$ is entropy, and $\langle \Delta E_{MM} \rangle$ is the sum of electrostatic and van der Waals

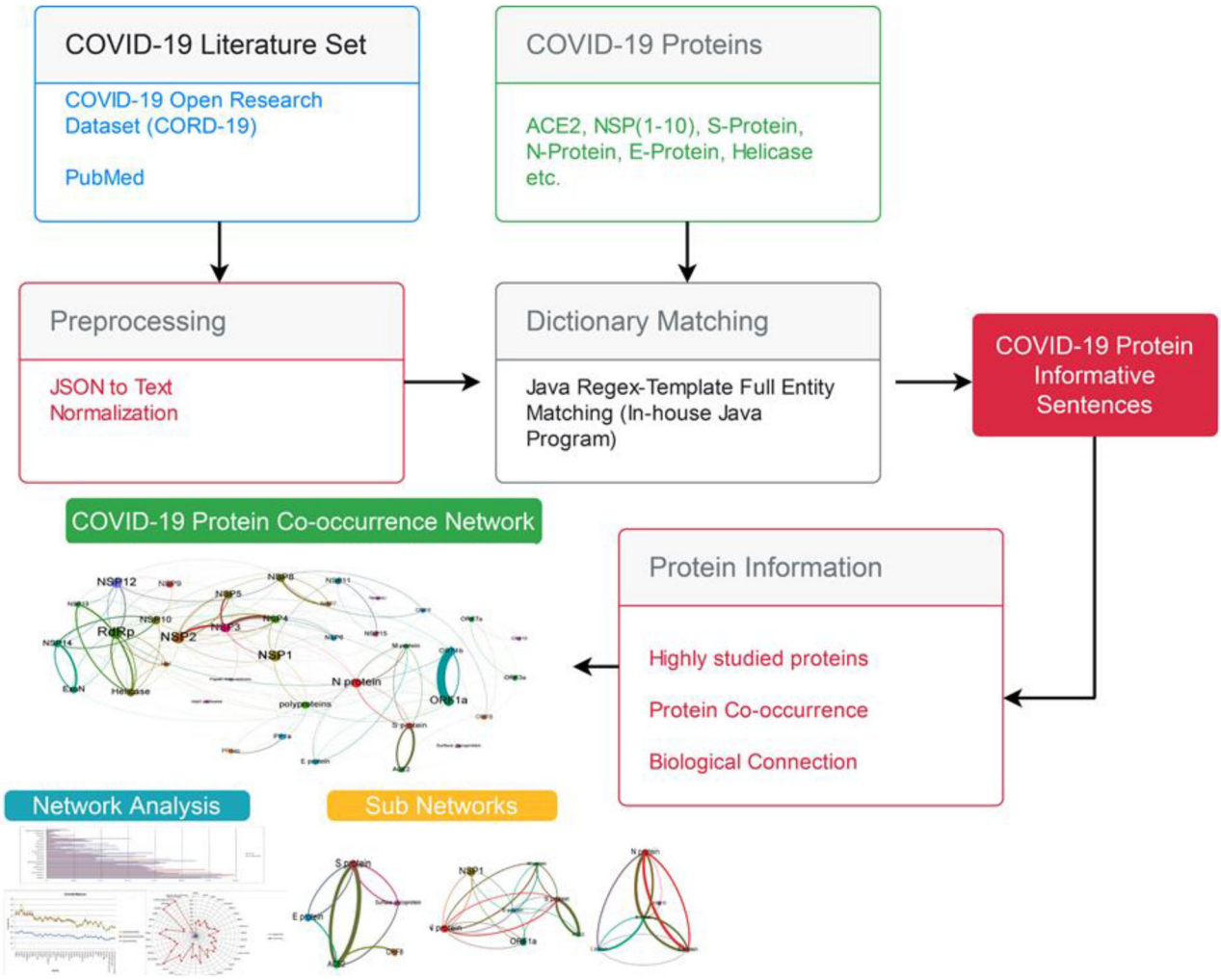


Figure 1. The schematic architecture of the text mining approach for the identification of critical proteins and their biological co-occurrence on COVID-19. The text mining approach employed the CORD-19 database as the literature source. A set of highly studied COVID-19 protein list was manually constructed and a dictionary matching procedure was applied to find protein co-occurrence with frequency. Finally, a protein co-occurrence network was created and analyzed to find the functional association among critical proteins in COVID-19.

interaction energy. Polar contributions were calculated using the PB model and nonpolar energy is estimated by solvent accessible surface area (SASA); TΔS is considered as negligible.

Results and discussion

Here we have adopted molecular docking followed by molecular dynamics simulation to reveal the binding mechanism of procyanidin with the proteins/enzymes associated with COVID 19. The proteins or enzymes associated with COVID 19 were identified through the named entity recognition method using an in-house developed Java script.

Extraction of co-occurring genes/proteins pairs

Dictionary-based named entity recognition (NER) approach was used to reveal the occurrence of protein/genes associated with COVID 19. We used in-house developed Java regular expression-based matching for the identification and extraction of gene and protein entities mentioned, and co-occurring genes/protein pairs in the COVID-19scientific literature set. The NER phase identified a total of 45,984 mentions with 38 major proteins with high reference. A detailed representation of the highly referenced genes is given in Figure 2a and liner frequency of the important proteins related to COVID 19 and the top 10 proteins related are depicted in Figure 2c& d, respectively. In the co-occurring extraction phase, we identified the genes that are co-occurring in high degrees up-to eleven from a single sentence (Figure 2b). Also, highly interacting benchmarking gene pairs were identified and the top 10 results are given in Table 1.

Text mining on procyanidin

There is a total of 2,327 entries are connected with procyanidin in PubMed and these abstracts are collected in text format. In this collection, we searched for various biological entities and activities of the molecule. The results revealed that procyanidin regulates oxidative stress and connected with the abnormal inflammatory response. The major diseases connected with procyanidin are inflammation, cancer, diabetes, obesity, atherosclerosis, cardiovascular disease, and hypertension. The major genes/proteins extracted from the

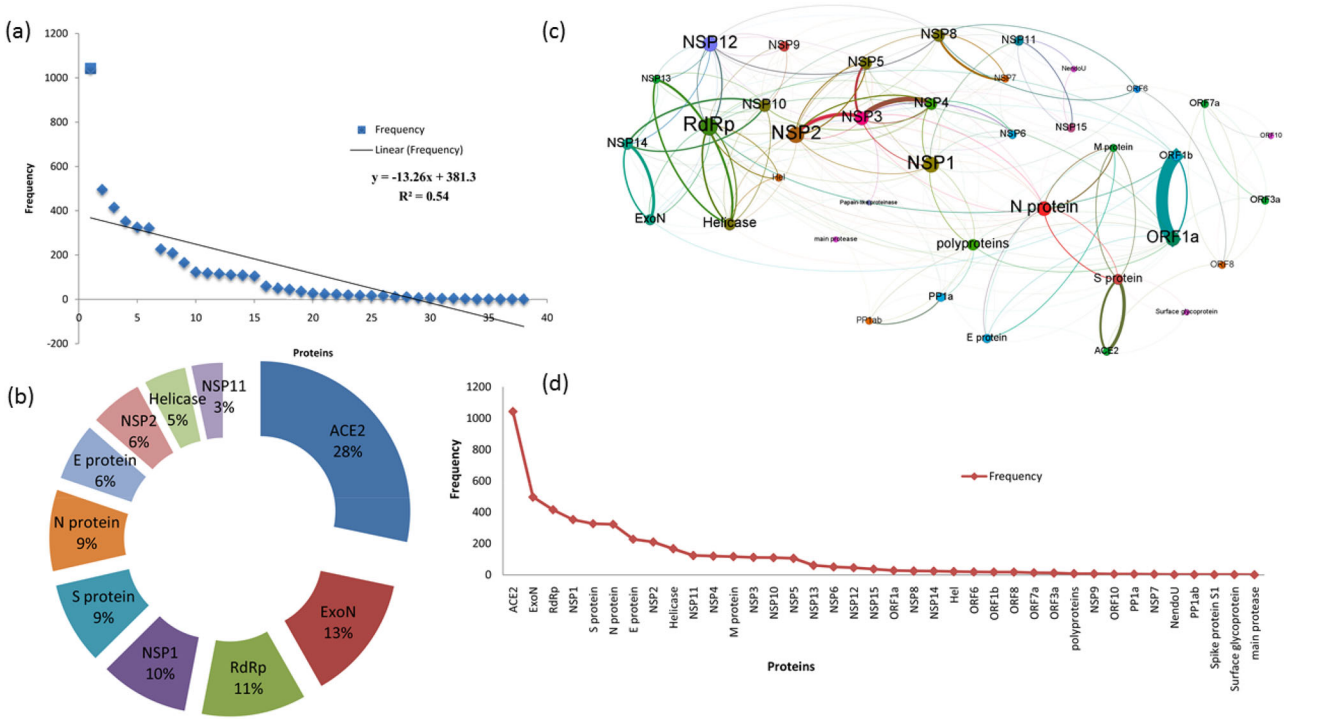


Figure 2. (a) Represents the linear frequency of the important proteins related to COVID-19 (b) Represents the top 10 critical proteins connected with COVID-19 (c) Protein co-occurrence network of COVID-19 (d)Highly reported proteins related to COVID-19.

Table 1. Highly interacting benchmarking gene-gene interaction pairs.

Gene A	Gene B	Frequency
ORF1a	ORF1b	303
ACE2	S protein	144
NSP3	NSP4	124
NSP13	Helicase	104
NSP2	NSP3	93
NSP14	ExoN	76
NSP3	NSP5	60
NSP7	NSP8	56
NSP14	NSP10	46
Helicase	RdRp	45

procyanidin literature set are, TNF, GSPB2, Nrf2, NF-kappaB, IL-6, CASP3, INS, ACE, COX-2, and p53.

Gene-Gene interaction network analysis

The co-occurrence-based score was used to generate a gene-gene interaction network. The network (Figure 2c) consists of 38 nodes (genes) and 209 edges(interactions). Analysis of the network has revealed the 13 nodes as the major nodes with higher interlink to other nodes. The network degree of each node is given in Figure 1 of supporting information. From the network, it is evident that the connection exists between ACE2 and S-protein and likewise proteins as shown in Table 1. Various network analysis properties of the COVID 19gene-gene interaction network also given in Figure 3. It is further revealed that a strong association between ORF1a and ORF1b genes/proteins along with the proteins that are associated with SARS-CoV-2. We also found that RNA-directed RNA polymerase (RdRp) is strongly connected with all the NSPs. The top ten proteins from various network analyses are found to be RdRp, NSP2, NSP1, NSP12, N protein, ORF1a, NSP3, NSP5, Helicase, and NSP8.

Subnetwork extraction and analysis

To evaluate these highly dense nodes, sub-networks were extracted from the main network. Specifically, sub-networks for genes such as ACE2, RDRp, S-protein, M-protein, N-protein, and E-protein were extracted. The extracted sub-networks are given as ACE2, Helicase S-protein, E-protein, M-protein and N-protein ((Figure 2 (a-f) of supporting information) and the corresponding connected nodes in these sub-networks are given in Table 2 with network analysis measurements.

Molecular docking

The procyanidin structure was obtained from PubChem and optimized using B3LYP functional and 6-31 G(d,p) basis set of DFT formalism(Becke, 1988; Kumari et al., 2014; Lee et al., 1988; Xu & Goddard, 2004). The optimized geometry of procyanidin was used to perform molecular docking on selected proteins as receptors. The proteins PDB structure is obtained from PDB databank and the remaining proteins were modeled by homology modeling implemented in I-TASSER webserver. The modeled structure was solvated using the TIP3P water model and subjected to a 10 ns MD simulation to remove steric hindrances and unfavorable interactions. The 3-D structure for the modeled proteins are given in Figure 3(i)-(iv) of supporting information. Molecular docking revealed a stronger affinity of procyanidin towards these proteins (Table 3). It is evident from previous experimental

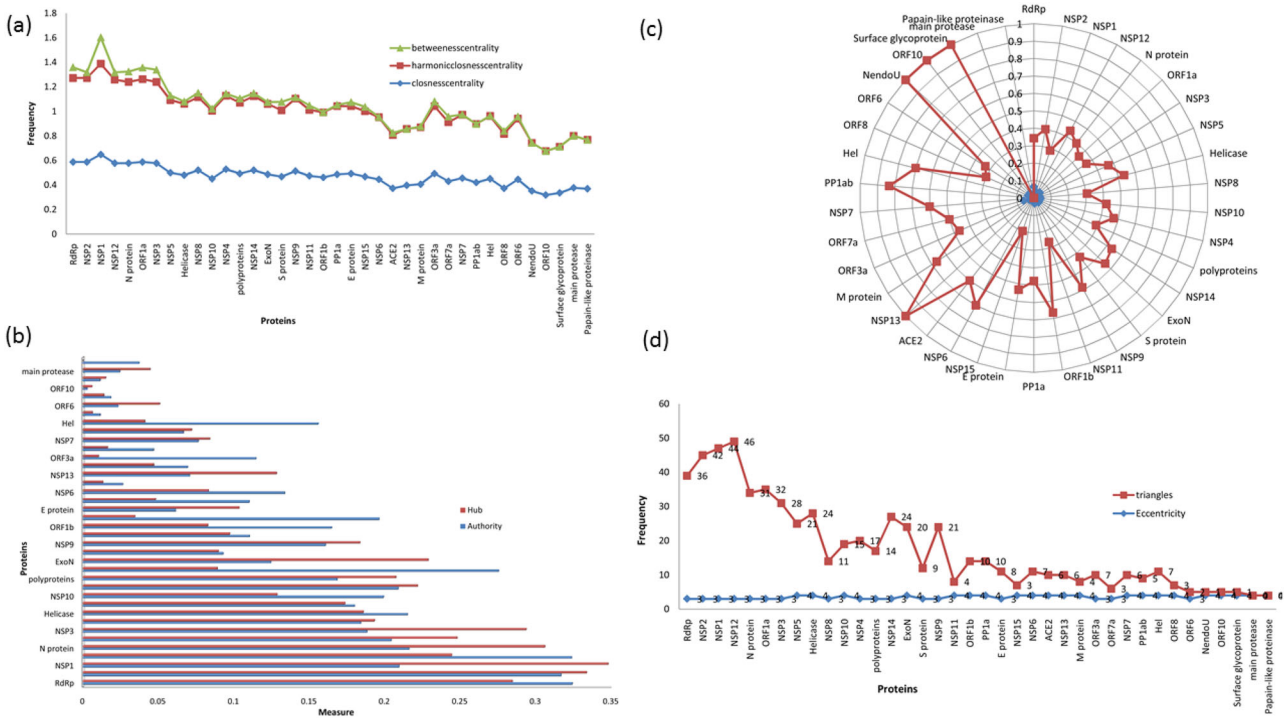


Figure 3. Network analysis properties of the protein co-occurrence network of COVID-19 **(a)** Centrality measures (closeness (blue colour), betweenness (green colour), harmonic closeness (red colour) of each protein from the co-occurrence network **(b)** Network Hub (red colour) and Authority (blue colour) values of each protein from the co-occurrence network **(c)** Page Rank (blue colour) and Clustering (red colour) statistics of each protein from the co-occurrence network **(d)** Triangles (red colour) and Eccentricity(blue colour) measures of each protein from the co-occurrence network.

studies that the inhibition of NSP's help to reduce the host translation and further reduction in virus replication. The strong interaction energy of the procyanidin molecule help to inhibit these proteins by changing their structural and functional properties. In all the docked complexes strong hydrogen and hydrophobic interactions networks are identified and provided in Table 3. The interaction network of the proteins is provided in Figures 3 and 4 of supporting information. The hydrogen bonds between residues and procyanidin disrupt the normal structure of the proteins that fail the protein or enzyme to perform its native functions. The binding site of procyanidin with all proteins rich in hydrogen bonds and hydrophobic interaction apart from conventional electrostatic and van der walls interactions. Among the 24 screened proteins we have considered ACE2, Mpro, and spike protein separately as it is the crucial proteins that directly linked to the entry of SARS-CoV-2. Molecular docking has revealed that procyanidin forms hydrogen bonds with residues Ser44, Ser47, Asp350, Asp382, Tyr385, Arg393, Asn394, His401 and hydrophobic interaction with Phe40, Phe390 in ACE2 (Figure 4). The higher number of these interactions depicts the strong affinity of procyanidin with ACE2. The residues Asn394, Gly395, Ser43, Leu351, Met62, Ser47, Asn51, His378, Ala348, Trp69, and Leu391 shows van der Waals interaction with the oxygen atom and the benzene ring of the procyanidin at the binding site. Further, our docked pockets are in agreement with reported binding sites of ACE2 reported through experimental studies (Akif et al., 2010, 2011; Natesh et al., 2003; Towler et al., 2004). Also, the binding site we reported is overlapping with the several sites reported from the PDB database. The active site of ACE2 with carboxypeptidase shows the residues Tyr510, Cys344, His345, Pro346, Met360, Lys363, Arg273, Asp368, Thr371, Tyr515 and Arg514 which is in the same site where procyanidin binds. Moreover, here we only emphasized the active sites which shows the highest binding energy for procyanidin. Mpro with procyanidin shows eight hydrogen bonds with residues Ser44, Ser47, Asp350, Asp382, Tyr385, Arg393, Asn394, His401, and hydrophobic interaction with Phe40, Phe390 at the binding site (Figure 5). Also, Met49 forms a pi-sulfur bond with the oxygen atom of procyanidin and pi-alky interaction between the benzene ring and Cys145. Further, the procyanidin interaction with spike protein shows hydrogen bond with Ser375, Thr376, Gly404, Asp405, Arg408, Ile410, and hydrophobic interaction with Thr376, Val407, Arg408 residues (Figure 6). The Lys378 and Asp405 residue at the binding site shows pi-cation and pi-anion interaction with the ring structure of procyanidin. The docked position and interaction energy of this protein indicate the possibilities of procyanidin as a potent drug candidate against the novel SARS-CoV-2. To understand the structural and dynamical features of the protein with the procyanidin a 100 ns molecular dynamics simulation in water was performed. The best-docked poses based on the docking score were used as initial conformation for the simulation.

Molecular dynamics simulation

100 ns molecular dynamics simulation of apo and procyanidin bound form of ACE2, Mpro, and spike proteins have been performed. The structural stability of the proteins was assessed using RMSD, RMSF, and principal component

Table 2. Connected proteins in each subnetwork with network analysis measurements.

Hub Protein	Connected Components	Closness centrality	Betweeness centrality	Eigen centrality	Page ranks	Hub
ACE2	S protein	0.468354	0.069328	0.309572	0.022098	0.089565
	E protein	0.493333	0.033867	0.314617	0.012692	0.103421
	ORF8	0.373737	0.023807	0.082985	0.036943	0.00629
	Surface glycoprotein	0.336364	0	0.04398	0.007195	0.015058
Helicase	RdRp	0.587302	0.087633	0.890853	0.065624	0.284112
	NSP2	0.587302	0.046212	0.968646	0.036751	0.333196
	N protein	0.578125	0.083752	0.790443	0.02517	0.305663
	NSP10	0.45122	0.017466	0.523294	0.039623	0.12841
	ExoN	0.486842	0.014993	0.606035	0.036957	0.228533
	NSP12	0.578125	0.059686	0.946454	0.036738	0.243958
	NSP13	0.397849	0	0.272434	0.01722	0.128044
	NSP9	0.513889	0.010905	0.629844	0.025474	0.183429
	Hel	0.45122	0.002244	0.355033	0.021652	0.040965
S Protein	E protein	0.493333	0.033867	0.314617	0.012692	0.103421
	M protein	0.406593	0.005986	0.16895	0.014046	0.046801
	N protein	0.578125	0.083752	0.790443	0.02517	0.305663
	ORF1a	0.587302	0.094328	0.78173	0.047027	0.247667
	ACE2	0.373737	0.017162	0.093084	0.012052	0.013174
	RdRp	0.587302	0.087633	0.890853	0.065624	0.284112
N Protein	E protein	0.493333	0.033867	0.314617	0.012692	0.103421
	M protein	0.406593	0.005986	0.16895	0.014046	0.046801
	NSP1	0.649123	0.215808	1	0.026757	0.347485
	NSP12	0.578125	0.059686	0.946454	0.036738	0.243958
	NSP15	0.468354	0.035962	0.262739	0.015383	0.047781
	NSP2	0.587302	0.046212	0.968646	0.036751	0.333196
	NSP3	0.578125	0.099375	0.767396	0.0163	0.293389
	NSP4	0.528571	0.017165	0.619071	0.017897	0.221376
	Helicase	0.480519	0.018757	0.649594	0.036368	0.185347
	ORF1a	0.587302	0.094328	0.78173	0.047027	0.247667
	RdRp	0.587302	0.087633	0.890853	0.065624	0.284112
	S protein	0.468354	0.069328	0.309572	0.022098	0.089565
M Protein	E protein	0.493333	0.033867	0.314617	0.012692	0.103421
	S protein	0.468354	0.069328	0.309572	0.022098	0.089565
	NSP15	0.468354	0.035962	0.262739	0.015383	0.047781
	N protein	0.578125	0.083752	0.790443	0.02517	0.305663
E Protein	S protein	0.468354	0.069328	0.309572	0.022098	0.089565
	N protein	0.578125	0.083752	0.790443	0.02517	0.305663
	ACE2	0.373737	0.017162	0.093084	0.012052	0.013174
	M protein	0.406593	0.005986	0.16895	0.014046	0.046801
	NSP1	0.649123	0.215808	1	0.026757	0.347485
	ORF1a	0.587302	0.094328	0.78173	0.047027	0.247667

analysis. The secondary and tertiary structural changes of the proteins due to the interaction of procyanidin also analyzed from the trajectories.

Structural stability of receptors

The structural stability of the proteins was assessed using root mean square deviation (RMSD) and fluctuations (RMSF) of backbone and Cα atoms. The fluctuations of backbone and Cα atoms in both complexed and apo-protein structures are evaluated from 100 ns simulation trajectories and depicted in Figure 4 of supporting information. The procyanidin bound ACE2 shows an average backbone fluctuation of 0.4 nm whereas the apo structure shows 0.2 nm. The Mpro and spike proteins show an average of 0.3 and 0.19 nm in backbone RMSD whereas procyanidin bound form shows an average of 0.4 nm with the highest fluctuation at 0.9 nm for Mpro and 0.6 nm for spike protein. The higher fluctuation of the protein while interacting with procyanidin indicates its lower stability that will lead to the non-functional state of the protein. It is essential to maintain the native structure of the ACE2 receptor for the strong attachment of spike protein and further progression of the virus replication. To understand the essential dynamics of these proteins we have performed the principal component analysis of Cα atoms from the 100 ns simulation trajectories. The PCA based Cα atoms of the proteins represent the eigenvectors of the covariance matrix that is argued by its coincident eigenvalues and total concerted motion of the protein. As a result, the PCA based RMSF calculation help to reveal the dynamics of the critical residues that take part in the inhibition of the protein. The PC1's and PC2's of three proteins show lower fluctuations at the native form than the procyanidin-bound state (Figure 7). The higher RMSF fluctuation of residues at the binding site indicates the conformational changes of the residues that destabilize the protein from its native structure. The binding site is identified as the target location for the spike protein and other virus proteins.

The hydrogen bonds and interaction energy

The number of hydrogen bond forming between procyanidin and surrounding residues of each protein has been calculated and provided in Figure 8b. An average of 5-12 hydrogen bonds was present in all the system throughout the simulation time, which is an indication of strong binding of procyanidin with protein. The hydrogen bonds are calculated based on the cut-off criteria: (a) donor – acceptor distance $\leq$

Table 3. Molecular docking results of procyanidin with different proteins that associated with novel COVID 19.

Protein	Binding energy (kcal/mol)	hydrogen bonds	Hydrophobic interaction	Function of the protein
Host translation inhibitor nsp1	−8.5	9Asn, 12Thr,13His, 81His,83His,134His	14Val, 134His	Inhibits host translation by interacting with the 40S ribosomal subunit
Non-structural protein 2 (nsp2)	−8.8	125Pro,127Ala,129Pro, 263Ser,263Ser	129Pro	play a role in the modulation of host cell survival signaling pathway
Papain-like proteinase	−9.7	178Asp,181Thr,211Ser, 215Lys,218Asp,807Asp, 1301Leu,1313Asn	1301Leu,1316Pro	Responsible for the cleavages located at the N-terminus of the replicase polyprotein
Non-structural protein 4 (nsp4)	−9.7	177Glu,225Thr,295Thr, 429Lys	177Glu,299Tyr, 428Leu	Participates in the assembly of virally-induced cytoplasmic double-membrane vesicles necessary for viral replication.
Proteinase 3CL-PRO	−7.5	131Arg,197Asp,199Thr, 238Asn,272Leu	239Tyr, 272Leu	Cleaves the C-terminus of replicase polyprotein at 11 sites
Non-structural protein 6 (nsp6)	−8.9	174Asn, 179Val,255Asn	171Val, 175Tyr, 1 84Phe	Plays a role in the initial induction of autophagosomes from host reticulum endoplasmic.
Non-structural protein 7 (nsp7)	−8.1	5Asp,34Gln,37Asn	7Lys,37Asn,41Leu	Forms a hexadecamer with nsp8
Non-structural protein 8 (nsp8)	−8.7	101Asp,105Asn,108Asn, 109Asn,140Asn,148Thr	104Asn,148Thr	Forms a hexadecamer with nsp7
Non-structural protein 9 (nsp9)	−8.4	11Gln,26Asp,29Leu, 47Asp,86Lys	29Leu	participate in viral replication by acting as a ssRNA-binding protein.
Non-structural protein 10 (nsp10)	−8.3	28Lys,85Asn,87Asp,91Asp, 92Leu,116Val	112Leu	Plays a pivotal role in viral transcription by stimulating both nsp14 3′-5′ exoribonuclease and nsp16 2′-O-methyltransferase activities
RNA-directed RNA polymerase (RdRp)	−8.9	456Tyr,556Thr,622Cys, 624Arg,682Ser	555Arg	Responsible for replication and transcription of the viral RNA genome.
Helicase (Hel)	−9.5	139Lys,142Glu,178Arg, 179Asn,339Arg,361Asn, 383Asp,410Thr	178Arg,179Asn, 410Thr	Multi-functional protein with a zinc-binding domain in N-terminus displaying RNA and DNA duplex-unwinding activities with 5′ to 3′ polarity.
Guanine-N7 methyltransferase (ExoN)	−9.6	290Val,292Trp,354Gln, 286Asn,422Asn,426Phe	290Val,292Trp, 335Pro,336Lys, 424His	an exoribonuclease activity acting on both ssRNA and dsRNA in a 3′ to 5′ direction and a N7-guanine methyltransferase activity.
Uridylate-specific endoribonuclease (NendoU)	−8.2	62Asn,64Lys,82Asn, 83Thr,124Asp	15Phe,60Lys,63Ile, 64Lys,83Thr	Uridylate-specific endoribonuclease
2′-O-methyltransferase (2′-O-MT)	−10.1	30Tyr,46Lys,74Ser, 130Asp,170Lys,198Asn,203Glu	76Lys	Uridylate-specific endoribonuclease
Surface glycoprotein/Spike protein (S)	−10.3	801Asn,925Asn,929Ser, 935Gln,936Asp,1138Tyr, 1180Gln,1184Asp	928Asn,1140Pro	Binding to human ACE2 and other possible receptors
ORF3a	−7.9	10Ile,12Thr,131Trp, 134Arg,135Ser,227His	9Thr	Forms homotetrameric potassium sensitive ion channels (viroporin) and may modulate virus release
E Protein	−9.0	2Tyr,32Ala	25Val,28Leu, 74Leu	Plays a central role in virus morphogenesis and assembly
M Protein	−7.8	4Ser,42Arg,107Arg, 125His,126Gly	8Ile,34Leu,97Ile, 107Arg,110Trp,128Ile	Component of the viral envelope that plays a central role in virus morphogenesis and assembly via its interactions with other viral proteins.
ORF6	−6.7	13Glu,38Lys	12Ala,38Lys	stimulate cellular DNA synthesis
ORF7a	−8.0	60Ser,61Thr,62Gln	109Phe,110Ile	Suppression of host tetherin activity.
ORF8	−8.9	48Arg,53Lys	6Phe,7Leu,10Ile,52Arg	
N Protein	−9.7	3Asp,17Phe,19Gly, 21Ser,23Ser,25Gly, 281Gln,393Thr	394Leu,395Leu,403Phe	Packages the positive strand viral genome RNA into a helical ribonucleocapsid (RNP) and plays a fundamental role during virion assembly through its interactions with the viral genome and membrane protein M.
ORF10	−7.2	15Ser,20Arg	11Phe,13Ile,16Leu	Exact function is not known but identified in the pathway
ACE2	−8.9	44Ser,47Ser,350Asp, 382Asp,385Tyr, 393Arg,394Asn,401His	40Phe,390Phe	Receptor that binds the virus protein
Mpro	−9.2	26Thr,46Ser, 143Gly,144Ser,145Cyc	142Asn	
Spike	−9.5			375Ser,376Thr,404Gly,405Asp,408Arg,410Ile

0.35 nm and (b) donor − H− acceptor bond angle $\leq 120^{\circ}$. At the beginning of the simulation, ACE2 shows seven hydrogen bonds with a total energy of 68.82 kJ/mol and end of the simulation five hydrogen bonds with 64 kJ/mol, however, we have observed an average of seven hydrogen bonds during the simulation time. At 10 ns, Mpro shows three hydrogen bonds with the energy of 35.45 kJ/mol and the last frame of the simulation shows five hydrogens both with a total of 53.64 kJ/mol. The first frame of spike protein having ten hydrogen atoms with the total energy of 100.83 kJ/mol

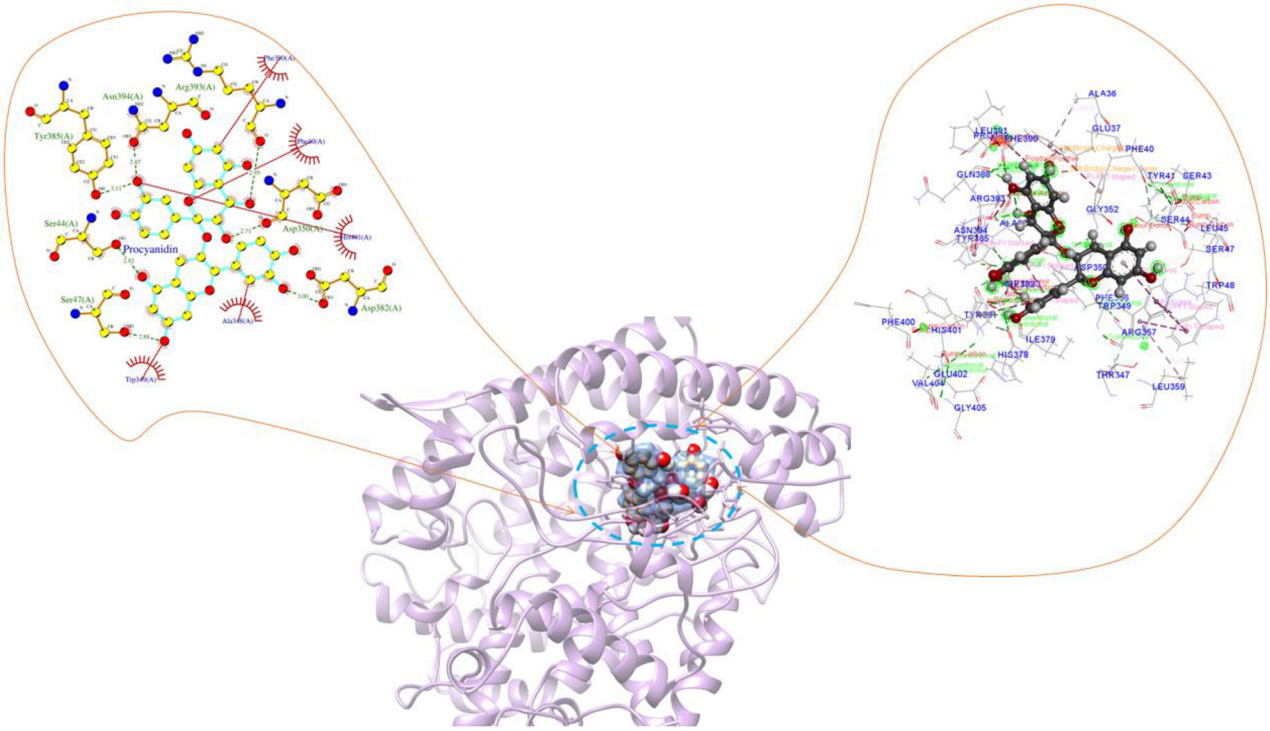


Figure 4. ACE2- procyanidin docked complex and representation of different interactions such as hydrogen bonds, hydrophobic interaction, van der Waals and electrostatic. The hydrogen bonds are marked in green line and red colour on the left panel represents the hydrophobic interactions.

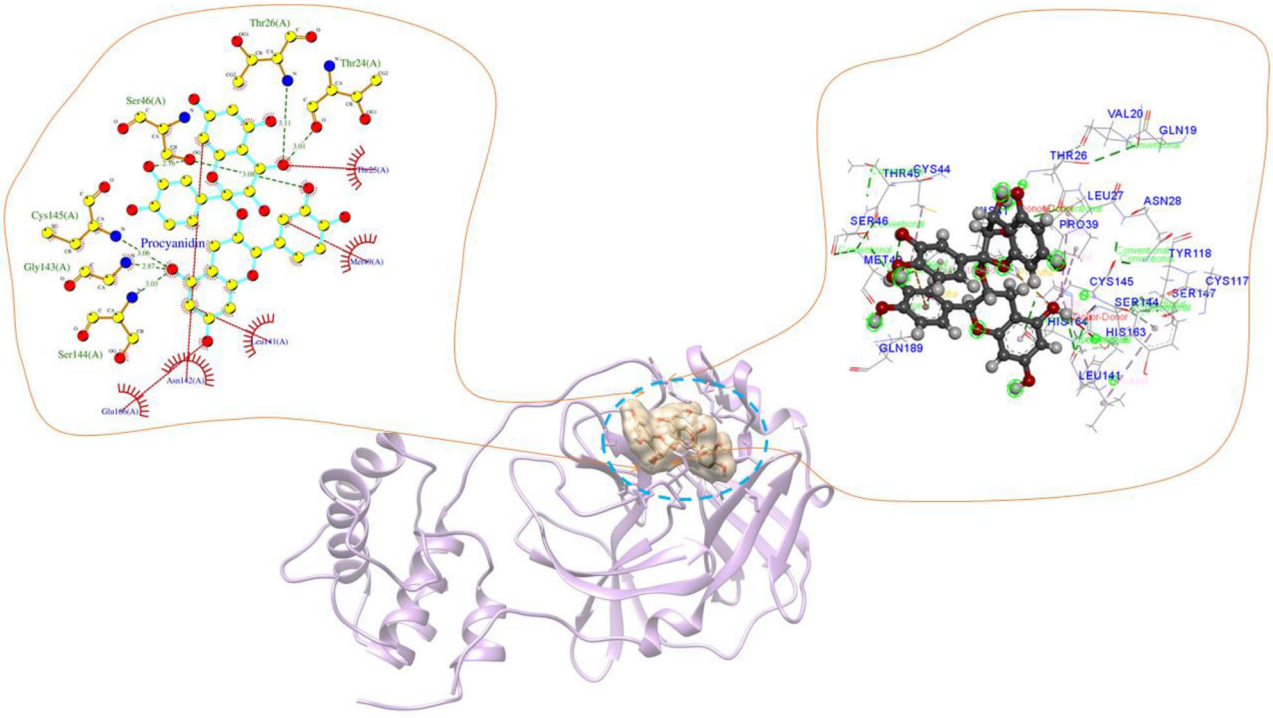


Figure 5. Mpro- procyanidin docked complex and representation of different interactions such as hydrogen bonds, hydrophobic interaction, van der Waals and electrostatic. The hydrogen bonds are marked in green line and red colour on the left panel represents the hydrophobic interactions.

and last frame shows 54.98 kJ/mol with four hydrogen bonds. Here, the starting frame and last frame of the simulation shows the presence of hydrogen bond which stabilize the procyanidin at the binding site. Further, the interaction energy such as electrostatic and van der Waals at the binding site have been calculated from the simulation trajectories (Figure 8c &d). The van der Waals interaction energy is dominant at the binding site of all proteins than electrostatic interaction energy. ACE2 shows an average of −145.06 kJ/mol and −78.51 kJ/mol of van der Waals and electrostatic interaction energy respectively. The binding site of ACE2 consists of residues such as Ser44, Ser47, Asp350, Asp382,

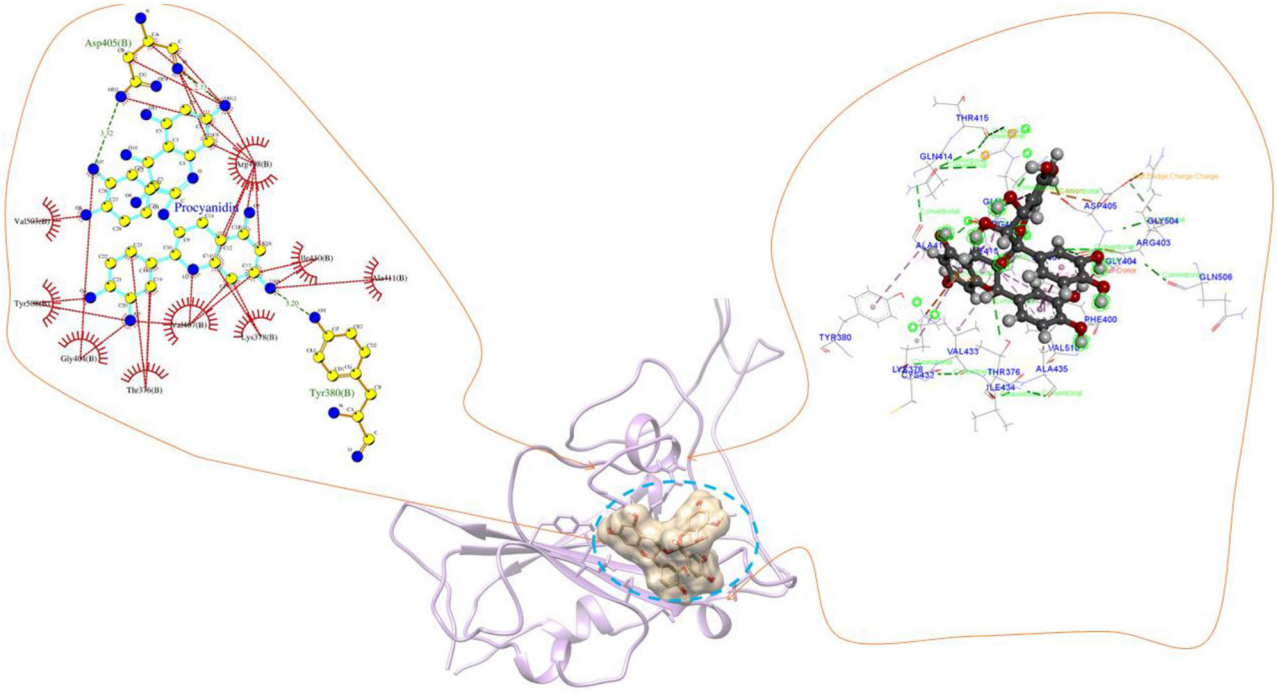


Figure 6. Spike protein- procyanidin docked complex and representation of different interactions such as hydrogen bonds, hydrophobic interaction, van der Waals and electrostatic. The hydrogen bonds are marked in green line and red colour on the left panel represents the hydrophobic interactions.

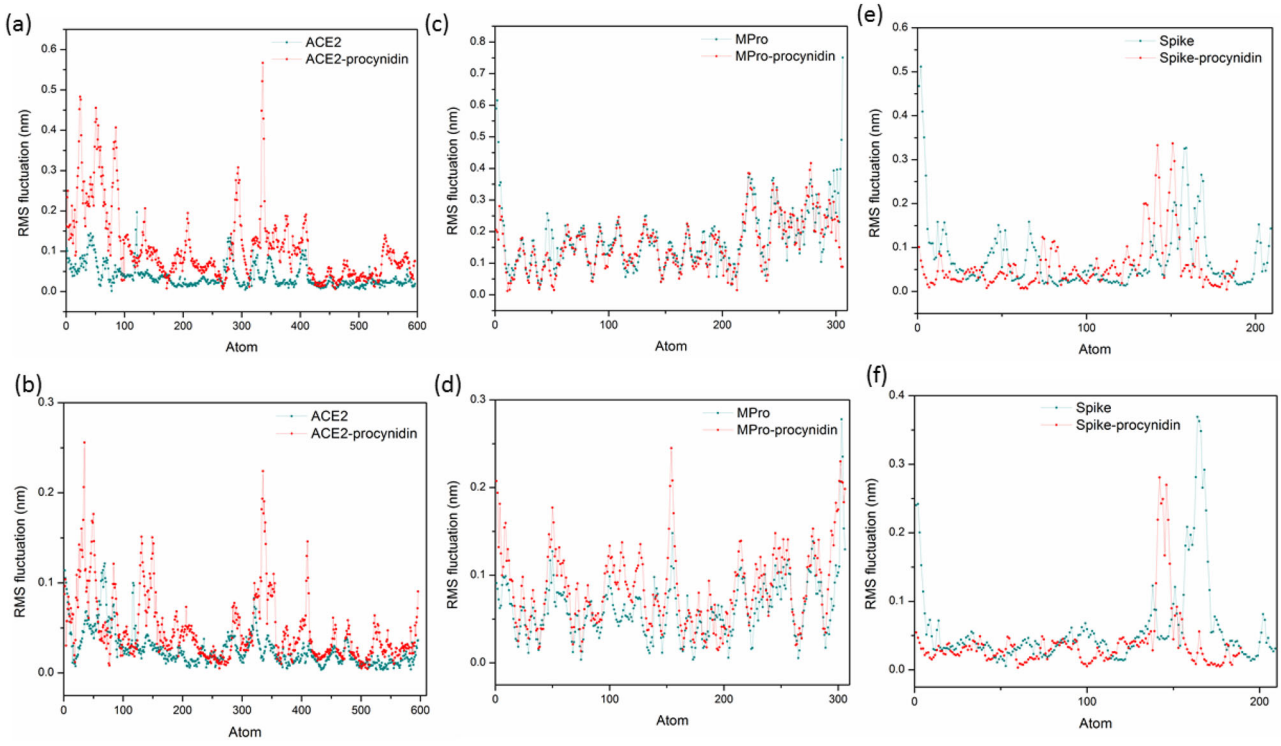


Figure 7. The Cα based RMSF fluctuations calculated through principal component analysis. The top panel represents the RMSF fluctuations along the first eigenvector and the bottom panel represents RMSF fluctuations along the second eigenvector. The native proteins fluctuations are shown in cyan and red indicates the RMSF fluctuations of procyanidin bound proteins. (a) native ACE2 (b) ACE2-procyanidin (c) Mpro(d) Mpro- procyanidin (e) spike protein (f) spike protein- procyanidin. Higher RMSF fluctuations on procyanidin bound proteins can be seen in all the proteins.

Tyr285, Arg293, Asn394, His401, Phe40, Phe390, Ala36, Glu37, Phe40, Tyr41, Ser43, Ser44, Trp48, Gly352, and Phe400 that are more favorable towards the van der Waal interaction. The Mpro and spike protein show an average of −163.05 kJ/mol,-106.7 kJ/mol of van der Waals −108.9 kJ/mol and −84.50 kJ/mol of electrostatic interaction energy respectively. The binding site of Mpro consists of Leu67, Val68, Gln69, Leu57, Tyr54, Pro39, Val20, Gln19, Ser144, Ser147, Leu141,

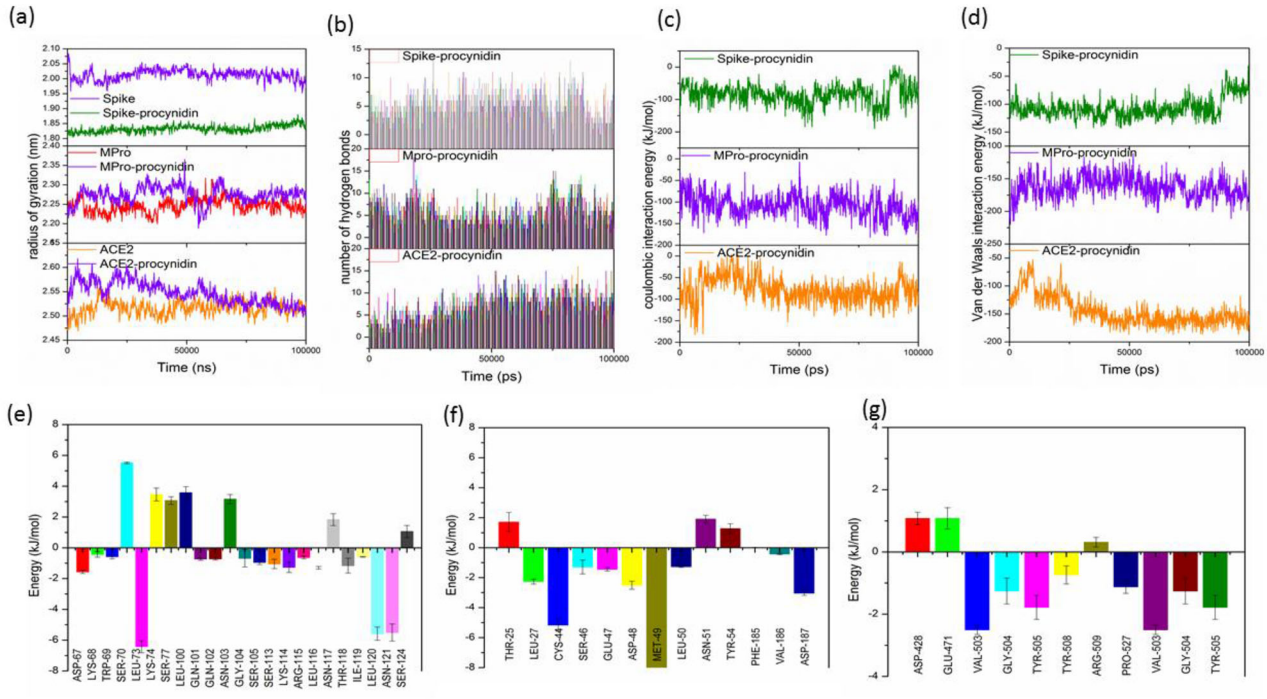


Figure 8. (a) represents the radius of gyration of the protein and procyanidin-protein complex. Native proteins show stable and lower values than the procyanidin bound proteins. (b) The number of hydrogen bonds formed between the procyanidin and proteins. (c) The electrostatic interaction energy between procyanidin and proteins (d) van der Waals interaction energy between procyanidin and proteins. The van der Waals interaction energies are found to be higher in all the proteins. (e)-(g) represents the energy contribution of residues at the binding site. (e) ACE2 (f) Mpro and (g) spike protein.

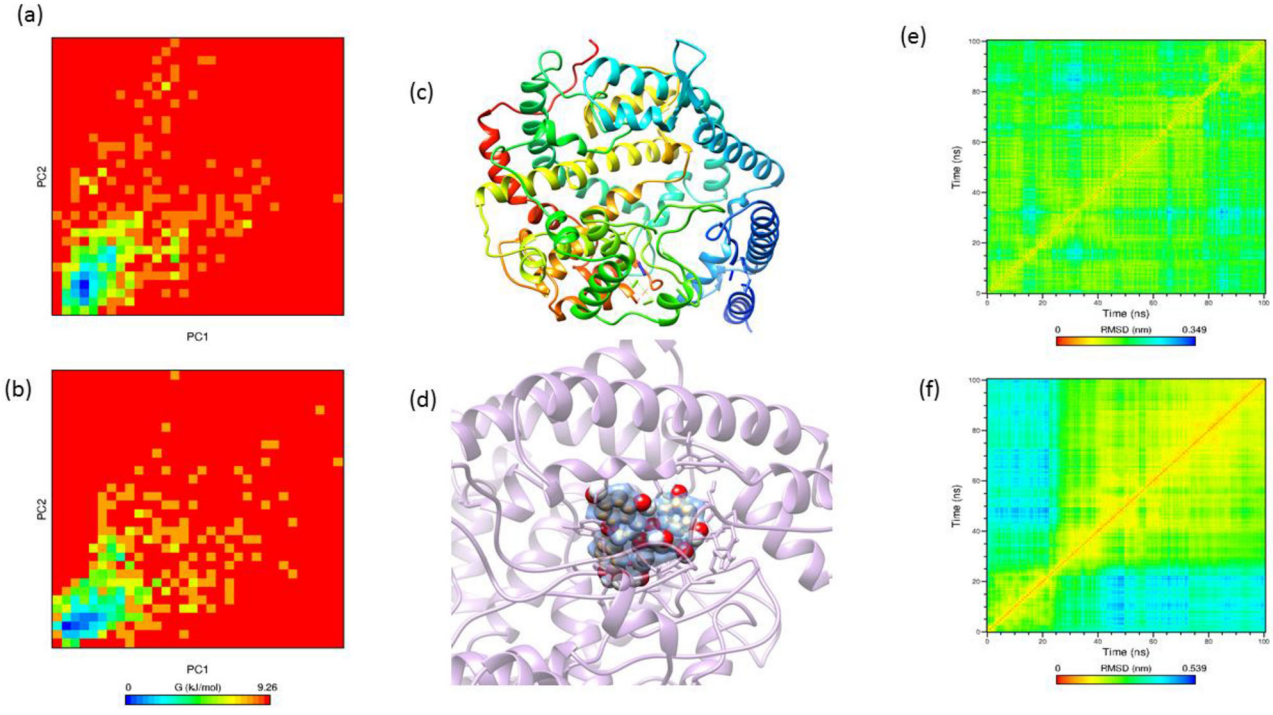


Figure 9. The two-dimensional representation of the free energy landscape calculated from principal components. (a) native ACE2 (b) procyanidin-ACE2 complex. (c) and (d) represents the apo and procyanidin docked complex of ACE2. (e) and (f) represents the RMSD matrix of ACE2 backbone atoms in native and procyanidin-ACE2 docked complex respectively. The RMSD matrix derived from the backbone atoms depicts the tertiary structure of the proteins. The procyanidin shows significant changes in the RMSD matrix, the red line indicates the lowest distance and blue colour indicates the highest distance. The presence of procyanidin in the proteins has changed the tertiary structure of the protein by inducing the structural fluctuations in the backbone atoms.

Cys117, His163, and His41. The presence of polar and non-polar residues at the binding site provides strong affinity energy towards the procyanidin. The spike protein binding sites consist of residues such as Ile410, Ala31, Arg408, Val407, Gly404, Thr376, Lys378, Tyr380, Gln414, Thr376. Altogether, strong interaction energy and positioning of the procyanidin at the deep inside the proteins make them a strong inhibitor.

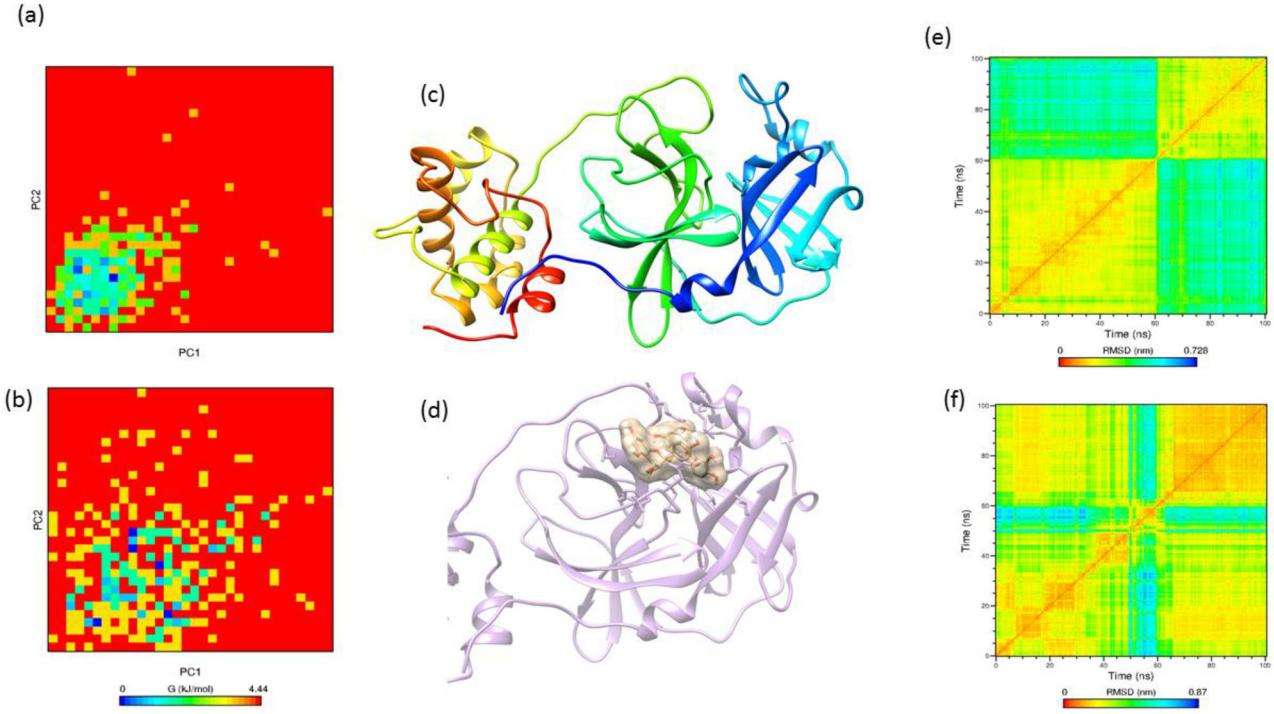


Figure 10. The two-dimensional representation of the free energy landscape calculated from principal components. (a) native Mpro (b) procyanidin-Mprocomplex. (c) and (d) represents the apo and procyanidin docked complex of Mpro. (e) and (f) represents the RMSD matrix of Mpro backbone atoms in native and procyanidin-Mpro docked complex respectively. The RMSD matrix derived from the backbone atoms depicts the tertiary structure of the proteins. The procyanidin shows significant changes in the RMSD matrix, the red line indicates the lowest distance and blue colour indicates the highest distance. The presence of procyanidin in the proteins has changed the tertiary structure of the protein by inducing the structural fluctuations in the backbone atoms.

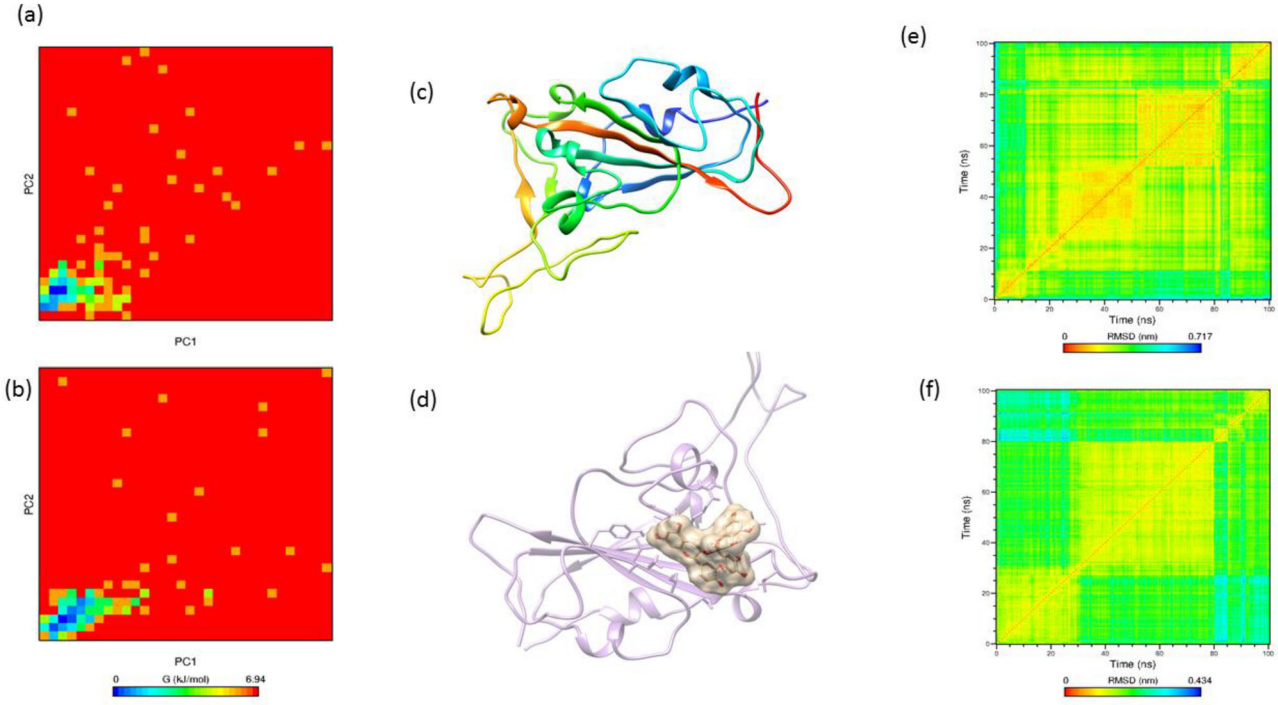


Figure 11. The two-dimensional representation of the free energy landscape calculated from principal components. (a) native spike protein (b) procyanidin-spike protein complex. (c) and (d) represents the apo and procyanidin docked complex of the spike protein. (e) and (f) represents the RMSD matrix of spike protein backbone atoms in native and procyanidin-spike protein docked complex respectively. The RMSD matrix derived from the backbone atoms depicts the tertiary structure of the proteins. The procyanidin shows significant changes in the RMSD matrix, the red line indicates the lowest distance and blue colour indicates the highest distance. The presence of procyanidin in the proteins has changed the tertiary structure of the protein by inducing the structural fluctuations in the backbone atoms.

Secondary and tertiary structure of the receptors

The initial structure of the ACE2 consists of 60.6% helix, 3.2% sheet, 8.1% turn and 28.2% coil whereas at the end of the 100 ns simulation with procyanidin shows a transformation to 53.7% helix, 0.7% sheet, 8.7% turn, 36.2% coil and 0.7% 3-10 helix. The Mpro consists of 23.9% helix, 29.4% sheet,

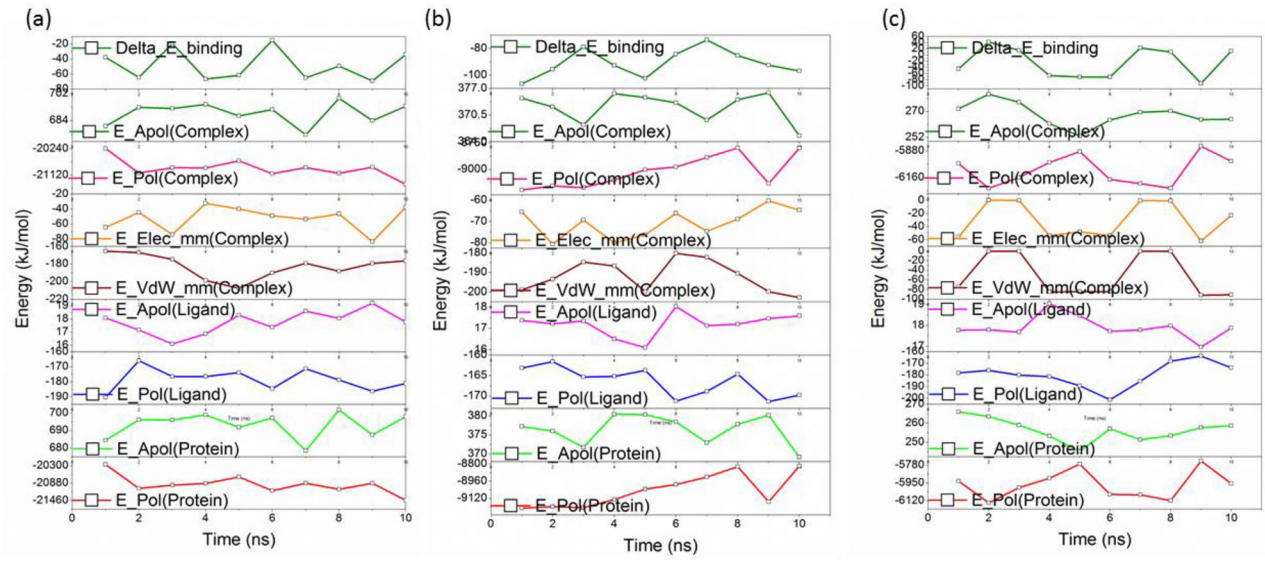


Figure 12. The binding energy decomposition obtained from MM-PBSA calculations. (a) ACE2 (b) Mpro (c) spike protein. In all the cases high van der Waals interaction energies are observed. E_{pol} represents the polar solvation energy, E_{apo} represents apolar solvation energy E_{ele} and E_{vdw}represents the electrostatic and van der Waals interaction energy. Delta-E represents the total binding energy.

Table 4. The binding energy calculated from MM-PBSA Method.

	van der Waal energy (kJ/mol)	Electrostatic energy (kJ/mol)	Polar solvation energy (kJ/mol)	SASA energy (kJ/mol)	Binding energy (kJ/mol)
ACE2	−184.16 ± 3.1	−52.72 ± 5.2	208.61 ± 3.03	−21.94 ± 1.03	−50.21 ± 6.3
Mpro	−193.13 ± 1.3	−69.83 ± 3.06	194.74 ± 1.06	−21.28 ± 3.02	−89.50 ± 6.32
Spike	−53.53 ± 1.6	−31.47 ± 2.03	68.94 ± 2.36	−7.00 ± 3.12	−23.06 ± 4.39

14.1% turn, 32.7% coilin its native state, and end of the simulation with procyanidin shows a transformation of 26.8% helix, 30.1% sheet, 8.8% turn, 34.3% coil. The spike protein is rich in β-sheets with 2.9% helix, 27.3% sheet, 16.7% turn, 50.7% coil, 2.4% 3-10 helix and final structure shows 6.9% helix, 30.7% sheet, 2.1% turn, 58.2% coil, 2.1% 3-10 helix. The time-dependent transformation of the secondary structure of the proteins shown in Figure 5 of supporting information. The transformation between residues in the range of 40-50, 280-350, and 400-410 indicates the binding site where these residues are directly interacting with the proteins which are consistent with the principal components' analysis. The secondary structure of the protein is a key factor that controls the function of the proteins or enzymes. The transformations of the secondary structure are directly correlated to the functional changes of the proteins. Further, the radius of gyration of the proteins was calculated for apo and procyanidin bound state and depicted in Figure 8a. The radius of gyration of the protein determines the compactness of the protein, the higher fluctuations in gyration values were observed for molecule bound state, which is in agreement with time-dependent secondary structure analysis. Further, the tertiary structure of the protein is assessed by calculating the mean square distance between the backbone atoms (Figure 9, 10, 11 e & f). The backbone RMSD matrix of all the protein shows a significant deviation from the apostate to the procyanidin-bound state. The red color indicates the shorter distances between backbone atoms and blue color indicates the larger distances equivalent to 0.54 nm. When the procyanidin interacts with the protein blue color appeared on the matrix map is an indication of the larger pairwise distance between the backbone atoms. Further, we have superimposed the tertiary structure of the protein in both apo and bound state and observed the transformation from its initial frame (Figure 9, 10, 11 c& d). The first frame to the last frame of ACE2 apo shows a 0.3 nm of RMSD over all atoms. Whereas, apo to bound state and first frame- of bound state to the last frame shows an RMSD of 0.5 nm and 0.45 nm respectively. The first to last frame of apo to procyanidin-bound (last frame) and procyanidin-bound first to last shows an RMSD of 0.37 nm, 0.47 nm, and 0.39 nm respectively. The spike protein shows the highest RMSD value in the superposed structure (0.74 nm) with apo to the procyanidin-bound frame. Also, the native spike protein with the last frame shows 0.34 nm and procyanidin-bound first frame to the last frame shows 0.4 nm, respectively.

Binding free energy and decomposition of energy

The binding affinities of the molecule with three proteins were calculated using the MM-PBSA method implemented in the g_mmpbsa tool. The contribution from electrostatic, van der Walls, polar, and SASA indicates the strong affinity of the molecule with protein. Also, calculated binding energy through the MM-PBSA method is agreed with the calculated non-bonded interaction energy. All the protein systems showed a higher contribution from van der Waals energy which is favorable to the binding energy. The energy components that contribute to the total binding energy shows favorable interactions through van der Waals, electrostatic

and SASA energies whereas polar solvation energy show repulsive or unfavorable towards the binding energy. The favorable interaction energies arise from the hydrophobic and polar residues at the binding site. The total binding energy for ACE2, Mpro, and spike proteins are found to be stronger and in agreement with other analyses. The contribution from residues at the binding site of each protein is depicted in Figure 12a, b &c. The residues such as Lue73, Leu120, Asn121, and Ser124 residues of ACE2 show higher contribution ($\sim$ 7 kJ/mol) towards the binding energy. However, we have observed many residues contribute between 1-5 kJ/mol (e.g.: Cys44, Met49, and Asp187) which stabilize the procyanidin in the binding pocket of ACE2, Mpro, and spike protein. The residues Val503, Tyr505 were identified for the Mpro protein and Val503, Tyr 505 for spike proteins (Table 4).

Conclusion

In this study, we have constructed the critical proteins network in the COVID 19 pathways through the named entity recognition method. The selected proteins were used to target againstprocyanidin, and molecular docking revealed the high affinity towards the selected major receptors. Molecular dynamics simulation revealed changes in the dynamics of receptors like ACE2, Mpro and spike in the presence of procyanidin. The structural changes induced by the procyanidin are sufficient to inhibit the proteins to prevent the progression of the disease. The strong van der Waals interaction at the binding site are the major reason for the high affinity of the procyanidin in all the receptors. Also, strong hydrogen bonds and hydrophobic interactions stabilize the procyanidin in the binding cavity. Further, site specific targeted study of these proteins will help to understand the activity of the proteins. Our findings help to design an in vitro and in vivo studies of procyanidin as a potent therapeutic agent in near future.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Authors contribution

Nikhil Maroli (NM) designed work performed molecular modelling and simulation and prepared the manuscript. BaluBhasuran (BB) performed text mining and PonmalaiKolandaivel (PK) and Jeyakumar Natarajan (JN) supervised the work. All the authors reviewed the manuscript.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, *1-2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Actis-Goretta, L., Ottaviani, J. I., Keen, C. L., & Fraga, C. G. (2003). Inhibition of angiotensin converting enzyme (ACE) activity by flavan-3-ols and procyanidins. *FEBS Letters*, *555*(3), 597–600. https://doi.org/10.1016/s0014-5793(03)01355-3

Akif, M., Georgiadis, D., Mahajan, A., Dive, V., Sturrock, E. D., Isaac, R. E., & Acharya, K. R. (2010). High-resolution crystal structures of Drosophila melanogaster angiotensin-converting enzyme in complex with novel inhibitors and antihypertensive drugs. *Journal of Molecular Biology*, *400*(3), 502–517. https://doi.org/10.1016/j.jmb.2010.05.024

Akif, M., Schwager, S. L., Anthony, C. S., Czarny, B., Beau, F., Dive, V., Sturrock, E. D., & Acharya, K. R. (2011). Novel mechanism of inhibition of human angiotensin-I-converting enzyme (ACE) by a highly specific phosphinic tripeptide. *The Biochemical Journal*, *436*(1), 53–59. https://doi.org/10.1042/BJ20102123

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO Journal*, *21*(13), 3213–3224. https://doi.org/10.1093/emboj/cdf327

Bae, J., Kumazoe, M., Murata, K., Fujimura, Y., & Tachibana, H. (2020). Procyanidin C1 Inhibits Melanoma Cell Growth by Activating 67-kDa Laminin Receptor Signaling. *Molecular Nutrition & Food Research*, *64*(7), 1900986. https://doi.org/10.1002/mnfr.201900986

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks* [Paper presentation]. Third International AAAI Conference on Weblogs and Social Media.

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review. A, General Physics*, *38*(6), 3098–3100. https://doi.org/10.1103/physreva.38.3098

Bhasuran, B., & Natarajan, J. (2018). Automatic extraction of gene-disease associations from literature using joint ensemble learning. *PloS One*, *13*(7), e0200699https://doi.org/10.1371/journal.pone.0200699

Bhasuran, B., Murugesan, G., Abdulkadhar, S., & Natarajan, J. (2016). Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases. J Biomed Inform, *64*, 1–9. https://doi.org/10.1016/j.jbi.2016.09.009

Bhasuran, B., Subramanian, D., & Natarajan, J. (2018). Text mining and network analysis to find functional associations of genes in high altitude diseases. *Computational Biology and Chemistry*, *75*, 101–110. https://doi.org/10.1016/j.compbiolchem.2018.05.002

Boopathi, S., Poma, A. B., & Kolandaivel, P. (2020). Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *Journal of Biomolecular Structure and Dynamics*, 1–10.

Brian, D. A., & Baric, R. S. (2005). Coronavirus genome structure and replication. In *Coronavirus replication and reverse genetics* (pp. 1–30). Springer.

Brielle, E. S., Schneidman-Duhovny, D., & Linial, M. (2020). The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. Viruses, *12*(5), 497. https://doi.org/10.3390/v12050497

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, *395*(10223), 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577–8593. https://doi.org/10.1063/1.470117

Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., Hogan, R. J., Tripp, R. A., & Pegan, S. D. (2020). Characterization and noncovalent inhibition of the deubiquitinase and delSGylase activity of SARS-CoV-2 papain-like protease. *ACS Infectious Diseases*, *6*(8), 2099–2109. https://doi.org/10.1021/acsinfecdis.0c00168

Frisch, M., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., & Petersson, G. & others. (2009). *Gaussian 09, Revision d. 01*, Gaussian. Inc. 201.

Guo, Y., Korteweg, C., McNutt, M. A., & Gu, J. (2008). Pathogenetic mechanisms of severe acute respiratory syndrome. *Virus Research*, *133*(1), 4–12. https://doi.org/10.1016/j.virusres.2007.01.022

Gupta, M. K., Vemula, S., Donde, R., Gouda, G., Behera, L., & Vadde, R. (2020). In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *Journal of Biomolecular Structure and Dynamics*, 1–11.

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hilgenfeld, R. (2014). From SARS to MERS: Crystallographic studies on coronaviral proteases enable antiviral drug design. *The FEBS Journal*, *281*(18), 4085–4096. https://doi.org/10.1111/febs.12936

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., & Cheng, Z. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 395(10223), 497-506.

Huang, J., & MacKerell, A. D. Jr, (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, *34*(25), 2135–2145. https://doi.org/10.1002/jcc.23354

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C. & Others, (2020). Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, *582*, 289–293.

Kumari, R., Kumar, R., Consortium, O. S. D. D., & Lynn, A, Open Source Drug Discovery Consortium (2014). g_mmpbsa-a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, *37*(2), 785–789. https://doi.org/10.1103/PhysRevB.37.785

Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annual Review of Virology*, *3*(1), 237–261. https://doi.org/10.1146/annurev-virology-110615-042301

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature, *426*(6965), 450–454. https://doi.org/10.1038/nature02145

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L. … (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

Mark, P., & Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*, *105*(43), 9954–9960. https://doi.org/10.1021/jp003020w

Maroli, N., Jayakrishnan, A., Ramalingam Manoharan, R., Kolandaivel, P., & Krishna, K. (2019). Combined Inhibitory Effects of Citrinin, Ochratoxin-A, and T-2 Toxin on Aquaporin-2. *The Journal of Physical Chemistry. B*, *123*(27), 5755–5768. https://doi.org/10.1021/acs.jpcb.9b03829

Maroli, N., Kalagatur, N. K., Bhasuran, B., Jayakrishnan, A., Manoharan, R. R., Kolandaivel, P., Natarajan, J., & Kadirvelu, K. (2019). Molecular Mechanism of T-2 Toxin-Induced Cerebral Edema by Aquaporin-4 Blocking and Permeation. *Journal of Chemical Information and Modeling*, *59*(11), 4942–4958. https://doi.org/10.1021/acs.jcim.9b00711

McBride, R., Van Zyl, M., & Fielding, B. C. (2014). The coronavirus nucleocapsid is a multifunctional protein. Viruses, *6*(8), 2991–3018. https://doi.org/10.3390/v6082991

Mubarak, A., Alturaiki, W., & Hemida, M. G. (2019). Middle east respiratory syndrome coronavirus (MERS-CoV): Infection, immunological response, and vaccine development. *Journal of Immunology Research*, *2019*, 6491738. https://doi.org/10.1155/2019/6491738

Muralidharan, N., Sakthivel, R., Velmurugan, D., & Gromiha, M. M. (2020). Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 Protease against COVID-19. *Journal of Biomolecular Structure and Dynamics*, 1–6.

NANO-D, I (2016). SAMSON–Software for adaptive modeling and simulation of nanosystems. *Website:* https://Samson-Connect.Net.

Natesh, R., Schwager, S. L. U., Sturrock, E. D., & Acharya, K. R. (2003). Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. Nature, *421*(6922), 551–554. https://doi.org/10.1038/nature01370

Pacheco-Palencia, L. A., Mertens-Talcott, S., & Talcott, S. T. (2008). Chemical composition, antioxidant properties, and thermal stability of a phytochemical enriched oil from Acai (Euterpe oleracea Mart.). *Journal of Agricultural and Food Chemistry*, *56*(12), 4631–4636. https://doi.org/10.1021/jf800161u

Paraskevis, D., Kostaki, E. G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., & Tsiodras, S. (2020). Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, Genetics and Evolution : journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *79*, 104212https://doi.org/10.1016/j.meegid.2020.104212

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Souquet, J.-M., Cheynier, V., Brossaud, F., & Moutounet, M. (1996). Polymeric proanthocyanidins from grape skins. Phytochemistry, *43*(2), 509–512. https://doi.org/10.1016/0031-9422(96)00301-9

Stobart, C. C., Sexton, N. R., Munjal, H., Lu, X., Molland, K. L., Tomar, S., Mesecar, A. D., & Denison, M. R. (2013). Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *Journal of Virology*, *87*(23), 12611–12618. https://doi.org/10.1128/JVI.02050-13

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., & Du, L. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: Implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology*, *17*(6), 613–620. https://doi.org/10.1038/s41423-020-0400-4

Te Velthuis, A. J. W., Van Den Worm, S. H. E., & Snijder, E. J. (2012). The SARS-coronavirus nsp7 + nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Research*, *40*(4), 1737–1747. https://doi.org/10.1093/nar/gkr893

Towler, P., Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M., Williams, D., Dales, N. A., Patane, M. A., & Pantoliano, M. W. (2004). ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *The Journal of Biological Chemistry*, *279*(17), 17996–18007. https://doi.org/10.1074/jbc.M311191200

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Vivas, N., Nonier, M.-F., Pianet, I., de Gaulejac, N. V., & Fouquet, É. (2006). Proanthocyanidins from Quercus petraea and Q. robur heartwood: Quantification and structures. *Comptes Rendus Chimie*, *9*(1), 120–126. https://doi.org/10.1016/j.crci.2005.09.001

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. "*Protein Engineering, Design and Selection*"", *8*(2), 127–134. https://doi.org/10.1093/protein/8.2.127

Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R., Liu, Z., & Merrill, W. & others. (2020). CORD-19: The Covid-19 Open Research Dataset. ArXiv.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. Nature, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Xu, X., & Goddard, W. A. (2004). The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proceedings of the National Academy of Sciences*, *101*(9), 2673–2677. https://doi.org/10.1073/pnas.0308730100

Xue, G., Li, S., Zhang, W., Du, B., Cui, J., Yan, C., Huang, L., Chen, L., Zhao, L., Sun, Y., Li, N., Zhao, H., Feng, Y., Wang, Z., Liu, S., Zhang, Q., Xie,

X., Liu, D., Yao, H., & Yuan, J. (2020). A Reverse-Transcription Recombinase-Aided Amplification Assay for Rapid Detection of the 2019 Novel Coronavirus (SARS-CoV-2). *Analytical Chemistry*, *92*(14), 9699–9705. https://doi.org/10.1021/acs.analchem.0c01032

Yang, J., & Xiao, Y.-Y. (2013). Grape phytochemicals and associated health benefits. *Critical Reviews in Food Science and Nutrition*, *53*(11), 1202–1225. https://doi.org/10.1080/10408398.2012.692408

Zhang, X. M., Herbst, W., Kousoulas, K. G., & Storz, J. (1994). Biological and genetic characterization of a hemagglutinating coronavirus isolated from a diarrhoeic child. *Journal of Medical Virology*, *44*(2), 152–161. https://doi.org/10.1002/jmv.1890440207

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*, *6*(1), 14–18. https://doi.org/10.1038/s41421-020-0153-3