In silico Molecular Docking Analysis Targeting SARS-CoV-2 Spike Protein and Selected Herbal Constituents

Anbazhagan Subbaiyan1, Karthikeyan Ravichandran2, Shiv Varan Singh3, Muthu Sankar3, Prasad Thomas1, Kuldeep Dhama4, Yashpal S. Malik5, Raj Kumar Singh6 and Pallab Chaudhuri1*

1Division of Bacteriology and Mycology, 2Division of Veterinary Public Health and Epidemiology, 3Division of Parasitology, 4Division of Pathology, 5Division of Biological Standardization, 6Division of Veterinary Biotechnology, ICAR-Indian Veterinary Research Institute, Izzatnagar, Bareilly - 243 122, Uttar Pradesh, India.

*Correspondence: pallab.chaudhuri@gmail.com

(Received: May 14, 2020; accepted: May 22, 2020)


© The Author(s) 2020. Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License which permits unrestricted use, sharing, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
Abstract

In modern drug discovery, molecular docking analysis is routinely used to understand and predict the interaction between a drug molecule and a target protein from a microbe. Drugs identified in this way may inhibit the entry and replication of pathogens in host cells. The SARS-CoV-2 associated coronavirus disease, COVID-19, has become the most contagious and deadly pandemic disease in the world today. In abeyance of any specific vaccine or therapeutic against SARS-CoV-2, the burgeoning situation urges a need for effective drugs to treat the virus-infected patients. Herbal medicines have been used as natural remedies for treating various infectious diseases since ancient times. The spike (S) protein of SARS-CoV-2 is important for the attachment and pathogenesis of the virus. Therefore, this study focused on the search of useful ligands for S protein among active constituents present in common herbs that could serve as efficient remedies for COVID-19. We analysed the binding efficiency of twelve compounds present in common herbs with the S protein of SARS-CoV-2 through molecular docking analysis and also results are validated with two different docking tools. The binding efficiency of ligands was scored based on their predicted pharmacological interactions coupled with binding energy estimates. In docking analysis, compound “i” (Epigallocatechin gallate (eGCG)) was found to have the highest binding affinity with the viral S protein, followed by compounds, “F” (Curcumin), “D” (Apigenin) and “E” (Chrysophanol). The present study corroborates that compound “i” (eGCG) mostly present in the integrants of green tea, shows the highest potentiality for acting as an inhibitor of SARS-CoV-2. Further, characterization of the amino acid residues comprising the viral binding site and the nature of the hydrogen bonding involved in the ligand-receptor interaction revealed significant findings with herbal compound “i” (eGCG) binding to the S protein at eight amino acid residues. The binding sites are situated near to the amino acids which are required for virus pathogenicity. The findings of the present study need in vivo experiments to prove the utility of “i”, “F”, “D” and “E” compounds and their further use in making herb-based anti-SARS-CoV-2 product in near future. This analysis may help to create a new ethno-drug formulation for preventing or curing the COVID-19.

Keywords: SARS-CoV-2, COVID-19, spike protein, molecular docking, in-silico, ligands, herbal medicine

INTRODUCTION

The World Health Organization (WHO) defines traditional medicine as: “the sum of total knowledge, practices, and skills based on the historical theories, beliefs, and experiences... in indigenous to various cultures that are used to maintain the human or animal health and to prevent, diagnose, improve, or treat physical/mental illnesses” 1.

Herbal remedies are widely used in both developed and developing world countries to treat various illnesses indispensable 2. The WHO reported, about 80% of the world’s population depends primarily on traditional medicine to treat their illnesses. Traditional medicine is often considered to be a kind of complementary or alternative medicine (CAM)3. Herbal medicines include herbs, herbal preparations, and finished herbal products (tea varieties), as well as additives derived from different kinds of herb/plant parts (ginger, garlic, lemon, and so on), which are used when preparing food in many Asian countries, including India and China. The active components of these herbs have many advantages, like lower toxicity and allergenicity than some commercial medications, regulating immunological responses, and causing viral destruction4. Various common herbs have been used to prevent viral infections, and their efficacy has been demonstrated in research trials5-6. Herbal plants like Bupleurum spp., Heteromorpha spp., and Scrophularia scorodonia have been used in the treatment of coronaviruses in China7, and Azadirachta indica, Carica papaya, and Hippophae rhamnoides have been scientifically proven to be effective in treating or preventing Dengue fever in India8. Therefore, identifying and documenting the herbs that are effective in treating contagious diseases is vital for future disease control programs.

Since December 2019, novel coronavirus disease (COVID-19) has been spreading globally from its initial epicenter Wuhan, China. It causes severe respiratory problems more often in children and older people, who have weaker immune defences. The agent, responsible for the infection,
severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), belongs to the Betacoronavirus genus in the family Coronaviridae, the members of which infect a wide range of hosts\(^9,10\). Most of the people affected with SARS-CoV-2 experience moderate respiratory illness, from non-pneumonia to mild pneumonia in nature, along with headache, runny nose, and fever\(^11\). However, older and comorbid people who suffer from cardiovascular diseases, diabetes, and chronic respiratory diseases are more likely to develop severe symptoms (dyspnea, respiratory failure, septic shock, and multiple organ dysfunction/failure). Recently, some other symptoms like bluish spots on the feet, clotting, and stroke also noticed in COVID-19 positive patients\(^12\).

SARS-CoV-2 is a single-stranded, positive-sense RNA virus; its genome is 29,891 kb in size, enclosed by a 5′-cap and 3′poly-A tail, and has a G + C content of 38%. The virion is encircled with an envelope containing viral nucleocapsid and arranged with helical symmetry\(^13\). The SARS-CoV-2 genome encodes four major structural proteins, specifically the spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. Among the proteins of similar coronaviruses, the S proteins of Severe acute respiratory syndrome-related coronavirus (SARS-CoV) and Middle East respiratory syndrome-related coronavirus (MERS-CoV) play an important role in binding with host cellular receptor, angiotensin-converting enzyme 2 (ACE2) and subsequent membrane fusion. Unlike non-structural proteins, the S protein is the major antigen responsible for inducing effective neutralizing antibodies to block viruses from binding to the host cells, and thus, inhibit viral infection\(^14,15\).

Molecular docking studies are primarily carried out to estimate how two or more molecules interact with each other and can depict the best-fit orientation of a ligand (a small drug-like molecule) that binds to a target protein. The most significant use of this approach is in determining the protein-ligand interaction due to its applications in drug discovery. Molecular docking analysis is a bioinformatics-based approach to analyse the fit, binding, and interactions between protein and ligand, based on their binding energy. The results of such interaction studies are presumed to allow the prediction of suitable ligands for drug production, which can then be used in a particular pharmaceutical treatment approach\(^16,17\). The current study aimed to compare the docking fit of various selected compounds from indigenous food additives and herbal constituents with the SARS-CoV-2 S protein and find the best-fitting components. Also, we characterized the amino acid residues comprising the viral binding site and the nature of the hydrogen bonding involved in the ligand-receptor interaction. This analysis may help to create a new ethno-drug formulation for preventing or curing COVID-19.

**MATERIALS AND METHODS**

**Selection of viral target protein**

Because of its importance in viral pathogenesis, S protein was targeted in this study. We attempted to predict the major ligands for the viral S protein using indigenous food additives and herbal constituents. Information on the 3D structure of the viral S protein (PDB ID: 6VYB) was obtained from the protein data bank (PDB) and used in the analyses. Protein structures and docking were visualized using RasMol 2.7.3 software\(^18\).

**Selection and preparation of ligands**

Information on the different herbal ligands selected in this study was based on the review of previous researches, and the details of the compounds in them and the herbs from which they were retrieved are given in Table 1. A total of 12 ligands were selected for testing. Information on the 3D structures of all 12 ligands were downloaded from PubChem\(^19\) in structure-data file (SDF) format and converted into MDL Molfile (MOL) format with the help of Open Babel Server\(^20\), and data in this format was used as input for Generic Evolutionary Method for Molecular Docking (iGEMDOCKv2.1) analyses\(^21\).

**Molecular docking analyses with iGEMDOCK**

The graphical-automatic drug design system iGEMDOCK was used for in silico docking, screening, and post-docking analyses. The S protein structure file was uploaded to the Prepare Binding Site server, and the “By Current File” option was selected so that the overall uncut protein surface could be checked for binding. The ligands were loaded using the “Prepare Compounds” option, and a library of ligands was then prepared for the analyses.
Default docking parameters were used for testing the docking performance of the ligands with the S protein of SARS-CoV-2 (population size=200, number of generations =70, and number of solutions =2). Standard docking was performed with the iGEMDOCK scoring function with ligand and electrostatic preferences set at 1.00. To get accurate and speed up the process, standard docking was used as a default setting. To avoid false positive or false negative results, four rounds of docking was carried out with the same protein and ligands as reported earlier41,42. In these analyses, a lower energy profile indicated a more stable interaction, and the lowest energy profile represented the most likely binding interaction between the protein and ligand tested. After the docking process, the best docking position for each of the individual ligands relative to the S protein was analysed. The outputs obtained from the docking analyses, including the binding position, binding energy, van der Waals force, and hydrogen bonding energy values, were then retrieved, and the best-fitting 3D conformation was analysed with RasMol Viewer18,21.

**Molecular docking analysis with AutoDock software**

This software composed of two different tools such as AutoDock Tool 1.5.6.
and AutoDockVina 1.1.2 and works based on the Lamarckian genetic algorithm. Initially, AutoDockTool was used to optimize and prepare the target protein and ligands in a format of pdbqt files. The Actual docking process was carried out with AutoDockVina software, in which binding site of protein was adjusted with the use of grid box and x, y, z axis values changed to 220 X 210 X 180. The grid point spacing was adjusted to 0.400 Å. The conformational similarity of docked poses was estimated by calculating the root mean square deviations (RMSD) values. RMSD values and the lowest binding energy of conformations is the most stable docking pose43. The output files from AutoDockVina were analysed with PyMOL 2.3.4 trial version software44. Finally, the molecular docking results of the iGEMDOCK software were compared with AutoDockVina software outputs.

**RESULTS**

In this study, the 3D structure model of the SARS-CoV-2 S protein was optimized and 12 ligands from previous studies and online resources were selected to test them as binding ligands for the S protein. The 3D structures of these ligands were also retrieved and optimized for docking analysis. The total binding energy for all the herbal ligands was calculated using iGEMDOCK software, and the binding conformations of the tested ligands with the S protein were also evaluated. From the docking analyses, the binding affinities of twelve compounds to the S protein were estimated based on their estimated ligand binding energy; the results are listed in Table 2. The binding position for each ligand molecule relative to the S protein was analysed, and the one with the lowest ligand binding energy with the S protein, among the various positions tested (four rounds of docking runs), was identified as the most probable binding position. The lowest energy score indicated the ligand for which the protein-ligand binding affinity was the highest, while higher energy values indicated lower binding affinity. Among the 12 herbal ligands, four compounds, “I”, “F”, “D”, and “E” were found to have lowest binding energy values than the other ligands, and thus, had the highest binding affinity (Highlighted in Table 2). Compound “I” (EGCG) had the lowest binding energy value with the S protein (-130.566 kcal/mol), followed by compounds “F” (Curcumin), “D” (Apigenin) and “E” (Chrysophanol) with binding energy values -115.198 kcal/mol, -108.614 kcal/mol and -107.385 kcal/mol, respectively. The other compounds tested, such as “H”, “L”, “J” and “K” had a moderate binding affinity with the S protein ranging from -105.462 kcal/mol for Emodin (H) to -89.949 kcal/mol for Ursolic acid (K).

<table>
<thead>
<tr>
<th>Code</th>
<th>Compound name</th>
<th>Total binding energy (kcal/mol)</th>
<th>van der Waal's force (kcal/mol)</th>
<th>H-bond energy (kcal/mol)</th>
<th>Electrostatic force (kcal/mol)</th>
<th>AverConPair (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ajoene</td>
<td>-74.2819</td>
<td>-66.2819</td>
<td>-6</td>
<td>0</td>
<td>32.4615</td>
</tr>
<tr>
<td>B</td>
<td>Allicin</td>
<td>-62.4326</td>
<td>-62.4326</td>
<td>0</td>
<td>0</td>
<td>40.3333</td>
</tr>
<tr>
<td>C</td>
<td>Aloe emodin</td>
<td>-69.2503</td>
<td>-69.2503</td>
<td>0</td>
<td>0</td>
<td>29.2308</td>
</tr>
<tr>
<td>D</td>
<td>Apigenin</td>
<td>-108.614</td>
<td>-82.1108</td>
<td>-26.503</td>
<td>0</td>
<td>33.9</td>
</tr>
<tr>
<td>E</td>
<td>Chrysophanol</td>
<td>-107.385</td>
<td>-90.5916</td>
<td>-16.7935</td>
<td>0</td>
<td>36.7895</td>
</tr>
<tr>
<td>F</td>
<td>Curcumin</td>
<td>-115.198</td>
<td>-87.5695</td>
<td>-27.6288</td>
<td>0</td>
<td>26.963</td>
</tr>
<tr>
<td>G</td>
<td>Diallytrisulfide</td>
<td>-53.2872</td>
<td>-53.2872</td>
<td>0</td>
<td>0</td>
<td>39.2222</td>
</tr>
<tr>
<td>H</td>
<td>Emodin</td>
<td>-105.462</td>
<td>-87.2314</td>
<td>-18.2303</td>
<td>0</td>
<td>32.85</td>
</tr>
<tr>
<td>I</td>
<td>Epigallocatechin-gallate</td>
<td>-130.566</td>
<td>-91.7244</td>
<td>-38.8417</td>
<td>0</td>
<td>25.7273</td>
</tr>
<tr>
<td>J</td>
<td>Gingerol</td>
<td>-98.0333</td>
<td>-84.1818</td>
<td>-13.8515</td>
<td>0</td>
<td>27.6667</td>
</tr>
<tr>
<td>K</td>
<td>Ursolic acid</td>
<td>-89.9499</td>
<td>-72.2658</td>
<td>-17.6841</td>
<td>0</td>
<td>23.5714</td>
</tr>
<tr>
<td>L</td>
<td>Zingerone</td>
<td>-102.184</td>
<td>-77.9523</td>
<td>-24.2321</td>
<td>0</td>
<td>26.9524</td>
</tr>
</tbody>
</table>

Table 2. The fit and interaction profiles of the tested ligands with the viral spike protein. The total binding energy, van der Waal’s force, H-bond energy, electrostatic force, and AverConPair of the binding of each ligand with the S protein are shown. The values shown below for each compound correspond to the lowest binding energy obtained among the four rounds of docking runs.
Besides, the analyses also revealed the characteristics of binding H-bonds and the involved energy of the 12 herbal compounds with the target S protein. Docking position analyses showed the amino acids within the S protein involved in binding with each of the ligands “I”, “F”, “D”, and “E”. The Compound “I” is forming a hydrogen bond with spike protein at eight sites (GLN1010, THR1009). The 3D image of protein-ligand (“I”) interaction shown in Fig. 2. The energy of H-bonds typically ranges from -8.7 to -1.1 kcal/mol. The compounds “F”, “D”, “E” formed hydrogen bonds at six, eight, and four sites, with H-bond energies from -7 to -1.1 kcal/mol, respectively. The details of the binding energy of each ligand, associated bonds, energy, and involved amino acid positions in S protein CDS are shown in Table 3 and Fig. 1.

AutoDockVina docking tool was used to compare and validate the results of iGEMDOCK output. The same spike protein and four compounds, which are having a high binding affinity with target protein based on iGEMDOCK were selected for comparison. Based on RMSD values, best poses were selected from the group of poses. The conformation which is having RMSD value lesser that 2.0Å is considered as the best pose. The RMSD value is used to measure the distance between two atoms in various protein-ligand conformations. Therefore best pose (<2.0Å) with the lowest binding energy poses are selected for further analyses. The result obtained from the AutoDockVina is mentioned in the Table No. 4. The AutoDockVina software results showed compound “I” having better affinity (-9.2 Kcal/mol) to the target protein. The other three compounds also showed a similar binding affinity like iGEMDOCK docking analysis.

Table 3. H-bond and Amino acid position profile table for the spike protein with tested ligands. The total binding energy, H-bond nature, position of amino acids, and H-bond energy of each bond are shown.

<table>
<thead>
<tr>
<th>Code</th>
<th>Compound name</th>
<th>Total binding energy (kcal/mol)</th>
<th>H-bond energy (kcal/mol)</th>
<th>Amino acid positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ajoene</td>
<td>-74.2819</td>
<td>H-S</td>
<td>GLN1010,THR1009</td>
</tr>
<tr>
<td>B</td>
<td>Allicin</td>
<td>-62.4326</td>
<td>H-S</td>
<td>LYS1038,ASP-1041</td>
</tr>
<tr>
<td>C</td>
<td>Aloe emodin</td>
<td>-69.2503</td>
<td>H-M</td>
<td>GLY908,ASP-1041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HIS-1048,TRP-886,ALA-890</td>
</tr>
<tr>
<td>D</td>
<td>Apigenin</td>
<td>-108.614</td>
<td>H-S</td>
<td>ASP1041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>H-M</td>
<td>ASP1041,ALG1044,ALG1046</td>
</tr>
<tr>
<td>E</td>
<td>Chrysophanol</td>
<td>-107.385</td>
<td>H-S</td>
<td>ASN-544,ARG-567,ASN-978,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>H-M</td>
<td>ASP-979</td>
</tr>
<tr>
<td>F</td>
<td>Curcumin</td>
<td>-115.198</td>
<td>H-M</td>
<td>ALA-522,THR-547</td>
</tr>
<tr>
<td>G</td>
<td>Diallyltrisulfide</td>
<td>-53.2872</td>
<td>H-S</td>
<td>LYS1038</td>
</tr>
<tr>
<td>H</td>
<td>Emodin</td>
<td>-105.462</td>
<td>H-S</td>
<td>GLY908,His1048</td>
</tr>
<tr>
<td>I</td>
<td>Epigallocatechin gallate</td>
<td>-130.566</td>
<td>H-S</td>
<td>GLN314,ASN317,ASP737,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>H-M</td>
<td>ASN764,THR859</td>
</tr>
<tr>
<td>J</td>
<td>Gingerol</td>
<td>-98.0333</td>
<td>H-M</td>
<td>THR315,VAL736,ASP737</td>
</tr>
<tr>
<td>K</td>
<td>Ursolic acid</td>
<td>-89.9499</td>
<td>H-M</td>
<td>HIS-1058</td>
</tr>
<tr>
<td>L</td>
<td>Zingerone</td>
<td>-102.184</td>
<td>H-M</td>
<td>HIS-1058,LEU861,HIS1058</td>
</tr>
</tbody>
</table>

[Table 3]
COVID-19 has spread to nearly 210 countries with nearly 5 million confirmed cases and 325,000 deaths. Presently, the case fatality rate caused by the contagion seems to be lower in Asian region than European, American, or the world. Given this, we considered people’s food habits and identified a few common medicinal ingredients that are currently used in food preparation in India and some neighboring countries. Several drug candidates have been evaluated for their antiviral activity against SARS-CoV-2 virus. But recent studies used chemical libraries to screen out better drug candidates for SARS-CoV-2 with limited success. So with this sense, we started to search herbs and herbal components to inhibit viral infections. Based on the literature, we selected twelve compounds that are currently used in food preparation in India and some neighboring countries. Among 12 ligands, iGEMDOCK tool identified the best fitting ligand as EGCG (“I”) followed by Curcumin (“F”), Apigenin (“D”), and Chrysophanol (“E”) towards spike protein.

All the results based on iGEMDOCK were cross-verified and validated by comparing with another molecular modeling simulation software AutoDockVina which applies a different algorithm. Similar to iGEMDOCK tool, EGCG (“I”) showed highest binding affinity (-9.2 kcal/mol) as compared to other compounds. Thus the superiority of EGCG (“I”) was corroborated based on two docking approaches.

These compounds are having a better binding with spike protein in the form of H-bonding and van der Waal’s force. The compound EGCG is binding to spike protein with eight amino acids through H-bonding. The amino acids involved in the binding of proteins to the ligand are mostly polar. And the binding sites of ligands to S protein is essential for the virus binding to the ACE2 receptors. Babcock and co-workers reported that the amino acids positioned at the 270-510 are essential for the SARS-CoV-2 virus to attach with the host cell ACE2 receptor. Three amino acids from this region (GLN-314; THR-315; ASN-317) were found to be involved in ligand binding (Table 3; Fig. 2). Since the compounds, we tested block these amino acids, which may limit the invasion of host cells by SARS-CoV-2.

Binding energy estimates and binding site analyses of all ligands showed that binding positions of these amino acids mostly differ in their sites at which they formed hydrogen bonds. These hydrogen bonds are necessary to maintain the structural stability of the protein-ligand complex. Additionally, the H-bond energy of all compounds was nearly always negative, which indicated that they had a relatively high binding affinity with the S protein. Compound “I” which is mostly present in the components of green tea, has a high probability of blocking the virus attachment and entry into host cells. Similarly, previous studies have also shown that this compound can bind with the influenza virus, Zika virus, and porcine circovirus outer proteins, thereby inhibiting viral entry into the host cells. This compound has the highest binding affinity towards SARS-CoV-2 spike protein compared to the antiviral drugs. Calligari and colleagues investigated the binding affinity of antiviral drugs with SARS-CoV-2 spike protein. In comparison to antiviral drug affinity, these herbal compounds showed better binding affinity.
Currently, many researchers are working on identifying candidate drugs via in silico analyses. The utilization of available antiviral medications might prove to be an effective method of inhibiting SARS-CoV-2 through the binding of these herbal ligands with the S glycoprotein, as well as the 3CL protease. Rane and coworkers analysed the potential utility of various phytochemicals as ligands for the viral S protein, as determined by molecular docking study. The present study is perhaps the first to apply a molecular docking approach to predict both the potential ligand binding efficiency among the common herb/food constituents and to further elucidate the involved amino acids of SARS-CoV-2 S protein during ligand binding. This study divulged that among the tested ingredients, based on having the lowest estimated ligand binding energy, EGCG (the active ingredient in Camellia sinensis) was the most potent ligand against the S protein of SARS-CoV-2. Global scientific community mainly focused on developing antiviral drugs rather than finding any compounds which are upregulating the immune system. This green tea compound EGCG also has many health benefits particularly modulating both adaptive and innate immune system functions. Therefore, compounds that are used for treating illness, should have some additional immune regulatory functions which are worthwhile.

**CONCLUSION**

The molecular docking analyses helped to explore the probable binding modes of twelve ligands with the S protein of SARS-CoV-2. Among the tested compounds, EGCG ("I"), a principal constituent of green tea, had the highest binding affinity to this protein. Therefore, including green tea in the diets of people might help to reduce the occurrence of COVID-19. However, more in vivo experimental research is required to validate our results and for developing more potent drugs for the prevention and control of COVID-19.

**ACKNOWLEDGMENTS**

All the listed author(s) are thankful to their representative universities/institutes for providing the related support to compile this work.
CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORS’ CONTRIBUTION
All the listed author(s) have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING
None.

ETHICS STATEMENT
This article does not contain any studies with human participants or animals performed by any of the authors.

DATA AVAILABILITY
Not applicable.

REFERENCES
22. Babcock GJ, Esshaki DJ, Thomas WD Jr., Ambrosino DM. Amino acids 270 to 510 of the severe acute respiratory


37. Yusof YA. Gingerol and its role in chronic diseases. In Drug discovery from mother nature. 2016;177-207. https://doi.org/10.1007/978-3-319-41342-6_8


