Taylor & Francis
Taylor & Francis Group

Check for updates

In silico virtual screening, characterization, docking and molecular dynamics studies of crucial SARS-CoV-2 proteins

Meshari Alazmi[a] (iD) and Olaa Motwalli[b]

[a]Department of Computer Science, College of Computer Science and Engineering, University of Ha'il, Ha'il, Kingdom of Saudi Arabia; [b]College of Computing and Informatics, Saudi Electronic University (SEU), Madinah, Kingdom of Saudi Arabia

Communicated by Ramaswamy H. Sarma

ABSTRACT

The ongoing pandemic COVID-19 (COrona Virus Immuno Deficiency-2019) which is caused by SARS-CoV-2 (Severe Acute Respiratory Syndrome–CoronaVirus-2) has emerged as a pandemic with 400,000 plus deaths till date. We do not have any drug or vaccine available for the inhibition of this deadly virus. The expedition for searching a potential drug or vaccine against COVID-19 will be of massive potential and favor. This study is focused on finding an effective natural origin compound which can put a check on the activity of this virus. We chose important proteins from the SARS-CoV-2 genome such as NSP4, NSP15 and RdRp along-with the human ACE2 receptor which is the first point of contact with the virus. Virtual screening and followed up molecular docking resulted in Baicalin and Limonin as the final lead molecules. 200 ns of MD simulation for each protein-ligand complex provides the insights that Baicalin has a potential to target NSP4, NSP15 and RdRp proteins. Limonin which is largely used in traditional Indian medicine system is found to inhibit the human ACE2 receptor (making it inefficient in binding to the receptor binding domain of SARS-CoV-2). Our studies propose Baicalin and Limonin in combination to be studied *in vitro* and *in vivo* against COVID-19.
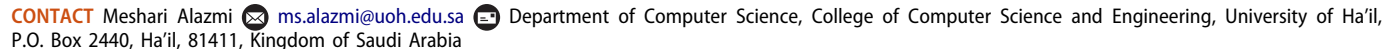
Introduction

SARS-CoV-2 is a single-stranded RNA virus discovered in the late 2019 in the city of Wuhan in China. This particular disease has been named as COVID-19, with common symptoms such as loss of smell, muscle ache, shortness of breath and fever with cough (Aanouz et al., 2020; Alagaili et al., 2014; Bhakkiyalakshmi et al., 2016; Bonkowski & Sinclair, 2016). The difficulty in breathing is caused by the virus attacking the T-lymphocytes in the lungs destroying the protecting proteins. This destruction builds up fluid in the lungs causing breathing difficulties (Wang et al., 2020). The disease has now spread globally with numbers as high as 8,466,256 humans infected taking up 451,943 deaths until date (https://www.worldometers.info/coronavirus/?utm_campaign=homeAdUOA?Si). World Health Organization (WHO) has declared this disease as an ongoing pandemic. COVID-19 has the capability to spread human to human through droplet exchange while coughing, talking, and sneezing (Chan et al., 2020). People with other ailments such as diabetes, heart problems, cancer, asthma, etc. are reported to show critical symptoms (Velavan & Meyer, 2020). Coronaviruses have affected humans in the past in the form of SARS-CoV and MERS-CoV, but SARS-CoV-2 has proven to be much invasive with higher rates of infection and deaths than the previous ones (Li et al., 2020).

The novel SARS-CoV-2 is a spherical shaped enveloped particle having single-stranded positive-sense or sense-stranded RNA, i.e.

their RNA encodes for mRNA (Fehr & Perlman, 2015). They have a very large genome of 26–30 kilo base pairs which encodes for variety of proteins involved in the mechanism of action (Fehr & Perlman, 2015). The two-third part of the RNA genomes is covered by the ORF1a/b, which produces the two viral replicase proteins that are polyproteins (PP1a and PP1ab). Sixteen mature nonstructural proteins (NSPs) arise from further processing of these two PPs. These NSPs take part in different viral functions including the formation of the replicase-transcriptase complex. The remaining genome part of the virus encodes the mRNA which produces the structural proteins, i.e. spike, envelope, membrane, and nucleocapsid, and other accessory proteins (Prajapat et al., 2020). The entry of virus in human body is facilitated by the interacting phenomenon between the Spike protein (S) of SARS-CoV-2 and the human ACE2 (Angiotensin-Converting enzyme-2) receptor (Yan et al., 2020). The trimeric Spike has a strong polar contact between ASP442 and ARG509 in the Receptor Binding Domain (RBD) of each monomer and for the human ACE2, strong H-bonds (TYR237.OH-VAL485.O, THR516.OH-THR229.OG1, TYR252.OHLEU156.O, THR567.OG1-TYR215.O) plays important role in Spike protein interaction with ACE2 protein. Other interactions are also reported in (Zhang, 2020).

In this current study we work on 4 proteins which are found to be very important when comes to virulence of the covid-19 disease. These 4 proteins are discussed as given below:
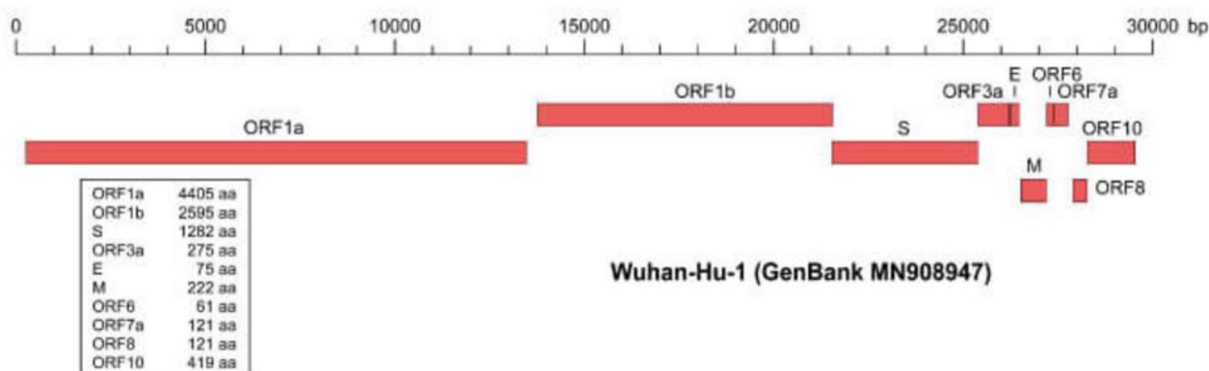
Figure 1. Proteins distributed in the SARS-CoV-2 genome. Figure taken from (https://thepathologist.com/subspecialties/the-covid-19-pandemic-a-summary).

1. Non-structural protein 4 (nsp4): Participates in the assembly of virally-induced cytoplasmic double-membrane vesicles necessary for viral replication (https://swissmodel.expasy.org/repository/species/2697049).

2. The RNA-dependent RNA polymerase [(RdRp), also named nsp12] is the central component of coronaviral replication and transcription machinery, and it appears to be a primary target for the antiviral drug Remdesivir (Gao et al., 2020).

3. Uridylate-specific endoribonuclease (NendoU/nsps15) an Mn^{2+}-dependent, uridylate-specific enzyme, which leaves 2'-3'-cyclic phosphates 5' to the cleaved bond (https://swissmodel.expasy.org/repository/species/2697049).

4. Human ACE2 (Angiotensin-Converting enzyme-2) receptor: ACE2 is a membrane-associated aminopeptidase expressed in vascular endothelia, renal and cardiovascular tissue, and epithelia of the small intestine and testes (Jia et al., 2005).

Various reports came in due course of time claiming to be the inhibitory molecules against SARS-CoV-2 (with limited knowledge of mechanism of action). Hydroxychloroquine has been reported as a potential drug acting against this disease (Liu et al., 2020) [13], where recently some groups have also suggested Remdesivir to be the factor of inhibition (Shannon et al., 2020). This study is devoted to find some potent compounds through natural sources which can act against this deadly disease and will be available in abundance as well as act with least side-effects. Structural biology and computer aided drug designing approach has been very effective in the past for various diseases. These important proteins (NSP4, NSP15, and RdRp) which are well-distributed in the SARS-CoV-2 genome (Figure 1) (https://thepathologist.com/subspecialties/the-covid-19-pandemic-a-summary) needs detailed study of their structure and function which will provide novel insights to design an effective, cost friendly drug with least side effects.

Methodology

Protein preparation

Proteins used in this study are the ones which are majorly involved in the mechanism of action of the novel SARS-CoV-2.

Experimentally solved 3D structures of the proteins non-structural protein 15 (nsp15) (PDB ID:6VWW) and RNA-dependent polymerase (RdRp) (PDB ID: 6YYT), human angiotensin receptor (ace2) (PDB ID:6M1D, chain B) were used for this study. 6M1D_B, 2KNC_A, 2KNC_B are three templates used for preparing ace2 protein in full (as experimental structure has some missing residues and loops). Fourth protein of our interest is Non-structural protein 4 (nsp4) whose experimental structure is not available. This protein is very important in viral genome replication. The sequence of nsp4 is downloaded from Uniprot (Boutet et al., 2016) (UniProtKB/Swiss-Prot: P0DTC1.1). This particular structure was produced using Phyre2 webserver (Kelley et al., 2015).

Saves server (Eisenberg et al., 1997; Laskowski, MacArthur, & Moss, 1993; Laskowski, MacArthur, Moss, et al., 1993; Laskowski et al., 1994; Laskowski & MacArthur, 2006) is used to verify the homology model of NSP4. Errat Quality Factor (Colovos & Yeates, 1993) for NSP4 is 80.4878. Only 45.40% of the residues have averaged 3D-1D score $>= 0.2$ when verified using Verify3D (Eisenberg et al., 1997). Ramachandran plot (Carugo & Djinovic-Carugo, 2013; Ramachandran et al., 1963) analysis revealed that out of 500 residues 65.9% residues are falling under the core region, 26.6% in allowed region, 4.9% generously allowed region, and 2.7% in disallowed region. Transmembrane helix prediction for nsp4 is also carried out using phyre2 webserver and 5 transmembrane helices (255–271, 275–305, 311–334, 338–367, and 371–395) were calculated. All the proteins were bring down to the least energy state using YASARA energy minimization server (Krieger et al., 2002) before further analyses. Binding site of all the proteins were obtained from (Pal & Talukdar, 2020) and given in the Table 1.

Virtual screening and ligand preparation

The natural compound library (containing approx. 100,000 compounds) was obtained from ZINC database (Sterling & Irwin, 2015) for the best optimal hit against these mentioned targets. PyRx (Dallakyan & Olson, 2015) from MGLTools (https://ccsb.scripps.edu/mgltools/) is used for the virtual screening using default settings and 10 best hits has been obtained. These 10 compounds were subjected to site-specific docking to finally select 5 compounds with the best docking score for further analysis. The respective PUBCHEM

Table 1. AutoDock mediated docking parameters like box type and grid box information for 4 protein-ligand complexes. The active site residue information for the respective proteins is also given as per the literature.

Complex	Box type	X	Y	Z	Active site residues
NSP15_Baicalin	Cube	24.03	24.03	24.03	Gln245, Gly248, Lys290, His235, His250, Thr341
NSP4_Baicalin	Cube	24.07	24.07	24.07	Ile494, Asp459, Thr461, Leu417, Cys418, Pro489
ACE2_Limonin	Cube	27.09	27.09	27.09	Pro426, Gln442, Ile439, Leu440
RdRp_Baicalin	Cube	40	40	40	Asp499, Lys545, Ser682, Thr687, Arg553, Lys551, Asn691, Leu786, Ser759, Asp618, Asp761, Lys783

identification numbers for these compounds further taken for site specific docking are Baicalin (64982), Kaempferol (5280863), Limonin (179651), Nimbolide (12313376) and Quercetin (5280343). The drugs were downloaded in the 3D-SDF format and geometry optimization was performed using UCSF-Chimera (Pettersen et al., 2004). Further energy minimization of all the compounds was performed using Avogadro (Hanwell et al., 2012).

Molecular docking studies

The protein-ligand binding mechanism of the chosen protein-ligand complexes was performed using Autodock 4.2 (Bikadi & Hazai, 2009; Forli et al., 2016). The docking analyses were performed using semi-flexible docking approach. In this study proteins are kept rigid and ligands were kept flexible. The allowed degrees of freedom for ligand molecules are 10. The steps involving conversion of molecules into pdbqt format, box type, grid box generation, etc. are specified by AutoDock. The box type and grid box parameters for all 4 complexes along-with the active site residue information is given in Table 1. The grid box was made keeping active site in the center of the box. Exhaustiveness of 100 was used to get the best output which takes more computational power and time for the analyses. However, the larger exhaustiveness gives better output. The docking poses with the least energy were analyzed using Discovery studio visualizer (Biovia, 2017).

Toxicity prediction

We also performed toxicity prediction of the selected 5 compounds to check and verify the drugs being least toxic for human use. The analyses were performed using ProTox-II, a virtual lab for the prediction of toxicities of small molecules (Banerjee et al., 2018). The drugs were uploaded to the server which yielded results showing the toxicity prediction in comparison to the already reported drugs such as Hydroxychloroquine and Remdesivir.

Molecular dynamics simulations

Molecular dynamics simulations to confirm and calculate the stability, fold and interactions of the docked proteins, 200 ns all-atom simulations were set-up for each complex. The parameters for the ligands were produced using Antechamber module of AmberTools18 (Case et al., 2005). The topology parameters were then converted to Gromacs input. The Amber forcefield ff14sb was used for to test the interactions. Intramolecular as well as intermolecular interactions were taken into consideration. The ions were neutralized and

simulations were performed in the presence of 150 mM NaCl. The energy minimization was performed and the system was equilibrated with NVT and NPT ensembles for 10 ns each. The temperature was set to 300 K. Production run was then started for all the complexes at 200 ns each with a threshold set up at 8 Å. The system stability was calculated using Gromacs 2019 (Van Der Spoel et al., 2005) and VMD analysis scripts (Humphrey et al., 1996). The molecular dynamics simulations including minimization, equilibration and production runs were carried out in triplicates to ensure stable conformations over systems.

Analysis for MD simulations

Molecular dynamics simulations were analysed using Gromacs analysis scripts for RMSD, RMSF, total energy, SASA, radius of gyration, h-bonds, and secondary structure prediction while xmgrace was used to plot the figures.

MMPBSA.py module was used to calculate the free energy and interaction energy of the ligand. The mathematical formula used to calculate the energies was:

$$\Delta G_{bind.solv} = \Delta G_{bind.vaccum} + \Delta G_{solv.complex} - \left(\Delta G_{solv.ligand} + \Delta G_{solv.receptor}\right)$$

The solvation energy for all the states was calculated using Generalized Born (GB) and Poisson Boltzman (OB). This analysis revealed the electrostatic contribution of the solvation state.

Results and discussion

Protein modeling

The protein modeling for the Non-Structured Protein (NSP4) performed using Phyre2 webserver (Kelley et al., 2015) yielded a high confidence full length model. Figure 2 shows the cartoon representation and surface representation of the NSP4 protein. Rests of all the proteins were also corrected via Phyre2 webserver. Supplementary Figure 1 shows the transmembrane embedded structure of NSP4 made by using MEMEMBED webserver (Nugent & Jones, 2013).

The other protein targets used were experimentally solved structures deposited in the PDB. The protein Nsp15 is responsible for the interference of protein with the innate immune response (Fährrolfes et al., 2017). Further studies have indicated that the mechanism of action of NSP15 is independent of the endonuclease activity (Liu et al., 2019). The independence of the endonuclease activity marks our third protein target which is RdRp, known as RNA-dependent polymerase (PDB ID: 6M17). RdRp is a very important target as this particular enzyme drives the replication of the viral

genome (Venkataraman et al., 2018). The protein on the human side (ACE2 receptor – PDB ID 6M1D) is the first entry point for the SARS-CoV-2 to playing the crucial role for the

host cells allowing the entry of the viral genome (Yan et al., 2020).

Virtual screening and ligand selection

Initial virtual screening of 100,000 natural compounds obtained from Zinc database yielded 10 best compounds. Supplementary Table 1 comprises top 10 hits obtained after virtual screening with respective binding energies in kcal/mol. These 10 best hits were then subjected to site specific molecular docking. The compounds screened were mostly phytochemicals from natural sources which showed capability of biological activities such as anti-oxidants, anti-viral, anti-inflammatory, anti-bacterial, anti-cancer and anti-microbial.

Toxicity prediction and ligand characteristics

The ligands were then subjected to drug likeliness and toxicity prediction using online tool named ProTox-II (Banerjee et al., 2018). In ProTox-II (Banerjee et al., 2018) there are 6 classes for toxicity (1 to 6) in which class 1 has $LD50 \leq 5$ which is fatal in nature on the other hand class 6 shows $LD50 > 5000$ which means compound is non-toxic. This study indicated how likely and effective a drug could be with least side effects and also informs us with a prediction score. The analysis revealed that Baicalin was found to be the best
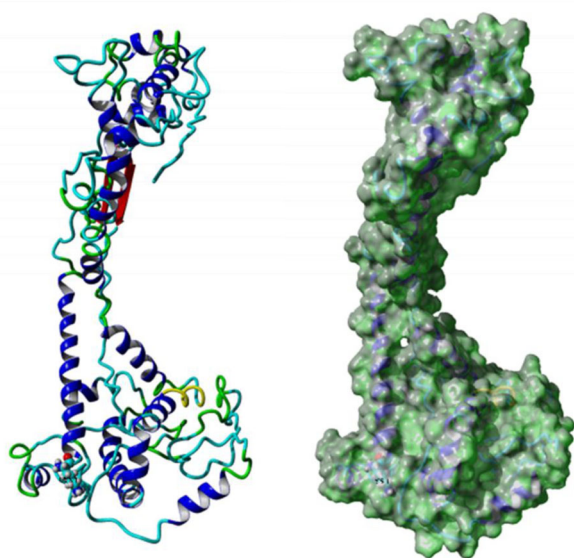

Figure 2. The NSP4 model from Phyre2 server showing cartoon representation (left) and molecular surface representation (right). Residue Lys1 is shown in ball model (in bottom) in cartoon figure. Figure is made by YASARA view.

Table 2. Shortlisted compounds with their drug parameters and toxicity report carried out using ProTox-II server (Banerjee et al., 2018).
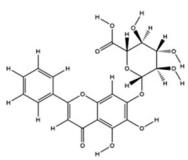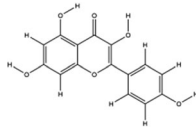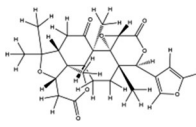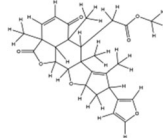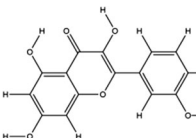
Compound	Structure	Mol. Wt (g/mol)	XLogP3-AA	H-bond Donor	Hbond Acceptor	Rotatable Bonds	Predicted LD50 (mg/kg)	Toxicity Class
Baicalin		446.4	1.1	6	11	4	5000	5
Kaempferol		286.2	1.9	4	6	1	3919	5
Limonin		470.5	1.8	0	8	1	244	3
Nimbolide		466.5	2.2	0	7	4	1000	4
Quercetin		302.2	1.5	5	7	1	159	3

Table 3. Molecular docking score of the final four protein-ligand complexes as obtained from *in-silico* docking experiment done using AutoDock 4.2 (Bikadi & Hazai, 2009). Binding energy values are defined in kcal/mol and Kd values are in micromolar range.

| Protein-ligand complex	Binding Energy (Kcal/mol)	Dissociation constant [Kd] (μM)	Polar contacts	Non-polar contacts
NSP4-Baicalin	−6.8 ± 0.78	9.009 ± 1.5	Ser496	Ile494, Leu417, Lys86
NSP15-Baicalin	−7.4 ± 0.52	3.489 ± 0.91	Lys290, Thr341, Leu346, Ser294, Tyr343	Pro344, Lys345, Trp333
RdRp-Baicalin	−8.7 ± 0.35	0.419 ± 0.25	Asp167, Lys624, Thr622, Asp621, Lys801, Ser798	Pro623, Tyr458, Asp626, and Arg627
ACE2-Limonin	−11.0 ± 0.18	0.007 ± 0.003	−	Phe532, Ile416, Phe597, Tyr596, Leu419, Phe447
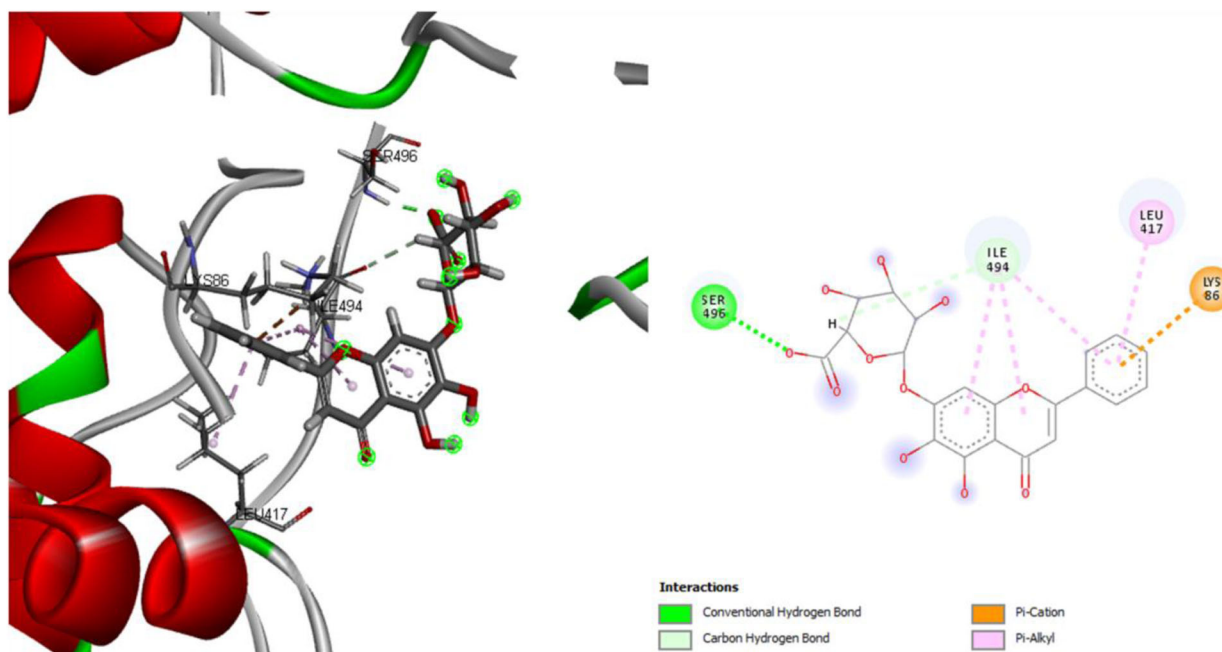
Figure 3. Docking pose (left) and Ligand interaction diagram (right) for protein-ligand interaction with NSP4 and baicalin shown in green sticks while interacting residues are labeled and shown in red sticks.

possible drug amongst the shortlisted compounds. The molecular weight and other parameters for the drugs were also found to be fitting well with the Lipinski rule of 5 for drug likeliness. The drug likeliness parameters with toxicity prediction are displayed in Table 2.

Molecular docking studies

Molecular docking performed with AutoDock 4.2 (Bikadi & Hazai, 2009; Danish, 2013; Forli et al., 2016; Seeliger & de Groot, 2010) further strengthened our study in finding an effective drug against this deadly disease. The drug binding scores in the form of kcal/mol for all drugs and targets are mentioned in Table 3.

Docking analyses and drug interactions

The best compound obtained after docking which binds effectively with NSP4. NSP15, and RdRp is Baicalin (the high-est-ranked bound compound in all poses made as a cluster at the ligand binding sites of these 3 target proteins). While Limonin binds best with human ACE2 receptor. The analyses

were performed target by target to check the efficiency of the ligand and the state of interaction using Discovery studio visualizer (Biovia, n.d.). as mentioned here below:

Target 1 – NSP4 (non structured protein 4)

NSP4 was the unstructured protein modeled using Phyre2 web-server (Kelley et al., 2015) and then further minimized using YASARA energy minimization server to be used to per-form docking studies. The docking NSP4 gave us the best hit with Baicalin. Analysis of docking results revealed that the drug binds in the active site of the protein, interacting with import-ant residues with polar and non-polar interactions. Key interac-tions involved the presence of H-bond with Ser496, hydrophobic interactions with Ile494 and Leu417. A Pi-cation interaction with Lys86 is also observed. The docking pose and ligand interaction diagram are displayed in Figure 3. The minimum binding energy for this interaction was found to be −6.8 kcal/mol. The interaction inhibiting the NSP4 protein will help in the crucial and important step in the replication mechanism and translation of the genome upon entry to the host organism.
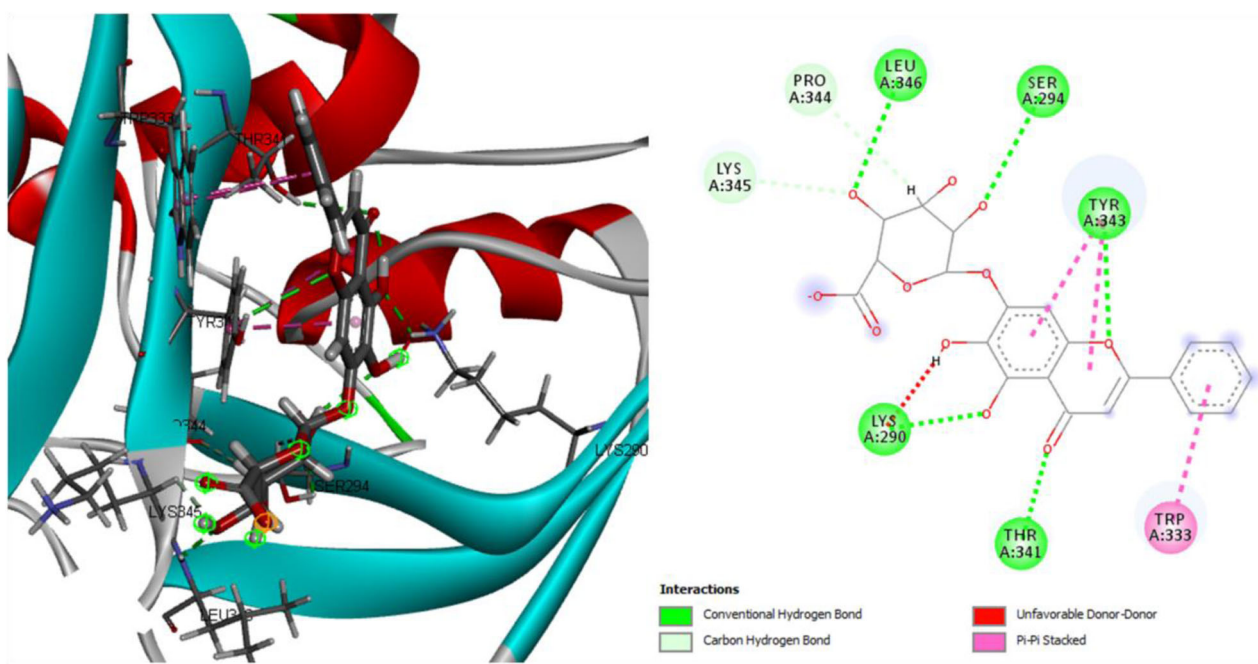
Figure 4. Docking pose (left) and Ligand interaction diagram (right) for protein-ligand interaction with NSP15 and baicalin shown in green sticks while interacting residues are labeled and shown in red sticks.
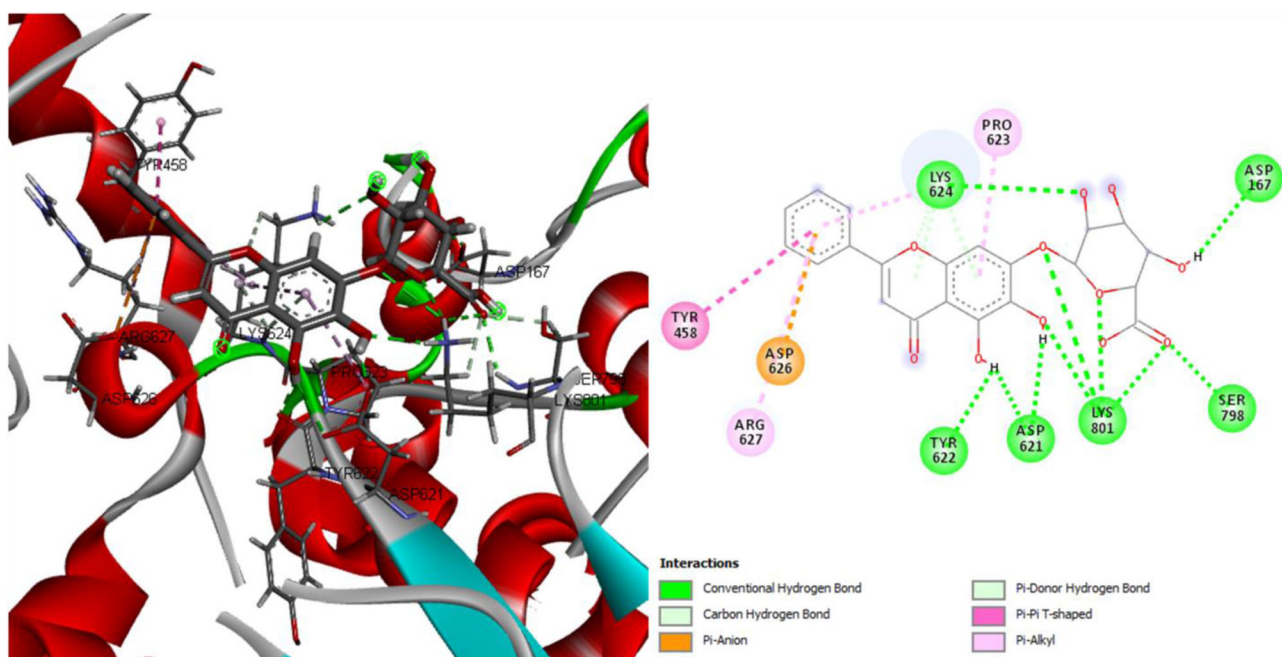

Figure 5. Docking pose (left) and Ligand interaction diagram (right) for protein-ligand interaction with RdRp and baicalin shown in green sticks while interacting residues are labeled and shown in red sticks.

Target 2 – NSP15 (uridylate-specific endoribonuclease)

Nsp15 is responsible for the interference of protein with the innate immune response. Further studies have indicated that the mechanism of action of NSP15 is independent of the endonuclease activity. The docking result of NSP15 with Baicalin revealed a docking score of −7.4 kcal/mol in the least energy conformation. Further ligand interaction analyses revealed NSP15 having important interactions with active site residues such as Lys290, Thr341 both forming H-bonds. Apart these two important residues, Baicalin forms H-bonds with Leu346, Ser294, Tyr343. Three non-polar interactions

observed are Pro344, Lys345, and Trp333. Docking pose and ligand interaction diagram is shown in the Figure 4. The mechanism of action of NSP15 which is independent of any endonuclease activity can successfully predicted to be inhibited with this interaction.

Target 3 – RdRp (RNA-dependent RNA polymerase)

RdRp being a catalytic RNA replication protein which contains important cavity surrounding hydrophobic residues. The docking with RdRp gave us the best hit with Baicalin
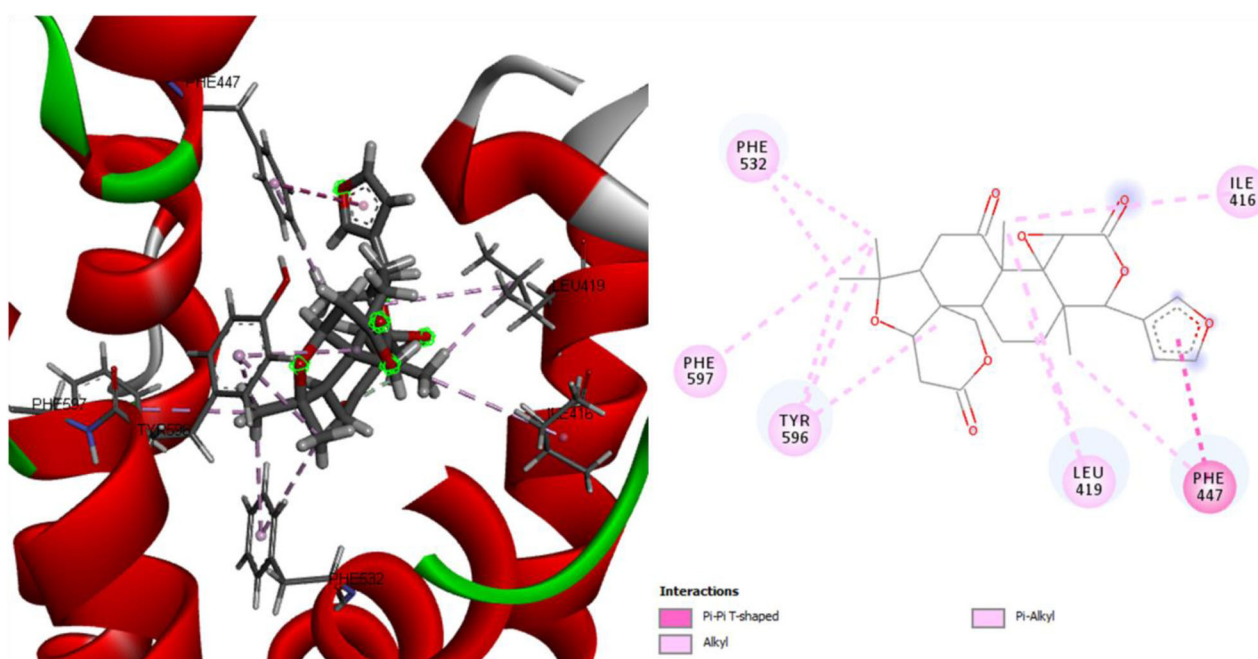
Figure 6. Docking pose (left) and Ligand interaction diagram (right) for protein-ligand interaction with ACE2 and Limonin shown in green sticks while interacting residues are labeled and shown in red sticks.

which binds in the very close vicinity of the active site. Although neighboring residues of the exact active site residues are interacting with the ligand, no active site residue is interacting. Therefore, docked placement of the ligand can be considered in the active site. Asp167, Lys624, Thr622, Asp621, Lys801, Ser798 residues form H-bonds with the ligand. Pro623, Tyr458, Asp626, and Arg627 form non-polar interactions.

All these interactions suggest that Baicalin binds very close to the active site and hence can be very potent and active in inhibiting the enzyme. The docking pose and ligand interaction diagram are shown in the Figure 5. The least energy state of this interaction revealed the binding energy to be −8.7 kcal/mol. This particular interaction will be important in inhibiting the complex mechanism by which RdRp drives the replication of the viral genome.

Target 4 – human ACE2 (angiotensin-converting enzyme 2)

Finally, the last target of our study was with an important protein on the host side, the ACE2 receptor (which is an important first contact of the novel CoV-2). Inhibiting this particular receptor will successfully make the SARS-CoV-2 devoid on attaching to the host cell and transferring its genome for replication and translation. Herewith the best interaction found to be acting against this particular protein was with limonin also known to cause the citrus bitterness with lots of biological activities. In the case of ACE2, we are getting results similar to the RdRp. Here also the ligand limonin is binding in the exact vicinity of the exact binding site and no active site residue is observed to have any interaction with the ligand. All non-polar interactions are observed. Residues Phe532, Ile416, Phe597, Tyr596, Leu419, and

Phe447 participated in the ligand binding. Docking pose and ligand interaction diagram is shown in Figure 6. The binding energy obtained is −11.0 kcal/mol. The target ACE2 used was in the open conformation, and once bound to Limonin, it will not be possible for the ACE2 inhibitor to transit in its closed conformation and have the binding capacity to the attacking virus.

Molecular dynamics simulations

Post-docking the selected compounds for interaction against respective targets, we moved on to perform molecular dynamics simulations. The experiment was performed in triplicates to minimize the error and artifacts while achieving better stability over systems. This step is vital to test the stability, performance, and optimization of the bound ligands to proteins. All production run trajectories of 200 ns each were analyzed using Gromacs and VMD analysis tool. The RMSD and RMSF were calculated using gmx_rms and gmx_rmsf commands. The system was found to reach equilibrium around 50 ns time frame. RMSD graphs for backbone and alpha carbon as shown in Figure 7 tells us that all trajectories reach equilibrium. Slight deviation around 70 ns has been observed in the case of NSP4-baicalin complex. As the protein is membrane embedded and elongated, such deviation is natural to be observed. This change in the RMSD may also be because NSP4 having unstructured regions with loops making the structure more dynamic in comparison to the globular proteins. Rest of the trajectories of other proteins are quite stabilized and equilibrated showing less deviation and dynamics.

As we run the MD simulation for all 4 protein-ligand complexes (200 ns each) in triplicates, the average RMSD over three runs for C-alpha atom and also for backbone suggests
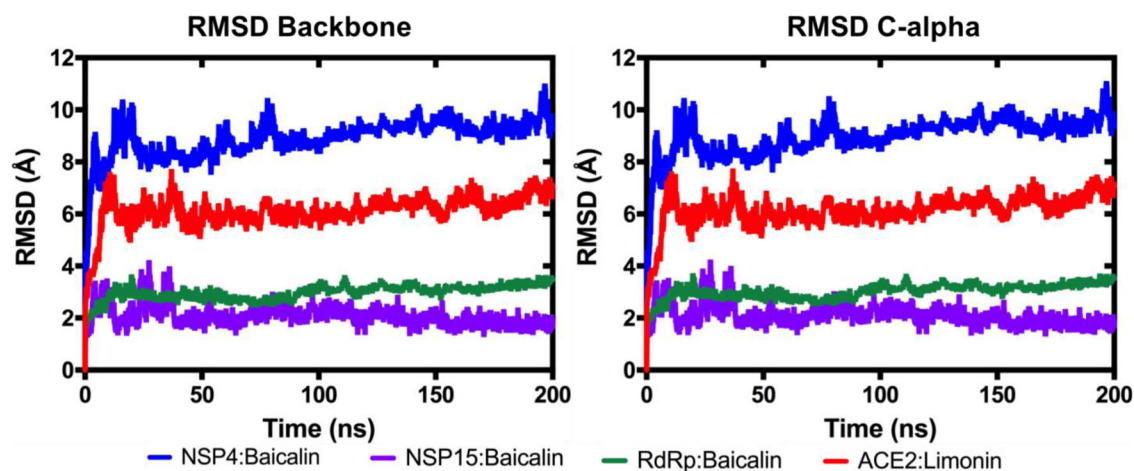
Figure 7. Molecular dynamics simulations for the protein: ligand complexes. RMSD plots showing structural stability for **(A)** NSP4: Baicalin complex, **(B)** NSP15:Baicalin complex, **(C)** RdRp:Baicalin complex and **(D)** human ACE2:Limonin complex.
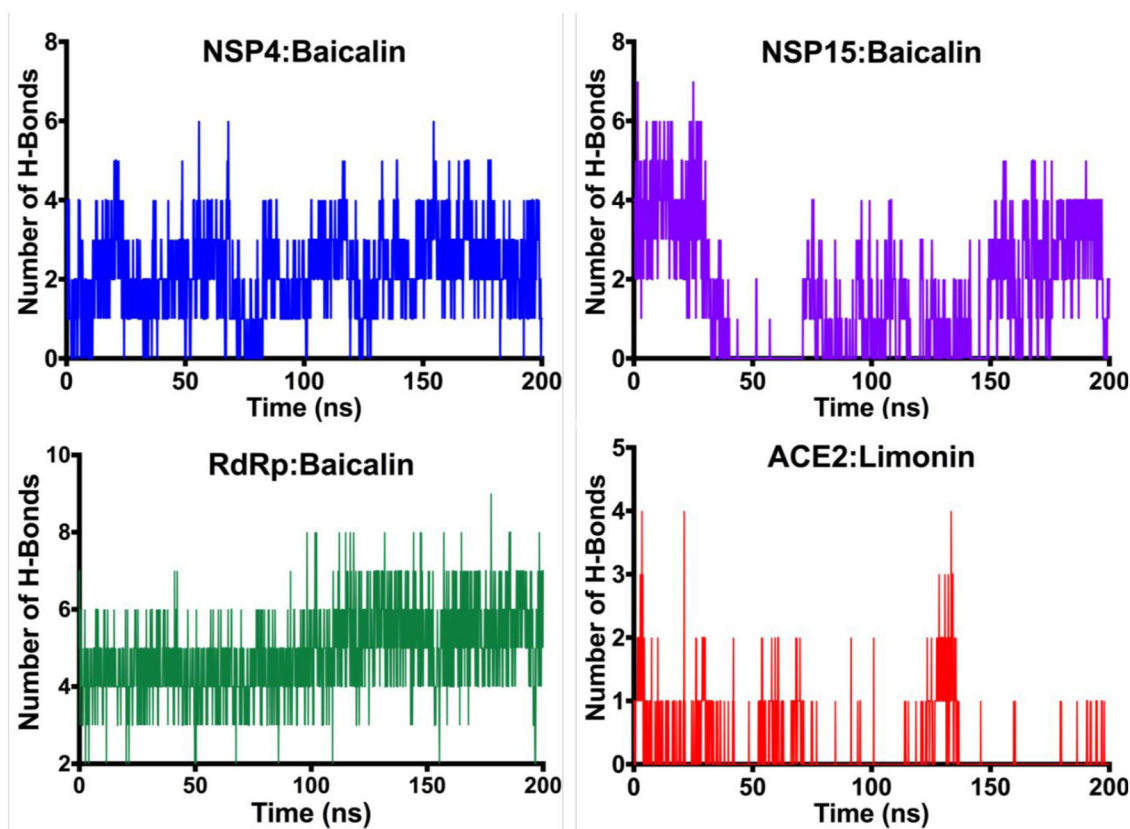

Figure 8. The number of H-bonds forming between protein and respective ligand in due course of simulation time of 200 ns. Average number of H-bonds formed between NSP4, NSP15 and RdRp with Baicalin are 4, 5 and 8 respectively. This number for ACE2-Limonin complex is around 4.

that all the runs are almost same. Using different initial velocities, the protein-ligand complexes achieve their global minima of equilibrium. This adds on that the simulation is free from errors and the ligand binding phenomenon with the respective protein is not by chance but an actually happening event. The values of average RMSD for C-alpha and Backbone are presented in the form of Supplementary Table 2.

We also checked the per-residue fluctuation using gmx_rmsf tool in GROMACS (Abraham et al., 2015; Berendsen et al., 1995; Van Der Spoel et al., 2005). RMSF (for alpha carbon atom) shown in Supplementary Figure 2B

analysis reveals that NSP15, RdRp, and ACE2 have more residues participating in bringing more fluctuation into the system. NSP4 residues are not playing that much part in increased fluctuation of the overall system while the RMSF plot per residue for backbone of the all protein complexes is shown in Supplementary Figure 2A.

Apart from RMSD and RMSF, number of Hydrogen bonds formed between protein and ligand during the simulation time is also calculated. Figure 8 shows the plots of these H-bonds forming between protein and respective ligand in due course of simulation time of 200 ns. Average number of H-bonds formed between NSP4, NSP15 and RdRp with Baicalin
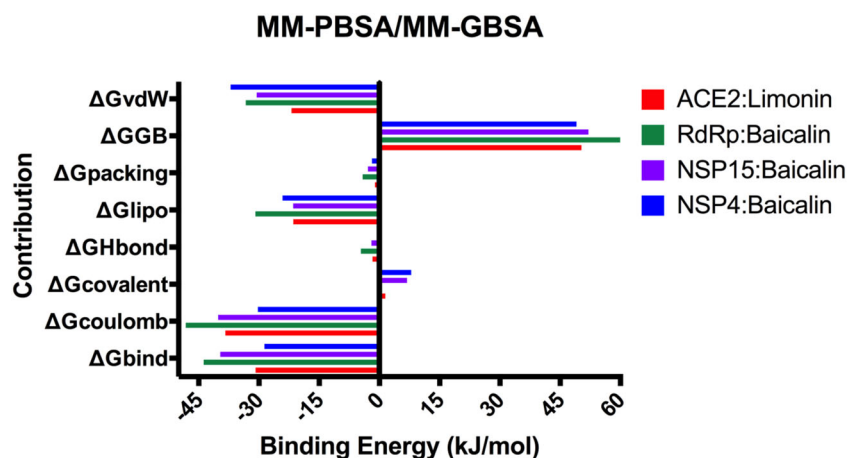
Figure 9. Values for all of the calculated binding energies and contributions for all 4 complexes were performed using MMPBSA.py.

Table 4. MM-PBSA/MM-GBSA results for all the 4 final complexes as obtained after 200 ns of simulation run using Gromacs.

Contribution	ACE2:Limonin	RdRp:Baicalin	NSP15:Baicalin	NSP4:Baicalin
ΔG_{bind}	−30.814	−43.776	−39.619	−28.625
$\Delta G_{coulomb}$	−38.357	−48.216	−40.154	−30.247
$\Delta G_{covalent}$	1.456	0.472	6.843	7.895
ΔG_{Hbond}	−1.708	−4.621	−2.011	−0.085
ΔG_{lipo}	−21.462	−30.884	−21.493	−24.125
$\Delta G_{packing}$	−1.135	−4.173	−2.856	−1.875
ΔG_{GB}	50.272	59.941	52.048	49.054
ΔG_{vdW}	−21.881	−33.294	−30.548	−37.051

are 4, 5 and 8 respectively. This number for ACE2-Limonin complex is around 4. Supplementary Figure 3A depicts the Radius of Gyration plots for all the 4 complexes over 200 ns of simulation time. As we can observe that Rg is decreasing in all cases over the time. It suggests that binding of ligands helps in the stabilization and compactness of the protein. The radius of gyration (Rg) of a particle is the root-mean-square distance of all electrons from their center of gravity. It is an important parameter and is often useful as an indicator for structural changes of a substance. Changes studied through the use of the radius of gyration are, for instance, association and dissociation effects, conformational changes by denaturation, binding of coenzymes, and temperature effects (O. Kratky, P. Laggner, in Encyclopedia of Physical Science and Technology (Third Edition), 2003).

Solvent accessible surface area for all the proteins was also calculated to check the effect of ligand binding on the residue profiling of the surface of the protein. Supplementary Figure 3B shows Solvent accessible surface area (SASA) plots for all 4 protein-ligand complexes. NSP4 and ACE2 shows decreasing SASA on the other hand, RdRp and NSP15 shows more or less same SASA profiling even after ligand binding. This suggests that NSP4 and ACE2 undergo crucial structural changes upon binding of the ligand.

The total energy of the complexes and individual energy components are depicted in Supplementary Figure 3C and Figure 9. Individual energy components like Vander walls forces, coulomb, H-bonds are calculated using MM-PBSA/MM-GBSA tool in Gromacs. A table (Table 4) representing the values is also included in the text. The reason being NSP4 is an unstructured protein with high number of loops and dynamics. The fluctuations were however mostly found in

the regions for the ligand binding, corroborating our previous analyses. These analyses help us conclude that Baicalin as well as Limonin exhibit as potential targets for SARS-CoV-2 and human ACE2 binding proteins. Supplementary Figure 4 shows the secondary structure change in the protein upon ligand binding in due course of 200 ns simulation run. Overall secondary structure change is nevertheless same except in NSP4 as it is a homology model and a membrane protein too having lots of elongated helical structures.

Conclusion

COVID-19 is the biggest pandemic world is facing right now after the Spanish Flu. SARS-CoV-2 emerged as the deadliest pathogen of the century. Genome of corona virus 2 and the resultant proteins are quite complex to study. An urgent check on this virus is the hottest topic of the hour. However, structural biology approach which has been working with all diseases in the past has been found to be the fastest, cheapest and reliable to discover drug against deadly diseases. We performed computer-aided drug discovery process against the important proteins involved in the mechanism of action for SARS-CoV-2. Three proteins (NSP4, NSP15 and RdRp) were taken from the SARS-CoV-2 genome, while ACE2 being the first point of contact upon virus entry into humans was also studied. Our results reveal that Baicalin, a herbal supplement and a plant based natural compound has the capability to inhibit SARS-CoV-2 target proteins (NSP4, NSP15 and RdRp) in the least energy conformation. Baicalin has been known as an important flavonoid with anti-cancer and anti-xiolytic effects. The human ACE2 receptor was however found to be interacted best with Limonin, which is found in the citrus peels giving bitterness. It has been largely used as traditional Indian medicine with a lot of biological activities, though some people have disliked it due to its bitter and pungent taste. After conducting this work, we suggest Baicalin and Limonin (separate or in combination) as Natural lead compounds to act against the novel SARS-CoV-2. Present study can be very helpful in stopping the COVID-19 pandemic to grow more. This *in-silico* analysis can be quickly combined with expedited research on the experimental side to unveil

the importance of naturally occurring compounds acting against this deadly disease.

Disclosure statement

The authors declare no conflict of interest.

Funding

ORCID

Meshari Alazmi 🔵 http://orcid.org/0000-0001-9074-1029

References

Aanouz, I., Belhassan, A., El-Khatabi, K., Lakhlifi, T., El-Ldrissi, M., & Bouachrine, M. (2020). Moroccan medicinal plants as inhibitors against SARS-CoV-2 main protease: Computational investigations. *Journal of Biomolecular Structure and Dynamics*, 1–9. https://doi.org/10.1080/07391102.2020.1758790

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

Alagaili, A. N., Briese, T., Mishra, N., Kapoor, V., Sameroff, S. C., Burbelo, P. D., de Wit, E., Munster, V. J., Hensley, L. E., Zalmout, I. S., Kapoor, A., Epstein, J. H., Karesh, W. B., Daszak, P., Mohammed, O. B., & Lipkin, W. I. (2014). Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *MBio*, 5(2), 14. https://doi.org/10.1128/mBio.01002-14

Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research*, 46(W1), W257–W263. https://doi.org/10.1093/nar/gky318

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E

Bhakkiyalakshmi, E., Sireesh, D., Sakthivadivel, M., Sivasubramanian, S., Gunasekaran, P., & Ramkumar, K. M. (2016). Anti-hyperlipidemic and anti-peroxidative role of pterostilbene via Nrf2 signaling in experimental diabetes. *European Journal of Pharmacology*, 777, 9–16. https://doi.org/10.1016/j.ejphar.2016.02.054

Bikadi, Z., & Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *Journal of Cheminformatics*, 1, 15https://doi.org/10.1186/1758-2946-1-15

Biovia, D. (2017). *BIOVIA Discovery Studio 2017 R2: A comprehensive predictive science application for the life sciences*. BIOVIA Discovery Studio.

Bonkowski, M. S., & Sinclair, D. A. (2016). Slowing ageing by design: The rise of NAD + and sirtuin-activating compounds. *Nature Reviews. Molecular Cell Biology*, 17(11), 679–690. https://doi.org/10.1038/nrm.2016.93

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., Poux, S., Bougueleret, L., & Xenarios, I. (2016). UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: How to use the entry view. *Methods in Molecular Biology*, 1374, 23–54. https://doi.org/10.1007/978-1-4939-3167-5_2

Carugo, O., & Djinovic-Carugo, K. (2013). Half a century of Ramachandran plots. *Acta Crystallographica. Section D, Biological Crystallography*, 69(Pt 8), 1333–1341. https://doi.org/10.1107/S090744491301158X

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16), 1668–1688. https://doi.org/10.1002/jcc.20290

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C.-Y., Poon, R. W.-S., Tsoi, H.-W., Lo, S. K.-F., Chan, K.-H., Poon, V. K.-M., Chan, W.-M., Ip, J. D., Cai, J.-P., Cheng, V. C.-C., Chen, H., Hui, C. K.-M., & Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*, 395(10223), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society*, 2(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Methods in Molecular Biology (Clifton, N.J.)*, 1263, 243–250. https://doi.org/10.1007/978-1-4939-2269-7_19

Danish, S. (2013). A simple click by click protocol to perform docking: AutoDock 4.2 made easy. *Excli Journal*, 12, 831.

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, 277, 396–404. https://doi.org/10.1016/s0076-6879(97)77022-8

Fährrolfes, R., Bietz, S., Flachsenberg, F., Meyder, A., Nittinger, E., Otto, T., Volkamer, A., & Rarey, M. (2017). ProteinsPlus: A web portal for structure analysis of macromolecules. *Nucleic Acids Research*, 45(W1), W337–W343. https://doi.org/10.1093/nar/gkx333

Fehr, A. R., & Perlman, S. (2015). Coronaviruses: An overview of their replication and pathogenesis. *Methods in Molecular Biology (Clifton, N.J.)*, 1282, 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1

Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols*, 11(5), 905–919. https://doi.org/10.1038/nprot.2016.051

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., Ge, J., Zheng, L., Zhang, Y., Wang, H., Zhu, Y., Zhu, C., Hu, T., Hua, T., Zhang, B., … Rao, Z. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (Science)*, 368(6492), 779–782. https://doi.org/10.1126/science.abb7498

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1), 17.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Jia, H. P., Look, D. C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C., Perlman, S., & McCray, P. B. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *Journal of Virology*, 79(23), 14614–14621. https://doi.org/10.1128/JVI.79.23.14614-14621.2005

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Krieger, E., Koraimann, G., & Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA-a self-parameterizing force field. *Proteins*, 47(3), 393–402. https://doi.org/10.1002/prot.10104

Laskowski, R., MacArthur, M., Moss, D., & Thornton, J. M. (1993). PROCHECK: A program to produce both detailed and schematic plots of proteins. *Journal of Applied Crystallography*, 24, 946–956.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. scripts.iucr.org

Laskowski, R. A., MacArthur, M. W., Smith, D. K., & Jones, D. T. (1994). PROCHECK v. 3.0. Program to check the stereochemistry quality of protein structures. Operating instructions.

Laskowski, R. A., & MacArthur, M. W. (2006). PROCHECK: Validation of protein structure coordinates.

Li, G., Fan, Y., Lai, Y., Han, T., Li, Z., Zhou, P., Pan, P., Wang, W., Hu, D., Liu, X., Zhang, Q., & Wu, J. (2020). Coronavirus infections and immune responses. *Journal of Medical Virology*, 92(4), 424–432. https://doi.org/10.1002/jmv.25685

Liu, J., Cao, R., Xu, M., Wang, X., Zhang, H., Hu, H., Li, Y., Hu, Z., Zhong, W., & Wang, M. (2020). Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. *Cell Discovery*, *6*, 16. https://doi.org/10.1038/s41421-020-0156-0

Liu, X., Fang, P., Fang, L., Hong, Y., Zhu, X., Wang, D., Peng, G., & Xiao, S. (2019). Porcine deltacoronavirus nsp15 antagonizes interferon-β production independently of its endoribonuclease activity. *Molecular Immunology*, *114*, 100–107. https://doi.org/10.1016/j.molimm.2019.07.003

Nugent, T., & Jones, D. T. (2013). Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics*, *14*, 276. https://doi.org/10.1186/1471-2105-14-276

Pal, S., & Talukdar, D. A. (2020). Compilation of potential protein targets for SARS-CoV-2: Preparation of homology model and active site determination for future rational antiviral design. *ChemRxiv*, 1–42. https://doi.org/10.26434/chemrxiv.12084468.v1

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

Prajapat, M., Sarma, P., Shekhar, N., Avti, P., Sinha, S., Kaur, H., Kumar, S., Bhattacharyya, A., Kumar, H., Bansal, S., & Medhi, B. (2020). Drug targets for corona virus: A systematic review. *Indian Journal of Pharmacology*, *52*(1), 56–65. https://doi.org/10.4103/ijp.IJP_115_20

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, *7*, 95–99. https://doi.org/10.1016/S0022-2836(63)80023-6

Seeliger, D., & de Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-Aided Molecular Design*, *24*(5), 417–422. https://doi.org/10.1007/s10822-010-9352-6

Shannon, A., Le, N. T.-T., Selisko, B., Eydoux, C., Alvarez, K., Guillemot, J.-C., Decroly, E., Peersen, O., Ferron, F., & Canard, B. (2020). Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 exonuclease active-sites. *Antiviral Research*, *178*, 104793. https://doi.org/10.1016/j.antiviral.2020.104793

Sterling, T., & Irwin, J. J. (2015). ZINC 15–ligand discovery for everyone. *Journal of Chemical Information and Modeling*, *55*(11), 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. https://doi.org/10.1002/jcc.20291

Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine & International Health: TM & IH*, *25*(3), 278–280. https://doi.org/10.1111/tmi.13383

Venkataraman, S., Prasad, B. V. L. S., & Selvarajan, R. (2018). RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses*, *10*(2), 76. https://doi.org/10.3390/v10020076

Wang, X., Xu, W., Hu, G., Xia, S., Sun, Z., & Liu, Z. (2020). SARS-CoV-2 infects T lymphocytes through its spike protein-mediated membrane fusion. *Cellular and Molecular Immunology*, 1–3. https://doi.org/10.1038/s41423-020-0424-9

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, *367*(6485), 1444–1448. https://doi.org/10.1126/science.abb2762

Zhang, J. (2020). Molecular dynamics studies on the COVID-19 trimeric spike glycoprotein, the human ACE2 ectodomain, and their bindings: Preliminary results.