

COMPARATIVE ANALYSIS OF SMALL MOLECULES AND NATURAL PLANT COMPOUNDS AS THERAPEUTIC INHIBITORS TARGETING RdRp AND NUCLEOCAPSID PROTEINS OF SARS COV 2: AN *IN SILICO* APPROACH

PAVAN KUMAR POLEBOYINA^{ORCID}, SMITA C PAWAR*

Department of Genetics and Biotechnology, University College of Science, Osmania University, Hyderabad, Telangana, India.

*Correspondence author: Smita C Pawar; Email: smita.prof@gmail.com

Received: 15 April 2023, Revised and Accepted: 04 June 2023

ABSTRACT

Objective: Coronavirus disease 2019 (COVID-19) is a virus-borne infection caused by the severe acute respiratory syndrome coronavirus disease-2 (SARS-CoV-2) virus. Nucleocapsid protein and RNA-dependent RNA polymerase (RdRp) activity in viral structural membrane, transcription, and replication have been identified as desirable targets for the development of novel antiviral strategies. The SARS-COV-2 N protein binds to the viral genome to promote the precise folding of the hammerhead ribozyme, preventing ineffective RNA confirmations, and directs them into a helical capsid shape or ribonucleoprotein complex, which is vital for viability. RNA synthesis requires RdRp to form phosphodiester bonds based on the RNA template. SARS-CoV-2 RNA synthesis, transcription, and replication depend on RdRp's complex with nsp7 and nsp8.

Methods: Our study targeted SARS-COV-2 RdRp and N proteins with natural plant compounds and small molecules. Hyperchem software optimized their structures geometrically and energetically. Based on MolDock, Rerank, and H-bonding energy, the best ligands were selected using the Molegro virtual docker.

Results: In our analysis, we have identified nine compounds against N protein and seven compounds against RdRp protein that had more potent inhibitory effects with the lowest MolDock scores. The top 6 (Alpha solanine, Betanin, cairicoside I, Ginsenoside rb 1, Naringin, Polyphyllin I) compounds that have better inhibitory effects against both proteins.

Conclusion: We conclude that the top six compounds have greater inhibitory efficacy against N and RdRp protein than other compounds. However, *in vitro* and *in vivo* experimental studies, as well as clinical trials, are required to achieve the desired result.

Keywords: RNA-dependent RNA polymerase, N protein, Severe acute respiratory syndrome coronavirus disease-2, Coronavirus disease, Natural plant compounds, Small molecules, Molecular docking.

© 2023 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2023v16i10.48095>. Journal homepage: <https://innovareacademics.in/journals/index.php/ajpcr>

INTRODUCTION

In the Wuhan City of China, pneumonia-infected patients with an unknown origin were identified on December 31, 2019 [1]. The current pandemic severe acute respiratory syndrome coronavirus disease-2 (SARS-Cov-2) reasons a public health emergency of international concern and significantly broken the worldwide economy. Coronavirus disease 2019 (COVID-19) is caused due to the newly discovered and deadly coronavirus is a highly contagious disease, named SARS-Cov-2. The name "coronavirus" is coined due to the crown-like projections on its surface as in Latin. "Corona" means "halo" or "Crown." On January 13, 2020, entire genome analysis becomes achieved found that a novel coronavirus (Gen Bank No. MN908947) officially called SARS-COV-2 previously known as SARS-COV-1 [2].

COV is a positive-sense single-stranded RNA (ssRNA) virus that consists of 29,903 nucleotides and un-translated sequences of 229 and 254 nucleotides on the 3'- and 5'-ends, respectively. It is included in the β coronavirus genus intently associated with the genomic organization of SARS-COV identified in 2003 [3]. The open reading frame (ORF) region of coronaviruses contains specific genes which encode for the spike, capsid, and replicative proteins [4]. The two ORF (ORF1a and ORF1b) and polyproteins split and encode structural and non-structural proteins. Structural proteins include E protein (envelope), M protein (membrane), S protein (spike), and N (Nucleocapsid) proteins, while non-structural proteins are in a range from nsp1 to nsp16 [5].

The replicase-transcriptase complex (RTC) consists of 16 nsps in which multiple enzymes, including nsp3 (papain-like protease), 3 CL

protease nsp5 (chymotrypsin-like main protease), primase complex, RNA-dependent RNA polymerase (RdRp) (nsp12), nsp13 (helicase) and nsp14 (exoribonuclease) [6]. The nsp12 is the main content of the replication machinery which involves the transcriptional activity of RdRp [7]. It is a popular target for a selective antiviral strategy against SARS-COV-2 because RNA synthesis by RdRp does not occur in mammalian cells [8]. The RdRp is present in all ssRNA viruses except in retroviruses. The structural analyses of RNA viruses have shown that their RdRp protein contains thumb, finger, and palm domains, which are referred to as a closed, right-handed polymerase [9]. RdRp complex (nsp7-nsp8-nsp12) is the core component that is responsible for viral RNA replication and this complex is also inclusive of nsp7-nsp8 (nsp8-1) heterodimer and an additional subunit nsp8 (nsp8-2) [10]. Inhibition of viral replication by targeting RdRp activity can become a powerful therapeutic approach [11].

SARS-COV-2 N protein is firmly associated and interacts with the viral genomic RNA structure and function. It contains phosphorylation sites in unstructured regions, and it plays a major role in the assembly and transcription of the virus while functioning as an RNA chaperone [12,13]. The nucleocapsid protein was composed of an alpha helix (21.24%), beta-fold (16.71%), beta-turn (6.92%), and random coil (55.13%). It has two main domains the RNA binding domain - N-terminal domain (NTD) and the C-terminal dimerization domain. The NTD domain plays an important role in viral replication and transcription by binding to the 3' end of the viral RNA genome due to positive amino acid electrostatic interactions. The major role of the N protein is to bind to the viral RNA to promote the exact folding of the hammerhead ribozyme avoiding

unproductive RNA conformations. It makes them into a helical capsid structure or ribonucleoprotein complex. Because of this function, it is suggested that the coronavirus N protein is an RNA chaperone and performs the main role in viral genomic RNA replication [14,15]. Our present study has been undertaken to compare the binding affinity analysis of various natural plant compounds and small molecules against RdRp and nucleocapsid proteins of SARS-COV-2 using molecular docking methods [16].

METHODS

Our present study involves retrieval of the 3D structure of target proteins and small molecules, natural plant compounds from Protein Data Bank [17], and Pubchem database [18], respectively. The proposed ligands are designed and optimized in Hyperchem8 [19]. Molecular docking experiments were performed by Molegro Virtual Docker (MVD2019.7.0.0) [20] and analyzed its data by Molegro Data modeler [21].

Retrieval of target proteins

3D structures of RdRp Cryo-EM structure of the apo nsp12-nsp7-nsp8 complex protein (PDB ID: 7BV1) and crystal structure of SARS-COV-2 nucleocapsid protein N-terminal RNA binding domain (PDB ID: 6M3M) was retrieved from Protein Data Bank. Before docking the water molecules, unwanted hetero atoms and other ligand compounds were removed by MVD.

Retrieval of natural plants compounds and small molecules

A total 56 compounds were virtually screened and the top ligand for each protein was taken for consideration. The structures of natural plant compounds and small molecules were retrieved from PubChem and some structures were drawn on the Hyperchem8 package; they were pre-optimized using the molecular mechanics force field (MM+, AMBER) procedure. To obtain the conformers of ligands with the lowest energy and semi-empirical method AM-1 was applied to the molecular structures. To avoid the local stability, each molecular structure was optimized several times with different starting points using the Polak-Rebier algorithm, until the root-mean-square gradient is equal to $0.001 \text{ kcal } \text{Å}^{-1} \text{ mol}^{-1}$ [22]. Then, the energy minimizations of all ligand structures were done using Hyperchem8. The first step was by calculating a single point that was used to determine the total molecular energy of the structure, the second step to determine the energy minimization algorithms that locate the flexible structure was used geometric optimization calculation (MM+, AMBER force field) [23]. Then, a total of 56 structures was used for molecular docking calculations.

Prediction of drug binding cavities

The drug-binding cavities in RdRp and Nucleocapsid proteins of SARS-COV-2 for natural plant compounds and small molecule ligands are not well characterized. The amino acid residues responsible for cavity formation in RdRp and nucleocapsid proteins were identified through the MVD cavity detection algorithm. The program generally identified five different cavities. MVD identified five different cavities in both RdRp and N proteins.

Protein and ligand preparation

Protein

RdRp Cryo-EM structure of the apo nsp12-nsp7-nsp8 complex protein (PDB ID: 7BV1) and crystal structure of SARS-COV-2 nucleocapsid protein N-terminal RNA binding domain (PDB ID: 6M3M) was selected for the molecular docking studies using MVD. Using the utilities provided in Molegro Virtual Docker, all necessary H atom addition, valency checks, and protein preparation (protonation) were done, repair, and rebuilt it. For precise docking, it is important that the imported structures must be prepared accurately, that is the atom connectivity and bond orders are correct, and partial atomic charges are assigned. PDB file often has poor or missing assignments of explicit hydrogens, and the PDB file format cannot accommodate bond order

information. Then, the repair and rebuilt protein were saved in *.Mol format. The final structure was visualized and analyzed with SPDBV 4.10 [24].

Ligand

The selected compounds were downloaded from the Pubchem, zinc database, and drug bank. Some torsion and peptide bond-containing ligands were drawn and minimized its energy using Hyperchem 8 software and imported into the MVD workspace in *.Mol format. Before import, the small molecules and all natural-plant compounds undergo a series of steps that generate variation and optimization of the structure.

Molecular docking studies

Molegro Virtual Docker (MVD2019_7_0_0) program was used for the validation of molecular docking. By comparing, the accuracy of MVD is higher than the other software such as Glide, Surflex, and FlexX. A total of 56 compounds were tested against RdRp and N proteins of SARS-COV-2. The MolDock scoring function was also set with a grid resolution of 0.30 Å . It was set at a maximum iteration of 1500 with a simplex evolution size of 50 and a minimum of 10 runs were performed for each compound with threshold energy of 100. In addition, the simplex evolution was set for 300 steps with a neighbor distance factor of 1.00. In MVD MolDock scoring function works based on a piecewise linear potential (PLP) and it is proposed by Gehlhaar and Yang

The MolDock score docking scoring function, EMolDock Score, is defined by the following energy terms Equation 1:

$$E_{\text{score}} = E_{\text{inter}} + E_{\text{intra}} \quad (1)$$

Where, E_{inter} represents the ligand-protein interaction energy... Equation 2:

$$E_{\text{inter}} = \sum_{i=\text{ligand}} \sum_{j=\text{protein}} [E_{\text{PLP}}(r_{ij}) + 332.0 \frac{q_i q_j}{4r_{ij}^2}] \quad (2)$$

Where E_{PLP} is the piecewise-linear potential, the numerical value of 332.0 fixes the units of the electrostatic energy in kcal/mol. The second term describes the electrostatic interactions between charged atoms, and the internal energy of the ligand (E_{intra}) expressed by equation 3...

$$E_{\text{intra}} = \sum_{i=\text{ligand}} \sum_{j=\text{protein}} [E_{\text{PLP}}(r_{ij})] + \sum_{\text{flexible bond}} A[1 - \cos(m\theta - \theta_0)] + E_{\text{clash}} \quad (3)$$

The double summation contains all atom pairs in the ligand except those which are connected with two bonds or less. The second term is torsional energy where θ is the torsional angle of the bond. E_{clash} assigns a penalty of 1000 provided that the distance between two heavy atoms is $< 2.0 \text{ Å}$ [25]. The best pose of each compound was selected for subsequent ligand-protein interaction energy analysis [26]. Hydrogen bond interaction and its binding energy were observed between the amino acid residues on the target site with the functional group of small molecules and plant compounds.

ADMET properties

The Top scoring ligands of N protein (Pralsetinib) and RdRp protein (Forskolin) were further used to estimate pharmacokinetic properties, drug-likeness, and toxicity using the pkCSM [27], SwissADME tool [28], and molinspiration cheminformatics tools [29].

RESULTS

Target protein conformation

The RdRp Cryo-EM structure of the apo nsp12-nsp7-nsp8 complex protein (PDB ID: 7BV1) with resolution 2.80 Å and the crystal structure of SARS-COV-2 nucleocapsid protein N-terminal RNA binding domain (PDB ID: 6M3M) with resolution 2.70 Å were retrieved from Protein Data Bank (PDB) database. The retrieved proteins checked structural

conformations by ERRAT [30], PROCHEK [31], and PDBsum [32]. The validated 3D, secondary structural analysis, and Ramachandran plot are shown in Figs. 1 and 2.

The N and RdRp proteins of SARS-CoV-2 active sites were obtained by the MVD cavity detection algorithm. In this study, the top volume cavity of N protein with volume=14.336Å³, surface area=60.16Å²(X=2.94, Y=1.13, Z=-2.90, XYZ axis), and RdRp protein with volume=3206.66 Å³, surface area=629.2 Å²(X=88.71, Y=89.73, Z=107.25, XYZ axis) were used for docking procedure (Fig. 3).

Determination of ligand structures

The 56 structures of small molecules and natural plant compounds were used to find out their potential binding with the target proteins

RdRp and N (Nucleocapsid). The ligand's structural information and structural flexibility after single point energy and after geometrical optimization energy are shown in Table 1.

Validation of docking results

During the docking study, each compound is selected as the best pose to determine the MolDock score, Rerank score, interaction energy, torsion angle, and H-bonding against N and RdRp protein. The MolDock score, Rerank score, the total interaction energy between the pose and the target molecule, torsion angle, and H-bonding energy of 56 compounds against N protein and RdRp protein are represented in Tables 2 and 3, respectively. In our study, we compared the binding efficacy of ligands against protein with and without energy minimization. We found that energy-minimized compounds show better binding results [33]. Among

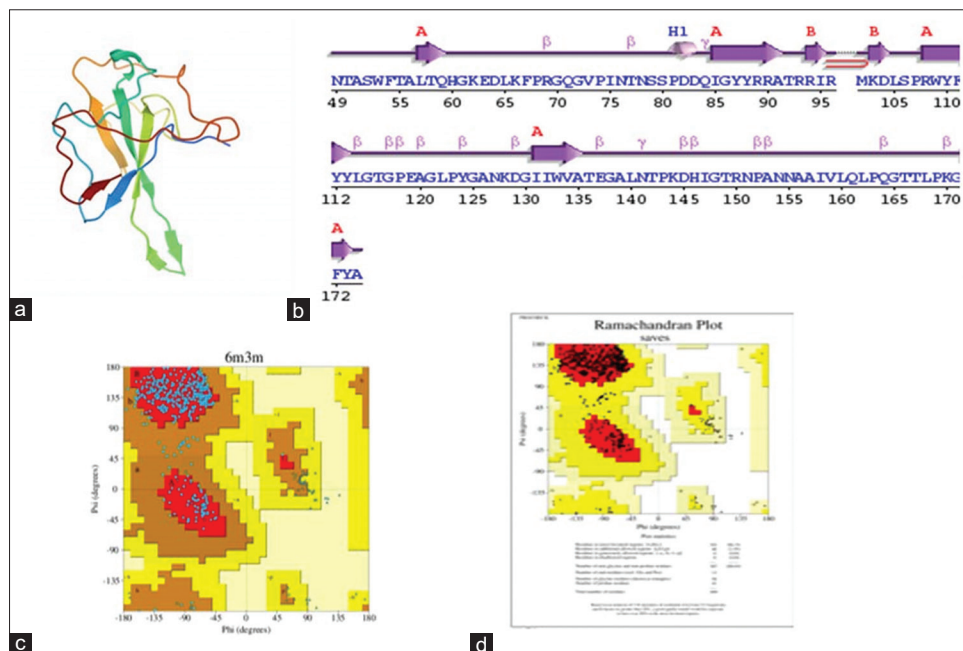


Fig. 1: Structural information of N protein (a) cartoon 3D structure of N protein (b) secondary structure of 3pty mapped obtained using PDBsum (c) and (d) Ramachandran plot

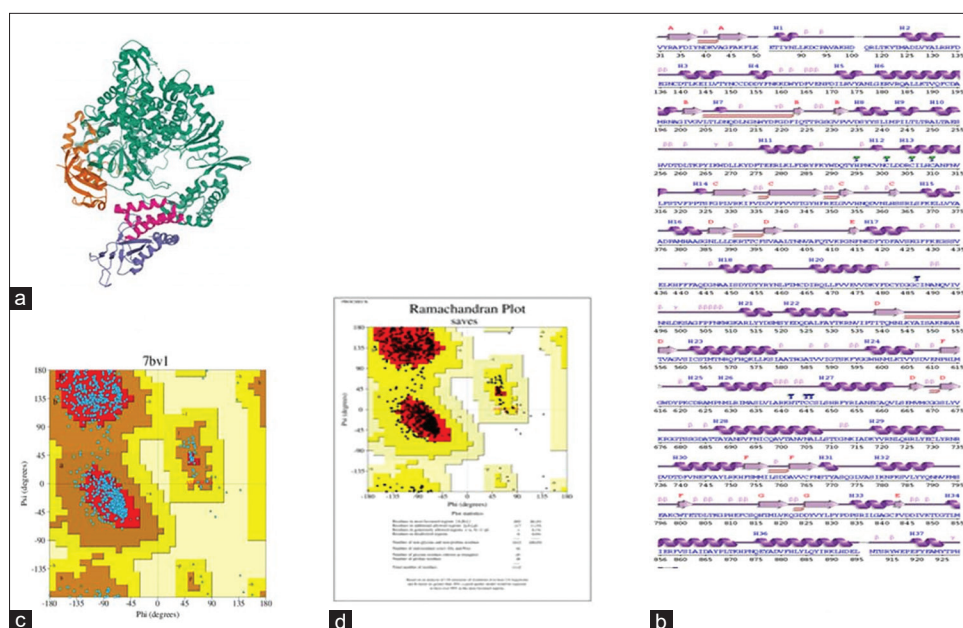
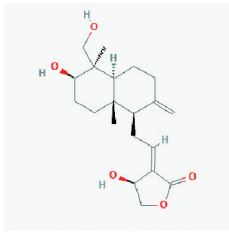
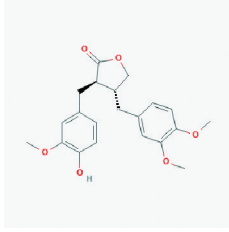
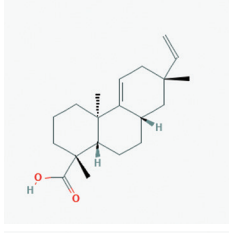
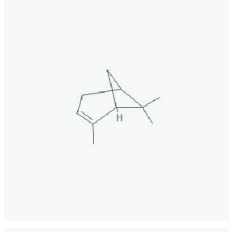
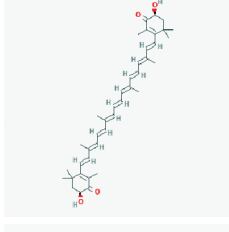
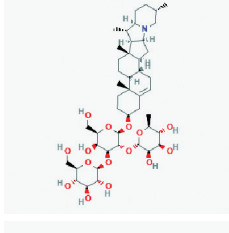
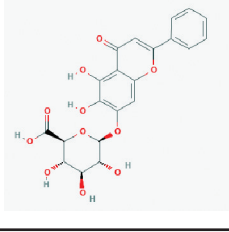


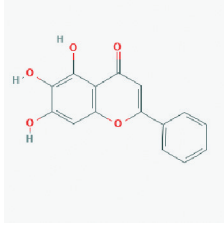
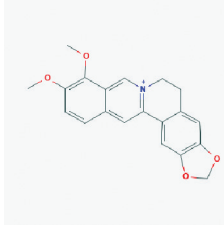
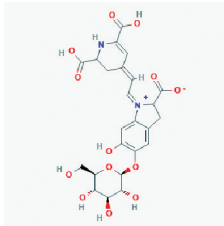
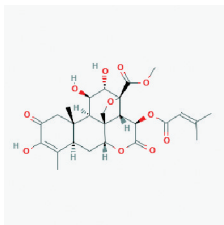
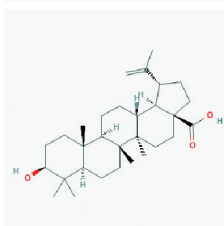
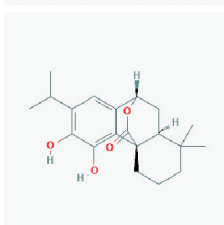
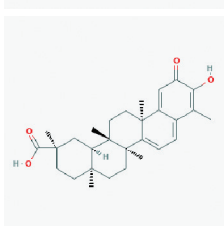
Fig. 2: Structural information of RdRp protein (a) cartoon 3d structure of RdRp protein (b) secondary structure of 3pty mapped obtained using PDBsum (c) and (d) Ramachandran Plot

Table 1: Ligand structural information and After single point and geometry optimization energy minimization calculations

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Andrographolide	5318517	C ₂₀ H ₃₀ O ₅ (350.4)		83.60 Kcal/mol	43.22 Kcal/mol
Arctigenin	64981	C ₂₁ H ₂₄ O ₆ (372.4)		81.56 Kcal/mol	49.26 Kcal/mol
Acanthoic acid	9817887	C ₂₀ H ₃₀ O ₂ (302.5)		77.38 Kcal/mol	37.14 Kcal/mol
Alpha pinene	6654	C ₁₀ H ₁₆ (136.23)		75.72 Kcal/mol	59.29 Kcal/mol
Astaxanthin	5281224	C ₄₀ H ₅₂ O ₄ (596.8)		174.15 Kcal/mol	40.71 Kcal/mol
Alpha solanine	9549171	C ₄₅ H ₇₃ NO ₁₅ (868.1)		174.75 Kcal/mol	96.90 Kcal/mol
Baicalin	64982	C ₂₁ H ₁₈ O ₁₁ (446.4)		116.23 Kcal/mol	22.71 Kcal/mol

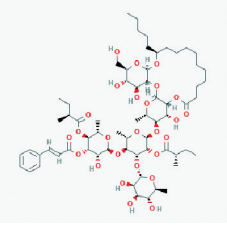
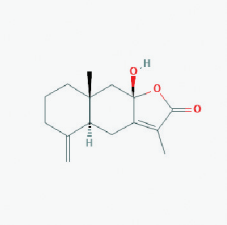
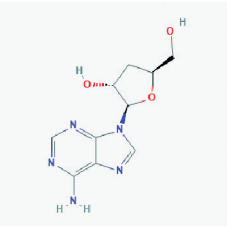
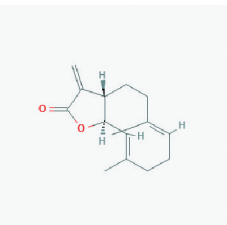
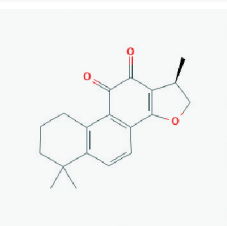
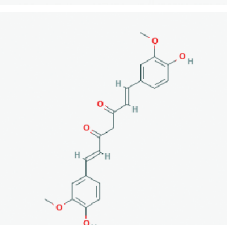
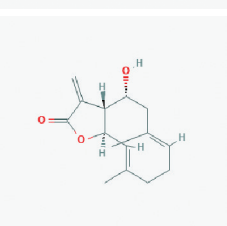
(Contd...)

Table 1: (Continued)

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Baicalein	5281605	$C_{15}H_{10}O_5$ (270.24)		84.86 Kcal/mol	30.62 Kcal/mol
Berberine	2353	$C_{20}H_{18}NO_4^+$ (336.4)		77.51 Kcal/mol	28.71 Kcal/mol
Betainin	12300103	$C_{24}H_{26}N_2O_{13}$ (550.5)		143.39 Kcal/mol	31.49 Kcal/mol
Brusatol	73432	$C_{26}H_{32}O_{11}$ (520.5)		156.10 Kcal/mol	55.88 Kcal/mol
Betulinic acid	64971	$C_{30}H_{48}O_3$ (456.7)		119.77 Kcal/mol	84.95 Kcal/mol
Carnosol	442009	$C_{20}H_{26}O_4$ (330.4)		92.79 Kcal/mol	61.67 Kcal/mol
Celastrol	122724	$C_{29}H_{38}O_4$ (450.6)		140.05 Kcal/mol	94.62 Kcal/mol

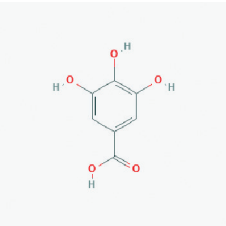
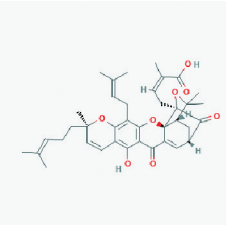
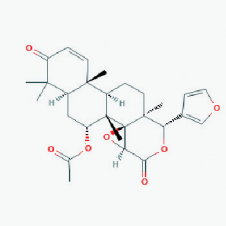
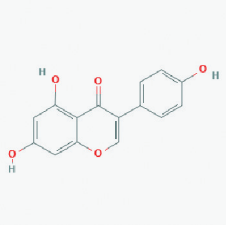
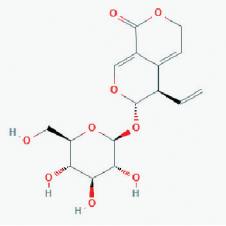
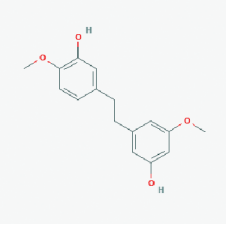
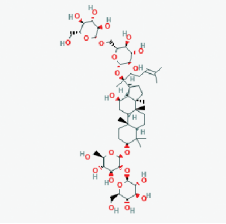
(Contd...)

Table 1: (Continued)

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Cairicoside I	122230626	$C_{65}H_{102}O_{26}$ (1299.5)		321.53 Kcal/mol	86.50 Kcal/mol
Codonolactone	155948	$C_{15}H_{20}O_3$ (248.32)		65.55 Kcal/mol	30.4 Kcal/mol
Cordycepin	6303	$C_{10}H_{13}N_5O_3$ (251.24)		56.38 Kcal/mol	28.73 Kcal/mol
Costunolide	5281437	$C_{15}H_{20}O_2$ (232.32)		63.24 Kcal/mol	17.16 Kcal/mol
Cryptotanshinone	160254	$C_{19}H_{20}O_3$ (296.4)		80.35 Kcal/mol	25.30 Kcal/mol
Curcumin	969516	$C_{21}H_{20}O_6$ (368.4)		112.18 Kcal/mol	10.98 Kcal/mol
Eupatolide	5281460	$C_{15}H_{20}O_3$ (248.32)		64.93 Kcal/mol	18.60 Kcal/mol

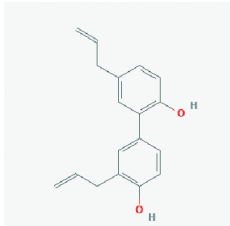
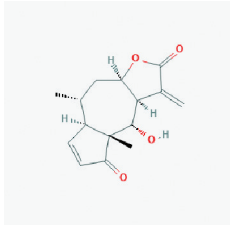
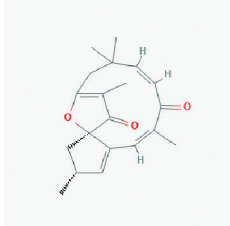
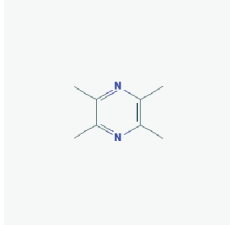
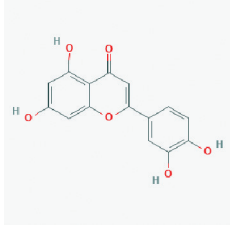
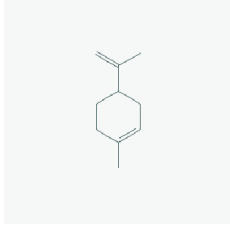
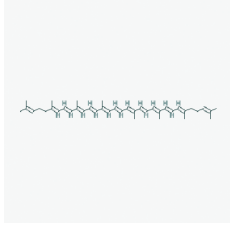
(Contd...)

Table 1: (Continued)

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Gallic acid	370	C ₇ H ₆ O ₅ (170.12)		46.89 Kcal/mol	5.58 Kcal/mol
Gambogic acid	9852185	C ₃₈ H ₄₄ O ₈ (628.7)		170.80 Kcal/mol	44.92 Kcal/mol
Gedunin	12004512	C ₂₈ H ₃₄ O ₇ (482.6)		309.07 Kcal/mol	225.87 Kcal/mol
Genistein	5280961	C ₁₅ H ₁₀ O ₅ (270.24)		85.22 Kcal/mol	37.07 Kcal/mol
Gentiopicroside	88708	C ₁₆ H ₂₀ O ₉ (356.32)		66.62 Kcal/mol	18.62 Kcal/mol
Gigantol	3085362	C ₁₆ H ₁₈ O ₄ (274.31)		55.08 Kcal/mol	7.44 Kcal/mol
Ginsenoside rb1	9898279	C ₅₄ H ₉₂ O ₂₃ (1109.3)		202.49 Kcal/mol	95.76 Kcal/mol

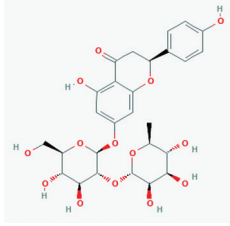
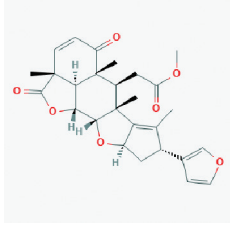
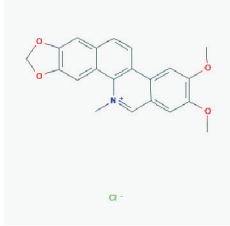
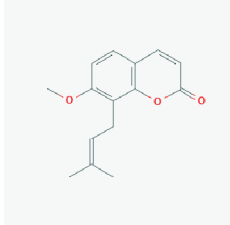
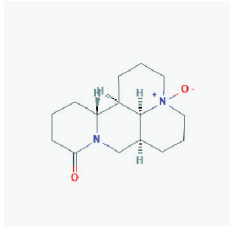
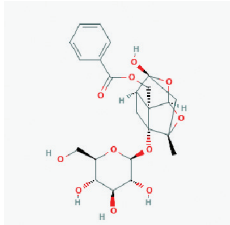
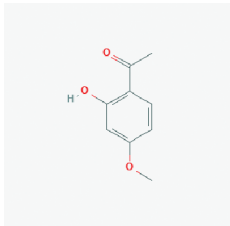
(Contd...)

Table 1: (Continued)

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Honokiol	72303	C ₁₈ H ₁₈ O ₂ (266.3)		63.78 Kcal/mol	28.46 Kcal/mol
Helenalin	23205	C ₁₅ H ₁₈ O ₄ (262.30)		96.23 Kcal/mol	46.14 Kcal/mol
Jatrophone	5281373	C ₂₀ H ₂₄ O ₃ (312.4)		107.12 Kcal/mol	33.06 Kcal/mol
Ligustrazine	14296	C ₈ H ₁₂ N ₂ (136.19)		20.66 Kcal/mol	1.87 Kcal/mol
Luteolin	5280445	C ₁₅ H ₁₀ O ₆ (286.24)		86.31 Kcal/mol	13.75 Kcal/mol
Limonene	22311	C ₁₀ H ₁₆ (136.23)		25.5 Kcal/mol	12.84 Kcal/mol
Lycopene	446925	C ₄₀ H ₅₆ (36.9)		175.8 Kcal/mol	38.98 Kcal/mol

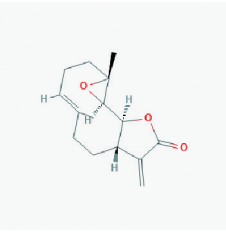
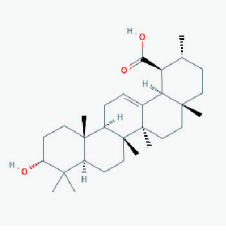
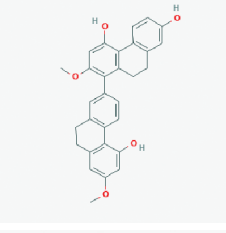
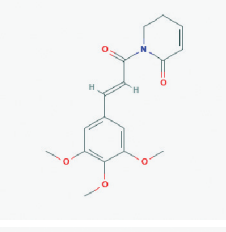
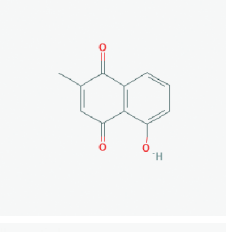
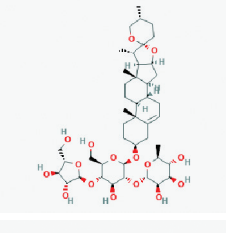
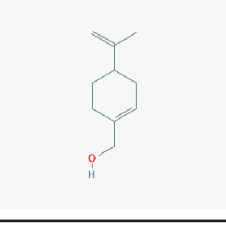
(Contd...)

Table 1: (Continued)

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Naringin	442428	$C_{27}H_{32}O_{14}$ (580.5)		99.91 Kcal/mol	33.08 Kcal/mol
Nimbolide	12313376	$C_{27}H_{30}O_7$ (466.5)		171.95 Kcal/mol	96.88 Kcal/mol
Nitidine chloride	25659	$C_{21}H_{18}ClNO_4$ (383.8)		106.24 Kcal/mol	31.38 Kcal/mol
Osthole	10228	$C_{15}H_{16}O_3$ (244.28)		61.29 Kcal/mol	8.65 Kcal/mol
Oxymatrine	114850	$C_{15}H_{24}N_2O_2$ (264.36)		48.43 Kcal/mol	28.91 Kcal/mol
Paeoniflorin	442534	$C_{23}H_{28}O_{11}$ (480.5)		150.59 Kcal/mol	87.94 Kcal/mol
Paeonol	11092	$C_9H_{10}O_3$ (166.17)		46.43 Kcal/mol	8.50 Kcal/mol

(Contd...)

Table 1: (Continued)

Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Parthenolide	7251185	C ₁₅ H ₂₀ O ₃ (248.32)		225.52 Kcal/mol	177.95 Kcal/mol
Plectranthiolic acid	154731119	C ₃₀ H ₄₈ O ₃ (456.7)		125.92 Kcal/mol	95.95 Kcal/mol
Phoyunnanin E	101380569	C ₃₀ H ₂₆ O ₅ (466.5)		102.00 Kcal/mol	26.41 Kcal/mol
Piperlogumine	101380569	C ₃₀ H ₂₆ O ₅ (466.5)		92.93 Kcal/mol	18.36 Kcal/mol
Plumbagin	10205	C ₁₁ H ₈ O ₃ (188.18)		68.49 Kcal/mol	8.15 Kcal/mol
Polyphyllin I	11018329	C ₄₄ H ₇₀ O ₁₆ (855.0)		177.11 Kcal/mol	97.87 Kcal/mol
Perillyl alcohol	10819	C ₁₀ H ₁₆ O (152.23)		25.96 Kcal/mol	13.50 Kcal/mol

(Contd...)

Table 1: (Continued)

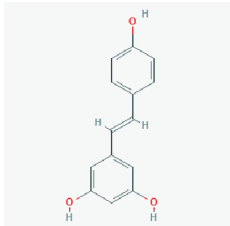
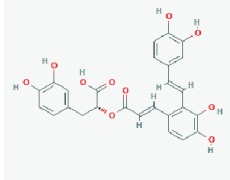
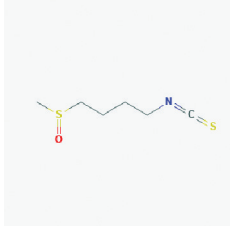
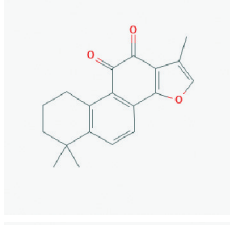
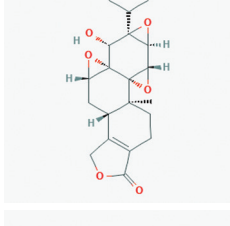
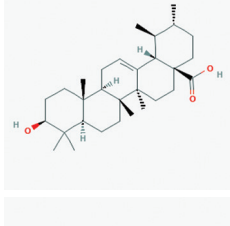
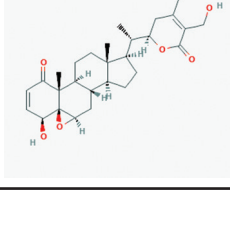
Name of the inhibitor	Pubchem CID	Molecular formula (Molecular weight g/mol)	STRUCTURE (2D)	After single point energy	After geometry optimization minimizing energy
Resveratrol	445154	C ₁₄ H ₁₂ O ₃ (228.24)		60.41 Kcal/mol	10.75 Kcal/mol
Salvianolic acid A	5281793	C ₂₆ H ₂₂ O ₁₀ (494.4)		141.70 Kcal/mol	13.73 Kcal/mol
Sulforaphane	5350	C ₆ H ₁₁ NOS ₂ (177.3)		270.60 Kcal/mol	3.35 Kcal/mol
Tanshinone IIA	164676	C ₁₉ H ₁₈ O ₃ (294.3)		86.22 Kcal/mol	31.05 Kcal/mol
Triptolide	107985	C ₂₀ H ₂₄ O ₆ (360.4)		559.67 Kcal/mol	497.19 Kcal/mol
Ursolic acid	64945	C ₃₀ H ₄₈ O ₃ (456.7)		113.94 Kcal/mol	82.21 Kcal/mol
Withaferin A	265237	C ₂₈ H ₃₈ O ₆ (470.6)		273.16 Kcal/mol	210.199 Kcal/mol

Table 2: MolDock score, rerank score, interaction, torsions, and Hbond energy of the docked compounds with N protein

Name of the inhibitor	MolDock Score, kcal/mol	Rerank Score, kcal/mol	Interaction energy, kcal/mol	Torsion angle, kcal/mol	H-Bond, kcal/mol
Andrographolide	-107.944	-82.44	-122.4	3	-10.89
Arctigenin	-110.939	-80.25	-113.42	7	-9.58
Acanthoic acid	-80.68	-51.46	-90.72	2	-4.05
Alpha pinene	-49.54	-15.53	-62.36	0	0
Astaxanthin	-133.88	-107.55	-180.85	19	-10.16
Alpha solanine	-159.99	-71.77	-172.51	8	-17.11
Baicalin	-105.37	-102.20	-148.59	4	-18.27
Baicalein	-75.30	-72.39	-98.96	1	-13.12
Berberine	-117.98	-57.97	-123.15	2	-2.6
Betanin	-160.78	-121.14	-169.77	8	-19.8
Brusatol	-122.35	-99.60	-124.69	6	-13.01
Betulinic acid	-105.32	-63.60	-109.40	2	-5.28
Carnosol	-64.60	-36.71	-82.93	1	-10.86
Celastrol	-100.725	-28.27	-121.27	1	-9.25
Cairicoside I	-285.68	-36.03	-245.10	24	-16.8
Codonolactone	-77.74	-63.24	-79.63	0	-8.29
Cordycepin	-90.37	-76.89	-97.15	2	-10.49
Costunolide	-92.71	-57.96	-76.99	0	-2.4
Cryptotanshinone	-102.47	-47.45	-108.00	0	-4.66
Curcumin	-133.59	-96.70	-122.56	10	-11.79
Eupatolide	-92.12	-56.51	-79.39	0	-9.15
Gallic acid	-79.03	-69.79	-86.67	1	-13.41
Gambogic acid	-164.94	-102.58	-147.09	11	-8.5
Gedunin	-119.97	-94.78	-121.98	3	-5.06
Genistein	-68.14	-55.04	-93.55	1	-11.26
Gentiopicroside	-109.99	-94.47	-129.60	4	-11.85
Gigantol	-112.45	-87.08	-111.03	5	-6.4
Ginsenoside rb1	-172.65	-114.14	-183.54	11	-18.50
Honokiol	-82.89	-50.03	-88.17	5	-7.4
Helenalin	-101.44	-44.15	-94.11	0	-6.9
Jatrophone	-107.70	-64.89	-85.10	0	-3.46
Ligustrazine	-45.03	-43.16	-54.79	0	-3.39
Luteolin	-100.55	-81.30	-108.48	1	-12.02
Limonene	-54.03	-47.23	-59.15	1	0
Lycopene	-150.00	-98.54	-149.19	29	0
Naringin	-145.45	-122.25	-171.67	6	-23.82
Nimbolide	-127.16	-90.66	-116.72	4	-12.05
Nitidine chloride	-122.18	-99.14	-124.62	2	-3.86
Osthole	-91.72	-70.25	-95.01	4	-3.9
Oxymatrine	-67.72	-64.05	-86.77	0	v3.48
Paeoniflorin	-114.76	-90.24	-132.29	3	-19.32
Paeonol	-71.70	-51.20	-72.66	2	-7.3
Parthenolide	-101.17	-70.32	-85.69	0	-5.0
Plectranthioc acid	-78.65	-7.22	-95.51	1	-6.36
Phoyunnanin E	-130.07	-66.91	-144.87	3	-10.75
Piperlogumine	-105.33	-81.94	-11.20	6	-6.46
Plumbagin	-67.63	-59.66	-78.48	0	-5.9
Polyphyllin I	-166.78	-53.30	-202.25	8	-18.28
Perillyl alcohol	-57.33	-49.25	-64.05	2	-5.0
Resveratrol	-102.06	-85.28	-105.42	2	-7.9
Salvianolic acid A	-143.71	-102.69	-120.18	9	-23.26
Sulforaphane	-69.60	-56.81	-69.23	6	-7.6
Tanshinone IIA	-91.01	-66.16	-96.79	0	-3.9
Triptolide	-108.18	-69.07	-116.04	1	-6.75
Ursolic acid	-111.21	-63.80	-115.39	1	-6.12
Withaferin A	-108.41	-84.51	-125.63	3	-10.33

all docked 56 ligands, nin compounds shown better binding energy with N protein and seven compounds with RdRp protein.

N protein results

Out of all 56 compounds Cairicoside I, Ginsenoside rb1, polyphyllin I, Gambogic acid, Betanin, and Alpha solanine revealed the most lowest MolDock score on N protein which is -285.68 kcal/mol, -172.65kcal/mol, -166.78kcal/mol, -164.94kcal/mol, -160.78kcal/mol, -160.01 kcal/mol, respectively, and Naringin, Salvianolic acid A, Betanin, Paeoniflorin, Ginsenoside rb1, Polyphyllin I, Baicalin shows excellent H-Bonding on N protein which is -23.82 kcal/mol, -23.26 kcal/mol,

-19.80 kcal/mol, -19.32 kcal/mol, -18.50 kcal/mol, -18.28 kcal/mol -18.27 kcal/mol respectively. By comparing entire parameters out of 56, nine compounds were selected for better inhibitory studies, which are alpha solanine binds into the active site of N protein with MolDock score -160.01 kcal/mol and binding site consists of amino acid residues suh as Asp129A, Lys62B, Glu68B, Lys128A, Asp64B, Arg89A, Glu119A, Pro118A, Asn154D, Ile131B, Ile132B, Trp132B, Trp133B, Ala126B, Gly125B, Asn127B, His146D, Ile147D, Trp53D, Asn78D, Asn151A, Asn49A, Asn155D, and Thr50A. It forms hydrogen bonds on Arg89A, Asn127B, Asn155D, Gly125B, Asn78D, Asn49A, and Ile131B amino acid residues with binding energy -17.11 kcal/mol. Baicalin binds into

Table 3: MolDock score, rerank score, interaction, torsions, and hbond energy of the docked compounds with RdRp protein

Name of the inhibitor	MolDock Score, kcal/mol	Rerank Score, kcal/mol	Interaction energy, kcal/mol	Torsion angle, kcal/mol	H-Bond, kcal/mol
Andrographolide	-78.57	-41.04	-96.44	1	-9.08
Arctigenin	-103.39	-58.12	-108.45	7	-6.7
Acanthoic acid	-82.50	-71.41	-94.94	2	-4.8
Alpha pinene	-52.19	-45.34	-65.00	0	0
Astaxanthin	-141.71	-89.26	-154.64	19	-8.09
Alpha solanine	-151.07	-54.29	-192.94	8	-16.47
Baicalin	-83.51	-88.97	-124.67	4	-11.9
Baicalein	-69.25	-64.81	-93.50	1	-10.31
Berberine	-92.68	-63.35	-97.33	2	-4.5
Betanin	-156.39	-124.42	-161.15	8	-17.11
Brusatol	-123.40	-20.73	-116.28	6	-8.82
Betulinic acid	-97.43	-21.71	-109.25	2	-4.05
Carnosol	-65.30	-54.71	-83.73	1	-8.58
Celastrol	-80.87	-66.33	-100.83	1	-10.03
Cairicoside I	-201.55	-38.06	-187.12	24	-7.29
Codonolactone	-74.66	-60.57	-76.58	0	-6.24
Cordycepin	-86.39	-70.50	-90.72	2	-9.16
Costunolide	-97.27	-72.13	-81.55	0	-2.5
Cryptotanshinone	-84.58	-69.34	v84.71	0	-2.0
Curcumin	-139.05	-84.62	-119.16	10	-1.36
Eupatolide	-87.92	-64.60	-75.19	0	-4.9
Gallic acid	-57.60	-32.54	-64.72	1	-13.04
Gambogic acid	-164.15	-96.60	-141.70	11	-10.46
Gedunin	-112.33	-83.70	-113.51	3	-3.6
Genistein	-70.18	-70.39	-98.21	1	-9.02
Gentiopicroside	-89.50	-56.30	-106.50	4	-20.08
Gigantol	-94.75	-72.94	-104.39	5	-7.08
Ginsenoside rb1	-147.04	-112.25	-175.82	11	-19.13
Honokiol	-82.84	-72.23	-90.51	5	-7.00
Helenalin	-98.80	-76.76	-91.48	0	-8.0
Jatrophone	-104.69	-72.77	-82.09	0	-4.74
Ligustrazine	-49.03	-44.10	-58.79	0	-2.5
Luteolin	-99.88	-72.35	-96.76	1	-10.15
Limonene	-53.51	-45.84	-58.64	1	0
Lycopene	-159.57	-77.27	-131.20	29	0
Naringin	-134.00	-94.96	-152.97	6	-18.25
Nimbolide	-138.56	-102.70	-136.41	4	-15.61
Nitidine chloride	-106.14	-76.65	-101.11	2	-5.19
Osthole	-80.39	-55.40	-84.35	4	-5.30
Oxymatrine	-66.42	-62.34	-85.47	0	-2.0
Paeoniflorin	-104.34	-79.60	-120.53	3	-13.28
Paeonol	-61.62	-52.00	-62.32	2	-8.1
Parthenolide	-93.43	-67.25	-77.96	0	-5.0
Plectranthioc acid	-95.26	-77.56	-108.66	1	-4.7
Phoyunnanin E	-125.92	-56.65	-134.39	3	-11.57
Piperlogumine	-107.52	-82.01	-109.57	6	-9.30
Plumbagin	-55.15	-51.81	-65.99	0	-5.5
Polyphyllin I	-140.21	-103.76	-161.01	8	-18.45
Perillyl alcohol	-53.41	-43.87	-64.93	2	-3.0
Resveratrol	-93.44	-64.43	-96.54	2	-6.3
Salvianolic acid A	-150.83	-100.05	-126.55	9	-11.02
Sulforaphane	-65.40	-49.94	-65.87	6	-8.17
Tanshinone IIA	-85.99	-54.53	-91.76	0	-3.5
Triptolide	-97.82	-76.40	-105.53	1	-5.27
Ursolic acid	-110.67	-60.04	-116.76	1	-6.0
Withaferin A	-105.21	-83.42	-118.83	3	-7.40

the active site of N protein with MolDock score -105.37 kcal/mol and binding site consists of amino acid residues such as Trp109B, Lys66B, Lue65B, Asp64B, Ile132B, Arg90B, Gly130B, Ile131B, Trp133B, Asn49A, Asp129B, Lys128B, Asn154B, Asn151D, Asn127B, Asn155B, Arg150D, Trp53B, Thr149D. it forms hydrogen bonds on Asn151D, Thr149D, Asn127B, Asn155D, Asn154D, Gly130B, Asp129B, and Asp64B With binding energy -18.27 kcal/mol. Betanin binds into the active site of N protein with MolDock score -160.78 kcal/mol and binding site consists of amino acid residues such as Trp133B, Lys128B, Asn127B, Trp53D, Asn78D, Asn49A, Thr50A, Ala51A, Arg89A, Ala91A, Arg90A, Thr92A, Lys66B, Glu63B, Pro169B, and Lys170B. It forms hydrogen

bonds on Asn127B, Arg89A, Arg90A, Thr92A, Glu63B, and Lys66B With binding energy -19.8 kcal/mol. Cairicoside I binds into the active site of N protein with MolDock score -285.68 kcal/mol and binding site consists of amino acid residues such as Trp53D, Ile147D, Ile158D, Asn78D, Asn155D, Asn154D, Asn127B, Asn151A, Thr50A, Pro118A, Tyr112A, Ser52A, Ala51A, Gly125B, Ile131B, Trp133B, Ile132B, Arp69B, The67B, Pro68B, Val159C, Tyr110A, Asp64B, Trp109B, Lue65B, The67B. it forms hydrogen bonds on Asn49A, Asn127B, Asn154D, Asn78D, Thr50A, Pro68B, and Lys66B With binding energy -16.8 kcal/mol. Ginsenoside rb1 binds into the active site of N protein with MolDock score -172.65 kcal/mol and binding site consists of

amino acid residues like Arg108A, Arg93A, Thr92A, Ana91A, Tyr110A, Ser52A, Tyr112A, Ala51A, Arg89A, Thr50A, Tro118A, Asn49A, Lys66B, Asn155D, Asn154D, Asn151D, Asn127B, Gly125B, Trp133B, Ile132B, Ile131B, Lys128B, Ala126B, Thr149D, Asn151D, Arg150D, Trp53D. it forms hydrogen bonds on Asn155D, Asn151D, Thr149D, Asn127B, Asn49A, Lys128B, Ile131B, Thr50A, and Tyr112A With binding energy -18.50 kcal/mol. Naringin binds into the active site of N protein with MolDock score -145.45 kcal/mol and binding site consists of amino acid residues such as Arg90A, Arg89A, Thr92A, Tyr110A, Tyr112A, Ser52A, Pro118A, Glu63B, Ala51A, Thr50A, Asn49A, Lys66B, Trp153B, Ile132B, and Phe67B. It forms hydrogen bonds on Trp133B, Thr50A, Thr92A, Glu63B, Lys66B, Arg89A, Ser52A, and Tyr122A With binding energy -23.82 kcal/mol. Paeoniflorin binds into the active site of N protein with MolDock score -114.76 kcal/mol and binding site consists of amino acid residues such as Trp53D, Asn155D, Thr50A, Asn49A, Asn127B, Asp129B, Lys128B, Ala126B, Gly125B, Gly130B, Tle131B, Trp133B, and Lys66B. It forms hydrogen bonds on Asn127B, Thr50A, Ile131B, Ala126B, Lys128B, and ASN 49A With binding energy -19.32 kcal/mol. Polyphyllin I binds into the active site of N protein with MolDock score -166.78 kcal/mol and binding site consists of amino acid residues such as Thr50D, Thr149D, Gly148D, Tle147D, Trp53D, Asn51D, Asn155D, Asn154D, Asp129D, lys128D, Asn127B, gly130B,

Thr50A, Ala51A, Lys66B, Pro152A, Asn49A, Ile132B, Trp133B, Phe67B, Pro68B, Arg69B, and Gln161C. It forms hydrogen bonds on Arg69B, Trp133B, and Phe67B With binding energy -18.28 kcal/mol. Salvianolic acid A binds into the active site of N protein with MolDock score -143.71 kcal/mol and binding site consists of amino acid residues such as Tyr112A, Arg89A, Pro118A, Thr50A, Asn154D, Asn155D, Trp53D, Asn127B, Asn49A, Lys66B, Asp64B, Lys128B, Asp129B, Ile131B, Gly130B, and Arg90B. It forms hydrogen bonds on Asn49A, Thr50A, Arg89A, Gly130B, Asp129B, Ile131B, Lys128B, Asn127B, and Tyr112A With binding energy -23.26 kcal/mol. Fig. 4 illustrates hydrogen bonding with amino acids of coronavirus protein N.

RdRp protein results

Alpha solanine, Betanin, Cairicoside I, Gambogic acid, Ginsenoside rb I, Salvianolic acid A, and Lycopene revealed the most lowest MolDock score on RdRp protein which is -151.07 kcal/mol, -156.39 kcal/mol, -201.55 kcal/mol, -164.15 kcal/mol, -147.04 kcal/mol, -150.83 kcal/mol, -159.57 kcal/mol, respectively and Polyphyllin I, Naringin, Ginsenoside rb 1, Gentiopicroside, Betanin, and Alpha solanine shows excellent H-Bonding on RdRp protein which is -18.45 kcal/mol, -18.25 kcal/mol, -19.13 kcal/mol, -20.08 kcal/mol, -17.11 kcal/mol and -16.47 kcal/mol, respectively. By comparing entire parameters out of 56 compounds, seven compounds were selected for better inhibitory studies, which are Alpha solanine binds into the active site of RdRp protein with MolDock score -151.07 kcal/mol and binding site consists of amino acid residues such as Val557A, Ser682A, Lys545A, Thr556A, Arg555A, Asp62A, Asp760A, Asp452A, Arg624A, Cys622A, Arg553A, Tyr455A, Lys551A, Lys621A, Pro620A, Lys798A, Tyr619A, and Ala554A. It forms hydrogen bonds on Lys621A, Cys622A, Tyr619A, Asp760A, and Asp623A amino acid residues with binding energy -16.47 kcal/mol. Betanin binds into the active site of RdRp protein with MolDock score -156.39 kcal/mol and binding site consists of amino acid residues such as Lys545A, Lys500A, Val557A, Gly683A, Thr556A, Ser685A, Ala554A, Asp452A, Arg553A, Tyr455A, Lys621A, Arg624A, Asp623A, Asn691A, Ser759A, Thr687A, Ala688A, Ser759A, and Lys545A. It forms hydrogen bonds on Ala688A, Thr556A, Arg624A, Arg555A, Asp452A, Lys621A, Asp623A, and Lys545A amino acid residues with binding energy -17.11 kcal/mol.

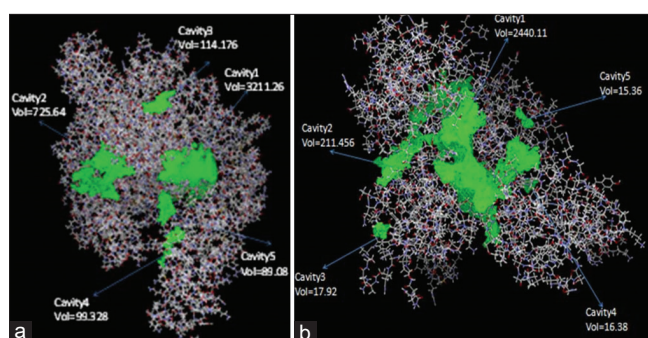


Fig. 3: Active sites of proteins (a) Nucleocapsid (b) RdRp

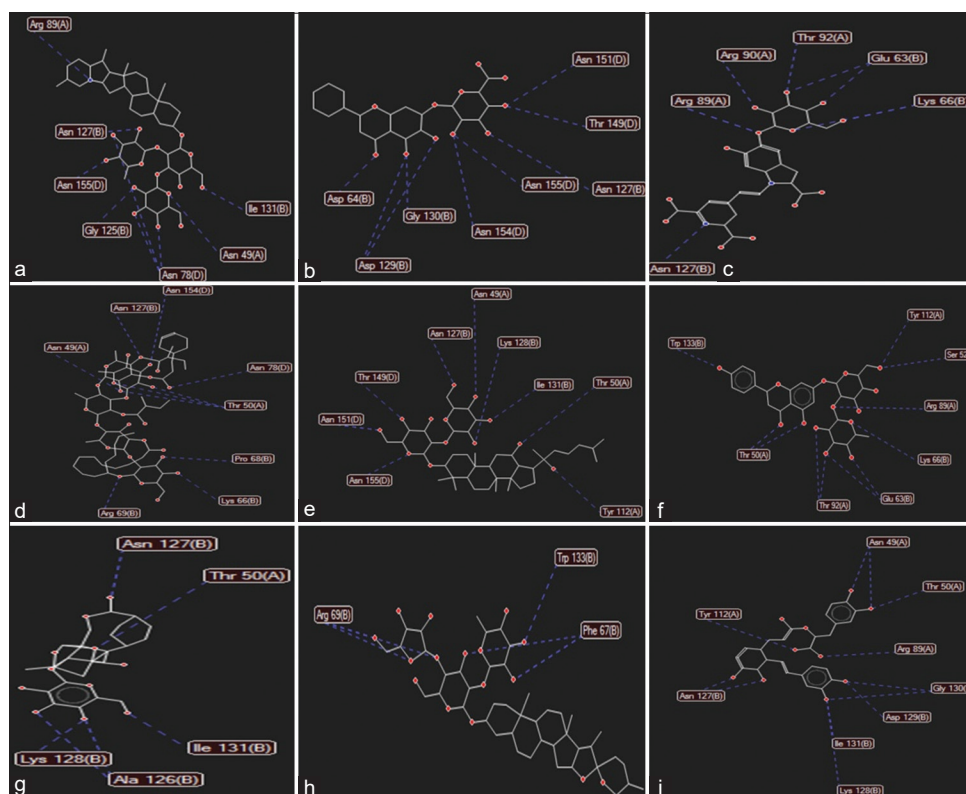


Fig. 4: Hydrogen Bond interactions of ligands with N protein (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g) Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

Gentiopicroside binds into the active site of RdRp protein with MolDock score -89.50 kcal/mol and binding site consists of amino acid residues like Tyr619A, Cys622A, Pro620A, Lys621A, Tyr458A, Asp623A, Arg624A, Tyr455A, Tyr458A, Arg553A, Asp452A, Ala554A, Arg555A and Tyr556A. It forms hydrogen bonds on Tyr619A, Cys622A, Asp623A, Lys621A, Arg555A, and Ala554A amino acid residues with binding energy -20.08 kcal/mol. Ginsenoside rb1 binds into the active site of RdRp protein with MolDock score -147.04 kcal/mol and binding site consists of amino acid residues such as Ala688A, Thr687A, Ser759A,

Asn691A, Asp760A, Ser682A, Asp623A, Tyr619A, Cys622A, Arg555A, Thr556A, Asp618A, Pro620A, Lys621A, Asp452A, Ala554A, Tyr455A, Lys798A and Arg553A. It forms hydrogen bonds on Lys798A, Asp618A, Arg555A, Asp452A, Ala554A, Lys621A, Arg553A, Asp760A, and Ser759A amino acid residues with binding energy -19.13 kcal/mol. Naringin binds into the active site of RdRp protein with MolDock score -134.00 kcal/mol and binding site consists of amino acid residues such as Ala550A, Ser549A, Lys551A, Ala554A, Arg555A, Arg553A, Asp452A, Thr556A, Tyr455A, Arg624A, Lys621A, Ser682A, Asp623A, Cys622A,

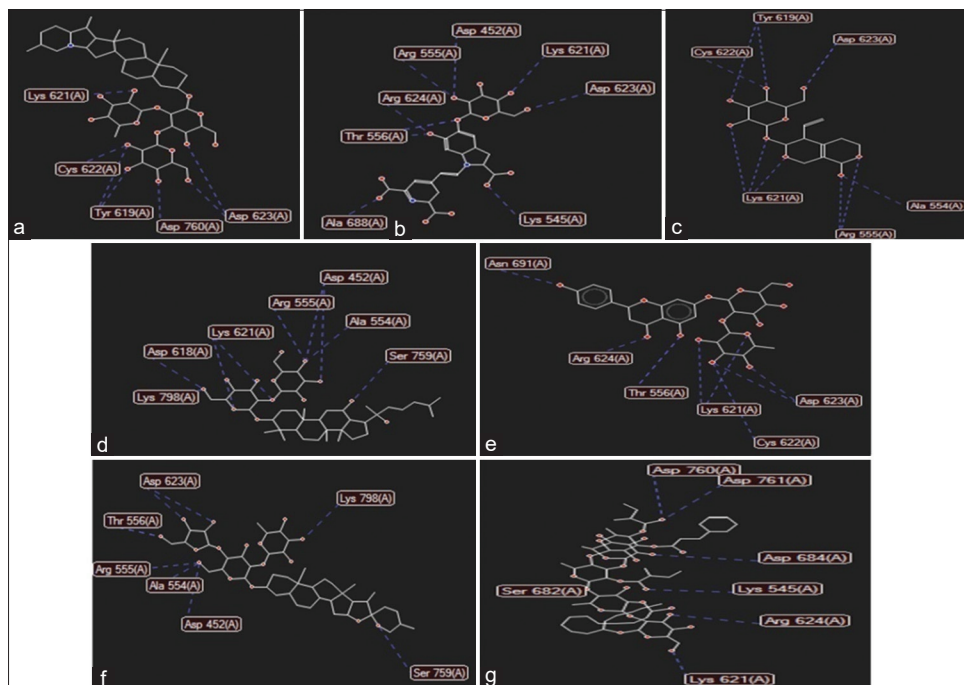


Fig. 5: Hydrogen bond interactions of ligands with RdRp protein (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g) Cairicoside I

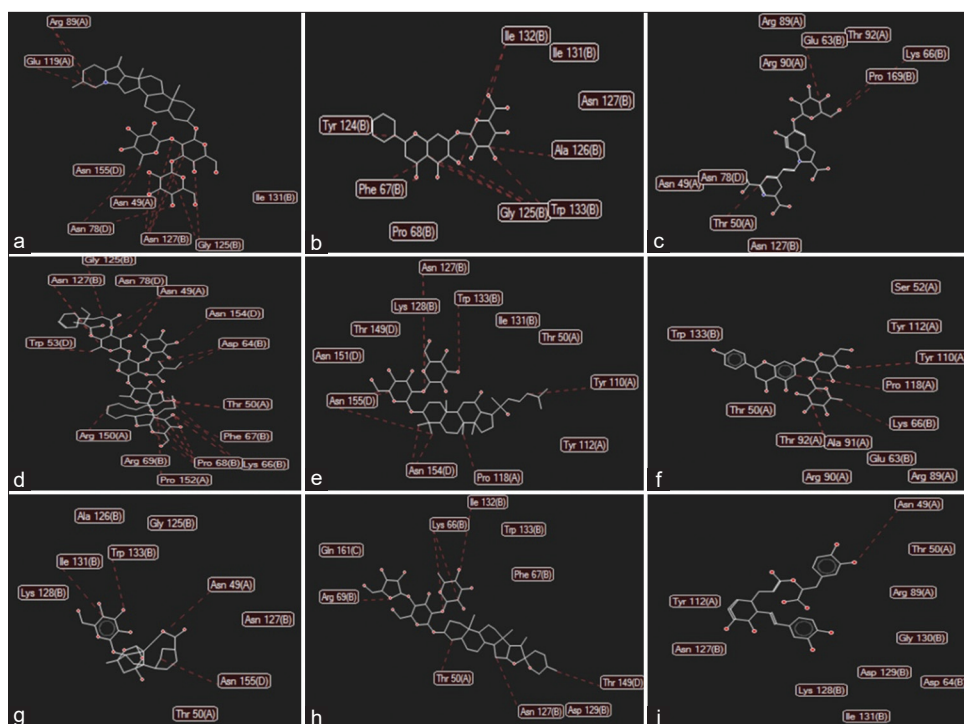


Fig. 6: Steric interactions of ligands with N protein (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g) Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

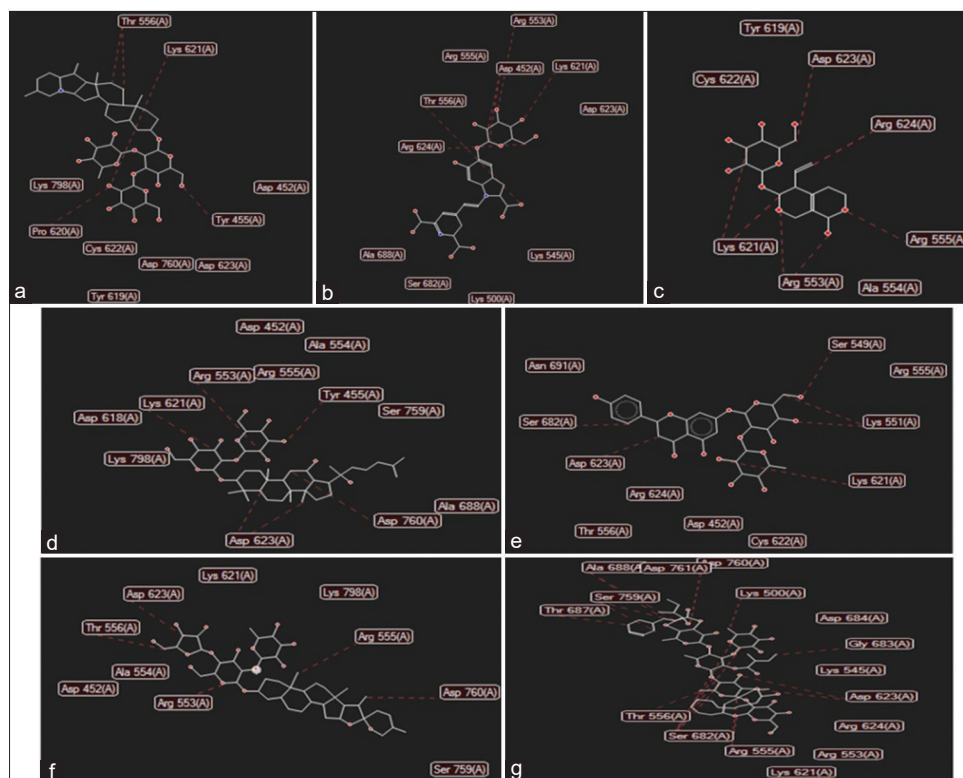


Fig. 7: Steric interactions of ligands with RdRp protein (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g) Cairicoside I

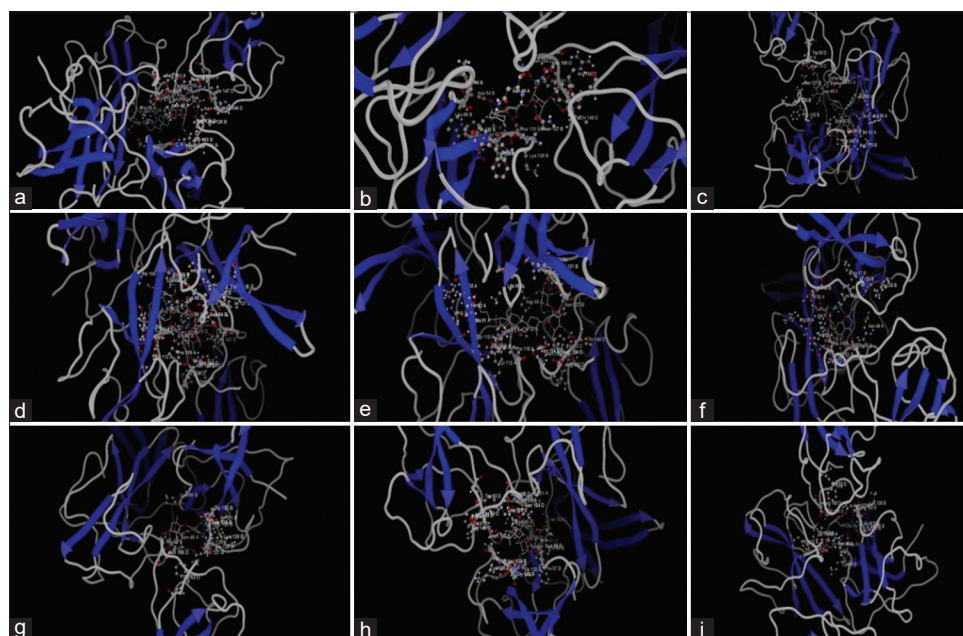


Fig. 8: Ligand docked against the crystal protein structure of COVID-19 Nucleocapsid (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g) Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

Thr687A, and Asn691A. It forms hydrogen bonds on ASN 691A, Arg624 A, Thr556A, Lys621A, Cys622A, and Asp623A amino acid residues with binding energy -18.25 kcal/mol. Polyphyllin I binds into the active site of RdRp protein with MolDock score -140.21 kcal/mol and binding site consists of amino acid residues such as Thr556A, Ala554A, Asp452A, Ser682A, Thr687A, Arg553A, Tyr455A, Arg624A, Asp623A, Ser759A, Leu758A, Asp760A, Cys622A, Lys798A, Pro620A, Asp648A, Tyr619A, Lys798A, and Lys551A. It forms hydrogen bonds on Asp623A, Thr556A,

Arg555A, Ala554A, Asp452A, Ser759A, and Lys798A amino acid residues with binding energy -18.45 kcal/mol. Cairicoside I binds into the active site of RdRp protein with MolDock score -201.55 kcal/mol and binding site consists of amino acid residues like Lys500A, Ala685A, Asp684A, Gly683A, Ala558A, Val557A, Ser682A, Thr687A, Ala688A, Ser759A, Leu758A, Asp761A, Cys813A, Ser814A, Asp760A, Tyr619A, Lys545A, Asp623A, Cys622A, Thr556A, Arg624A, Asp452A, Tyr455A, Arg553A, Lys621A, and Pro620A. It forms hydrogen bonds on Asp684A,

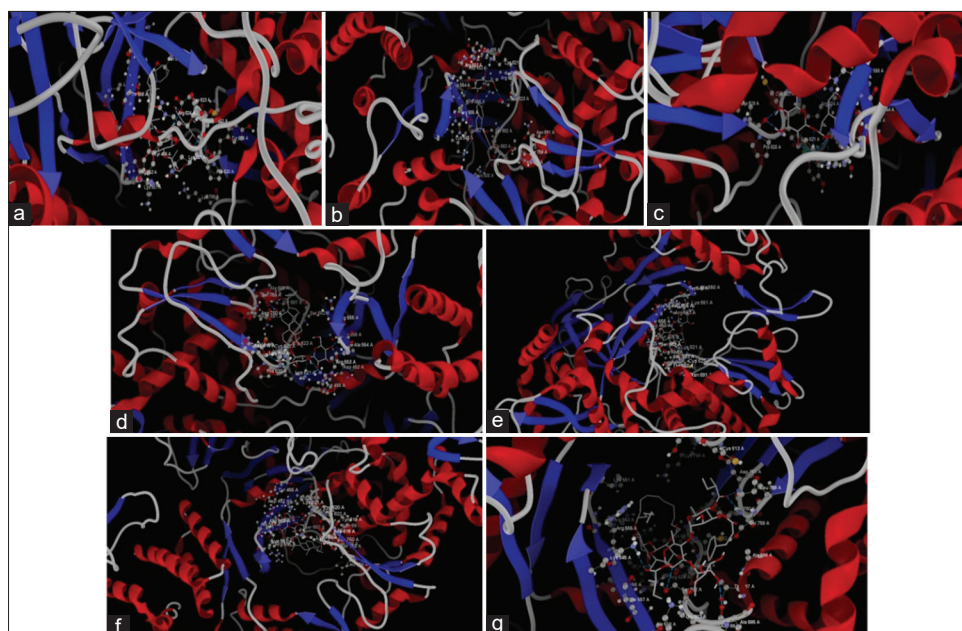


Fig. 9: Ligand docked against the crystal protein structure of COVID-19 RdRp (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g) Cairicoside I

Table 4: Amino acid residues around active site and docked against N protein

Name of the Inhibitor	Amino acid residues around active site	Ligand binding amino acids
Alpha solanine	Asp129A, Lys62B, Glu68B, Lys128A, Asp64B, Arg89A, Glu119A, Pro118A, Asn154D, Ile131B, Ile132B, Trp132B, Trp133B, Ala126B, Gly125B, Asn127B, His146D, Ile147D, Trp53D, Asn78D, Asn151A, Asn49A, Asn155D, Thr50A, Trp109B, Lys66B, Lue65B, Asp64B, Ile132B, Arg90B, Gly130B, Ile131B, Trp133B, Asn49A, Asp129B, Lys128B, Asn154B, Asn151D, Asn127B, Asn155B, Arg150D, Trp53B, Thr149D	Arg89A, Asn127B, Asn155D, Gly125B, Asn78D, Asn49A, Ile131B
Baicalin	Trp133B, Lys128B, Asn127B, Trp53D, Asn78D, Asn49A, Thr50A, Ala51A, Arg89A, Ala91A, Arg90A, Thr92A, Lys66B, Glu63B, Pro169B, Lys170B, Trp53D, Ile147D, Ile158D, Asn78D, Asn155D, Asn154D, Asn127B, Asn151A, Thr50A, Pro118A, Tyr112A, Ser52A, Ala51A, Gly125B, Ile131B, Trp133B, Ile132B, Arp69B, The67B, Pro68B, Val159C, Tyr110A, Asp64B, Trp109B, Lue65B, The67B	Asn151D, Thr149D, Asn127B, Asn155D, Asn154D, Gly130B, Asp129B, Asp64B
Betanin	Trp133B, Lys128B, Asn127B, Trp53D, Asn78D, Asn49A, Thr50A, Ala51A, Arg89A, Ala91A, Arg90A, Thr92A, Lys66B, Glu63B, Pro169B, Lys170B, Trp53D, Ile147D, Ile158D, Asn78D, Asn155D, Asn154D, Asn127B, Asn151A, Thr50A, Pro118A, Tyr112A, Ser52A, Ala51A, Gly125B, Ile131B, Trp133B, Ile132B, Arp69B, The67B, Pro68B, Val159C, Tyr110A, Asp64B, Trp109B, Lue65B, The67B	Asn127B, Arg89A, Arg90A, Thr92A, Glu63B, Lys66B
Cairicoside I	Trp53D, Ile147D, Ile158D, Asn78D, Asn155D, Asn154D, Asn127B, Asn151A, Thr50A, Pro118A, Tyr112A, Ser52A, Ala51A, Gly125B, Ile131B, Trp133B, Ile132B, Arp69B, The67B, Pro68B, Val159C, Tyr110A, Asp64B, Trp109B, Lue65B, The67B	Asn49A, Asn127B, Asn154D, Asn78D, Thr50A, Pro68B, Lys66B
Ginsenoside rb1	Arg108A, Arg93A, Thr92A, Ana91A, Tyr110A, Ser52A, Tyr112A, Ala51A, Arg89A, Thr50A, Tro118A, Asn49A, Lys66B, Asn155D, Asn154D, Asn151D, Asn127B, Gly125B, Trp133B, Ile132B, Ile131B, Lys128B, Ala126B, Thr149D, Asn151D, Arg150D, Trp53D	Asn155D, Asn151D, Thr149D, Asn127B, Asn49A, Lys128B, Ile131B, Thr50A, Tyr112A
Naringin	Arg90A, Arg89A, Thr92A, Tyr110A, Tyr112A, Ser52A, Pro118A, Glu63B, Ala51A, Thr50A, Asn49A, Lys66B, Trp153B, Ile132B, Phe67B	Trp133B, Thr50A, Thr92A, Glu63B, Lys66B, Arg89A, Ser52A, Tyr122A
Paeoniflorin	Trp53D, Asn155D, Thr50A, Asn49A, Asn127B, Asp129B, Lys128B, Ala126B, Gly125B, Gly130B, Tle131B, Trp133B, Lys66B	Asn127B, Thr50A, Ile131B, Ala126B, Lys128B, ASN 49A
Polyphyllin I	Thr50D, Thr149D, Gly148D, Tle147D, Trp53D, Asn51D, Asn155D, Asn154D, Asp129D, lys128D, Asn127B, gly130B, Thr50A, Ala51A, Lys66B, Pro152A, Asn49A, Ile132B, Trp133B, Phe67B, Pro68B, Arg69B, Gln161C	Arg69B, Trp133B, Phe67B
Salvianolic acid A	Tyr112A, Arg89A, Pro118A, Thr50A, Asn154D, Asn155D, Trp53D, Asn127B, Asn49A, Lys66B, Asp64B, Lys128B, Asp129B, Ile131B, Gly130B, Arg90B	Asn49A, Thr50A, Arg89A, Gly130B, Asp129B, Ile131B, Lys128B, Asn127B, Tyr112A

Asp761A, Asp760A, Lys545A, Lys621A, Arg624A, and Ser628A amino acid residues with binding energy -7.29 kcal/mol. Fig. 5 illustrates hydrogen bonding with amino acids of coronavirus protein RdRp.

Figs. 6 and 7 shows, steric interactions with residues of COVID-19 proteins N and RdRp, respectively.

Figs. 8 and 9 represents ligand docked against the crystal protein structures of COVID-19 Nucleocapsid and RdRp Proteins, respectively.

The amino acid residues around active site and Docked against N and RdRp protein are illustrated in Tables 4 and 5 with structures at Figs. 10 and 11, respectively.

ADMET results

In the docking study, more negative BE corresponded to the strong binding of selected compounds to the target proteins. It is a fact that weaker binding will ultimately have a rapid dissociation rate [34]. In this study, the selected compounds exhibited lower binding energy

Table 5: Amino acid residues around active site and docked against RdRp protein

Name of the Inhibitor	Amino acid residues around active site	Ligand-binding amino acids
Alpha solanine	Val557A, Ser682A, Lys545A, Thr556A, Arg555A, Asp62A, Asp760A, Asp452A, Arg624A, Cys622A, Arg553A, Tyr455A, Lys551A, Lys621A, Pro620A, Lys798A, Tyr619A, Ala554A	Lys621A, Cys622A, Tyr619A, Asp760A, Asp623A
Betanin	Lys545A, Lys500A, Val557A, Gly683A, Thr556A, Ser685A, Ala554A, Asp452A, Arg553A, Tyr455A, Lys621A, Arg624A, Asp623A, Asn691A, Ser759A, Thr687A, Ala688A, Ser759A, Lys545A	Ala688A, Thr556A, Arg624A, Arg555A, Asp452A, Lys621A, Asp623A, Lys545A
Gentiopicroside	Tyr619A, Cys622A, Pro620A, Lys621A, Tyr458A, Asp623A, Arg624A, Tyr455A, Tyr458A, Arg553A, Asp452A, Ala554A, Arg555A, Tyr556A	Tyr619A, Cys622A, Asp623A, Lys621A, Arg555A, Ala554A
Ginsenoside rb1	Ala688A, Thr687A, Ser759A, Asn691A, Asp760A, Ser682A, Asp623A, Tyr619A, Cys622A, Arg555A, Thr556A, Asp618A, Pro620A, Lys621A, Asp452A, Ala554A, Tyr455A, Lys798A, Arg553A	Lys798A, Asp618A, Arg555A, Asp452A, Ala554A, Lys621A, Arg553A, Asp760A, Ser759A
Naringin	Ala550A, Ser549A, Lys551A, Ala554A, Arg555A, Arg553A, Asp452A, Thr556A, Tyr455A, Arg624A, Lys621A, Ser682A, Asp623A, Cys622A, Thr687A, Asn691A	ASN 691A, Arg624 A, Thr556A, Lys621A, Cys622A, Asp623A
Polyphyllin I	Thr556A, Ala554A, Asp452A, Ser682A, Thr687A, Arg553A, Tyr455A, Arg624A, Asp623A, Ser759A, Leu758A, Asp760A, Cys622A, Lys798A, Pro620A, Asp648A, Tyr619A, Lys798A, Lys551A	Asp623A, Thr556A, Arg555A, Ala554A, Asp452A, Ser759A, Lys798A
Cairicoside I	Lys500A, Ala685A, Asp684A, Gly683A, Ala558A, Val557A, Ser682A, Thr687A, Ala688A, Ser759A, Leu758A, Asp761A, Cys813A, Ser814A, Asp760A, Tyr619A, Lys545A, Asp623A, Cys622A, Thr556A, Arg624A, Asp452A, Tyr455A, Arg553A, Lys621A, Pro620A,	Asp684A, Asp761A, Asp760A, Lys545A, Lys621A, Arg624A, Ser628A

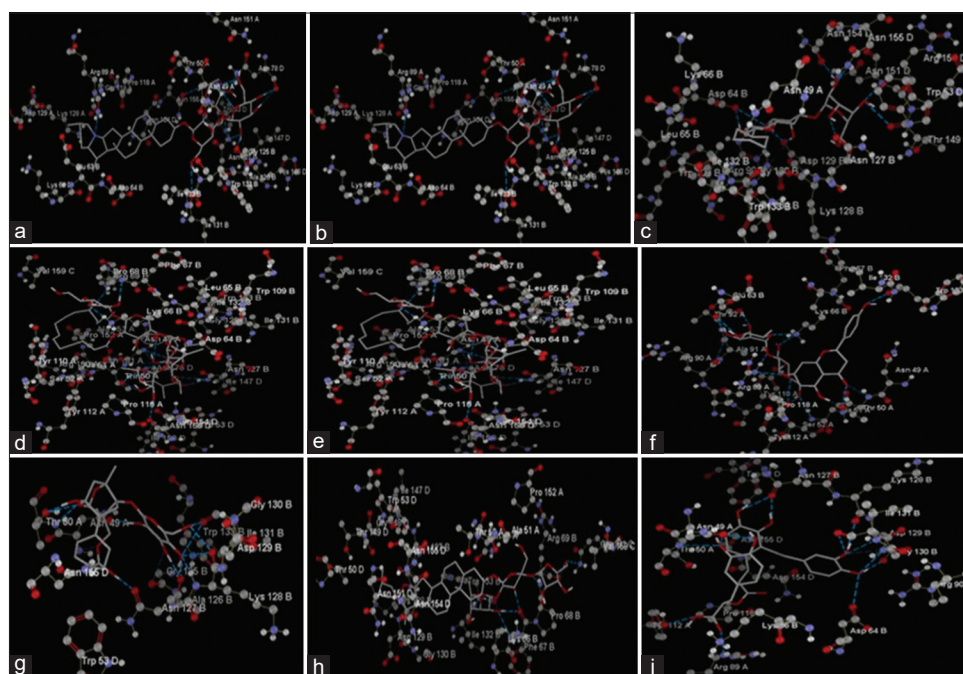


Fig. 10: The amino acid residues around active site and docked against N protein (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g) Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

which means strong binding with the N and RdRp proteins, suggesting that these compounds could be utilized as an inhibitor of the RdRp and Nucleocapsid proteins to combat COVID-19. According to the Rule of Five, a molecule might be no longer orally active if it violates or is greater of the four rules. However, the rules are not suitable for natural products because they are complicated and they have been deduced from relatively simple small molecules [35]. However, we have filtered the compounds based on the binding efficacy on both proteins of COVID-19. The results of the pkCSM, SwissADME tool, and molinspiration cheminformatics tool show the top-scored compound's pharmacokinetic properties (N and RdRp protein inhibitory compounds) (Table 6).

DISCUSSION

Developing drugs to decrease the symptoms of SARS-Cov-2 infection is the main task now in worldwide. Nowadays, new research is published on SARS-CoV-2 and proposes a new drug among drug repurposing medicines for the possible treatment of COVID-19. In our study, we targeted two essential proteins (N and RdRp) for identifying small molecules and natural plant compounds as therapeutic agents, using *in silico* analysis. The three-dimensional proteins structures of N and RdRp were retrieved from RCSB Protein Data Bank. According to our results, some natural plant compounds were detected against N protein

including Alpha solanine, Baicalin, Betanin, Cairicoside I, Ginsenoside rb1, Naringin, Paeoniflorin, Polyphyllin I, and Salvianolic acid A; among them, Alpha solanine, Betanin, Cairicoside I, Ginsenoside rb1, Naringin, and Polyphyllin I were identified to have interaction with RdRp too. This study has identified 56 natural plant compounds, some small molecules with higher binding affinity and interaction with N and RdRp proteins active pocket residues. The result of this and other similar *in silico* studies on natural plant compounds and small molecules is promising for further *in vitro*, *in vivo*, and clinical trial studies of SARS-CoV-2 treatment.

CONCLUSION

The present study involves the capability natural plant compounds and small molecules from medicinal plants in forming stable interactions with N and RdRp proteins of the SARS-CoV-2. Developing of herbal drugs with minimum side effects is a better opportunity to explore the medicinal and other biological properties of natural plant compounds. To identify its inhibitory effects and usefulness, it is mandatory to focus on visualization and identification of unused natural plant compounds in a particular disease over the world. Then, it is far emphasized on extraction, isolation, and characterization of natural plant compounds and its small molecules is a gift of nature in a rational and scientific way. In this study, two main COVID-19 viral proteins, N and RdRp were selected to dock against natural plant compounds and small molecules. MolDock score, Rerank score, H-Bonding interactions, and free energy calculations suggest favorable binding of ligands such as Alpha solanine, Betanin, Cairicoside I, Ginsenoside rb1, Naringin, and polyphyllin. These compounds shows harmonious inhibitory effect on both N and RdRp proteins and several amino acid residues were observed to participate in such interactions. Overall, the findings in the article proposes the natural plant compounds and small molecules as stable N and RdRp protein catalytic active site binders, synergistic with the experimentally known drug-N and RdRp interactions. However, many of them to be further investigate experimentally against COVID-19 *in vitro* and *in vivo*.

DATA AVAILABILITY

Data supporting the productivity of this investigation are available from the corresponding author on request.

ACKNOWLEDGMENT

Thankful CSIR-UGC for providing Fellowship to PAVAN KUMAR POLEBOYINA, reference Nos. 1621 and thankful to the staff of the Department of Genetics and Biotechnology, Osmania University for encouragement and support in Bioinformatics/software facilities.

AUTHOR'S CONTRIBUTIONS

Smita C Pawar and Pavan Kumar Poleboyina conceptualized, designed interpreted data and edited manuscript. Pavan Kumar Poleboyina conducted the study. All authors have approved the manuscript in the current form.

CONFLICTS OF INTERESTS

The authors declare that there are no conflicts of interest. No additional benefits will be received from a third party directly or indirectly by the authors.

FUNDING

The authors did not receive support from any organization for the submitted work.

ETHICAL APPROVAL

Not applicable.

CONSENT TO PARTICIPATE

Not applicable.

CONSENT TO PUBLISH

Not applicable.

REFERENCES

- Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-COV-2; coronavirus Disease-19). *Clin Exp Pediatr* 2020;63:119-24. doi: 10.3345/cep.2020.00493, PMID 32252141, PMC7170784
- Gajjar ND, Dhameliya TM, Shah GB. In search of RdRp and Mpro inhibitors against SARS CoV-2: Molecular docking, molecular dynamic simulations and ADMET analysis. *J Mol Struct* 2021;1239:130488. doi: 10.1016/j.molstruc.2021.130488, PMID 33903778
- Mandal M, Chowdhury SK, Khan AA, Baildya N, Dutta T, Misra D, et al. Inhibitory efficacy of RNA virus drugs against SARS-CoV-2 proteins: An extensive study. *J Mol Struct* 2021;1234:130152. doi: 10.1016/j.molstruc.2021.130152, PMID 33678903
- COVID-19 Infection: Origin, Transmission, and Characteristics of Human Coronaviruses. Elsevier Enhanced Reader. Available from: <https://reader.elsevier.com/reader/sd/pii/S2090123220300540?token=6d8bb062f469ed88128e0bea7af62a17089cb020930ab21f56295339e534532e592a1609bc5109784E9655941e42ac33&originregion=eu-west-1&origincreation=20210606200221> [Last accessed on 2021 Jun 07].
- Mahrosh HS, Mustafa G. An *in silico* approach to target RNA-dependent RNA polymerase of COVID-19 with naturally occurring phytochemicals. *Environ Dev Sustain* 2021;23:16674-87. doi: 10.1007/s10668-021-01373-5, PMID 33841038
- Zhu W, Chen CZ, Gorshkov K, Xu M, Lo DC, Zheng W. RNA-dependent RNA polymerase as a target for COVID-19 drug discovery. *SLAS Discov* 2020;25:1141-51. doi: 10.1177/2472555220942123, PMID 32660307
- Biswal M, Diggs S, Xu D, Khudaverdyan N, Lu J, Fang J, et al. Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Res* 2021;49:5956-66. doi: 10.1093/nar/gkab370
- Ramanathan A, Pullara F, Kumar TD, Shaikh N, Kumar US, Priya Doss GC, et al. Structure-based virtual screening to identify novel potential compound as an alternative to remdesivir to overcome the RdRp protein mutations in SARS-CoV-2. *Front Mole Biosci* 2021;8:645216. doi: 10.3389/fmolb.2021.645216
- Machitani M, Yasukawa M, Nakashima J, Furuichi Y, Masutomi K. RNA-dependent RNA polymerase, RdRP, a promising therapeutic target for cancer and potentially COVID-19. *Cancer Sci* 2020;111:3976-84. doi: 10.1111/cas.14618, PMID 32805774
- Jiang Y, Yin W, Xu HE. Since January 2020 Elsevier has Created a COVID-19 Resource Centre with Free Information in English and Mandarin on the Novel Coronavirus COVID-19. The COVID-19 Resource Centre is Hosted on Elsevier Connect, the Company's Public News and Information. Netherlands: Elsevier; 2020.
- Aftab SO, Ghouri MZ, Masood MU, Haider Z, Khan Z, Ahmad A, et al. Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. *J Transl Med* 2020;18:275. doi: 10.1186/s12967-020-02439-0, PMID 32635935
- Tugaeva KV, Hawkins DE, Smith JL, Bayfield OW, Ker DS, Sysoev AA, et al. The mechanism of SARS-CoV-2 nucleocapsid protein recognition by the human 14-3-3 proteins. *J Mol Biol* 2021;433:166875. doi: 10.1016/j.jmb.2021.166875, PMID 33556408
- Batra M, Tian R, Zhang C, Clarence E, Sacher CS, Miranda JN, et al. Role of IgG against N-protein of SARS-CoV2 in COVID19 clinical outcomes. *Sci Rep* 2021;11:3455. doi: 10.1038/s41598-021-83108-0, PMID 33568776
- Gao T, Gao Y, Liu X, Nie Z, Sun H, Lin K, et al. Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. *BMC Microbiol* 2021;21:58. doi: 10.1186/s12866-021-02107-3, PMID 33618668
- Tatar G, Ozyurt E, Turhan K. Computational drug repurposing study of the RNA binding domain of SARS-CoV-2 nucleocapsid protein with antiviral agents. *Biotechnol Prog* 2021;37:e3110. doi: 10.1002/btpr.3110, PMID 33314794
- Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr Comput Aid Drug Des* 2011;7:146-57. doi: 10.2174/157340911795677602, PMID 21534921
- RCSB PDB: Homepage. Available from: <https://www.rcsb.org> [Last accessed on 2021 Jun 24].

18. Available from: <https://pubchem.ncbi.nlm.nih.gov> [Last accessed on 2021 Jun 24].
19. Daoud I, Melkemi N, Salah T, Ghalem S. Combined QSAR, molecular docking and molecular dynamics study on new acetylcholinesterase and butyrylcholinesterase inhibitors. *Comput Biol Chem* 2018;74:304-26. doi: 10.1016/j.compbiolchem.2018.03.021
20. Bitencourt-Ferreira G, de Azevedo WF Jr. Molegro virtual docker for docking. *Methods Mol Biol* 2019;2053:149-67. doi: 10.1007/978-1-4939-9752-7_10, PMID 31452104
21. Poleboyina SM, Poleboyina PK, Pawar SC, Guntuku G. Homology modeling, screening, and identification of potential FOXO6 inhibitors curtail gastric cancer progression: An in silico drug repurposing approach. *Appl Biochem Biotechnol* 2023;1-30. doi:10.1007/S12010-023-04490-1/TABLES/4
22. Sadeghi F, Afkhami A, Madrakian T, Ghavami R. Computational study to select the capable anthracycline derivatives through an overview of drug structure-specificity and cancer cell line-specificity. *Chem Pap* 2021;75:523-38. doi: 10.1007/s11696-020-01321-z
23. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712-25. doi: 10.1002/prot.21123, PMID 16981200
24. Swiss PD. Viewer - Home. Available from: <https://spdbv.vital-it.ch> [Last accessed on 2021 Jul 08].
25. Bouchagra S, Benamia F, Djeghaba Z. Docking studies of (-)-epigallocatechin-3-gallate: A potential noncompetitive pancreatic lipase inhibitor. *Res J Pharm Biol Chem Sci* 2016;7:2493-505.
26. Singh SP, Deb CR, Ahmed SU, Saratchandra Y, Konwar BK. Molecular docking simulation analysis of the interaction of dietary flavonols with heat shock protein 90. *J Biomed Res* 2015;30:67-74. doi: 10.7555/JBR.29.20130158, PMID 26423731, PMC4726836
27. PKCSM. Available from: <https://biosig.unimelb.edu.au/pkcsm/prediction> [Last accessed on 2021 Jun 07].
28. Swiss ADME. Available from: <https://www.swissadme.ch> [Last accessed on 2021 Jul 03].
29. Molinspiration. Cheminformatics. Available from: <https://www.molinspiration.com> [Last accessed on 2021 Jul 03].
30. ERRAT - Help - DOE-MBI Structure Lab UCLA. Available from: <https://servicesn.mbi.ucla.edu/errata/help> [Last accessed on 2021 Jun 07].
31. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283-91. doi: 10.1107/S0021889892009944
32. PDBsum Entry: 6m3m. Available from: <https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/getpage.pl> [Last accessed on 2021 Jun 07].
33. Thomsen R, Christensen MH. MolDock: A new technique for high-accuracy molecular docking. *J Med Chem* 2006;49:3315-21. doi: 10.1021/jm051197e, PMID 16722650
34. Copeland RA. Conformational adaptation in drug-target interactions and residence time. *Future Med Chem* 2011;3:1491-501. doi: 10.4155/fmc.11.112, PMID 21882942
35. Guan L, Yang H, Cai Y, Sun L, Di P, Li W, *et al.* ADMET-score-a comprehensive scoring function for evaluation of chemical drug-likeness. *Medchemcomm* 2019;10:148-57. doi: 10.1039/c8md00472b, PMID 30774861