Identifying Therapeutic Compounds Targeting RNA-Dependent-RNA-Polymerase of Sars-Cov-2

Muhammad Sohaib Roomi[1], Muhammad Khalid Mahmood[2], Yaser Daanial Khan[1]

F2018279051@umt.edu.pk, khalid.math@pu.edu.pk, Yaser.khan@umt.edu.pk

[1]Department of Computer Science, School of Systems and Technology, University of Management and Technology, Lahore, Pakistan

[2] Department of Mathematics, Faculty of Science, University of Punjab, Lahore, Pakistan

Abstract:

COVID-19 has emerged as the biggest threat of this century for mankind. This contagious disease was initially transmitted from animals (probably bats or pangolins) to humans and later it spread across the globe through human to human transmission. Scientists rushed to understand the structure and mechanism of the virus so that antiviral drugs or vaccines to control this disease can be developed. A key to stop the progression of the disease is to inhibit the replication mechanism of Sars-Cov-2. RNA-dependent-RNA polymerase protein also called RdRp protein is the engine of Sars-Cov-2 that replicates the virus using viral RNA when it gains entry into the human cell. The replication of the virus is the main process that acts as a catalyst in the progression of disease. RdRp is the main target of researchers working to develop antiviral drugs to inhibit the mechanism of the virus. Numerous drugs proposed for the treatment of COVID-19 such as Camostat Mesylate, Remdesivir, Famotidine, Hesperidin, etc. are under trial to analyze the aftermath of their medicinal use. Nature is enriched with compounds that have antiviral activities and can potentially play a pivotal role to inhibit this virus. This study focuses on the phytochemicals that have the potential to exhibit antiviral activities. A large number of compounds were screened and a cohort of most suitable ones are suggested via in-silico evidence that can inhibit the functionality of RdRp and hence the replication of Sars-Cov-2.

Introduction:

Coronavirus is a vast family of viruses. 7 known coronaviruses can enter into human cells. The first case of coronavirus in humans was reported in 1965, which had mild symptoms of flu and fever. Coronaviruses are significant pathogens for both humans and animals. These are medium-sized but can have a very large RNA genome. They can bind with the host cells and mutate when they transfer from one species to another. Subsequent mutation can lead to its transmission into humans. They can bind themselves to the respiratory tract causing an infection. The symptoms of coronavirus infection are: illness, flu, mild fever, diarrhea, and difficulty in breathing. Severe Acute Respiratory Syndrome named as Sars is an infectious disease caused by Sars-CoV that spreads swiftly and causes illness and flu at the initial stage. Sars-Cov-2 is just like Sars-Cov in its working and structure but more dangerous in terms of severity. It spreads from person to person through coughing or sneezing droplets and physical contact. In 2019, Sars-CoV-2 emerged from Wuhan, China, and took the world by storm. The world was not prepared for it and as a result, both humans and the world economy have suffered very adversely. At this point, over 9.5 million infected cases have been reported and the death toll has reached over 480,000. The onset of COVID-19

has led to a drastic reduction in social and economic activities throughout the world. At this point, doctors and researchers from every country are trying hard to devise an effective strategy for controlling the disease. To propose an effective and long-lasting solution, understanding of the structure of the virus and its action is very important. Recent studies have been able to develop an understanding of the mechanism and structure of the virus through 3D modeling.

Structure of SARS-COV-2:

To get an insight into the action of Sars-Cov-2 viruses and discover suitable antiviral compounds, it is very important to elucidate the proteomic buildup of Sars-CoV-2. Its proteomic data encompasses different proteins that form its makeup such as the Spike and RNA dependent RNA polymerase (RdRp) proteins. The entire Genome RNA structure inside the coronavirus is nearly 30000 bases long. As a whole, it contains 4 proteins that form the viral envelope which are the Spike protein, E protein, Hemagglutinin (M) protein, and N protein. Figure 1 illustrates the structure of Sars-Cov-2.

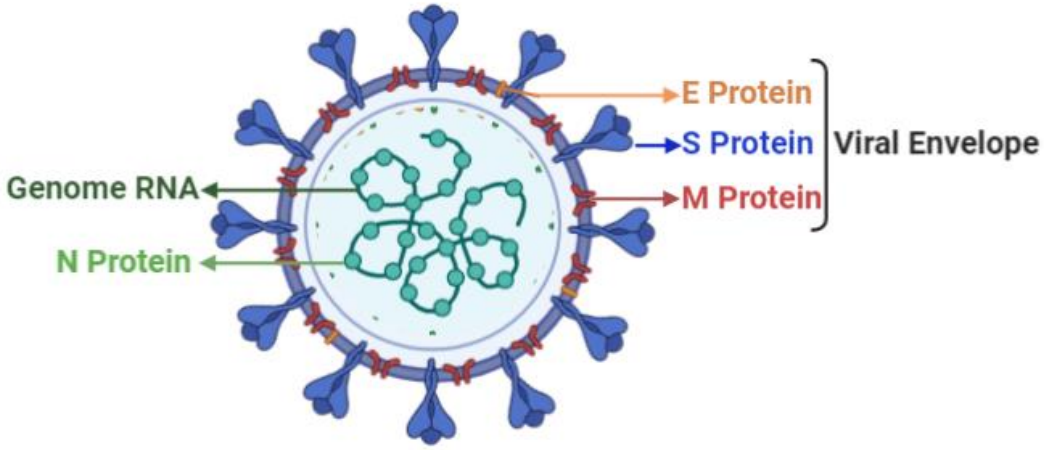


Figure 1 Structure of Sars-Cov-2

It is very important to understand how Sars-Cov-2 gains entry into the human cells. Angiotensinogen is a hormone found inside the liver which is also found in kidneys and different segments of the brain. This hormone is responsible for managing blood pressure. Angiotensinogen is converted into Angiotensin 1 also named AT-I, by an enzyme produced by a kidney called Renin. In the next step, this AT-I is converted into Angiotensin 2 which is named AT-II by an enzyme called ACE which is produced in the lungs. AT-II is a vasoconstrictor that means it narrows the blood vessels, as a result, aldosterone is produced which causes an increase in blood pressure. AT-II creates two states in the body, one is a low state in which ACE2 binds with Angiotensin Receptor I also named ATR-1 on the surface of the membrane. As a result of this ACE2-ATR-I binding, ACE2 creates Angiotensin 17 (AT-17) which is responsible for vasodilation and decreases inflammation which is good for the human body.

The second state is called high state in which due to the high level of AT-II, it does not allow to bind ATR-I with the sites of ACE2 resulting in a gap on the site of ACE2. Because of this gap, Spike protein at the surface of Sars-Cov-2 finds sites to attack, it binds with the sites of ACE2 where ATR-I did not bind and

Sars-Cov-2 anchors itself to an entry point into the human cell. It is worth mentioning here that Spike binds with Human Ace with an affinity of -21 kcal/mol, if the spike is to be targeted, ligand must have a binding affinity of more than -21 kcal/mol, which makes it almost impossible to find such a ligand which could bind with Spike protein with a higher binding affinity [1]. This leads to the conclusion that finding ligands that target Spike proteins may not prove fruitful. After entry, the virus needs to replicate itself so that it can propagate itself within its host cells. The RdRp protein plays a pivotal role during this replication process. RdRp is the most significant gene in the virus genome which is encoded inside the RNA of the virus, it speeds up the process of RNA replication from the RNA template and provides safe passage to the virus that is just entered into human cells.

Endoplasmic reticulum is a system of membrane that performs multiple functions i.e. Modification, folding, and transfer of proteins. After entering into the human cell, the virus contacts this system and persuades the development of a double-membrane vesicle by developing a complex with it. It generates a copy of genomic RNA. Further, it converts this Negative RNA to positive RNA which makes it mRNA. But this mRNA cannot replicate by itself and translate into a protein. The virus exploits the ribosome machinery of the human cell. The ribosome is tricked into working for the virus and translates the mRNA, creating viral proteins in thousands in each replication cycle. These viral proteins are received by the Golgi apparatus which pack them into vesicles and later send to different destinations. In this way, the whole protein creation apparatus of a human cell is used by the virus for its multiplication. Below figure 2 is the illustration of virus attachment and replication mechanism.

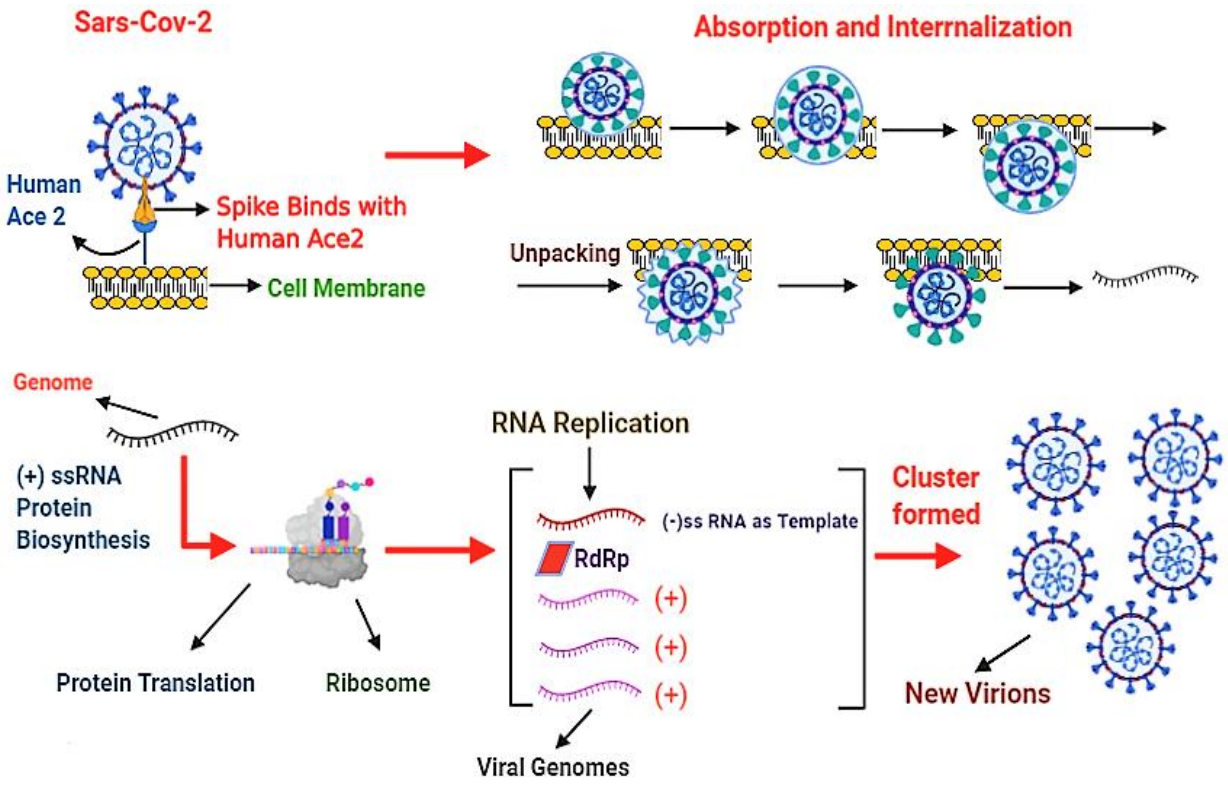


Figure 2 Working Mechanism of Sars-Cov-2

As depicted in figure 3 below, RdRp which is also Non-Structural Protein 12 (nsp12) illustrated in the complex with two small proteins nsp7 and nsp8 and has right-hand cup structure with palm subdomain, thumb subdomain, and fingers subdomain.

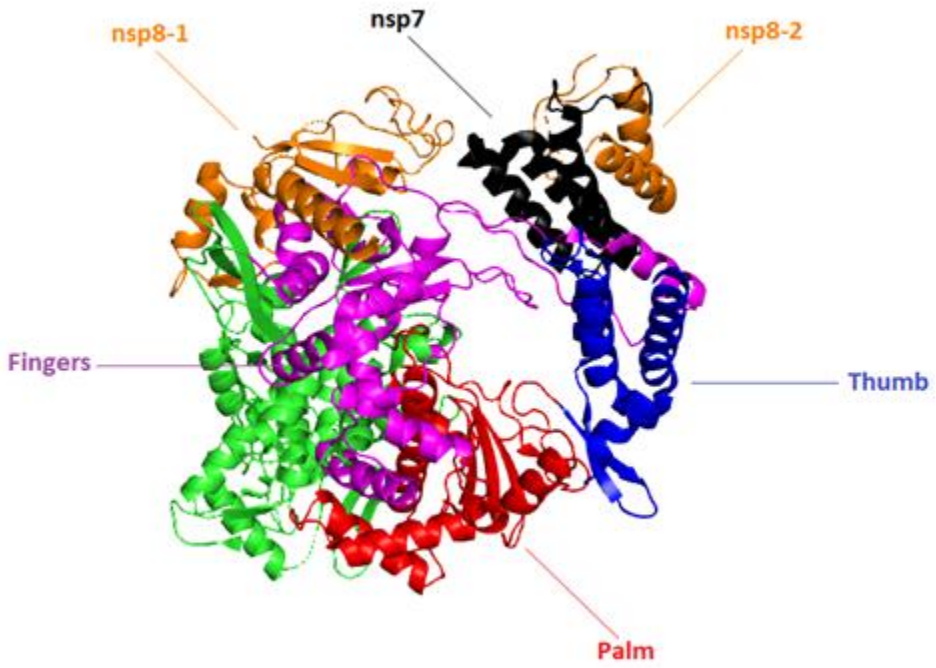


Figure 3 RdRp Protein Structure of Sars-Cov-2 in complex with cofactors (nsp-7, nsp-8)

Table 1 shows the range of residues that cover palm, fingers, and thumb subdomains in the structure of RNA-Dependent-RNA-Polymerase.

Table 1 Residue Analysis of RdRp Protein structure

Subdomain	Residue Range
Palm	**T582-P620 and T680-Q815**
Fingers	**L366-A581 and K621**
Thumb	**H816-E920**

The Amino acid sequence of the Sars-Cov-2 genome in many respects resembles Sars-Cov that caused the SARS outbreak in 2002-2003. One strategy to inhibit the progression of the disease is to find ligands that target the RdRp protein. Antiviral drugs that can considerably compromise the function of RdRp protein will be able to suppress the viral multiplication and hence disease progression. Researchers are working to discover an antiviral drug that targets its key residues by splitting the strands of RNA that cause replications for the virus. In this way, its replication and connection with the virus could be inhibited compromising its proper functioning. Table 2 shows different Motifs and the residues that cover each motif along with the type of residues.

Table 2 Residue Analysis of RdRp Protein structure

Motif	Residues	Type of Residues
A	T611-M626	Divalent-cations Residues
C	F753-N767	Catalytic Residues
F	K545, R553 and R555	Hydrophilic Residues
G	K500, S501	(+) ss-RNA entry tunnel

From the information of above Table 1 and Table 2, it is observed that residues D759-D761 in motif C set up catalytic active sites. D760 and D761 form a complex with magnesium ions inside the catalytic center [2]. Motifs F and G reside inside the finger subdomain and they form a link with the transcribed strand (where mRNA is created) and instruct this strand into an active site. Motif F also binds with primer RNA with residues K545 and R555 by interacting with the +1 base. While key residue 618 is a classic case of Divalent-cation-binding residue which exists in most of the viral enzymes together with the viruses of Hepatitis C and polio. Based on these properties, residues that play a pivotal role in replication can be identified. These key binding residues are: ASP 618, ASP 623, ASP 760, ASP 761, ASN 691, SER 682, THR 680, ARG 555, and VAL 557[3]. A drug therapy that targets these residues of RdRp protein will be able to produce an antiviral effect by inhibiting its function. Moreover, clinically proven drugs like Remdesivir binds to THR 680, SER 682, and VAL 557, pp-sofosbuvir binds to ASN 691, ARG 555, and ASP 623[3]. Binding details of both drugs will give an insight to discover potential compounds that can cover these binding residues as well as remaining residues that are not covered by Remdesivir and pp-sofosbuvir.

Phytochemicals are naturally occurring substances that can contain antiviral and antibiotic properties proving effective for the treatment of diseases. Several plants have therapeutic compounds for example compounds of Artemisia can inhibit tumor growth inside the body and can be used as anticancer substances. Compounds of Azadirachta are used as an antibacterial and for the treatment of skin diseases, stomach upsets, diabetes, fever, and different eye diseases [4]. Compounds of aconitum heterophyllum are antibacterial, antiviral, and anti-inflammatory, these are used for fever, flu, cough, upper tract respiratory diseases, and malaria. Many natural plants and herbs contain substances that have been used as antiviral, antioxidant, and antibacterial purposes for centuries. A large set of plants containing flavonoids, alkaloids, Vitamin C, Sennosides etc. have antiviral properties that can be effective for the treatment of disease. There are thousands of phytochemicals and natural substances whose structures are openly accessible in databases like PubChem, RCSB, chEMBL, and ChemSpider. Medicinal trials on these substances for a specific ailment can consume huge effort in terms of time and money and still required results may not be achieved. In-silico simulation techniques can considerably narrow down on the number of relevant substances through very accurate and meticulous modeling. These methods are capable of providing an insight into the compound structure, analyze physical and chemical properties, and predict the suitability of compounds against target receptors.

Related Work:

Different in silico methods have been used recently to simulate interactions and to evaluate the suitability of drugs for a specific disease. ModeBase was used to create the 3D model of Spike Protein to exhibit the binding of Angiotensin-Converting-Enzyme and Spike Protein. Docking was done by using different virtual

screening methods through software named Schrodinger. Grid Generator tool was used to create a grid [5]. To analyze correlation, Claudia Cava et al. performed an analysis between Human Ace2 and other proteins by TCGA-LUAD to get all the possible interactions while path enrichment analysis is performed using a Fisher's test [6]. T. Joshi et al. used virtual screening to screen 318 phytochemicals to get a suitable compound to analyze the interaction with Human Ace2. PyMol is used to remove ions and water molecules. Open babel is used to convert the SDF format ligand file to PDB format. Rigid docking method is performed to get different conformations of ligand at different binding sites and in the results observed in Lig-Plot+ software [7]. Ammar D. et al. used Computer-Aided Design (CADD) to show the interactions of ligands and receptors. The study also showed that molecular docking is done to evaluate the interactions between Human enzymes and potential ligands. Molecular docking study and ADMET profiling is used to analyze the inhibitors. Homology modeling is used to develop a structure of a protein by using its sequence and then to perform structure-based virtual screening from a large number of chemical compounds AutoDock Vina is used. For binding residues and pockets, AutoDock Vina 4.2 is used [8]. Manoj Kumar et al. used Molecular Dynamic Simulation to study the structure of protein. Further, the DrugMint server is used to prepare drugs like ligands for screening, CASTp is used to calculate the pockets in the protein. Subsequently, autodock is used for docking ligands with receptors and analyzing binding affinity of every compound to set the threshold. Additionally, the comparative analysis of sequences was performed by Multalin [9]. Several researchers have also applied machine learning and artificial intelligence-based models to study the genomic properties [10]–[24].

In this study, a method is proposed to carefully examine Sars-Cov-2 specific antiviral properties of substances. Irrelevant or ineffective chemicals are screened out. The selected compounds are further scrutinized by different docking and interaction techniques. Based on these results the most suitable compounds are proposed for the treatment of COVID-19 that can be the potential therapeutic candidates for the treatment of COVID-19 and open broad-spectrum treatment for other RNA viruses.

Material and Methods

Phytochemicals Preparation:

The selection is performed by analyzing the properties of numerous plants. Then 3D chemical structure of 4596 phytochemicals obtained from natural herbs was extracted from databases like PubChem [25], ChemSpider [26], chEMBL [27] and IMMPAT [28]. Compounds converted from SDF format to PDB format using Open Babel. Subsequently, the phytochemical library is prepared for further processing**.**

Receptor Protein Preparation:

The recent crystal structure of RdRp protein is retrieved from Protein Data Bank (PDB ID: 6M71). Molecule SARS-Cov-2 NSP 12 has one chain with 942 amino acids. Water molecules and hydrogen atoms were removed from the Receptor by using the MGL tools of Autodock Vina.

Virtual Screening:

Virtual screening of phytochemical compounds is performed by the RPBS webserver to narrow down the potential structures that are likely to bind with the receptor. This server uses the AutoDock Vina package which is accurate and yields good screening results [30]. Grid Center Coordinates were set to: X=-2.3,

Y=45.7, Z=28.6. The search space was set to: X=55, Y=55, Z=55. Listed compounds were uploaded on the server for virtual screening. Results with a binding energy of a vast number of compounds are analyzed and all ligands which had binding affinity numerically greater than -7 were discarded.

Molecular Docking:

Suitable compounds that were selected from the results of virtual screening were further docked with the target Protein using AutoDock Vina. Grid box parameters were set to: X=-3.27, Y=44.29, Z=-28.65 and Dimensions were set to: X=35, Y=35, Z=35 (Angstrom). Universal Force Field (UFF) method was used for minimization which is more effective in finding the minimized energy than any other method. A web-based tool named admetSAR was used for profiling and finding drug similarity.

Universal force field (UFF) Optimization:

After loading ligands in Autodock Vina, UFF optimization is used to carry out the optimization of molecular geometry with the help of molecular mechanics. Method of energy minimization is used before the process of docking. This ensures that ligand's length, structure, and angles of bonds are precise before performing the docking process. This method provides good results with organic and inorganic compounds.

Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method:

Broyden-Fletcher-Goldfarb-Shanno (BFGS) is used in autodock Vina for local optimization. This method helps to generate different conformers of ligands. Just like other optimization methods, BFGS also uses gradients with scoring function i.e. the derivative of functions with its arguments. In this situation, arguments contain position, orientation of ligands, and torsion values for effective bonds. This gradient is used to decide the direction of local optima. Before calculating the second derivative which may prove costly, BFGS estimates using top-level updates provided by gradient assessment. In the end, an optimized structure is chosen for selection and by using the Metropolis basis, the next iteration will start from this structure and if this structure scores better than the best available solution then this will be again optimized and will be used as the current best solution [31]. This search process continues until the limit of iterations is reached.

Visualization:

After the completion of the above process, the most suitable results and interaction of every Ligand are further analyzed to visualize the ligand-receptor binding sites in PyMol. PyMol is a very efficient visualization software that supports 2D and 3D structure of the complex and binding interactions as well as distance measurement with sequence information of protein and ligand.

Molecular Dynamic Simulation:

Molecular dynamic (MD) simulation is widely used to evaluate the structural behavior and stability of the protein. In this study, Nanoscale Molecular Dynamic (NAMD) software is used to perform simulation on RdRp protein and proposed compounds [32]. The temperature of 310K was set for simulation. Configuration files were generated using the CHARMM website [33]. Parameter files were obtained using

the CHARMM General Force Field (CGenFF) tool. Protein with complex was solvated using water molecules. The energy of the system was minimized for 2000 steps and dielectric was set to 1.0.

Electrostatic Potential Calculation:

Electrostatic potential simulation was performed by PyMol which uses the Poisson-Boltzmann method based on cubic spline charge discretization. Solute dielectric and solvent dielectric were set to 2.0 and 78.0 respectively. Temperature was set to 310K whereas the solvent probe radius was 1.400 Å. Table 3 shows grid values that were used for the calculation of electrostatic charge of RdRp.

Table 3 Grid details in Electrostatic Potential Calculation

Grid Dimension	193 x 225 x 193
Grid Spacing	0.754 x 0.638 x 0.745
Grid Lengths	143.018 x 143.018 x 143.018
Grid Center	-6.628 x 46.827 x -30.579

Figure 4 shows the flow chart of the above-described methodology and a brief overview of the screening and selection process of phytochemicals that are extracted from different databases.

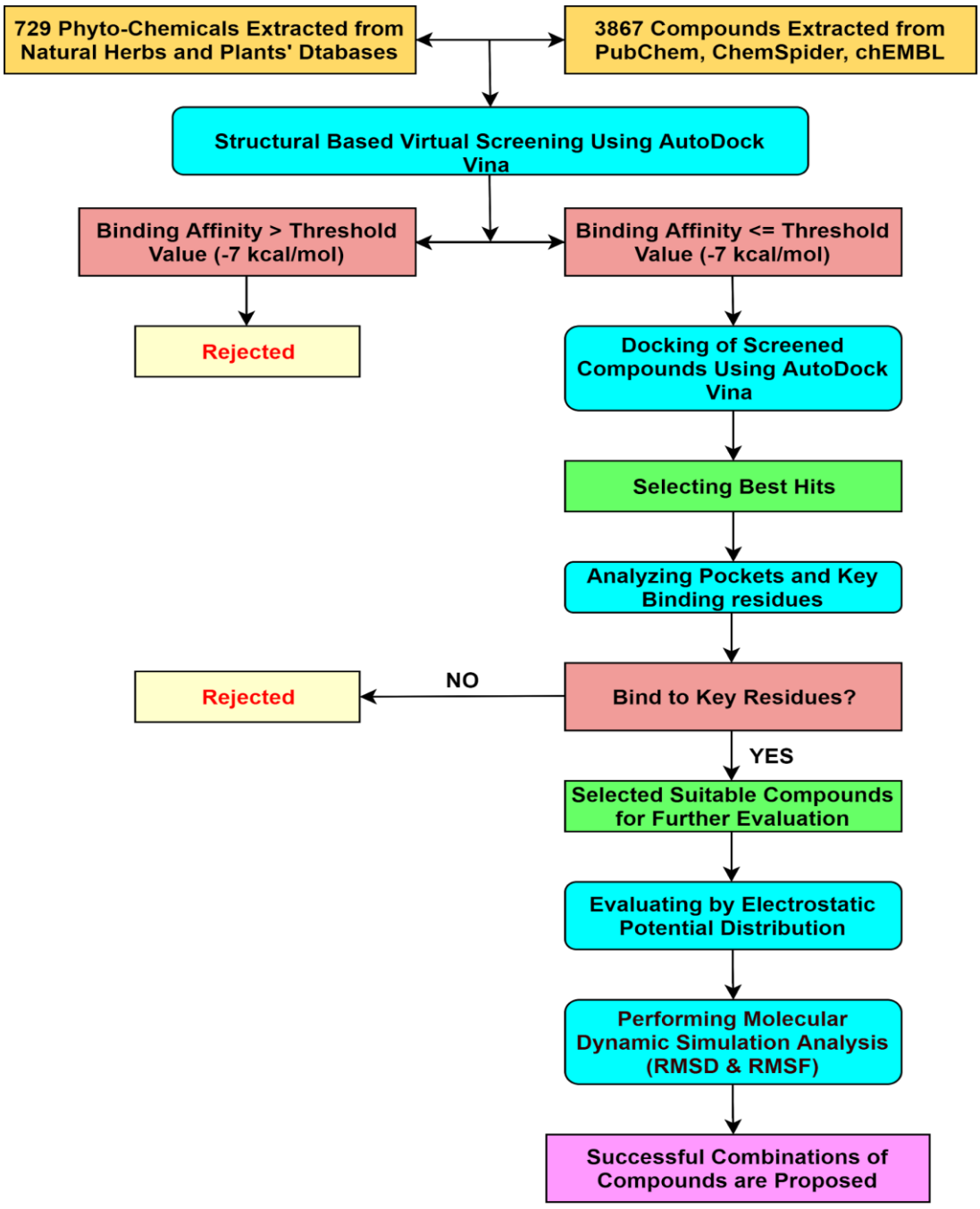


Figure 4 Workflow of the proposed Study

The next step is to identify key binding sites and evaluate ligands by fitting different conformations in RdRp protein. These residues are analyzed with the key residues of RdRp and then the compounds with the most suitable binding residues are analyzed in terms of their chemical and physical properties and the following compounds are listed to analyze the key binding residues.

Results:

Table 3 shows the names of selected ligands along with the binding affinity value of each ligand. For each ligand, the residue sites of RdRp with which it binds to are also listed. All screened compounds have at least -7 binding affinity. Mentioned compounds are mostly naturally occurring substances and they have shown promising antiviral activities and very good binding affinity values with RdRp during the docking process.

Table 4 Binding Analysis of RdRp Protein with compounds

Ligand Name	Binding Residues	Binding Affinity (kcal/mol)
Nimbinene	ARG 118	-13.5
Desacetylnimbinolide	SER 682	-13.4
	ARG 555	
	ARG 624	
	THR 556	
6-Deacetylnimbinene	ASP 623	-13.4
	ARG 624	
7-Deacetyl-7-oxogedunin	TYR 346	-13.3
Isovepaol	ARG 524	-13.3
	THR 626	
	THR 556	
2',3'-Dehydrosalannol	TYR 530	-12.9
	SER 343	
Deacetylsalannin	SER 518	-11.9
Salannin	ARG 733	-11.2
Salannol acetate	PHE 396	-10.7
	ARG 349	
Nimbolin A	ARG 132	-10
	LEU 207	
17-epi-17-Hydroxyazadiradione	PRO 323	-9.7
	ARG 349	
17-Hydroxyazadiradione	LEU 207	-9.6
	ASP 208	
	HIS 133	
.beta.-Amyrin	ARG 733	-9.4
	TYR 728	
Melianin A	ARG 733	-9.3
	TYR 728	
Cyclohexane-1- carboxylic acid	ARG 555	-9.0
	ASN 691	
	SER 682	
Fraxinellone	ASN 360	-8.8
	ASN 356	

17-Epiazadiradione	ARG 10	-8.7
24-Methylenecycloartan-3-one	ASN 691	-8.7
14-dien-7-yl acetate	ALA 688	-8.7
Naringin	ARG 555	-8.6
	ASP 623	
isomargosinolide	ASP 711	-8.6
Sennaglucosides	ASP 760	-8.4
	ASP 761	
	ASP 623	
	ASP 618	
Nimbidinin	TYR 728	-8.3
	SER709	
Margosinolide	TYR 728	-8.3
Nimbocinolide	HIS 725	-8.3
6beta-Hydroxystigmast-4-en-3-one	LYS 676	-8.2
8-difluoro-7-hydroxy chromen-4-one	THR 680	-7.5
	ARG 553	
6-methoxykaempferol	THR 394	-7.4
	PHE 396	
Axillarin	ILE 548	-7.2
	ARG 624	
	ARG 555	
Famotidine	ASP 618	-7.1
	ASP 623	
	ASP 760	
Famotidone	ASP 760	-7.0
	ASP 761	

Furthermore, only those ligands are shortlisted which exhibit binding interactions involving any of the key residues of RdRp protein discussed earlier i.e. ASP 618, ASP 623, ASP 760, ASP 761, ASN 691, SER 682, ARG 555, THR 680, and VAL 557 and also have high affinity with the RdRp protein. Table 4 shows a comparison of compounds selected under this criteria with some of those which are currently under trial. Ligands' names along with the binding details, distance of interaction from each binding site, and estimated inhibition constant is also shown. The comparison shows that the proposed compounds have very good binding affinity and bind to key residues to compromise the replication of RdRp. Below compounds (highlighted in Green) have better binding affinity and binding interaction with RdRp key residues than the compounds that are currently under trial i.e. Camostat Mesylate [34], Hesperidin [35] and Remdesivir [36].

Table 5 Binding Analysis of RdRp Protein with compounds

Ligand Name	Binding Residue with RdRp	Residue Distance	Binding Affinity (kcal/mol)	Inhibition Constant (nM)
Naringin	**ARG 555**	**2.9**	-8.6	488
	ASP 623	**2.3**		
Desacetylnimbinolide	THR 556	2.4	-13.4	0.146
	ARG 555	**2.3**		
	ARG 624	**2.6**		
	SER 682	**2.2**		
Sennaglucosides	**ASP 618**	**2.1**	-8.4	685
	ASP 760	**2.1**		
	SER 682	**2.1**		
	ASP 623	**2.3**		
	ASP 761	**2.6**		
Famotidine	**ASP-760**	2.5	-7.1	6164
	TRP-800	2.2		
Famotidone	**ASP 760**	2.3	-7.0	7299
	ASP 761	1.9		
Cyclohexane-1-Carboxylic acid	**ASN 691**	**2.5**	-9.0	248
	ARG 555	**2.7**		
	SER 682	**2.3**		
8-difluoro-7-hydroxy chromen-4-one	**THR 680**	**2.8**	-7.5	3135
	ARG 553	2.2		
Camostat Mesylate	TYR 346	2	-6.8	10233
Hesperidin	THR 394	2.5	-8.4	685
	PHE 396	2.7		
Ritonavir	ASP 208	2.5	-8	1347
	SER 682	**2.1**		
Remdesivir	ARG-249	2	-8.1	1137
	ARG-349	2.2		
Hydroxychloroquine	GLY-327	2.7	-6.3	23820
	HIS-347	2.3		

Further details of each of the proposed compounds are provided below. Almost, all of them are phytochemicals that have been in use for ages. Their usage, toxicity, metabolism, and side-effects are well studied and understood.

Naringin (PubChem ID: 442428):

Naringin is a bioflavonoid and it belongs to the family of flavonoids, it is found in citrus fruit and has exhibited antiviral, anti-inflammatory, and possesses antioxidant properties. It is used in the treatment of diabetes, hypertension, and metabolic syndrome. It has also shown anticancer effects as it behaves as suppressing or blocking agents in the treatment of cancer. It induces cell apoptosis and impedes cell proliferation in tumor cells of Bladder cancer, Breast cancer, and cervical cancer [37]. Figure 5 shows the chemical structure of Naringin.

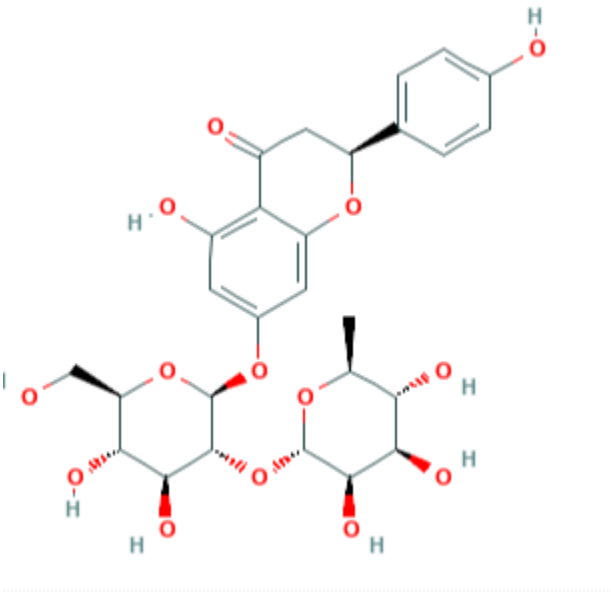


Figure 5 Chemical Structure of Naringin [25]

Desacetylnimbinolide (PubChem ID: 102285346)

This naturally occurring substance is extracted from plant Azadirachta Indica which is commonly known as *Neem* in the Indian subcontinent. It has been used in Chinese and Unani medicines for many years. It is enriched with antioxidants. It plays a vital role in anticancer management [38]. It is very safe for medicinal purposes and used in the treatment of diabetes, fever, and skin disease. Figure 6 depicts the structure of Desacetylnimbinolide.

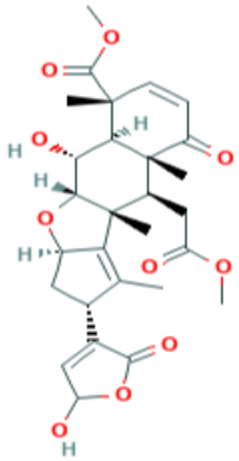


Figure 6 Chemical Structure of Desacetylnimbinolide [25]

Sennaglucosides (PubChem ID: 5199)

This is the most effective substance found in this research. This is also a naturally occurring compound that is extracted from a plant called Alexandria Senna. Its leaves are used for medicinal purposes. It is used to treat constipation and has strong laxative effects. It is also used to empty the stomach before surgery and its medication is taken by mouth. It prevents the reabsorption of water and electrolytes which results in increment of fluids in the intestine. It is safe and well-tolerated. Figure 7 is showing the structure of Sennaglucosides.

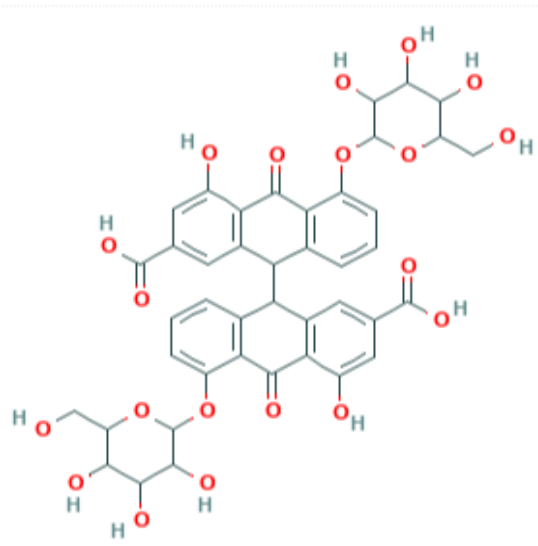


Figure 7 Structure of Sennaglucosides [25]

Famotidine (PubChem ID: 5702160)

Famotidine is used to decrease acids produced in the stomach and intestine. It is a histamine-2 inhibitor and used to treat ulcers in the stomach. It is also used for ZES (Zollinger-Ellison-Syndrome) in which the stomach produces excessive acids. It also prevents ulcers from coming back. It is available in the market with the name of Pepcid and its medication is taken by mouth. Figure 8 is showing the structure of famotidine.

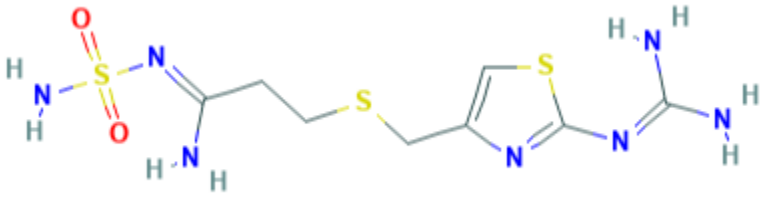


Figure 8 Chemical Structure of Famotidine [25]

Famotidone (PubChem ID: 129849878)

Famotidone can be used for hayfever, skin allergies, and itchy nose. It can also be used for the treatment of skin rashes for adults and children over 6 years. Figure 9 is showing the structure of famotidone.

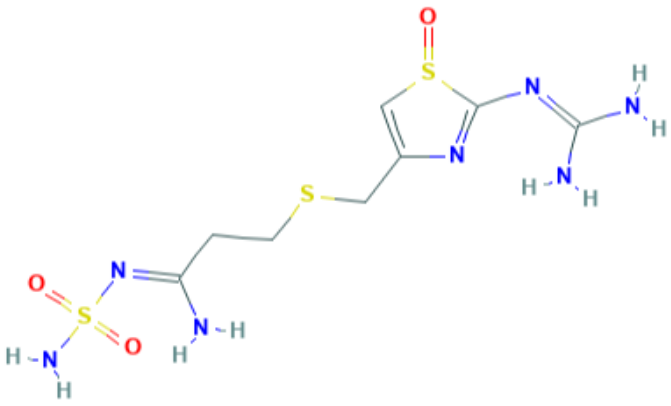


Figure 9 Chemical Structure of Famotidone [25]

4-({3,4-dihydroxy-5-[(3,4,5-trihydroxybenzoyl)oxy]benzoyl}oxy)-1-hydroxy-3,5-bis[(3,4,5-trihydroxybenzoyl)oxy]cyclohexane-1-carboxylic acid (PubChem ID: 442676)

This compound which is also named **Quinic acid** (multiplied acylated with galloyl moieties) is extracted from Eucalyptus bark. It is antiseptic and anti-inflammatory and used for the treatment of asthma. It contains a substance that kills bacteria. It is also used for skin diseases like skin ulcers and Gout.

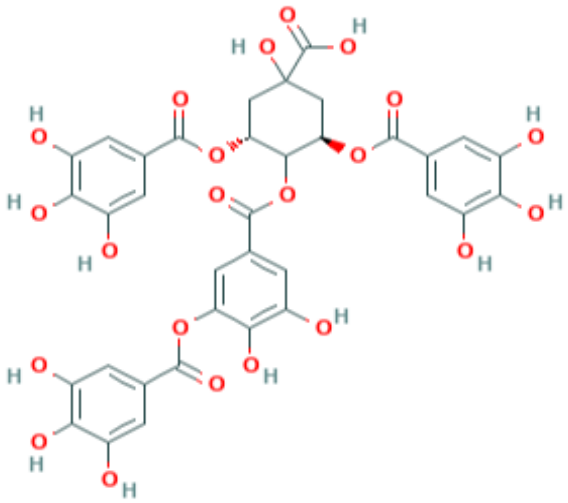


Figure 10 Structure of Cyclohexane-1-carboxylic acid (Quinic Acid) [25]

5-Amino-2-(4-amino-3-fluorophenyl)-6, 8-difluoro-7-[hydroxy-(3-hydroxypropylamino) methyl] chromen-4-one (PubChem ID: 67771200)

This compound is extracted from Rutaceae which belongs to the rue family of flowering plants. This is also found in citrus fruits like orange and lemon. It is used in many diseases like asthma, constipation, fever, and diarrhea.

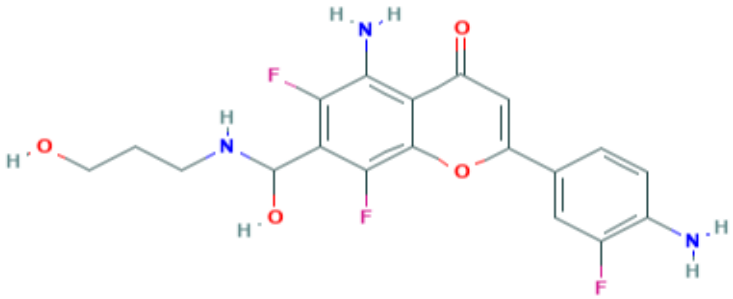


Figure 11 Chemical Structure of above compound [25]

Proposed compounds can be used in combinations of 2 or 3 to inhibit the working of RdRp most effectively. Now we shall visualize results in combinations.

1) ***Naringin & Sennaglucosides:***

Figure 12, shows that a combination of two promising compounds (Sennaglucosides and Naringin) binds to 5 key binding sites of RdRp. Magenta and Green Color represent Sennaglucosides and Naringin respectively. Yellow dots illustrate binding interactions between the combination of compounds and RdRp.

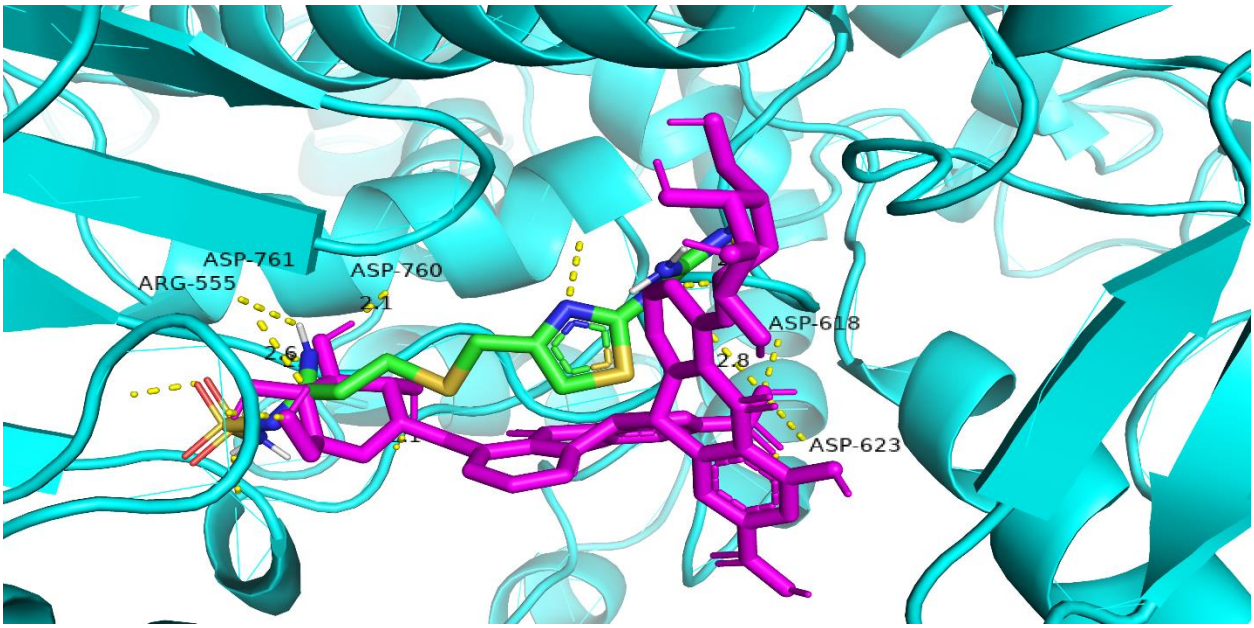


Figure 12 Combined Interaction of Sennaglucosides and Naringin

To further understand the results, table 5 shows that Sennaglucosides and Naringin have inhibition constants of 685 nM and 488 nM respectively, and interact with 5 key sites (both compounds bind with ASP 623 simultaneously, that's why this binding site is neglected for Naringin) of RdRp protein. It is also shown in the table that the combination of Sennaglucosides and Naringin interacts with the key binding residues of RdRp with a good binding affinity of over -8.3.

Table 6 Combination details with key residues, refer to figure 12

Ligand Name	Color	No. of Key Binding Residues	Binding Residues	Binding Affinity (kcal/mol)	Inhibition Constant (nM)
Sennaglucosides	Magenta	4	ASP 618, ASP 623, ASP 760, ASP 761	-8.4	685

Naringin	Green	2 (-1)	ARG 555, ASP 623	-8.6	488

2) ***Sennaglucosides and Desacetylnimbinolide:***

Figure 13 shows the conformations of Sennaglucosides (Magenta) and Desacetylnimbinolide (Green) in the combination which best fit the key residues and cover over 10 binding sites but our main focus is key residues. This combination covers 6 binding residues and binding interactions are shown in yellow dotted lines.

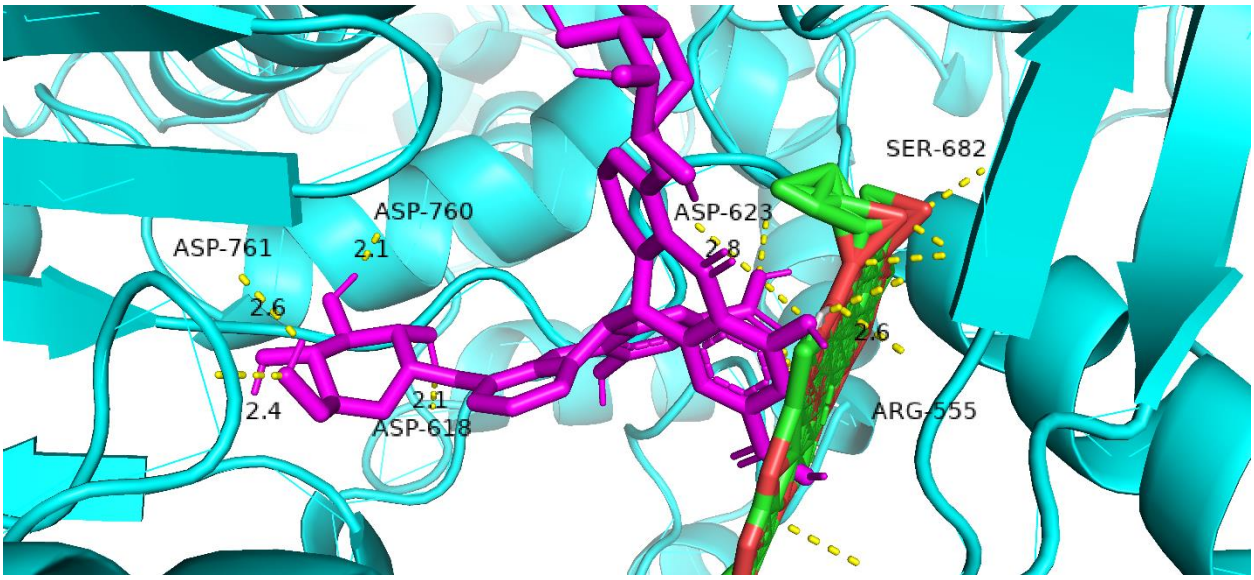


Figure 13 Combined Interaction of Sennaglucosides and Desacetylnimbinolide

Table 6 shows that Sennaglucosides binds to 4 key residues with an inhibition constant of 685 nM and Desacetylnimbinolide binds to 2 key residues with an inhibition constant of 0.146 nM and combination of both compounds can cover 6 key binding sites that are very important to inhibit the function of RdRp. This interaction with the key binding residues can halt the exponential growth of Sars-Cov-2 in human cells by compromising the function of RdRp.

Table 7 Combination details with key residues, refer to figure 13.

Ligand Name	Color	No. of Key Binding Residues	Binding Residues	Binding Affinity (kcal/mol)	Inhibition Constant (nM)
Sennaglucosides	Magenta	4	ASP 618, ASP 623, ASP 760, ASP 761	-8.4	685
Desacetylnimbinolide	Green	2	ARG 555, SER 682	-13.4	0.146

3) ***Desacetylnimbinolide and Naringin:***

In figure 14, the interaction of Desacetylnimbinolide (represented in Green) and Naringin (represented in Magenta) has been illustrated with key sites of RdRp. Binding interactions are shown in yellow dots.

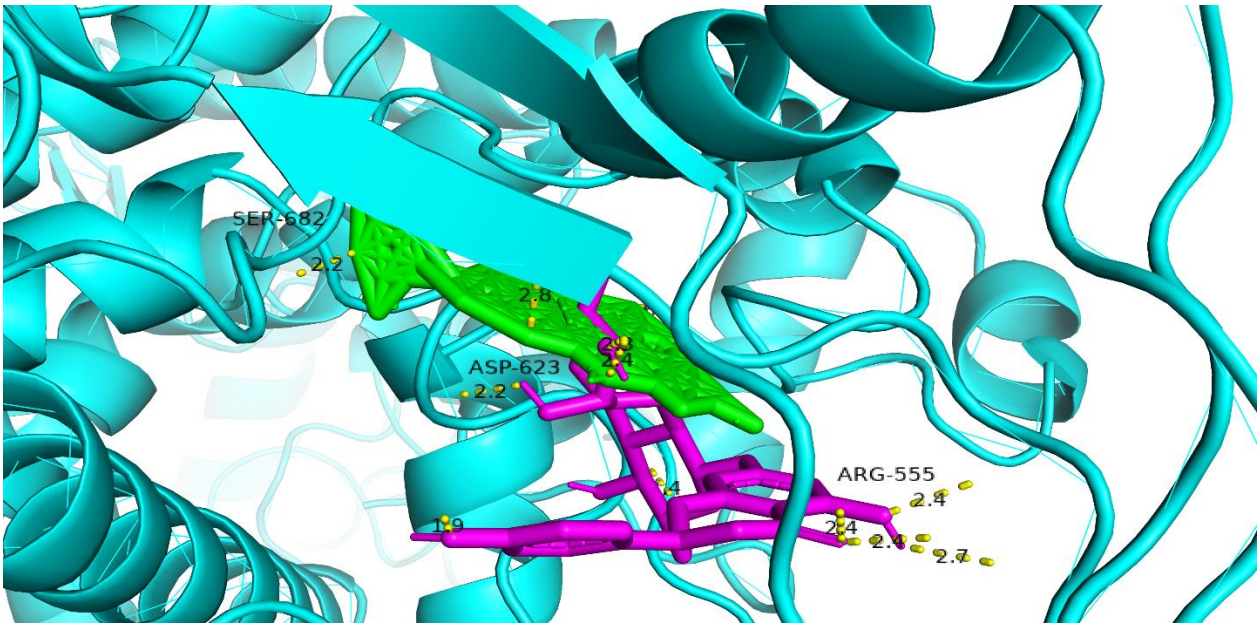


Figure 14 Combined Interaction of Desacetylnimbinolide and Naringin

Table 7 shows that the combination of these two compounds covers 3 key binding sites (ARG-555 is common in both compounds' interaction, ARG 555 from Naringin is not included in Figure 13 and neglected in Table 8). The details of binding residues that each compound cover along with the binding affinity and inhibition constant is also shown. These compounds can be very effective for the treatment of RNA-related and antiviral diseases.

Table 8 Combination details with key residues, refer to figure 14.

Ligand Name	Color	No. of Key Binding Residues	Binding Residues	Binding Affinity (kcal/mol)	Inhibition Constant (nM)
Desacetylnimbinolide	Green	2	ARG 555, SER 682	-13.4	0.146
Naringin	Magenta	2 (-1)	ARG 555, ASP 623	-8.6	488

4) ***Famotidine and Famotidone:***

In figure 15, Famotidine (represented in Green) is already under consideration for COVID-19 possible drug and combination of it with Famotidone (represented in Magenta) which is more stable and binds to key sites of RdRp, makes this combination worth looking at.

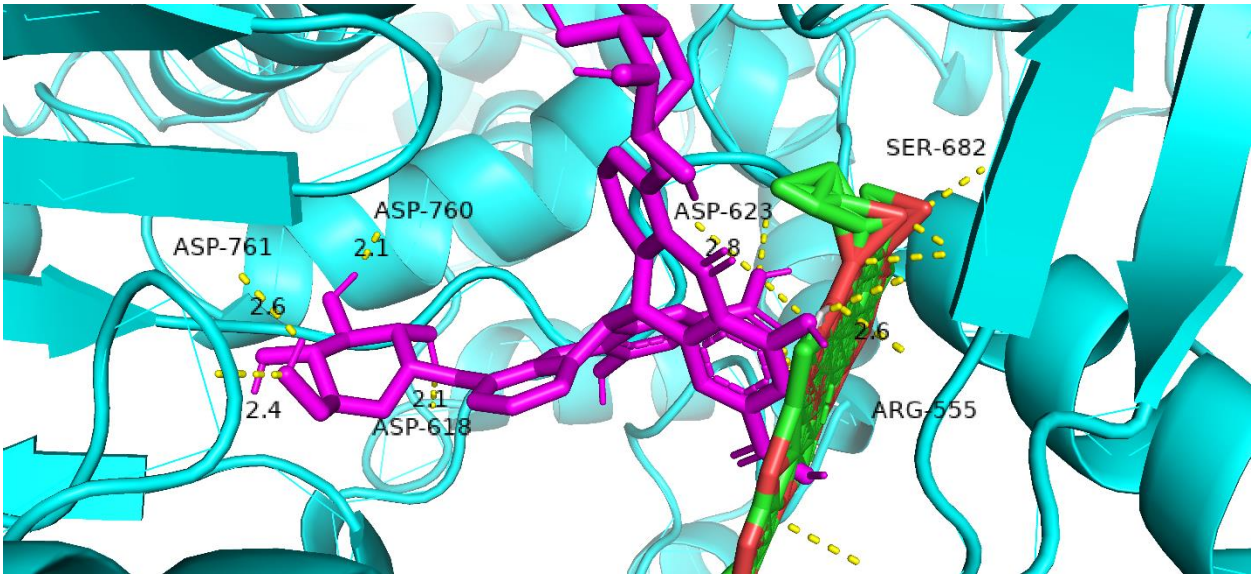


Figure 15 Combined Interaction of Famotidine and Famotidone

Table 8 shows the binding sites of RdRp that this combination binds to, along with the binding affinity and inhibition constants represented in kcal/mol and nM respectively. This combination covers 3 key residues (ASP 623, SER 682, and ASP 760).

Table 9 Combination details with key residues, refer to figure 15.

Ligand Name	Color	No. of Key Binding Residues	Binding Residues	Binding Affinity (kcal/mol)	Inhibition Constant (nM)
Famotidine	Green	1	ASP 760	-7.1	6164
Famotidone	Magenta	2	SER 682, ASP 623	-7.0	7299

5) ***Sennaglucosides, Cyclohexane-1- carboxylic acid and 8-difluoro-7-hydroxy chromen-4-one:***

In figure 16, a combination of Cyclohexane-1-carboxylic acid which is also known as Quinic Acid (represented in Blue), Sennaglucosides (represented in Magenta), and 8-difluoro-7-hydroxy chromen-4-one (represented in Orange) yields the best results in terms of binding to key residues.

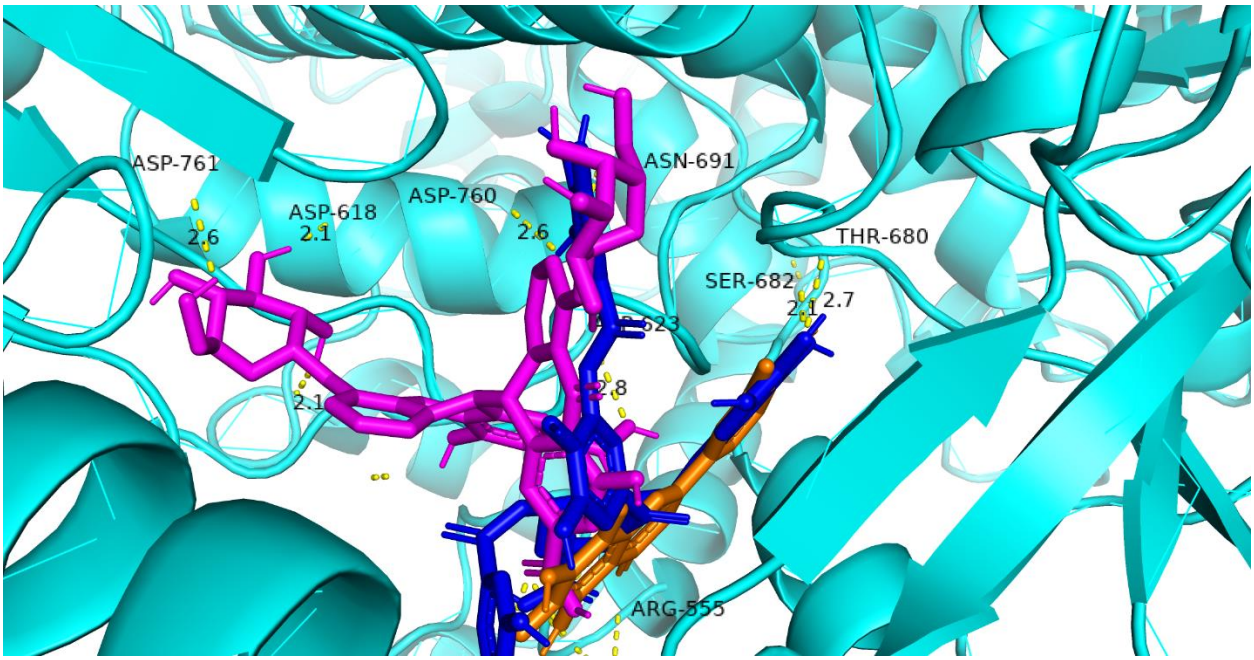


Figure 16 Combined Interaction of Sennaglucosides, Cyclohexane-1- carboxylic acid and 8-difluoro-7- hydroxy chromen-4-one

Table 9 shows that Cyclohexane-1-carboxylic acid (Quinic Acid) binds to 3 key residues, Sennaglucosides binds to 4 key residues, and 8-difluoro-7-hydroxy chromen-4-one binds to 1 key residue, making it an effective combination that binds to 8 key residues of RdRp collectively including ASN 691 and THR 680 which are very important and there are very rare compounds that bind to these two (ASN 691, THR 680) key residues. Binding affinity and inhibition constant of each compound is also mentioned.

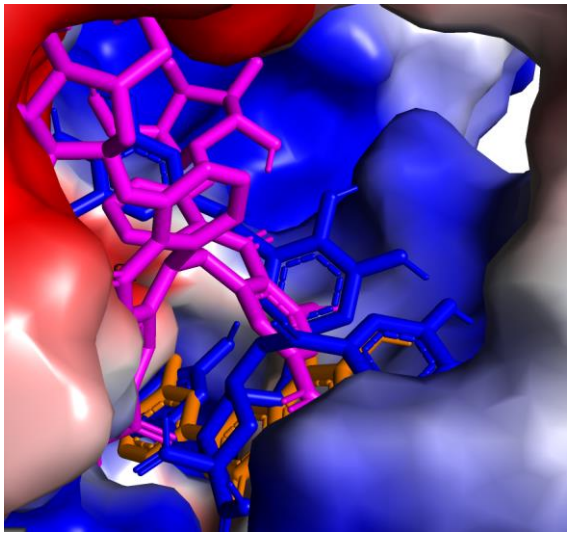
Table 10 Combination details with key residues, refer to figure 16.

Ligand Name	Color	No. of Key Binding Residues	Binding Residues	Binding Affinity (kcal/mol)	Inhibition Constant (nM)
Cyclohexane-1-carboxylic acid	Blue	3	ASN 691, SER 682, ARG 555	-9.0	248
Sennaglucosides	Magenta	4	ASP 618, ASP 623, ASP 760, ASP 761	-8.4	685
8-difluoro-7-hydroxy chromen-4-one	Orange	1	THR 680	-7.5	3135

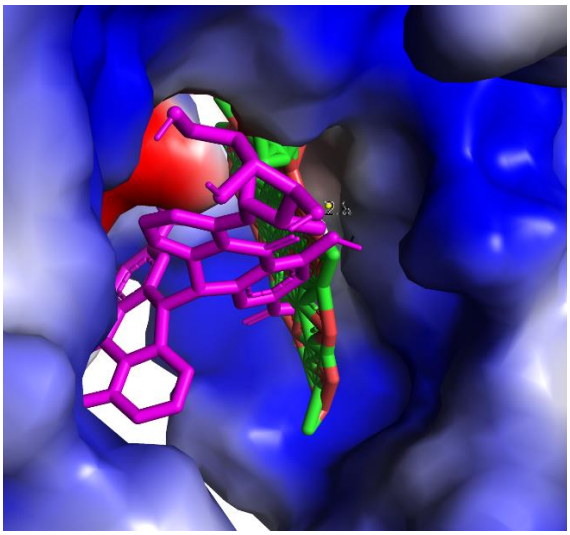
Electrostatic Potential Distribution:

Electrostatic potential is an effective way to understand the structural properties and characteristics of protein and ligands which bind to it. Electrostatic potential charges are mapped on the surface of the RdRp protein of Sars-Cov-2, to show the distribution of positive and negative charges and the intensity on the surface of the protein. Figure 17 shows the electrostatic potential distribution of the proposed combination of ligands. Prediction of electrostatic potential on the surface of RdRp protein of (a) RdRp_Quinic Acid, Sennaglucosides, 8-difluoro-7-hydroxy chromen-4-one, (b) RdRp_Sennaglucosides, Desacetylnimbinolide, (c) RdRp_Naringin, Sennaglucosides, (d) RdRp_Desacetylnimbinolide, Naringin and (e) RdRp_Famotidine_Famotidone. Blue and Red represent positive and negative charges respectively.

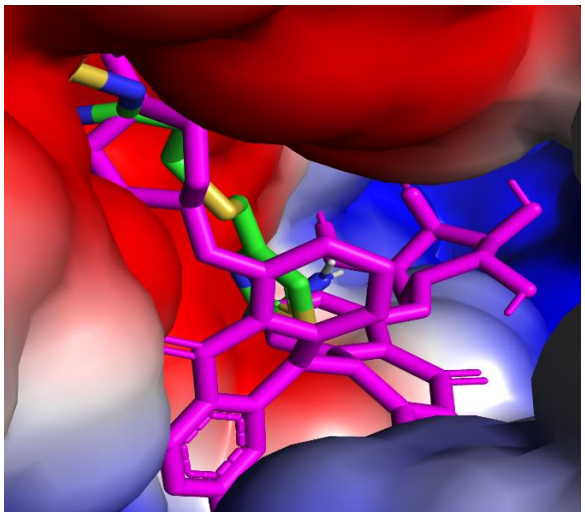
(a)



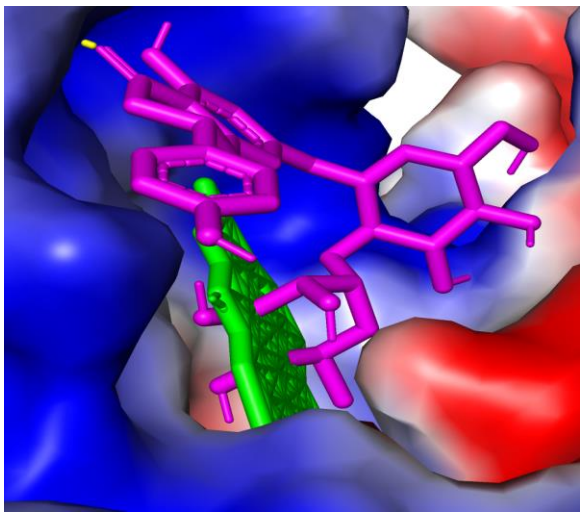
(b)

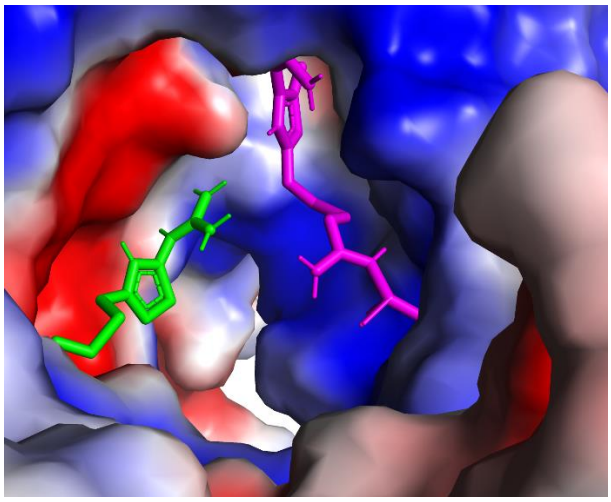


(c)



(d)





(e)

Figure 17 Electrostatic potential on the surface of RdRp protein with complex (a) Quinic Acid, Sennaglucosides, 8-difluoro-7-hydroxy chromen-4-one, (b) Sennaglucosides, Desacetylnimbinolide, (c) Naringin, Sennaglucosides, (d) Desacetylnimbinolide, Naringin and (e) Famotidine, Famotidone.

Distribution of Positive potential charges (Blue) covers the inner cavity of binding pockets of RdRp protein and the remaining surface is covered by the negative charges (Red). Above Figure 17 shows the location of combinations of ligands in the inner cavity of binding pockets of protein which is a clear indication of the fact that proposed compounds bind to key binding sites (Cavity of Binding pockets is shown in the blue) which are the main cause of replication and progression of Virus in the host cells. Thus covering these sites will inhibit the working of RdRp protein.

Molecular Dynamic Simulation Analysis:

In this study, Root Mean Square Deviation (RMSD) is measured to evaluate the distance between backbone atoms of superimposed molecules. As shown in Figure 18, RMSD of RdRp protein remained stable between 16ns to 25 ns timescale at 1.581 Å, then showed a slight upward deviation until 34ns and at 35ns it persisted at 1.581 Å till the end. The RMSD of RdRp_ Quinic Acid, Sennaglucosides, and 8-difluoro-7-hydroxy chromen-4-one showed rise until 20ns at 1.7 Å and after slight fluctuation it gained stability at 25ns at 1.76 Å. The RMSD of RdRp_Sennaglucosides and Desacetylnimbinolide showed stability at 15ns timescale at 1.55 Å and after slight upward fluctuation it system was balanced at 28na timescale at 1.77 Å. The RMSD of RdRp_Naringin and Sennaglucosides increased up to 15ns timescale at 1.62 Å and then fluctuated downward on timescale at 1.57 Å and the system was balanced at 34ns timescale at 1.66 Å. The RMSD of RdRp_Desacetylnimbinolide and Naringin gained stability at 21ns timescale at 1.61 Å. The RMSD of Famotidine and Famotidone ascended until 11ns and then the system was stable until 22ns timescale at 1.53 Å. Figure 18 shows the RMSD plots of protein with all suggested compounds.

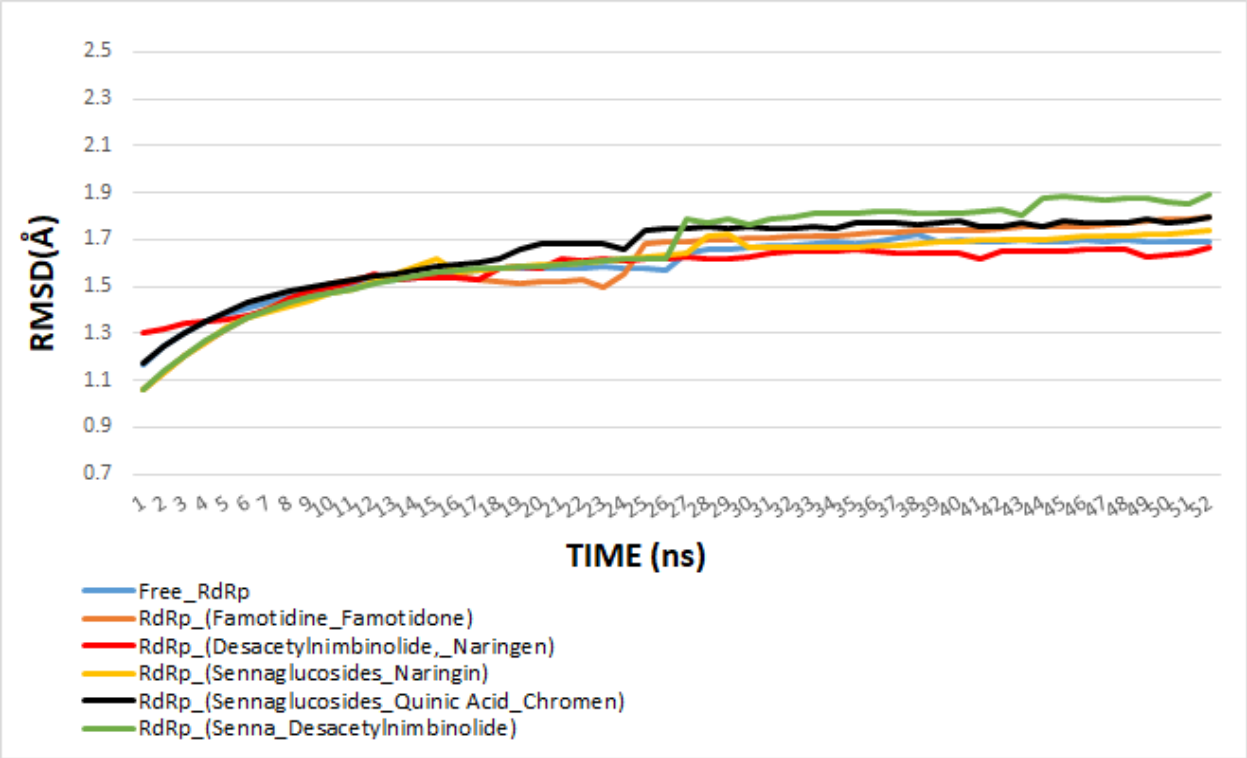


Figure 18 RMSD plot of free protein and complexes with 50ns MD-Simulation. Free RdRp (Blue), RdRp_Famotidine, Famotidone (Orange), RdRp_Desacetylnimbinolide, Naringin (Red), RdRp_Sennaglucosides, Naringin (Yellow), RdRp_Quinic Acid, Sennaglucosides, 8-difluoro-7-hydroxy chromen-4-one (Black), RdRp_Sennaglucosides, Desacetylnimbinolide (Green).

A brief analysis has shown the Root Mean Square Fluctuation (RMSF) of residues of RdRp protein with its complexes. In Figure 19, RdRp and its binding compounds have shown the fluctuations between 1.2 Å and 1.8 Å. This depicts that proposed compounds have maintained close binding contact with the binding residues during Molecular Dynamic simulation.

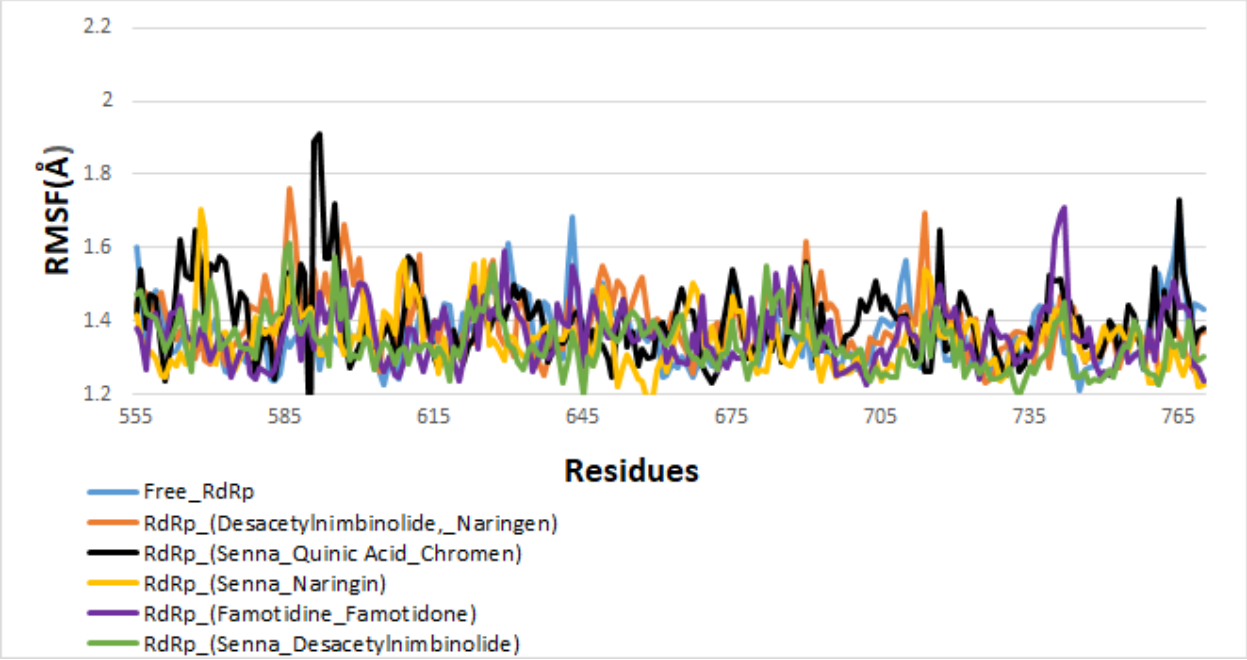


Figure 19 RMSF plot of free protein and complexes during the key binding residues. Free RdRp (Blue), RdRp_Desacetylnimbinolide, Naringin (Brown), , RdRp_ Quinic Acid, Sennaglucosides, 8-difluoro 7-hydroxy chromen-4-one (Black), RdRp_Sennaglucosides, Naringin (Yellow), RdRp_ Famotidine, Famotidone (Purple), RdRp_ Sennaglucosides, Desacetylnimbinolide (Green)

Suggested Combination:

By analyzing the above results, we can predict the most suitable combination according to the number of key residues that it covers. In table 11, combinations of the proposed compounds are listed according to the most suitable first. The table also shows the names of binding residues that these compounds cover along with the number of compounds in each combination. Combinations of compounds are selected according to the ability to cover the maximum key residues as well as binding affinity with RdRp.

Table 11 Compound Combination details with key residues

Compound Combination	Key Residues Covered	No. of Compounds in the combination	No. of Key Residues Covered
• Cyclohexane-1-carboxylic acid (Quinic Acid) • Sennaglucosides • 8-difluoro-7-hydroxy chromen-4-one	ARG 555, THR 680, SER 682, ASN 691, ASP 618, ASP 623, ASP 760, ASP 761	3	8
• Sennaglucosides • Desacetylnimbinolide	ASP 618, ASP 623, ASP 760, ASP 761, ARG 555, SER 682	2	6
• Naringin • Sennaglucosides	ASP 618, ASP 623, ASP 760, ASP 761, ARG 555	2	5

● Desacetylnimbinolide ● Naringin	ARG 555, SER 682, ASP 623	2	3
● Famotidine ● Famotidone	SER 682, ASP 623, ASP 760	2	3

Conclusion:

COVID-19 is a viral disease that has caused a pandemic in the modern era. Not only has it affected social life but it has imparted an impeding effect on world economies. People having an underlying health condition are at great risk. The only way to undo this threat is either by finding a vaccine or a potent antiviral therapy against the virus. Researchers all over the world have proposed numerous drug therapies for the disease. This study covers in-silico identification of phytochemicals that can prove effective in inhibiting the function of RdRp proteins of Sars-Cov-2. The study proposes 7 compounds that can prove effective as per in-silico evidence when used in combinations or individually. These compounds have shown promising signs towards the development of antiviral medications for the COVID-19. Most of them are naturally occurring substances with low toxicity, very few side effects, proven anti-pathogenic effects, and most importantly are easily available. They bind to the key sites of RdRp protein to inhibit its functioning and stop the replication of coronavirus. All the results have been carefully analyzed through the use of in silico methods and machine learning models. Their binding affinities and binding sites are thoroughly observed for result compilation. The most promising observation from the simulation is that a therapy based on the combination of Cyclohexane-1-carboxylic acid (Quinic Acid), Sennaglucosides, and 8-difluoro-7-hydroxy chromen-4-one can bind to eight out of nine key residue sites of RdRp protein of Sars-Cov-2. This is a strong indication that the combination of these compounds can significantly compromise the replication cycle of Sars-Cov-2 and hence alleviate the severity of the disease.

Future Work:

All the results shown in the study are obtained from in silico methods. The proper clinical trial and medical observation will reveal more crucial information about their effectiveness. If the proposed compounds make an impact in the development of the vaccine of COVID-19 then these compounds can also be used in further research of RNA-related viral and other contagious diseases.

References

[1] Y. Chen, Y. Guo, Y. Pan, and Z. J. Zhao, "Structure analysis of the receptor binding of 2019-nCoV," Biochem. Biophys. Res. Commun., 2020.

[2] W. Yin et al., "Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir," Science (80-.)., 2020.

[3] Y. Gao et al., "Structure of the RNA-dependent RNA polymerase from COVID-19 virus," Science (80-.)., 2020.

[4] Y. D. Khan and M. S. Roomi, "Promising Compounds for Treatment of Covid-19," VAWKUM Trans. Comput. Sci., vol. 17, no. 1, pp. 1–8, 2020.

[5] D. C. Hall Jr and H.-F. Ji, "A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease," Travel Med. Infect. Dis., p. 101646, 2020.

[6] C. Cava, G. Bertoli, and I. Castiglioni, "In Silico Discovery of Candidate Drugs against Covid-19," Viruses, vol. 12, no. 4, p. 404, 2020.

[7] T. Joshi et al., "In silico screening of natural compounds against COVID-19 by targeting Mpro and ACE2 using molecular docking.," Eur. Rev. Med. Pharmacol. Sci., vol. 24, pp. 4529–4536, 2020.

[8] A. D. Elmezayen, A. Al-Obaidi, A. T. Şahin, and K. Yelekçi, "Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes," J. Biomol. Struct. Dyn., no. just-accepted, pp. 1–12, 2020.

[9] M. K. Gupta, S. Vemula, R. Donde, G. Gouda, L. Behera, and R. Vadde, "In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel," J. Biomol. Struct. Dyn., no. just-accepted, pp. 1–17, 2020.

[10] M. Awais, W. Hussain, Y. D. Khan, N. Rasool, S. A. Khan, and K.-C. Chou, "iPhosH-PseAAC: Identify phosphohistidine sites in proteins by blending statistical moments and position relative features according to the Chou's 5-step rule and general pseudo amino acid composition," IEEE/ACM Trans. Comput. Biol. Bioinforma., 2019.

[11] O. Barukab, Y. D. Khan, S. A. Khan, and K.-C. Chou, "iSulfoTyr-PseAAC: Identify Tyrosine Sulfation Sites by Incorporating Statistical Moments via Chou's 5-steps Rule and Pseudo Components," Curr. Genomics, vol. 20, no. 4, pp. 306–320, 2019.

[12] A. Ehsan, K. Mahmood, Y. D. Khan, S. A. Khan, and K.-C. Chou, "A novel modeling in mathematical biology for classification of signal peptides," Sci. Rep., vol. 8, no. 1, pp. 1–16, 2018.

[13] A. Ehsan, M. K. Mahmood, Y. D. Khan, O. M. Barukab, S. A. Khan, and K.-C. Chou, "iHyd-PseAAC (EPSV): identifying hydroxylation sites in proteins by extracting enhanced position and sequence variant feature via chou's 5-step rule and general pseudo amino acid composition," Curr. Genomics, vol. 20, no. 2, pp. 124–133, 2019.

[14] A. W. Ghauri, Y. D. Khan, N. Rasool, S. A. Khan, and K.-C. Chou, "pNitro-Tyr-PseAAC: predict nitrotyrosine sites in proteins by incorporating five features into Chou's general PseAAC," Curr. Pharm. Des., vol. 24, no. 34, pp. 4034–4043, 2018.

[15] W. Hussain, Y. D. Khan, N. Rasool, S. A. Khan, and K.-C. Chou, "SPrenylC-PseAAC: A sequence-based

model developed via Chou's 5-steps rule and general PseAAC for identifying S-prenylation sites in proteins," J. Theor. Biol., vol. 468, pp. 1–11, 2019.

[16] W. Hussain, Y. D. Khan, N. Rasool, S. A. Khan, and K.-C. Chou, "SPalmitoylC-PseAAC: A sequence-based model developed via Chou's 5-steps rule and general PseAAC for identifying S-palmitoylation sites in proteins," Anal. Biochem., vol. 568, pp. 14–23, 2019.

[17] S. Ilyas, W. Hussain, A. Ashraf, Y. D. Khan, S. A. Khan, and K.-C. Chou, "iMethylK-PseAAC: Improving Accuracy of Lysine Methylation Sites Identification by Incorporating Statistical Moments and Position Relative Features into General PseAAC via Chou's 5-steps Rule," Curr. Genomics, vol. 20, no. 4, pp. 275–292, 2019.

[18] S. A. Khan, Y. D. Khan, S. Ahmad, and K. H. Allehaibi, "N-MyristoylG-PseAAC: sequence-based prediction of N-myristoyl glycine sites in proteins by integration of PseAAC and statistical moments," Lett. Org. Chem., vol. 16, no. 3, pp. 226–234, 2019.

[19] Y. D. Khan, A. Batool, N. Rasool, S. A. Khan, and K.-C. Chou, "Prediction of nitrosocysteine sites using position and composition variant features," Lett. Org. Chem., vol. 16, no. 4, pp. 283–293, 2019.

[20] Y. D. Khan, F. Ahmed, and S. A. Khan, "Situation recognition using image moments and recurrent neural networks," Neural Comput. Appl., vol. 24, no. 7–8, pp. 1519–1529, 2014.

[21] Y. D. Khan, N. Amin, W. Hussain, N. Rasool, S. A. Khan, and K.-C. Chou, "iProtease-PseAAC (2L): A two-layer predictor for identifying proteases and their types using Chou's 5-step-rule and general PseAAC," Anal. Biochem., vol. 588, p. 113477, 2020.

[22] Y. D. Khan, N. Rasool, W. Hussain, S. A. Khan, and K.-C. Chou, "iPhosT-PseAAC: Identify phosphothreonine sites by incorporating sequence statistical moments into PseAAC," Anal. Biochem., vol. 550, pp. 109–116, 2018.

[23] S. J. Malebary, M. S. ur Rehman, and Y. D. Khan, "iCrotoK-PseAAC: Identify lysine crotonylation sites by blending position relative statistical features according to the Chou's 5-step rule," PLoS One, vol. 14, no. 11, 2019.

[24] N. Rasool, W. Husssain, and Y. D. Khan, "Revelation of enzyme activity of mutant pyrazinamidases from Mycobacterium tuberculosis upon binding with various metals using quantum mechanical approach," Comput. Biol. Chem., vol. 83, p. 107108, 2019.

[25] S. Kim et al., "PubChem substance and compound databases," Nucleic Acids Res., vol. 44, no. D1, pp. D1202–D1213, 2016.

[26] H. E. Pence and A. Williams, "ChemSpider: an online chemical information resource." ACS Publications, 2010.

[27] A. Gaulton et al., "ChEMBL: a large-scale bioactivity database for drug discovery," Nucleic Acids Res., vol. 40, no. D1, pp. D1100–D1107, 2012.

[28] K. Mohanraj et al., "IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics," Sci. Rep., vol. 8, no. 1, pp. 1–17, 2018.

[29] J. Yang and Y. Zhang, "Protein structure and function prediction using I-TASSER," Curr. Protoc. Bioinforma., vol. 52, no. 1, pp. 5–8, 2015.

[30] C. M. Labbé et al., "MTiOpenScreen: a web server for structure-based virtual screening," Nucleic

Acids Res., vol. 43, no. W1, pp. W448–W454, 2015.

[31] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," J. Comput. Chem., vol. 31, no. 2, pp. 455–461, 2010.

[32] M. Aljofan and A. Gaipov, "COVID-19 Treatment: The Race Against Time. Electron J Gen Med. 2020; 17 (6): em227." 2020.

[33] R. C. Bernardi, L. F. Milles, and H. E. Gaub, "NAMD as a Tool for In Silico Force Spectroscopy," Biophys. J., vol. 118, no. 3, p. 144a, 2020.

[34] K. Vanommeslaeghe et al., "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," J. Comput. Chem., vol. 31, no. 4, pp. 671–690, 2010.

[35] L. Dong, S. Hu, and J. Gao, "Discovering drugs to treat coronavirus disease 2019 (COVID-19)," Drug Discov. Ther., vol. 14, no. 1, pp. 58–60, 2020.

[36] Y. Wang et al., "Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial," Lancet, 2020.

[37] R. Chen, Q.-L. Qi, M.-T. Wang, and Q.-Y. Li, "Therapeutic potential of naringin: an overview," Pharm. Biol., vol. 54, no. 12, pp. 3203–3210, 2016.

[38] M. A. Alzohairy, "Therapeutics role of Azadirachta indica (Neem) and their active constituents in disease prevention and treatment," Evidence-Based Complement. Altern. Med., vol. 2016, 2016.