Original Research Article

Development of SARS-CoV-2 Inhibitors Using Molecular Docking Study with Different Coronavirus Spike Protein and ACE2

Israa Mohamed Shamkh [1*]

Dina Pratiwi [2]

[1]Chemo and Bioinformatics Laboratory, Bio Search Research Institution, Giza, Giza Governorate, Egypt

[2]Department of Pharmacy, Sekolah Tinggi Farmasi Muhammadiyah Tangerang, Tangerang, Banten, Indonesia

*email:
esraa.m.ahmed@std.agr.cu.edu.eg

Abstract

The novel coronavirus SARS-CoV-2 is an acute respiratory tract infection that emerged in Wuhan city, China. The spike protein of coronaviruses is the main driving force for host cell recognition and is responsible for binding to the ACE2 receptor on the host cell and mediates the fusion of host and viral membranes. Recognizing compounds that could form a complex with the spike protein (S-protein) potently could inhibit SARS-CoV-2 infections. The software was used to survey 300 plant natural compounds or derivatives for their binding ability with the SARS-CoV-2 S-protein. The docking score for ligands towards each protein was calculated to estimate the binding free energy. Four compounds showed a strong ability to bind with the S-protein (neohesperidin, quercetin 3-*O*-rutinoside-7-*O*-glucoside, 14-ketostypodiol diacetate, and hydroxypropyl methylcellulose) and used to predict its docking model and binding regions. The highest predicted ligand/protein affinity was with quercetin 3-*O*-rutinoside-7-*O*-glucoside followed by neohesperidin. The four compounds were also tested against other related coronavirus and showed their binding ability to S-protein of the bat, SARS, and MERS coronavirus strains, indicating that they could bind and block the spike activities and subsequently prevent them infection of different coronaviruses. Molecular docking also showed the probability of the four ligands binding to the host cell receptor ACE2. The interaction residues and the binding energy for the complexes were identified. The strong binding ability of the four compounds to the S-protein and the ACE2 protein indicates that they might be used to develop therapeutics specific against SARS-CoV-2 and close related human coronaviruses.

INTRODUCTION

Coronaviruses (CoVs) belong to order nidovirales, family coronaviridae. Human coronaviruses cause respiratory infections associated with influenza-like illness ranging from the common cold to more severe symptoms[1]. The 21st century witnessed three outbreaks of human deadly pneumonia coronaviruses; Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) in 2003, Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in 2012, and SARS-like-CoV named 2019-nCoV (also known as SARS-CoV-2) in December 2019[2,3]. Genomic analysis on the coronavirus revealed that Bat coronavirus RaTG13 appears to be the closest relative to the SARS-CoV-2 compared to SARS-CoV[4]. The SARS-CoV-2, like other CoVs, are enveloped, positive-sense, long single-stranded RNA viruses and translate two groups of proteins, i.e., structural proteins such as Spike (S), Nucleocapsid (N), Matrix (M), and Envelope (E), as well as non-structural proteins, such as proteases and RNA-dependent RNA polymerase (RdRp)[5]. Coronaviruses depend on RdRp for the high frequency of RNA recombination and are among the

main factors that cause phenotypical and genotypical diversity of CoVs that make them capable of jumping across species[6].

The homotrimeric spike glycoprotein helps the virus initiate the infection by attaching to the host cell receptor, mediates virus fusion and genome entry into the host cell[7]. It is a large type I transmembrane protein composed of two subunits; the S1 subunit mainly contains a receptor-binding domain (RBD) responsible for recognizing the host cell surface receptor angiotensin-converting enzyme 2 (ACE2) and binding to it. The second subunit (S2) contains the basic elements required for the membrane fusion and entry into the host cells[8,9]. The SARS-CoV-2 S-protein and its interaction with the cell receptor ACE2 have been studied using cryo-EM, and results confirmed the function of the S1 and S2 subunits[10].

The 3D atomic scale of the SARS-CoV-2 S-protein was recently reported, and structural evidence that it binds to the ACE2 with 10- to 20-fold higher affinity than the SARS-CoV S-protein. Binding residues between the RBD in SARS-CoV-2 and ACE2 were determined and compared to the SARD-CoV[11,12]. Structural analysis showed highly conserved or shared similar side chain properties with those in the SARS-CoV RBD. The SARS-CoV-2 has an extended insertion containing short β5 and β6 strands, α4 and α5 helices, and loops, which represent the receptor-binding motif (RBM) containing most of the contacting residues of SARS-CoV-2 for ACE2 binding[13,14]. Two epitopes of two SARS-CoV antibodies targeting the RBD are also analyzed with the SARS-CoV-2 RBD, providing insights into the future identification of cross-reactive antibodies[15].

Scientists have focused on the SARS-CoV-2 S-protein as a key target for vaccines, therapeutic antibodies, and diagnostics. In fact, to discover a new vaccine and therapeutic antibody needs many years of laborious work[16]. The bioinformatics analysis approved a fast way to find potential molecules from the marketed drugs to develop a new drug against the SARS-CoV-2. Once the efficacy is determined, it can be approved by the Green Channel or approved by the hospital ethics committee for rapid clinical treatment[17]. Through this technology, several compounds, including natural plant compounds, have been screened and confirmed to directly inhibit the essential proteins responsible for viral entry and replication, such as S-protein of SARS or MERS coronavirus. Currently, commercial antiviral molecules and chemical compounds extracted from traditional Chinese medicinal herbs were investigated[18,19].

Molecular docking using AutoDock Vina is a popular tool used in the virtual screening of small molecules against proteins and is also used to investigate the interactions of natural products against the target protein[20]. Pharmacokinetic study and *in silico* absorption, distribution, metabolism, and excretion (ADME) modeling is used to speed up drug approval as it indicates if new compounds have side effects on human health[21]. In this study, a molecular docking system was performed to screen and select the binding affinity of the S-protein of SARS-CoV-2 as well as the host cell receptor ACE2 against some natural plant compounds or derivatives that might be used to block virion binding to host cells and subsequently prevent viral infection and spreading.

METHOD

Hardware and Software

The hardware used is a Notebook with specification Intel® Core™ i3-4005U CPU@1.70 GHz, 4 GB memory, 64-bit Windows 7 Operating System. Software used includes I-TASSER from University of Michigan (freeware, https://zhanglab.ccmb.med.umich.edu/I-TASSER/), SWISS-MODEL from Swiss Institute of Bioinformatics (freeware, https://swissmodel.expasy.org/), 3Drefine from University of Missouri (freeware, http://sysbio.rnet.missouri.edu/3Drefine/), SAMSON 2020 R3 1.0.0 from OneAngstrom (trial version, https://www.samson-connect.net/), Open Babel 3.1.1 from Open Babel development team (freeware, http://openbabel.org/wiki/Main_Page), SwissADME from Swiss Institute of Bioinformatics (freeware, http://www.swissadme.ch/), Discovery Studio Visualizer v20.1.0.19295 from Dassault Systemes BIOVIA (freeware, https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/visualization/), MEGA-X from Pennsylvania State University (freeware, https://www.megasoftware.net/), Clustal Omega 1.2 from European Bioinformatics Institute (freeware, https://www.ebi.ac.uk/Tools/msa/clustalo/), and

ESpript 3.0 from Lyon University (https://espript.ibcp.fr/ESPript/ESPript/).

Ligands

The ligands used in this study were quercetin 3-*O*-rutinoside-7-*O*-glucoside (PubChem ID 10190763), neohesperidin (232990), hydroxypropyl methylcellulose (57503849), and 14-ketostypodiol diacetate (SMILES CC1CCC2C3(C)CCC(=O)C(C)(C)C3CCC2(C)C14CC5=C(O4)C(=CC(=C5OC(C)=O)OC(C)=O)C).

Receptors

The receptor used were SARS-CoV-2 (QHD43416.1), SARS-GD01 (AAP51227.1), Bat coronavirus RaTG13 (QHR63300.2), MERS (QFQ59587.1), and ACE2 protein (PDB ID 6M0J).

Docking protocol

Preparation of SARS-CoV-2 S-protein structures

The sequence of the human CoVs spikes protein of the SARS-CoV-2 with GenBank accession no. QHD43416.1 was downloaded from National Centre for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/). The protein sequence was retrieved in the FASTA format as an amino-acid length sequence and used to build the 3D structure monomeric, trimeric, and trimeric binding models using I-TASSER. The 3D structure for the S-protein of other CoVs was built using the SWISS-MODEL. All water molecules and ligands were removed for pre-docking while hydrogen atoms were added to the target protein. In addition, affinity minimization was performed using the 3Drefine server. The docking system was built using SAMSON 2020.

Dataset and ligands selection

The 3D structure of 300 natural and synthetic compounds, which drive from the natural plants with drug-like properties and their derivatives, were selected. Sub-structural features of the ligand were carefully selected from references and separately downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/) in SDF format converted into MOL2 format using Open Babel.

Screening with SwissADME

Selected compound structures were converted to SMILES notations and submitted to the webserver for calculation and filtration by the SwissADME to identify the physicochemical features and predict the ADME parameters, drug-like nature, pharmacokinetic properties, and medicinal chemistry of the selected compounds. The ADME depends on collecting data and developing models to assess and predict pharmacokinetic properties. The compounds that become ready for docking with the target protein were reduced to 250 ligands using the SwissADME, depending upon their solubility and cytotoxicity to humans.

Spike protein-ligand docking

The SARS-CoV-2 S-protein model and other CoVs related to S-protein was docked against the test ligands using SAMSON 2020. This software used AutoDock Vina to maximize the accuracy of these predictions while minimizing the computer time. The program works based on quantum mechanics. It predicts the potential affinity, molecular structure, geometry optimization of the structure, vibration frequencies of coordinates of atoms, bond length, and bond angle[20]. Following the exhaustive search, 100 poses were analyzed, and the best scoring poses were used to calculate the binding affinity of the ligands. The ligands that tightly bind to a target protein with a high score were selected.

ACE2-ligand docking

The receptor protein ACE2 was downloaded from the RCSB protein data bank (https://www.rcsb.org/). Pre-docking was carried out as described using SAMSON 2020 and Discovery studio Visualizer. The 17 interacting amino acid residues (Gln24, Thr27, Phe28, Asp30, Lys31, His34, Glu35, Glu37, Asp38, Tyr41, Gln42, Leu79, Met82, Tyr83, Asn330, Lys353, and Gly354) that represent the bounded region of the ACE2 to the SARS-CoV-2 S-protein were previously reported[14] and were selected to study the interaction with the selected ligands using AutoDock Vina and SAMSON 2020 as described before.

Assessment

Virtual screening and docking parameter

Virtual screening utilized docking and scoring of each compound from the previous dataset. This technique was employed based on the prediction of binding modes and binding affinities of each compound using docking to four proteins structure (experimental protein and 3D structure models of the other proteins)[22]. The docking program behaves to get the docking parameter in the SAMSON 2020, in which the

program could make docking for a library of ligands with a single protein. By considering this, diverse compounds from plants and protein targets were evaluated. In general, it was important to visualize the docked poses of high-scoring compounds because many ligands were docked in different orientations. This kind of study becomes difficult when the size of the dataset increased. Therefore, it was important to eliminate unuseful compounds by SwissADME before docking by restricting the dataset to drug-like compounds and taking into considerations appropriate property, sub-structural features, solubility, and cytotoxicity to be deal with human use and eliminate the probability of side effect to get the best feature of the ligands then the docking was placed[23]. Hence, the bounded ligands were analyzed with Discovery Studio Visualizer, which was used to analyze and screen the ligand properties to reach the functional domain of protein in the human body.

Sequence alignment and phylogenetic tree

The amino acid sequences for 30 CoV S-proteins were obtained from the NCBI database. Alignments were applied to build a phylogenetic tree using the Mega X. For this alignment and constrict tree, the MUSCLE algorithm was used. The phylogenetic relationships among the 30 spike proteins were carried out using the nearest-neighbour interactions (NNI) with WAG+G+I substitution model and 500 bootstrap replicates. To reach a rational phylogenic tree, we eliminated proteins and repetitive sequences with the same species.

Multiple Sequence Alignment (MSA) of the phylogenetically with the closely related CoV S-proteins sequences were used to determine the conserved region of these sequences by Clustal Omega with default parameters. In addition, the ESpript 3.0 was used to align conserved sequences among the selected proteins and secondary structure of SARS-CoV-2 S-protein. Amino acid alignment of three related CoVs S-proteins was performed using the default parameters.

RESULTS AND DISCUSSION

Sequence structure analysis SARS-CoV-2 S-protein

Comparison of the amino acids of the SARS-CoV-2 S-protein for the modeling (GenBank accession no. QHD43416.1) and the experimental one (PDB ID 6VSB) showed that the sequences were identical in the N terminal for the two sequences (1208 amino acids) but different at the C terminal, as the modeling SARS-CoV-2 S-protein had a longer and different sequences of 82 amino acids from the corresponding 62 amino acids of the experimental one[24].

Protein model of SARS-CoV-2 S-protein

The predicted 3D structure of the SARS-CoV-2 S-protein model was built with I-TASSER (**Figure 1A**) using the published sequence at the NCBI. The built model was compared with the experimental one (Cryo-EM structure) downloaded from the Protein Data Bank (**Figure 1B**). The structure of the built model was divided into four regions to facilitate the comparison with the experimental one. The two models showed different configurations. The folding in the experimental one was more compact, and region "iv" was not presented. From these results, it was important to state that although there was a similarity between the two models, the differences between these two spike proteins could mislead the scientists when depending on the model folding in docking with any compounds, and hence, the best results were obtained when using the experimental data[25].

The homotrimer of the experimental spike-protein for the SARS-CoV-2 was used to design the ligand-protein interactions with the four selected ligands (**Figure 2**) using the SAMSON 2020. The resembled complexes were analyzed using Discovery Studio Visualizer to resemble the ligand with the interacting residues and how the ligand 3D structure allows the binding with the ligands.

A

B
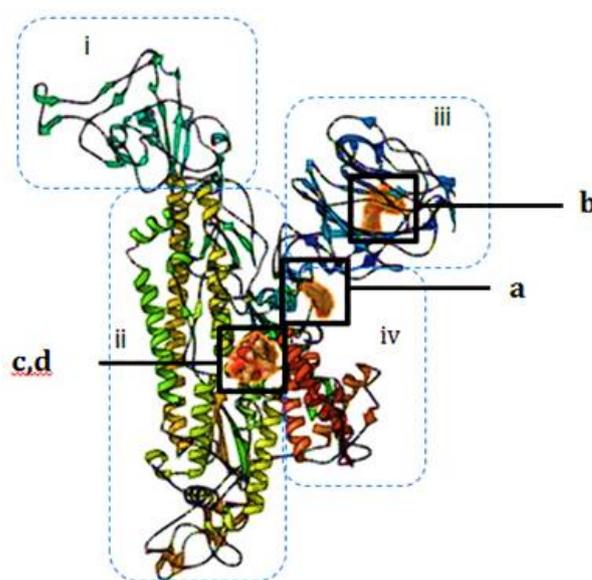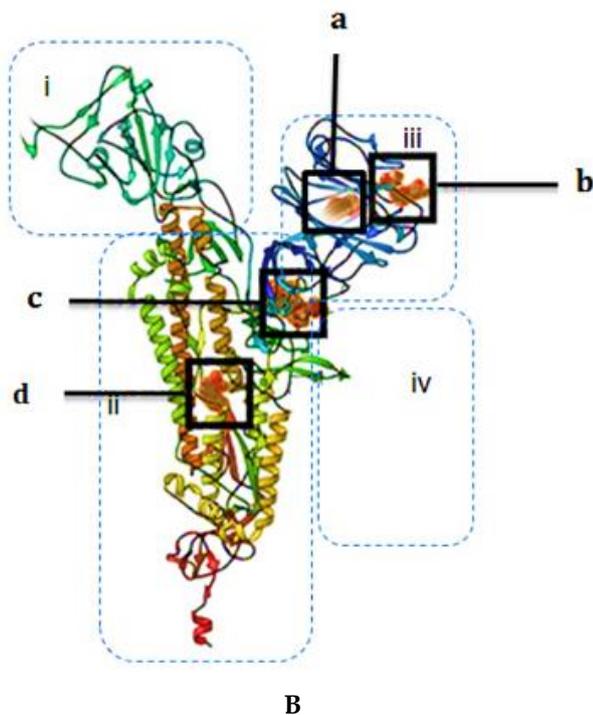
Figure 1. Comparison between the single protomer of SARS-CoV-2 S-protein for (**A**) built model compared to (**B**) the experimental one. Boxes were indicating the intermolecular interaction with (**a**) 14-ketostypodiol diacetate, (**b**) neohesperidin, (**c**) quercetin 3-*O*-rutinoside-7-*O*-glucoside, and (**d**) hydroxypropyl methylcellulose. The two ligands in the **iv** region were present in modeled one and not in the experimental one.

A

B
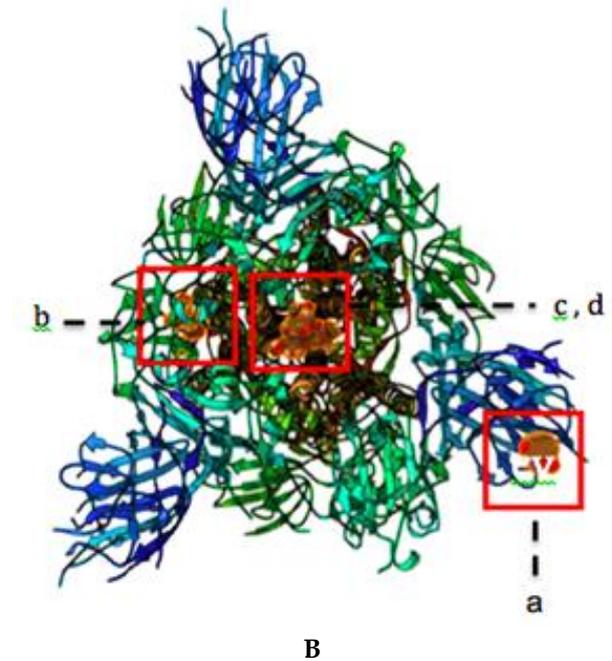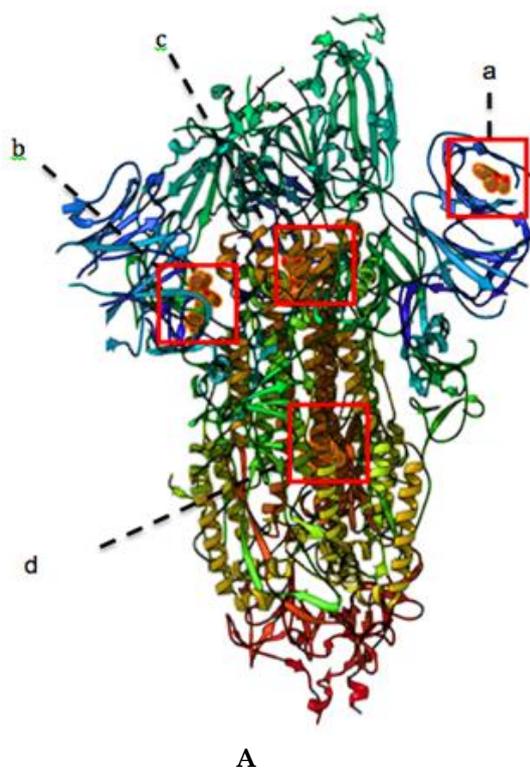
Figure 2. The proposed binding mode of the experimental SARS-CoV-2 S-protein for (**A**) top and (**B**) side views in its homotrimer state with (**a**) 14-ketostypodiol diacetate, (**b**) neohesperidin, (**c**) quercetin 3-*O*-rutinoside-7-*O*-glucoside, and (**d**) hydroxypropyl methylcellulose.

Comparison of the docking complexes of the SARS-CoV-2 S-protein/ligands for the built model and the experimental structure showed high similarities for binding affinity (ΔG) between the built model and the experimental structure, although the interacting residues were different. However, ΔG of quercetin 3-*O*-rutinoside-7-*O*-glucoside and neohesperidin (-16.7 and -15.2 kcal/mol, respectively) were lower than 14-ketostypodiol diacetate and hydroxypropyl methylcellulose with the same ΔG of -13.7 kcal/mol. The interacting residues for the phytochemicals and derivate compounds were identified in the two models, except for the 14-ketostypodiol diacetate, which could interact with the S-protein through Van der Waals interaction. The amino acids of the S-protein in alignment with the four ligands were presented in **Figure 3**.

The locations of the interacting residues for the experimental docking complexes (**Table I**) showed that the binding of the neohesperidin and 14-ketostypodiol diacetate might prevent the target protein from attaching with the host cell membrane (ACE2) as they were located at the S1 ectodomain subunit and might prevent infection process[26]. Furthermore, the active site of the spike protein S2 ectodomain subunit residue 686 to 1237 and spike

protein subunit S2' residue 816 to 1237 tightly react with quercetin 3-*O*-rutinoside-7-*O*-glucoside.

A

B

C

D
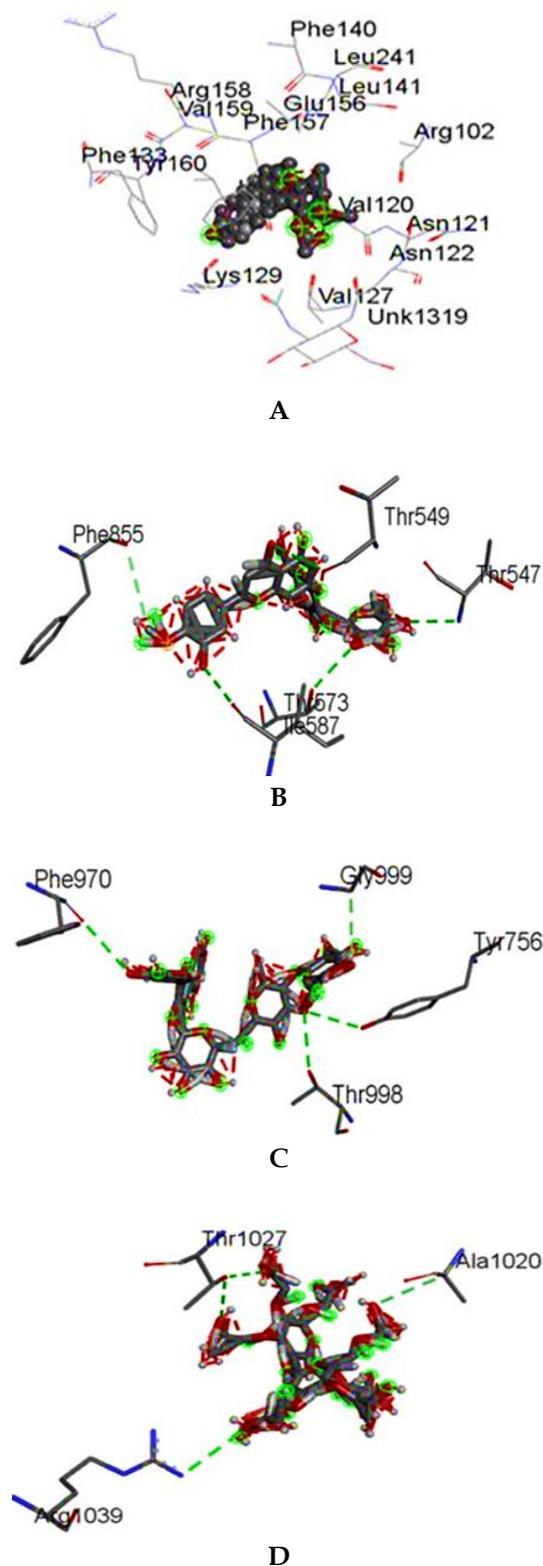
Figure 3. The binding mode of the experimental SARS-CoV-2 S-protein with (**A**) 14-ketostypodiol diacetate, (**B**) neohesperidin, (**C**) quercetin 3-*O*-rutinoside-7-*O*-glucoside, and (**D**) hydroxypropyl methylcellulose.

These active sites mediate fusion between the virus and cellular membranes and allow the insertion of virion genetic material into the host cell[27]. In comparison, quercetin 3-*O*-rutinoside-7-*O*-glucoside interact with S2 and S2' ectodomain subunits of the SARS-CoV-2 S-protein and might act as an inhibitor for S2 and S2' ectodomain subunits and prevent virus fusion and entry into the host cell. Therefore, these ligands might act as anti-SARS-CoV-2 if introduced into the target host cells[28,29].

Table I. Comparisons between the SARS-CoV-2 SARS-GD01, Bat-RaTG13, and MERS S-proteins

S-protein	Neohesperidin		Quercetin 3-*O*-rutinoside-7-*O*-glucoside		14-Ketostypodiol diacetate		Hydroxypropyl methylcellulose	
	Interacting residues	ΔG (kcal/mol)	Interacting residues	ΔG (kcal/mol)	Interacting residues	ΔG (kcal/mol)	Interacting residues	ΔG (kcal/mol)
SARS-CoV-2	Thr-54 Thr-549 Thr-587 Thr-673 Phe-855	-15.2	Thr-756 Phe-970 Thr-998 Gly-999	-16.7	-*	-13.7	Arg-139 Ala-1020 Thr-1027	-13.7
MERS	Asn-73 His-74 Asn-135 Phe-138 Gln-242	-14.7	Ser-279 Glu-285 Val-594 Thr-608 Gly-638	-16.4	Asp-849 His-1040	-15.2	Val-54 Pro-282 Asp-613	-12.5
Bat-RaTG13	Lys-850 Phe-851 Asp-852	-13.6	Ser-726 Thr-774 Pyp-859 Asp-863 His-1054 Gly-1055	-17.3	Gly-1055	-13.9	His-245 Thr-250	-12.8
SARS-GD01	Lie-143 Tyr-144 Pro-173 Gly-175 His-208 Ser-224 His-298	-16.3	Tyr-144 Cys-176 Gly-177 His-208 Pro-210 Asp-213 Cys-214 Glu-247	-18.1	Gly-177	-13.8	Unk-1 Pro-59 Gln-60 Tyr-71 Leu-74	-11.7

*Van der Waals interactions do not occur with amino acid residues

Identifying sequence related to SARS-CoV-2 S-protein
To identify if the four ligands were specific for docking with the SARS-CoV-2 S-protein only or other related viruses, the most related S-protein of the near CoVs that infect humans was selected for the ligands docking analysis. For selecting the CoV S-proteins, phylogenetic analysis was carried out with 30 CoV isolates (**Figure 4**). According to different species and their host, the constructed tree divided the CoVs sequences (human or bat). The S-protein was highly divergent from other CoVs with less than 77% identity with SARS-CoV-2, except the Bat-RaTG13 S-protein, which showed the close phylogenetic relationship to the SARS-CoV-2 S-protein, indicating that the virus might originate from bats[30,31].
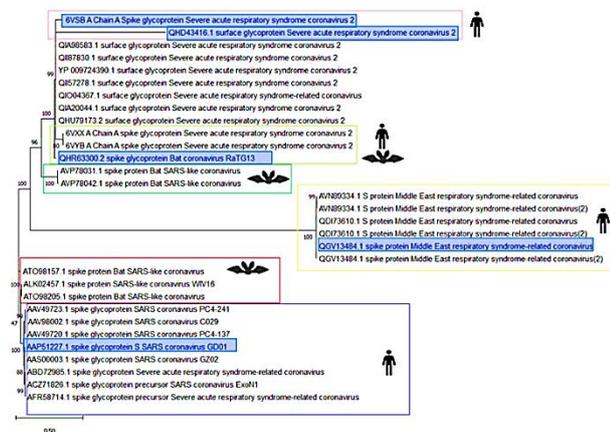
Figure 4. Phylogenetic tree based on amino acid sequences of CoVs S-proteins. The scale bars represent 0.1 substitutions per nucleotide position.

From human infection CoVs, one isolate was selected from each of the two most related clusters, i.e., Bat-RaTG13 (QHR63300.2), SARS-GD01 (AAP51227.1), and one from out-group MERS (QGV13484.1) to study their docking ability with the test ligands[32]. The amino acid sequence revealed that the Bat-RaTG13 S-protein sequence was the closest to SARS-CoV-2 S-protein with 97.41% identity. In contrast, SARS-GD01 and MERS show only 76.19% and 35% identity, respectively, against SARS-CoV-2 S-protein.

Docking test ligands with different CoVs S-proteins
Two close related S-proteins (Bat-RaTG13 and SARS-GD01) to SARS-CoV-2 and one out-group (MERS) were used to study the binding abilities with the test ligands[33]. The binding models of the four ligands with the target protein surface were presented in **Figure 5**. The binding ΔG that reflects the interaction between the test ligand and S-protein of each CoVs was

calculated and presented in **Table I**. Results show that four ligands might have the ability to bind with high affinities to the CoVs, indicating that those ligands may have a wide range of binding to the S-protein of CoVs[34,35].

For SARS-CoV-2 S-protein, the best ligand was quercetin 3-*O*-rutinoside-7-*O*-glucoside that bind to protein active side residues (Tyr-756, Phe-970, Thr-998, Gly-999) with a ΔG of -16.7 kcal/mol in the S2 subunit of the protein, followed by neohesperidin, which bind to S1, S2 subunit with ΔG of -15.2 kcal/mol and interact in the active side (Thr-547, Thr-549, Thr-587, Thr-673, Phe-855), then 14-ketostypodiol diacetate that binds in the S1 subunit with ΔG of -13.7 kcal/mol but without direct binding with the amino acids residues (structurally constrained binding), and hydroxypropyl methylcellulose which bind in the S1, S2 subunit with ΔG of -13.7 kcal/mol.

For SARS-GD01 S-protein, quercetin 3-*O*-rutinoside-7-*O*-glucoside interacts with four active site residues, including Tyr-144, Cys-176, Gly-177, His-208, Pro-210, Asp-213, Cys-214, and Glu-247 (**Figure 6**). Quercetin 3-*O*-rutinoside-7-*O*-glucoside showed the lowest ΔG (-18.1 kcal/mol), while neohesperidin, 14-ketostypodiol diacetate, and hydroxypropyl methylcellulose were showed ΔG of -16.3, -13.8, and -11.7 kcal/mol, respectively. All ligands bind in the S1 subunit of the SARS-GD01 S-protein, which was responsible for initiating infection of the virion to the host cell. The binding sites of hydroxypropyl methylcellulose and quercetin 3-*O*-rutinoside-7-*O*-glucoside were overlapped. Therefore, a complex of the S-protein with the bind two ligands would form (**Figure 6**).

In the Bat-RaTG13 S-protein, the highest ligand binding score was quercetin 3-*O*-rutinoside-7-*O*-glucoside, which interacts with six active site residues, including Ser-726, Thr-774, Pyp-859, Asp-863, His-1054, and Gly-1055 with ΔG of -17.3 kcal/mol. The 14-ketostypodiol diacetate and neohesperidin had ΔG of -13.9 and -13.6 kcal/mol, respectively, and bound to the S2 subunit. Meanwhile, hydroxypropyl methylcellulose had the highest ΔG of -11.7 kcal/mol) and showed binding interacting with His-245 and Thr-250. Results show the potential of all ligands fit within the S1 and S2 subunits and could work as inhibitors for Bat-RaTG13. Because the binding sites of quercetin 3-*O*-rutinoside-7-*O*-glucoside and 14-ketostypodiol diacetate were overlapped, the two ligands could bind together before binding to the S-protein (**Figure 6**).
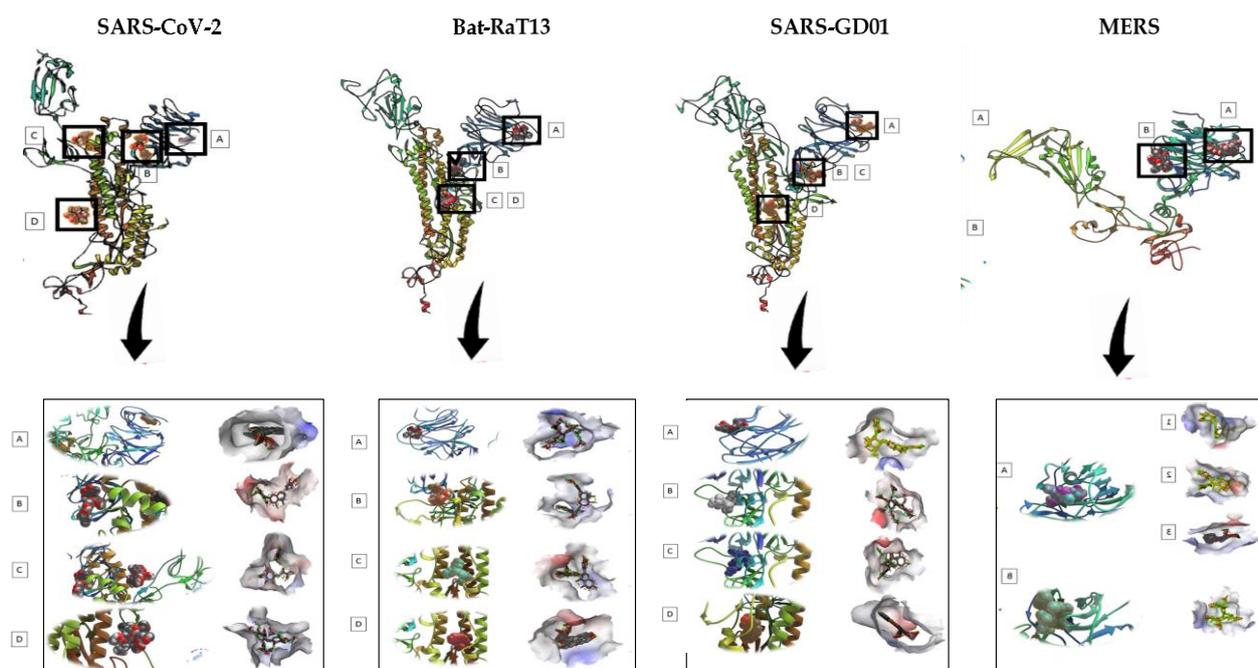
Figure 5. Upper: The 3D structure of the viral S-proteins for SARS-CoV-2, SARS-GD01, Bat-RaTG13, and MERS was showing the binding sites with (**A**) hydroxypropyl methylcellulose, (**B**) neohesperidin, (**C**) quercetin 3-*O*-rutinoside-7-*O*-glucoside, and (**D**) 14-ketostypodiol diacetate. **Lower**: The proposed regions of binding the four ligands with the S-proteins.
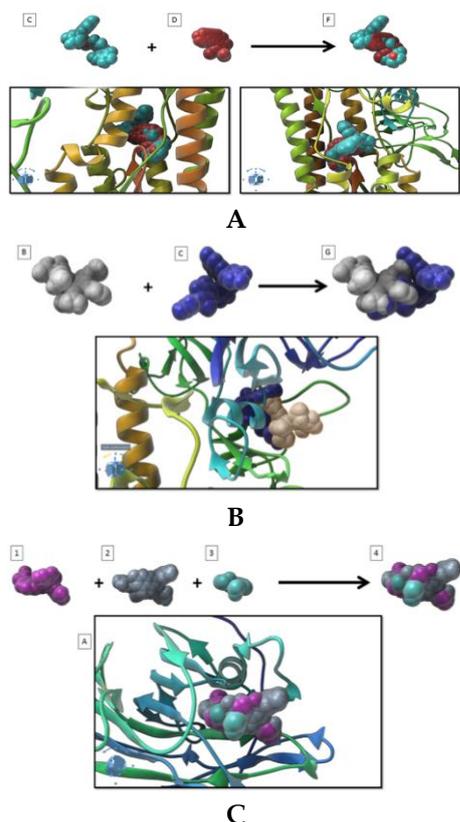
Figure 6. The binding models of ligands in S-protein. (**A**) Bat-RaTG13 with quercetin 3-*O*-rutinoside-7-*O*-glucoside and 14-ketostypodiol diacetate were bind in a complex to increase the stability. (**B**) SARS-GD01 with hydroxypropyl methylcellulose and quercetin 3-*O*-rutinoside-7-*O*-glucoside to determine the complex which were more relevant to the binding site. (**C**) MERS with neohesperidin, quercetin 3-*O*-rutinoside-7-*O*-glucoside, and 14-ketostypodiol diacetate in a complex of three ligand were more stable.

For the MERS S-protein, the hydroxypropyl methylcellulose interacts with three active site residues of the S-protein, including Val-54, Pro-282, and Asp-613. Hydroxypropyl methylcellulose binds with the ΔG of -12.8 kcal/mol. The quercetin 3-*O*-rutinoside-7-*O*-glucoside, neohesperidin, and 14-ketostypodiol diacetate bind to S-protein with ΔG of -16.4, -14.7, and -15.2 kcal/mol, respectively. The S-protein active sites of those three ligands were overlapped; adding them together to the S-protein of MERS would form tri-compounds and bind to the active site. All ligands fit with the active sites of the protein, which was responsible for initiating the infection processes[36]. The atomic-level structural of S-protein for Bat-RaT13, SARS-GD01, and MERS with the different compounds were presented in **Figure 7**.

Docking test ligands with ACE2
The molecular docking of four ligands with ACE2 was performed to determine whether they strongly bind to the S-protein only or its cell receptor as well. Results obtained from molecular docking showed that the four ligands were able to interact with high scores to the ACE2 receptor, as shown in **Figure 8**. **Table II** summarizes the number of contacting residues, domain residues, interacting residues, interaction mode, and ΔG.
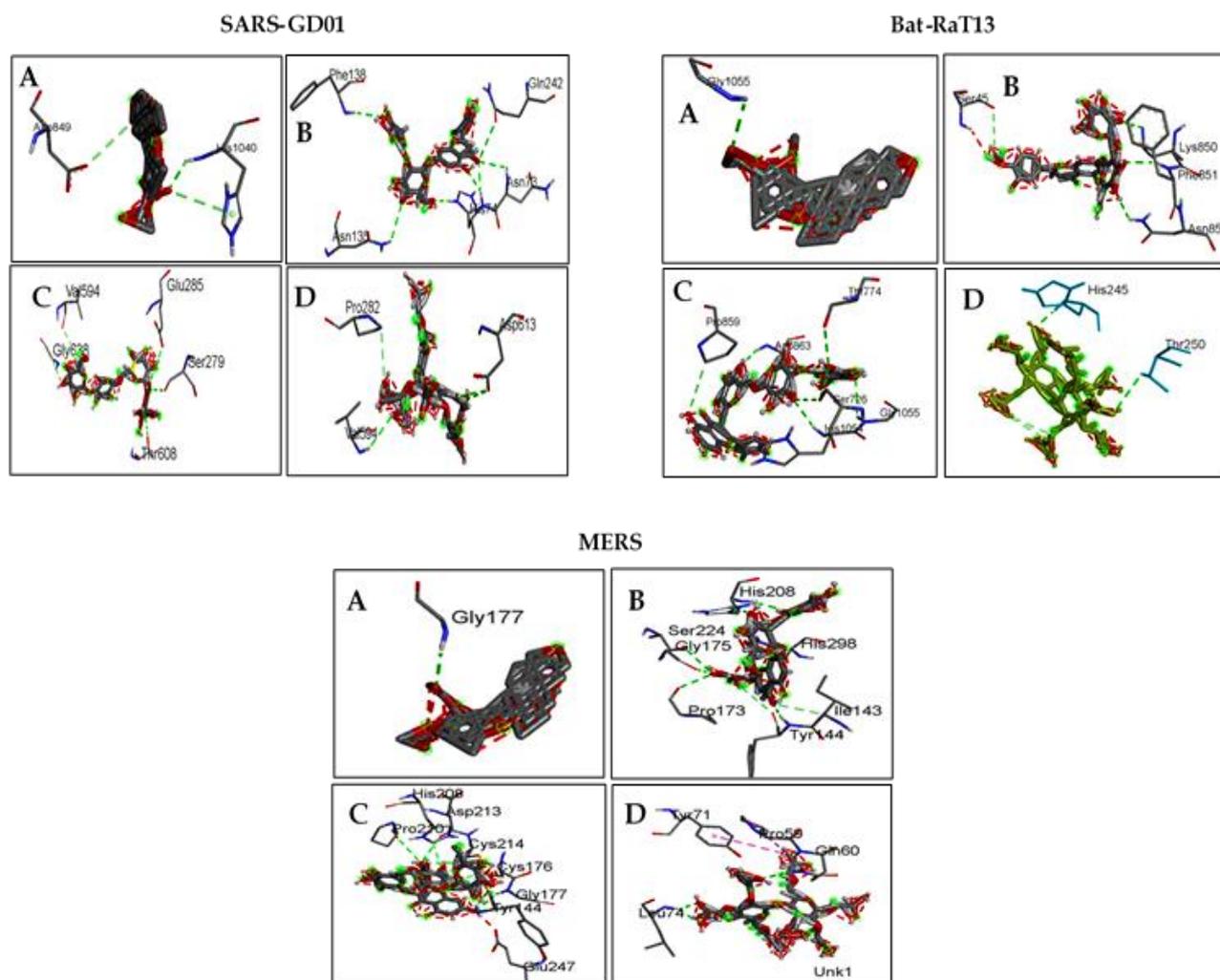
Figure 7. The atomic-level structural of S-protein for Bat-RaT13, SARS-GD01, and MERS with the same binding site, with the ligands of (**A**) 14-ketostypodiol diacetate, (**B**) neohesperidin, (**C**) quercetin 3-*O*-rutinoside-7-*O*-glucoside, and (**D**) hydroxypropyl methylcellulose.

The 14-ketostypodiol diacetate showed the lowest ΔG (-10.6 kcal/mol), binds to eight residues of ACE2 and interacted with residues in seven different modes ((Gln-24, Tyr-83), (Thr-27), (Phe-28), (Asp-30), (Glu-35), (Glu-37), (Tyr-41)). Hydroxypropyl methylcellulose tightly bound to nine residues of ACE2 and interacted with residues in four different modes ((Phe-28, Asp-30, Glu-35), (Asp-30, Lys-31, Glu-35), (Glu-37, Tyr-41, Glu-42), (Glu-37, Tyr-41, Glu-354)) with ΔG of -9.8 kcal/mol. Docking analysis showed that quercetin 3-*O*-rutinoside-7-*O*-glucoside tightly bound to 11 residues of ACE2 and interacted with residues in seven different modes ((Phe-28, His-34), (Asp-30, Glu-37), (Lys-31, Glu-35, Tyr-83), (Glu-37, Asn-330, Glu-354), (Asp-30, Lys-31), (Glu-37, Tyr-41), (Glu-37, Tyr-41, Glu-354), (Phe-28, Glu-42)) with ΔG of -9.4 kcal/mol. For the last, neohesperidin showed that it tightly binds to 14 residues of ACE2 and interacted with residues in 11 different modes ((Phe-28), (Glu-24,

Thr-27, Phe-28, Asp-30), (Phe-28, Glu-35), (Phe-28, Glu-37), (Asp-30, Lys-31), (Asp-30, His-34, Glu-37), (Glu-37, Tyr-41, Gln-42), (Gln-42, Gly-354), (Phe-28, Leu-79), (Gln-24, Phe-28, Asp-30, Tyr-38), (Gly354)) with ΔG of -8.2 kcal/mol.

The molecular docking showed that out of 300 natural plant and plant derivative compounds from the PubChem database, only four could bind with the SARS-CoV-2 S-protein with high affinity. These ligands were natural plant products, so they were considered to be safe for humans[37]. The SwissADME server was used to analyze the solubility and cytotoxicity of those compounds. Also, the isolation of these compounds from their plants was well established[38-40]. Previous publications also showed that plant phytochemicals were predicted to be a potent inhibitor of the SARS-CoV-2 protease using homology modeling[41]. Six citrus flavonoids (naringenin, naringin, hesperetin, hesperidin,

neohesperidin, and nobiletin) were used for molecular docking and predicting ΔG with the ACE2. However, the results show that the ΔG required for the binding between the receptor and the ligands was relatively high[42,43].

D
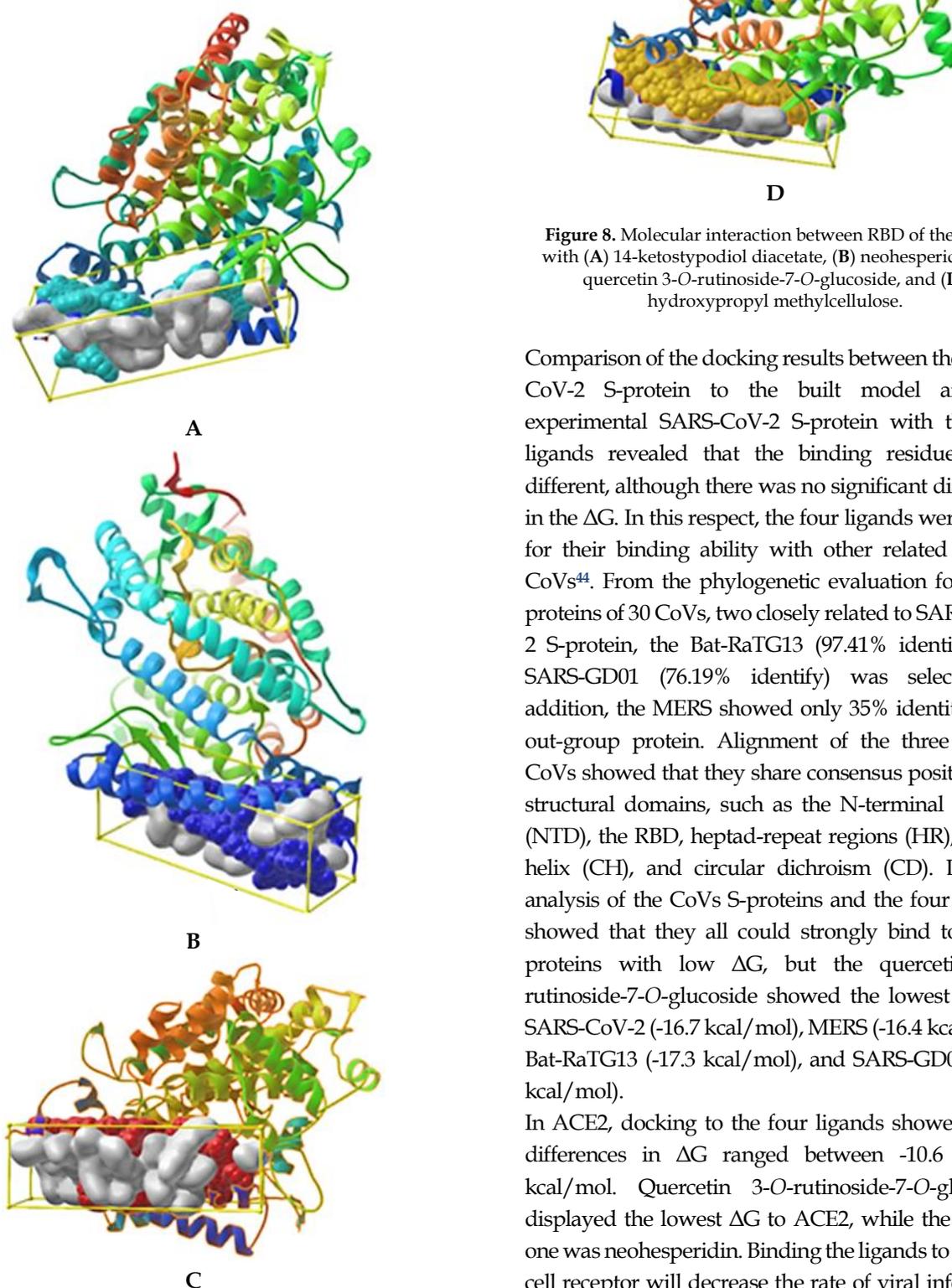
Figure 8. Molecular interaction between RBD of the ACE2 with (**A**) 14-ketostypodiol diacetate, (**B**) neohesperidin, (**C**) quercetin 3-*O*-rutinoside-7-*O*-glucoside, and (**D**) hydroxypropyl methylcellulose.

Comparison of the docking results between the SARS-CoV-2 S-protein to the built model and the experimental SARS-CoV-2 S-protein with the four ligands revealed that the binding residues were different, although there was no significant difference in the ΔG. In this respect, the four ligands were tested for their binding ability with other related human CoVs[44]. From the phylogenetic evaluation for the S-proteins of 30 CoVs, two closely related to SARS-CoV-2 S-protein, the Bat-RaTG13 (97.41% identity) and SARS-GD01 (76.19% identify) was selected. In addition, the MERS showed only 35% identity as an out-group protein. Alignment of the three related CoVs showed that they share consensus position and structural domains, such as the N-terminal domain (NTD), the RBD, heptad-repeat regions (HR), central helix (CH), and circular dichroism (CD). Docking analysis of the CoVs S-proteins and the four ligands showed that they all could strongly bind to the S-proteins with low ΔG, but the quercetin 3-*O*-rutinoside-7-*O*-glucoside showed the lowest ΔG for SARS-CoV-2 (-16.7 kcal/mol), MERS (-16.4 kcal/mol), Bat-RaTG13 (-17.3 kcal/mol), and SARS-GD01 (-18.1 kcal/mol).

In ACE2, docking to the four ligands showed slight differences in ΔG ranged between -10.6 to -8.2 kcal/mol. Quercetin 3-*O*-rutinoside-7-*O*-glucoside displayed the lowest ΔG to ACE2, while the highest one was neohesperidin. Binding the ligands to the host cell receptor will decrease the rate of viral infection[45]. The homotrimer of SARS-CoV-2 S-protein was modeled, and the 3D structure of the experimental one

A

B

C

was compared and showed slight differences between them. Although protein structure homology modeling had become a routine technique to generate 3D models for proteins, it was not accurate as of the experimental structures[46].

Table II. Interaction of ACE2 amino acid residues with the test ligand

Ligands	ACE2 contacting residues	ACE2 domain residues	ACE2 interacting residues	ACE2 interaction mode	ΔG (kcal/mol)
14-Ketostypodiol diacetate	(Gln-24, Tyr-83), (Thr-27), (Phe-28), (Asp-30), (Glu-35), (Glu-37), (Tyr-41)	17	7	8	-10.6
Neohesperidin	(Phe-28), (Phe-28, Glu-24, Asp-30, Thr-27), (Phe-28, Glu-35), (Phe-28, Glu-37), (Asp-30, Lys-31), (Glu-37, His-34, Asp-30), (Glu-37, Tyr-41, Gln-42), (Gln-42, Gly-354), (Leu-79, Phe-28), (Tyr-38, Gln-24, Phe-28, Asp-30), (Gly354)	17	11	14	-8.2
Quercetin 3-*O*-rutinoside-7-*O*-glucoside	(His-34, Phe-28), (Glu-37, Asp-30), (Tyr-83, Lys-31, Glu-35), (Glu-37, Glu-354, Asn-330), (Asp-30, Lys-31), (Tyr-41, Glu-37), (Glu-354, Tyr-41, Glu-37), (Phe-28, Glu-42)	17	7	11	-9.4
Hydroxypropyl methylcellulose	(Phe-28, Asp-30, Glu-35), (Glu-35, Lys-31, Asp-30), (Glu-42, Tyr-41, Glu-37), (Tyr-41, Glu-37, Glu-354)	17	12	9	-9.8

CONCLUSION

The four ligands, quercetin 3-*O*-rutinoside-7-*O*-glucoside, neohesperidin, 14-ketostypodiol diacetate, and hydroxypropyl methylcellulose are predicted to be potent inhibitors for both S-protein and ACE2. Based on satisfactory results obtained from the molecular docking for ligand/S-protein and ligand/ACE2, we demonstrate the ability of the four ligands to be used as a prophylactic medication in COVID-19 prevention.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare that are relevant to the content of this article.

FUNDING

DATA AVAILABILITY

All data are available from the authors.

ACKNOWLEDGMENTS

AUTHORS' CONTRIBUTIONS

Israa Mohamed Shamkh: conceptualization, data curation, formal analysis, investigation, software, visualization, writing – original draft. **Dina Pratiwi**: formal analysis, writing – review & editing.

REFERENCES

1. Pal M, Berhanu G, Desalegn C, Kandi V. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. Cureus. 2020;12(3):e7423. doi:10.7759/cureus.7423

2. Yi Y, Lagniton PNP, Ye S, Li E, Xu RH. COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci. 2020;16(10):1753-66. doi:10.7150/ijbs.45134

3. da Costa VG, Moreli ML, Saivish MV. The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. Arch Virol. 2020;165(7):1517-26. doi:10.1007/s00705-020-04628-0

4. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr Biol. 2020;30(11):2196-203. doi:10.1016/j.cub.2020.05.023

5. Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, et al. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. Biochim Biophys Acta Mol Basis Dis. 2020;1866(10):165878. doi:10.1016/j.bbadis.2020.165878

6. Sironi M, Hasnain SE, Rosenthal B, Phan T, Luciani F, Shaw MA, et al. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. Infect Genet Evol. 2020;84:104384. doi:10.1016/j.meegid.2020.104384

7. Astuti A, Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes Metab Syndr. 2020;14(4):407-12. doi:10.1016/j.dsx.2020.04.020

8. Mittal S, Manjunath K, Ranjan RK, Kaushik S, Kumar S, Verma V. COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. PLoS Pathog. 2020;16(8):e1008762. doi:10.1371/journal.ppat.1008762

9. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol. 2016;3(1):237-61. doi:10.1146/annurev-virology-110615-042301

10. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367(6485):1444-8. doi:10.1126/science.abb2762

11. Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res. 2020;30(4):343-55. doi:10.1038/s41422-020-0305-x

12. Othman H, Bouslama Z, Bradenburg JT, da Rocha J, Hamdi Y, Ghedira K, et al. Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: Similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism. Biochem Biophys Res Commun. 2020;527(3):702-8. doi:10.1016/j.bbrc.2020.05.028

13. Sternberg A, Naujokat C. Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination. Life Sci. 2020;257:118056. doi:10.1016/j.lfs.2020.118056

14. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215-20. doi:10.1038/s41586-020-2180-5

15. Tai W, Zhang X, He Y, Jiang S, Du L. Identification of SARS-CoV RBD-targeting monoclonal antibodies with cross-reactive or neutralizing activity against SARS-CoV-2. Antiviral Res. 2020;179:104820. doi:10.1016/j.antiviral.2020.104820

16. Kaur SP, Gupta V. COVID-19 Vaccine: A comprehensive status report. Virus Res. 2020;288:198114. doi:10.1016/j.virusres.2020.198114

17. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B. 2020;10(5):766-88. doi:10.1016/j.apsb.2020.02.008

18. Xian Y, Zhang J, Bian Z, Zhou H, Zhang Z, Lin Z, et al. Bioactive natural compounds against human coronaviruses: a review and perspective. Acta Pharm Sin B. 2020;10(7):1163-74. doi:10.1016/j.apsb.2020.06.002

19. Verma S, Twilley D, Esmear T, Oosthuizen CB, Reid AM, Nel M, et al. Anti-SARS-CoV Natural Products With the Potential to Inhibit SARS-CoV-2 (COVID-19). Front Pharmacol. 2020;11:561334. doi:10.3389/fphar.2020.561334

20. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem. 2010;31(2):455-61. doi:10.1002/jcc.21334

21. Yamashita F, Hashida M. In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet. 2004;19(5):327-38. doi:10.2133/dmpk.19.327

22. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. Curr Top Med Chem. 2014;14(16):1923-38. doi:10.2174/1568026614666140929124445

23. Pantsar T, Poso A. Binding Affinity via Docking: Fact and Fiction. Molecules. 2018;23(8):1899. doi:10.3390/molecules23081899

24. Herrera NG, Morano NC, Celikgil A, Georgiev GI, Malonis RJ, Lee JH, et al. Characterization of the SARS-CoV-2 S Protein: Biophysical, Biochemical, Structural, and Antigenic Analysis. ACS Omega. 2021;6(1):85-102. doi:10.1021/acsomega.0c03512

25. Robson B. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance. Comput Biol Med. 2020;121:103749. doi:10.1016/j.compbiomed.2020.103749

26. Liu M, Wang T, Zhou Y, Zhao Y, Zhang Y, Li J. Potential Role of ACE2 in Coronavirus Disease 2019 (COVID-19) Prevention and Management. J Transl Int Med. 2020;8(1):9-19. doi:10.2478/jtim-2020-0003

27. Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. Antiviral Res. 2020;178:104792. doi:10.1016/j.antiviral.2020.104792

28. Vishwakarma P, Yadav N, Rizvi ZA, Khan NA, Chiranjivi AK, Mani S, et al. Severe Acute Respiratory Syndrome Coronavirus 2 Spike Protein Based Novel Epitopes Induce Potent Immune Responses in vivo and Inhibit Viral Replication in vitro. Front Immunol. 2021;12:613045. doi:10.3389/fimmu.2021.613045

29. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. 2020;41(9):1141-9. doi:10.1038/s41401-020-0485-4

30. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;588(7836):E6. doi:10.1038/s41586-020-2951-z

31. Ge XY, Wang N, Zhang W, Hu B, Li B, Zhang YZ, et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virol Sin. 2016;31(1):31-40. doi:10.1007/s12250-016-3713-9

32. Zhu Z, Lian X, Su X, Wu W, Marraro GA, Zeng Y. From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. Respir Res. 2020;21(1):224. doi:10.1186/s12931-020-01479-w

33. Zheng J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. Int J Biol Sci. 2020;16(10):1678-85. doi:10.7150/ijbs.45053

34. Pandey P, Rane JS, Chatterjee A, Kumar A, Khan R, Prakash A, et al. Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. J Biomol Struct Dyn. 2020:1-11. doi:10.1080/07391102.2020.1796811

35. Hiremath S, Kumar HDV, Nandan M, Mantesh M, Shankarappa KS, Venkataravanappa V, et al. In silico docking analysis revealed the potential of phytochemicals present in Phyllanthus amarus and Andrographis paniculata, used in Ayurveda medicine in inhibiting SARS-CoV-2. 3 Biotech. 2021;11(2):44. doi:10.1007/s13205-020-02578-7

36. Vincent S, Arokiyaraj S, Saravanan M, Dhanraj M. Molecular Docking Studies on the Anti-viral Effects of Compounds from Kabasura Kudineer on SARS-CoV-2 3CLpro. Front Mol Biosci. 2020;7:613401. doi:10.3389/fmolb.2020.613401

37. Abubakar AR, Haque M. Preparation of Medicinal Plants: Basic Extraction and Fractionation Procedures for Experimental Purposes. J Pharm Bioallied Sci. 2020;12(1):1-10. doi:10.4103/jpbs.JPBS_175_19

38. Pereira DM, Cheel J, Areche C, San-Martin A, Rovirosa J, Silva LR, et al. Anti-Proliferative Activity of Meroditerpenoids Isolated from the Brown Alga Stypopodium flabelliforme against Several Cancer Cell Lines. Mar Drugs. 2011;9(5):852-62. doi:10.3390/md9050852

39. Vedpal, Jayaram U, Wadhwani A, Dhanabal SP. Isolation and characterization of flavonoids from the roots of medicinal plant Tadehagi triquetrum (L.) H.Ohashi. Nat Pod Res. 2020;34(13):1913-8. doi:10.1080/14786419.2018.1561679

40. Ju WT, Kwon OC, Kim HB, Sung GB, Kim HW, Kim YS. Qualitative and quantitative analysis of flavonoids from 12 species of Korean mulberry leaves. J Food Sci Technol. 2018;55(5):1789-96. doi:10.1007/s13197-018-3093-2

41. Qamar MTU, Alqahtani SM, Alamri MA, Chen LL. Structural basis of SARS-CoV-2 3CL pro and anti-COVID-19 drug discovery from medicinal plants. J Pharm Anal. 2020;10(4):313-9. doi:10.1016/j.jpha.2020.03.009

42. Bhowmik D, Nandi R, Prakash A, Kumar D. Evaluation of flavonoids as 2019-nCoV cell entry inhibitor through molecular docking and pharmacological analysis. Heliyon. 2021;7(3):e06515. doi:10.1016/j.heliyon.2021.e06515

43. Tutunchi H, Naeini F, Ostadrahimi A, Hosseinzadeh-Attar MJ. Naringenin, a flavanone with antiviral and anti-inflammatory effects: A promising treatment strategy against COVID-19. Phytother Res. 2020:[Epub ahead of print]. doi:10.1002/ptr.6781

44. Basu A, Sarkar A, Maulik U. Molecular docking study of potential phytochemicals and their effects on the complex of SARS-CoV2 spike protein and human ACE2. Sci Rep. 2020;10:17699. doi:10.1038/s41598-020-74715-4

45. Maginnis MS. Virus–Receptor Interactions: The Key to Cellular Invasion. J Mol Biol. 2018;430(17):2590-611. doi:10.1016/j.jmb.2018.06.024

46. Schwede T. Protein Modelling: What Happened to the "Protein Structure Gap"? Structure. 2013;21(9):1531-40. doi:10.1016/j.str.2013.08.007