

In silico screening of known small molecules to bind ACE2 specific RBD on Spike Glycoprotein of SARS-CoV-2 for repurposing against COVID-19

Bharath B R (✉ bharath.br@atrimed.com)

Atrimed Biotech LLP

Hrishikesh Damle

Atrimed Pharmaceuticals Pvt. Ltd.

Shiban Ganju

Atrimed Pharmaceuticals Pvt. Ltd.

LathaDamle

Atrimed Biotech LLP

Research Article

Keywords: SARS-CoV-2, Pathogen host interaction, SARS-CoV-2 Spike Glycoprotein, ACE2, Small Molecules

Posted Date: May 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28557/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Human coronavirus (SARS-CoV-2) is causing a pandemic with significant morbidity and mortality. Although, no effective novel drugs are available, drug repurposing is emerging as an effective strategy. In this study, we present an in silico drug repurposing study implementing successful concepts of computer aided drug design (CADD) technology with an objective to repurpose known drugs to interfere the viral cellular entry via the spike glycoprotein (SARS-CoV-2-S), which mediates the virus–cell receptor interaction. SARS-CoV-2-S uses ACE2 to enter cells. Totally, 4015 known and approved small molecules were screened for interaction with SARS-CoV-2 S through docking studies and 15 lead molecules were shortlisted. Further, three molecules streptomycin, ciprofloxacin and Glycyrrhizic acid (GA) were selected based on their reported anti-viral activity, safety, availability, affordability and subjected for Molecular Dynamics (MD) simulation. The MD simulation results indicate that GA from plant origin may be repurposed against SARS-CoV-2 and further studies are needed for validation.

1. Introduction

The complete genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is about 82% identical to the SARS coronavirus (SARS-CoV); both viruses share the common clade having the genus *Betacoronavirus* as root node [2-3]. There are currently no medications or vaccines proven to be effective for the treatment or prevention of the 2019 SARS-CoV-2 [4-7].

Drug repurposing as a commercially viable drug discovery strategy from existing drugs could significantly shorten the time and reduce the cost compared to de novo drug discovery [8-10]. Experimental approaches, however, at pre-clinical and clinical stages for drug repurposing require high cost and time [11]. Computational approaches can offer quick, considerable and novel testable hypotheses for systematic drug repositioning [9].

Current drugs in different phases of clinical trials are being investigating as inhibitor of drug targets playing significant role in coronavirus infection lifecycle. The drug targets might be involved in entry of SARS-CoV-2 into the host cell (Eg. umifenovir and chloroquine), replication (Eg. lopinavir/ritonavir) and RNA synthesis (Eg. remdesivir/favipiravir). Among these, targeting SARS-CoV-2 cellular entry via the spike glycoprotein has emerged as a major option for repurposing [10]. As the SARS-CoV-2 S glycoprotein is surface-exposed and mediates entry into host cells, it is the main target of neutralizing antibodies (Abs) upon infection and the focus of therapeutic and vaccine design [11].

The entry of SARS-CoV-2 into host cells is mediated by homotrimer, protruding and transmembrane SARS-CoV-2 S glycoprotein localised on viral surface comprised of two functional subunits individually responsible for interaction with host cell receptor (S1 subunit) and fusion of the viral and cellular membranes (S2 subunit). SARS-CoV-2 S is cleaved at the boundary between the S1 and S2 subunits, which remain non-covalently bound in the prefusion conformation [12-20].

The receptor-binding domain (RBD) site on distal S1 subunit contributes to the stabilisation of the prefusion state of the membrane-anchored S2 subunit that contains the fusion machinery [19-25]. For all CoVs, spike glycoprotein is further cleaved by host proteases at the so-called S20 site located immediately upstream of the fusion peptide [17,26]. This cleavage has been proposed to activate the protein for membrane fusion via extensive irreversible conformational changes [12,17,18,24,27]. As a result, coronavirus entry into susceptible cells is a complex process that requires the concerted action of receptor binding and proteolytic processing of the SARS-CoV-2 S protein to promote the fusion of SARS-CoV-2 cells [28].

Hence, in the current study, it is aimed to predict and validate the structure of SARS-CoV-2 S protein using computer aided homology modelling tools and screen a library of small molecules for their interaction with SARS-CoV-2 S protein.

2. Materials And Methods

2.1. Sequence analysis

The complete genome of coronavirus 2 SARS-CoV-2 (Acc. No : MT159721.1, Length: 29882 bp) was retrieved from NCBI and subjected for basic local alignment search using NCBI-BLAST [29]. The BLAST search has revealed 96.04%, 91.64% and 82.30% identity with Bat coronavirus RaTG13 (Acc. No : MN996532.1, Length: 29855 bp), Pangolin coronavirus (Acc. No : MT084071.1, Length: 27213 bp) and SARS-CoV (Acc. No : JX163927, Length: 29646bp) respectively.

2.2. Topological analysis of pathogen-host interactome (PHI) for target validation

The drug target identification and validation was carried out using network based topological analysis method using web based application Pathogen-Host Interaction Search Tool (PHISTO) available at the URL: <http://www.phisto.org/browse.xhtml#>. PHISTO is the most comprehensive pathogen-human protein-protein interaction database in the web used to explore the molecular connectivities between the pathways in SARS-CoV and Human interaction by means of topological analysis. Properties of nodes including connectivity degree (k), betweenness centrality (BC) and closeness centrality (CC) were adopted to evaluate the nodes in a network; especially k and BC are two fundamental parameters in the network theory [30,31].

2.3. Prediction of ligand binding site

The similarity between RBD domains of S-protein from SARS-CoV-2 (Acc.No.QII57328, Length:1273aa), SARS-CoV (Acc.No: AFR58728, Length:1255aa) and RatG13 (Acc.No: QHR63300, Length:1269aa) were aligned using multiple sequence alignment (MSA) using Clustal-Omega tool available in EBI. Conservation in ACE2 receptor interaction was seen among all the three sequences aligned, which have directed the active binding site prediction.

The protein-protein interaction between SARS-CoV-2 S and human ACE2 Receptor complex was studied using the crystal structure (PDB ID: 6CS2) available in PDB. The amino acids involved in the interaction were identified as ligand binding sites for inhibitor molecules.

2.4. Homology modelling of SARS-CoV-2 S-Protein

The homology modelling was performed for the protein sequence SARS-CoV-2 S having crystal structure of SARS-CoV spike glycoprotein and ACE2 complex (PDB ID: 6ACD) as template using Swissmodeller. The modelled protein was validated for quality and prepared for molecular docking studies using protein preparation wizard available in Schrodinger small molecule suite.

2.5. Protein Preparation

The modelled S-Protein was pre-processed for docking studies using the Protein Preparation Wizard available in Schrödinger suite 2019-2. Crystallographic water molecules (water molecules without 3 H bonds) were deleted and hydrogen bonds corresponding to pH 7.0 were added considering the appropriate ionization states for both the acidic and basic amino acid residues. Missing side chain atoms were added, and breaks present in the structure were built using Prime v4.0, Schrödinger 2019-2 [32]. Using the OPLS_2005 force field [33] energy of modelled structure was minimised.

2.6. Ligand Preparation

The three dimensional conformers of 4015 small molecule drugs in use to treat various diseases and as nutritional supplements were downloaded from DrugCentral database (<http://drugcentral.org/>) and subjected for ligand minimisation using Ligrip application provided in Schrödinger Maestro. The ligand minimisation was performed by assigning force field OPLS_2005 and stereo isomers were calculated retaining specific chiralities. The ADME predictions were done for all ligands using QikProp package [34].

2.7. Molecular Docking

The modelled structure of SARS-CoV-2 S protein, the known target for COVID-19 was prepared and active site was defined with a 10 Å radius around the selected residues (Arg426, Tyr436, Pro462, Thr486, Gly488 and Tyr491) present in the modelled structure and a grid box 20X20X20 Å was generated. The docking of small molecules over SARS-CoV-2 S was performed using Grid-Based Ligand Docking with Energetics Glide v7.8, Schrödinger 2019-2 [35] in 'High Throughput Virtual Screening' (HTVS), 'standard precision' (SP) and 'extra precision' (XP) modes consecutively. The final best docked conformer was selected using a Glide score function, Glide energy and Glide E model energy. Finally, the lowest-energy docked complex of three known molecules Streptomycin, Ciprofloxacin and Glycyrrhizic acid in complex with SARS-CoV-2 S were selected for molecular dynamic simulations.

2.8. Molecular Dynamics (MD) Simulations

The molecular dynamics [36] of the docked complexes of shortlisted molecules were studied using the OPLS_2005 force field [39] in an explicit solvent with the TIP3P water model [37] within the Desmond. The molecular system was solvated with water molecules in an orthorhombic box for Streptomycin/SARS-CoV-2 S, Ciprofloxacin SARS-CoV-2 S and Glycyrrhizic acid/SARS-CoV-2 S complexes allowing for a 10 Å buffer region between protein atoms and box sides. The systems were placed at a distance of 10 Å from the edge of the box and LBFGS minimisation was performed with 3 vectors and minimum 10 steepest descent steps until a gradient threshold of 25 kcal/mol/Å was reached. The maximum iterations during minimisation were 2000 and the convergence threshold was kept at 1.0 kcal/mol/Å. For long range electrostatic interactions Smooth Particle Mesh Ewald [38] method was used at a tolerance of 1e-09 and a cut-off radius of 9Å was selected for short range electrostatic interactions. Before equilibration and MD simulations, the systems were minimised and pre-equilibrated using the default relaxation routine implemented in Desmond. Then the systems were gradually heated in the NPT ensemble to 300° K with a time step of 2 fs. A 100 ns MD simulation in the NPT ensemble (T^{1/4} 300° K, thermostat relaxation time ^{1/4200} ps; P ^{1/41} atm; barostat relaxation time ^{1/4200} ps) was performed using a Nose–Hoover thermostat [39]. Data were collected every 100 ps during MD run. 3-D structures and trajectories were visually inspected using the Maestro graphical interface. An average structure obtained from the last 250 ps of MD simulation was refined by means of 1000 steps of steepest descent followed by conjugate gradient energy minimisation. The maximum number of cycle of minimisation was 5000 and the convergence criterion for the energy gradient was 0.001 kJ/mol/Å.

3. Results And Discussions

3.1. Multiple sequence alignment of complete genomes

The complete genome of coronavirus SARS-CoV-2, Bat coronavirus RaTG13, Pangolin coronavirus and SARS-CoV obtained as blast hits were aligned using Clustal Omega tool available in EBI tools and the phylogenetic tree was constructed (Figure 1).

The MSA has demonstrated the molecular similarities between the organisms. RaTG13 is identified as a neighbour genome for SARS-CoV-2 and justified the hypothesis that, the infection may be transfected from Bats. Meanwhile, the subsequent neighbours were Pangolin MP789 and SARS-CoV. This preliminary sequence alignment has enabled the understanding of sequence similarities and evolutionary information which is very much fundamental in the process of drug discovery.

3.2. Topological analysis and target validation

A thorough analysis of virus-host interactomes may reveal insights into viral infection and pathogenic strategies. Due to lac of interaction data on SARS-CoV-2, the significant identity between SARS-CoV and SARS-CoV-2 proteomes was considered and the SARS-CoV-Human interactome was built by screening domain interactions between SARS-CoV-Human protein-protein interactions (PPIs), and then the network distribution, topological and function analysis was performed (Figure 2). The circular shapes corresponds to proteins (Nodes) which are labeled by Uniprot_ID and the details on nodes are listed in Table 1.

Among 14 proteins on SARS-CoV, major proportion of SARS-CoV-Human interaction is reserved by 5 non-structural proteins (NS3B, NS6, NS7A, NS7B and NS8A with 4, 3, 8, 3 and 1 human proteins correspondingly), 3 Open Reading Frame (ORF) poly proteins (ORF9B, A7J8L3 and A7J8L2 with 4, 1 and 1 human proteins correspondingly), 2 Replicase proteins (R1A and R1AB with 2 and 1 human proteins correspondingly), Membrane protein (VME1 with human IKKB), Envelop membrane protein (VEMP with human B2CL1), Nucleoprotein (NCAP with 4 human proteins) and Spike glycoprotein (SPIKE with human ACE2). With this observations, we can find the higher specificity of Membrane, Envelop and Spike glycoprotein to interact with host through specific entry points. Hence, these three SARS-CoV proteins can be a potential targets to inhibit the pathogen-host interaction specifically while other interactions were versatile. To ensure the impact of inhibition of IKKB, B2CL1 and ACE2 mediated interaction, the landscape of SARS-CoV-Human interaction was further analysed for degree and betweenness centrality distributions of the host as shown in Table 2.

Degree measures how many neighbours a node direct connect to while betweenness measures how often nodes occur on the shortest paths between other nodes. In the PPIs network the nodes with high degree defined as hub protein and the nodes with high betweenness defined as bottleneck protein, both are key or important proteins. In the current topological analysis, the node with low degree of distribution 1 and betweenness centrality 0.0 was A2A3R6. But, the molecular function of A2A3R6 (Uniprot ID: A2A3R6) is not well understood in both human physiology or pathology. Hence, the second node ACE2 with degree of distribution 4 and betweenness centrality 2271 was next significant node as shown in Figure 3 and it is identified as a key node or key player in the SARS-CoV-host interaction. Hence, the SARS-CoV Spike protein interacting with human_ACE2 is identified as a potential drug target. As the information over SARS-CoV-2-Human interaction is not available, the SARS-CoV-Human interaction data can be used. We studied the similarity between SARS-CoV and SARS-CoV-2 by sequence analysis and RBD prediction.

3.3 Sequence analysis and RBD predictions

As depicted in Figure 4, the alignment between S-protein from SARS-CoV-2 and Bat coronavirus RaTG13 was seen closer when compare with SARS- CoV. The alignment at RBD site from 317 residue to 569 found more than 80 percent similar to SARS-CoV and RaTG13 particularly, at major residues, Tyr436, Thr486, Gly488 and Tyr491 except at Arg426 and Pro462 as shown in Figure 5. Considering the evolution, the available crystal structure of SARS- CoV in complex with ACE2 (PDB ID: 6CS2) was used as templet for homology modelling.

The residues involved in interaction of SARS-CoV with ACE2 were predicted using prime module available in Schrodinger small molecule suit and the major interactions are tabulated in table 3 and shown in Figure 6. Very strong interaction was seen between a smallest amino acid Gly488 with Lys353, Gly354 and Asp355 facilitated by 1 hydrogen bond and 94.9% of buried SASA. Remaining other residues have also shown significantly strong interactions with ACE2. Hence, the same residues were made centric to generate the grid.

3.4. Homology modelling and validation of SARS-CoV-2 S-Protein

The modelling of SARS-CoV-2 was performed using the crystal structure of SARS-CoV spike glycoprotein as template, which was 97% identical to the query. The modelled protein shown in Figure 7A was validated for quality and prepared for molecular docking studies using protein preparation wizard available in Schrodinger small molecule suite. The Ramachandran plot generated using protein preparation wizard has confirmed the quality of modelled structure by plotting >95% residues in allowed region, as shown in Figure 7B.

3.5. Molecular docking of small molecules with SARS-CoV-2 S

The repurposing of small molecules for the treatment of COVID-19 requires the knowledge on interaction of therapeutic with SARS-CoV-2 S. Initial HTVS screening suggested 142 molecules with reasonable interaction with SARS-CoV-2 S and further the shortlisted molecules were docked in SP mode where the accuracy of prediction is improved. The docking in SP mode has suggested 15 top molecules listed in table 4 as lead molecules. As hydroxychloroquine is identified to treat COVID-19, it was also subjected for subsequent docking in XP mode. All 15 molecules showed better interaction than hydroxychloroquine with SARS-CoV-2 S. The three molecules Streptomycin, Ciprofloxacin and Glycyrrhizic acid, with low interaction penalties and displaying better interactions with ACE2 binding site on RBD of SARS-CoV-2 S as shown in Figure 8A-C respectively. The three molecules were selected based on their reported anti-viral activity, safety, availability and affordability.

For SARS-CoV-2 S, from glide generated docking model, streptomycin could bind to the receptor, and the binding mode was highly similar to that of ACE2 interaction. The binding pocket of streptomycin was in the RBD site which was observed as an acceptor for ACE2. The streptomycin was well fitted with the shape of the pocket as shown in Figure 8A and 9A with XP score -6.5, where it formed total five hydrogen bonds, among them two hydrogen bonds were seen by donating electrons from N31 and N32 atoms to the Glu493 side chain atoms, simultaneously other two hydrogen bonds were observed between backbone atoms of Leu501 by receiving electrons from hydroxyl groups at 5th and 6th carbon atom of S1 six carbon ring on streptomycin. Remaining one hydrogen bond was formed between backbone atom of Ser503 and hydroxyl group at 6th carbon atom at G3 group of streptomycin. However, the stability of the interaction cannot be pronounced without molecular dynamic simulations.

The docking model of ciprofloxacin illustrated its binding mode on RBD site which was observed as a key site to interfere the virus-host interaction. The ciprofloxacin found reasonably fitted with steric complementarity on RBD pocket as shown in Figure 8B and 9B with XP score -5.31. The interaction of ciprofloxacin with SARS-CoV-2 S was facilitated by two hydrogen bonds with Val492 and Phe499 each by receiving and donating the electrons respectively from hydroxyl and ketone groups.

The docking model of Glycyrrhizic acid illustrated its binding mode on RBD site which was observed as a key site to interfere the virus-host interaction. The Glycyrrhizic acid fitted with steric complementarity on RBD pocket as shown in Figure 8C and 9C with XP score -7.474. The interaction of Glycyrrhizic acid with SARS-CoV-2 S was facilitated by three hydrogen bonds with Leu464, Val492 and Glu493 by receiving

electrons from hydroxyl groups of Glycyrrhizic acid. Additionally, the ketone group of Glycyrrhizic acid has formed hydrogen bond with backbone atoms of Phe499 receiving the electrons.

3.6. Molecular dynamics simulation of protein ligand complex

As the receptor SARS-CoV-2 S has 1273aa, it requires enormous computational time to perform MD simulation for whole range of protein, hence, we confined this study only to the RBD portion from 317th residue to 569th residue was trimmed for MD Simulation.

3.6.1. Root Mean Square Deviation (RMSD) analysis of protein ligand complex

The RMSD is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The plots in Figure 10 shows the RMSD evolution of a protein (left Y-axis). All protein frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the atom selection. Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation. RMSD analysis can indicate if the simulation has equilibrated its fluctuations towards the end of the simulation are around some thermal average structure. Changes of the order of 1-3 Å are perfectly acceptable for small, globular proteins. Changes much larger than that, however, indicate that the protein is undergoing a large conformational change during the simulation.

The RMSD plot for Streptomycin/SARS-CoV-2 S complex shown in Figure 10A has attained the equilibrium at 5ns and there after shown a stability with maximum RMSD of 1 Å (2.5Å-3. Å) upto 55ns. After 55ns the change in equilibrium state was observed, however the RMSD was seen with in 1.5 Å which is acceptable. Similarly, the Ligand RMSD (right Y-axis) indicates how stable the ligand is with respect to the protein and its binding pocket. The RMSD values for Streptomycin were observed significantly larger than the RMSD of the SARS-CoV-2 S at RBD domain, then it is likely that the Streptomycin is likely to diffuse away from its initial binding site after 48ns.

The RMSD plot for Ciprofloxacin/SARS-CoV-2 S complex shown in Figure 10B has attained the equilibrium at 2ns and there after shown a stability with maximum RMSD of 1.8 Å (2.4Å-4.2Å) upto 58ns. After 58ns the sudden change in equilibrium state was observed, however the RMSD was seen with in 2 Å which is acceptable. On the other hand, the RMSD values for Ciprofloxacin were observed significantly in align with the RMSD of the SARS-CoV-2 S at RBD domain, then it is likely that the Ciprofloxacin can retain in its initial binding site upto 100ns.

The RMSD plot for Glycyrrhizic acid/SARS-CoV-2 S complex shown in Figure 10C has attained the equilibrium till 100ns. Compared to SARS-CoV-2 S in complex with Glycyrrhizic acid to SARS-CoV-2 S in complex with Streptomycin and Ciprofloxacin it found stable till the end of the simulation without any drift in equilibrium. On the other hand, the RMSD values for Glycyrrhizic acid were observed significantly in align with the RMSD of the SARS-CoV-2 S at RBD domain in almost all the frames, Hence it is likely to retain in its initial binding site upto 100ns and predicted to inhibit SARS-CoV-2 S at RBD domain

comparatively better than Streptomycin and Ciprofloxacin for long duration but its contact with key ligands has to be confirmed through RMSF and protein ligand contact analysis.

3.6.2. Root Mean Square Fluctuations (RMSF) and secondary structure elements

The RMSF is useful for characterising local changes along the protein chain. In RMSF plot, peaks indicate areas of the protein that fluctuate the most during the simulation. Typically, the tails fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions. The secondary structure of RBD on SARS-CoV-2 S has the same secondary structure elements on RBD from SARS_CoV with 74% homologous residues. However, 2019-CoV has a distinct loop with flexible glycy residues replacing rigid prolyl residues in SARS-CoV. Molecular modelling revealed that 2019-CoV RBD has a stronger interaction with angiotensin converting enzyme 2 (ACE2). A unique phenylalanine Phe486 in the flexible loop likely plays a major role because its penetration into a deep hydrophobic pocket in ACE2. In the trimmed RBD structure this loop starts from 148th residue and ends at 172nd residue. Since the ligand binding site is located at this loop region higher RMSD was noticed. The protein residues that interact with the ligand are marked with green-colored vertical bars on RMSF plots. In the RMSF plot for RBD domain of Streptomycin/SARS-CoV-2 S complex shown in Figure 11A, the RMSF at loop region was 5.6Å with very high ligand contacts (green-colored vertical bars). This was in par with molecular docking interactions. In the RMSF plot for RBD domain of Ciprofloxacin/SARS-CoV-2 S complex shown in Figure 11B, the RMSF at loop region was 5.6Å with a few ligand contacts (green-colored vertical bars) and justifies the interactions seen in molecular docking. Further, In the RMSF plot for RBD domain of Glycyrrhizic acid/SARS-CoV-2 S complex shown in Figure 11C, the RMSF at loop region was 6.3Å with good number of ligand contacts (green-colored vertical bars) and justifies the interactions seen in molecular docking. Though the ligand contacts are seen in interactions, their simulation time coverage determines their stability.

3.6.3. Protein-ligand contacts

Protein interactions with the ligand can be monitored throughout the simulation. These interactions can be categorised by type and summarised, as shown in the plots given in Figure 12A-12C. Protein-ligand interactions are categorised into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. Each interaction type contains more specific subtypes, which can be explored through the 'Simulation Interactions Diagram' panel. The stacked bar charts are normalised over the course of the trajectory: for example, a value of 0.7 suggests that 70% of the simulation time the specific interaction is maintained. Values over 1.0 are possible as some protein residue may make multiple contacts of same subtype with the ligand.

In the protein ligand contact plot for Streptomycin/SARS-CoV-2 S complex shown in Figure 12A, residues Glu493 and Lys544 were seen with the maximum interactions fraction i.e 0.20 facilitated by hydrogen bonds and water bridges, which suggest that 20% of the simulation time the specific interaction is

maintained and such a interactions are not promising. Hence, Streptomycin cannot be a potential inhibitor of SARS-CoV-2 S to offer anti-COVID19 activity.

In the protein ligand contact plot for Ciprofloxacin/SARS-CoV-2 S complex shown in Figure 12B, residues Phe465, Tyr482, Tyr498 and Phe499 were seen with the interactions fractions 0.75, 0.6, 0.35 and 0.39 respectively facilitated by hydrophobic, hydrogen bonds and water bridges, which suggest that 70%,60%, 35% and 39% of the simulation time the specific interaction is maintained by respective residues and such a interactions are good. Hence, Ciprofloxacin may be a potential inhibitor of SARS-CoV-2 S to offer anti-COVID19 activity.

In the protein ligand contact plot for Glycyrrhizic acid/SARS-CoV-2 S complex shown in Figure 12C, residues Val492, Glu493, Asn496, Cys497 and Phe499 were seen with the interactions fractions 0.78, 1.12, 0.80, 0.60 and 0.80 respectively facilitated by hydrophobic, hydrogen bonds and water bridges, which suggest that 78%,100%, 80%, 60% and 80% of the simulation time the specific interaction is maintained by respective residues and such a interactions are excellent and promising. Hence, Glycyrrhizic acid can be a potential inhibitor of SARS-CoV-2 S to offer anti-COVID19 activity.

4. Conclusion

Through our topological analysis ACE2 with less degree of distribution is likely to be targeted by viruses like SARS-CoV. Hence, the viral protein SARS-CoV-2 S interacting with host ACE2 is as potential drug target for the repurposing of known drugs. Further, sequence alignment and domain analysis has suggested RBD as a ligand binding site and molecular docking studies have suggested Streptomycin, Ciprofloxacin and Glycyrrhizic acid as possible leads to inhibit SARS-CoV-2 S. Conclusively, molecular dynamic simulation analysis has indicated that Glycyrrhizic acid as a promising small molecule could be repurposed as potential inhibitor of SARS-CoV-2 S to offer anti-COVID19 activity.

Declarations

Competing interests: Latha Damle is the founder of Atrimed Biotech LLP and holds equity in Atrimed Pharmaceuticals. Shibam Ganju and Hrishikesh Damle hold equity shares in Atrimed Pharmaceuticals.

References

1. Tonew, M. Tonew, K. Eckardt, H Thrum, Gumpert B. Streptovirudins – new antibiotics with antiviral activity. The antiviral spectrum and inhibition of Newcastle disease virus in cell cultures. *Acta Virol.* 19(4):311-7. (1975).
2. Zhou, X.-L. et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273 (2020).
3. E. Gorbalenya et al., Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. *Nat. Microbiol.* (2020).

4. World Health Organization. Clinical management of severe acute respiratory infection when novel coronavirus (CoV) infection is suspected [https://www.who.int/publicationsdetail/clinical-management-of-severe-acute-respiratory-infection-when-novelcoronavirus-\(CoV\)-infection-is-suspected](https://www.who.int/publicationsdetail/clinical-management-of-severe-acute-respiratory-infection-when-novelcoronavirus-(CoV)-infection-is-suspected)
5. CDC Website: <https://www.cdc.gov/coronavirus/2019-CoV/hcp/clinical-guidancemanagement-patients.html>
6. FDA Website: <https://www.fda.gov/emergency-preparedness-and-response/mcmissues/coronavirus-disease-2019-covid-19>.
7. Korea Biomedical Review website: <http://www.koreabiomed.com/news/articleView.html?idxno=7428>.
8. Cheng, F. In silico oncology drug repositioning and polypharmacology. *Methods Mol. Biol.* 1878, 243–261 (2019).
9. Cheng, F., Hong, H., Yang, S. & Wei, Y. Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief Bioinformatics* 18, 682–697 (2017).
10. Nature Biotechnology Website :<https://www.nature.com/articles/d41587-020-00003-1> visited on 20th March.
11. Walls et al., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *Cell*, 281–292. (2020).
12. Belouzard, S., Chu, V.C., and Whittaker, G.R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. USA* 106, 5871–5876. (2009).
13. Bosch, B.J., van der Zee, R., de Haan, C.A., and Rottier, P.J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* 77, 8801–8811. (2003).
14. Burkard, C. Et al., Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog.* 10, e1004502. (2014).
15. Kirchdoerfer, R.N., et al., Pre-fusion structure of a human coronavirus spike protein. *Nature* 531, 118–121. (2016).
16. Millet, J.K., and Whittaker, G.R. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl. Acad. Sci. USA* 111, 15214–15219. (2014).
17. Millet, J.K., and Whittaker, G.R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* 202, 120–134. (2015).
18. Park, J.E. et al., Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proc. Natl. Acad. Sci. USA* 113, 12262–12267. (2016).
19. Walls, A.C. et al., Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 531, 114–117. (2016).

20. Walls, A.C. et al., Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 23, 899–905. (2016).
21. Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., and Wang, X. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* 27, 119–129. (2017).
22. Kirchdoerfer, R.N. et al., Pre-fusion structure of a human coronavirus spike protein. *Nature* 531, 118–121. (2016).
23. Pallesen, J. et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike anti-gen. *Proc. Natl. Acad. Sci. USA* 114, E7348–E7357. (2017).
24. Walls, A.C. et al., Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. USA* 114, 11157–11162. (2017).
25. Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* 8, 15092. (2017).
26. Madu, I.G., Roth, S.L., Belouzard, S., and Whittaker, G.R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol.* 83, 7411–7421. (2009).
27. Heald-Sargent, T., and Gallagher, T. Ready, The corona- virus spike protein and acquisition of fusion competence. *Viruses* 4, 557–580. (2012).
28. Morris DE, Cleary DW and Clarke SC. Secondary Bacterial Infections Associated with Influenza Pandemics. *Front. Microbiol.* 8:1041. (2017).
29. Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller. "A greedy algorithm for aligning DNA sequences", *J Comput Biol* 2000; 7(1-2):203-14. (2000).
30. Raman K, Construction and analysis of protein-protein interaction networks. *Autom Exp.* (2010) 2: 2-10.1186/1759-4499-2-2.
31. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, Lee D: A protein interaction network associated with asthma. *J Theor Biol.* 2008, 252: 722-731.
32. Jacobson M.P. et al., A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Structure, Function and Bioinformatics* 55: 351-367. (2004).
33. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* 6: 1509–1519. (2010).
34. Dagan-Wiener, A. et al., Bitter or not? BitterPredict, a tool for predicting taste from chemical structure. *Sci Rep* 7, (2017).
35. Friesner RA. Et al., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49: 6177–6196. (2006).

36. Guo Z, Mohanty U, Noehre J, Sawyer TK, Sherman W, Krilov G. Probing the α -helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis. *Chem. Biol. Drug Des.* 75: 348–359. (2010).
37. Jorgensen WL, Chandrasekhar J. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 926–935. (1983).
38. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103: 8577–8593. (1995).
39. Martyna GJ, Klein ML, Tuckerman M. Nose–Hoover chains—the canonical ensemble via continuous dynamics. *J. Chem. Phys.* 97: 2635–2643. (1992).

Tables

Due to technical limitations, the tables are only available as a download in the supplemental files section.

Figures

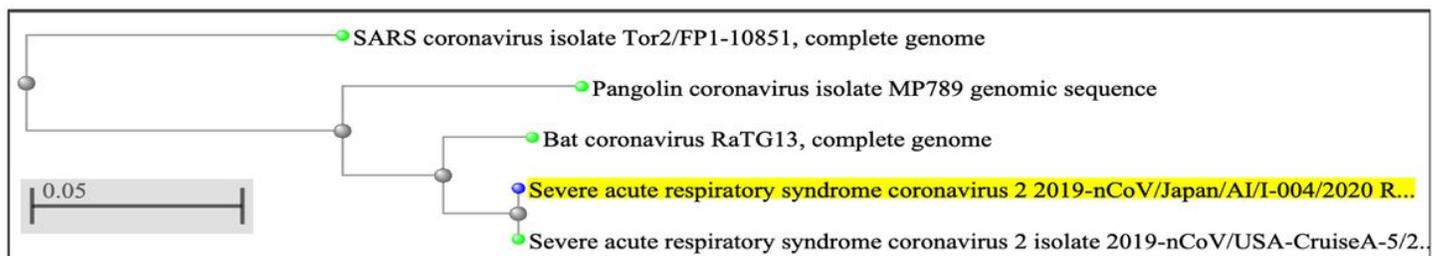


Figure 1

Cladogram constructed using neighbour joining tree construction method, depicting the distance between the complete genome sequences subjected for MSA.

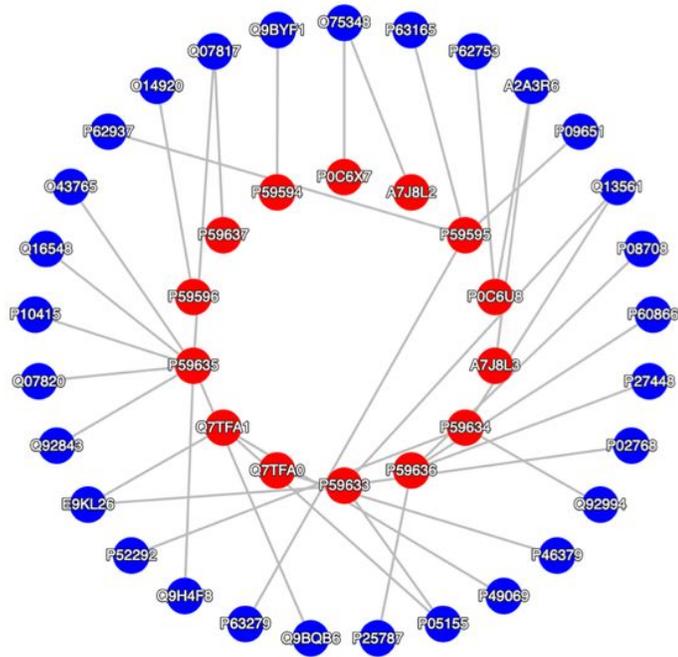
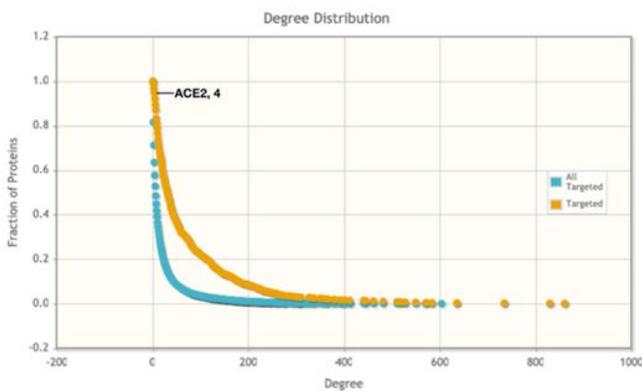


Figure 2

The PHI of SARS-CoV-Human illustrating the PPIs between proteins. The 28 blue nodes corresponds to human proteins and 14 red nodes are SARS-CoV-2 proteins. The nodes are connected by 35 edges.

A



B

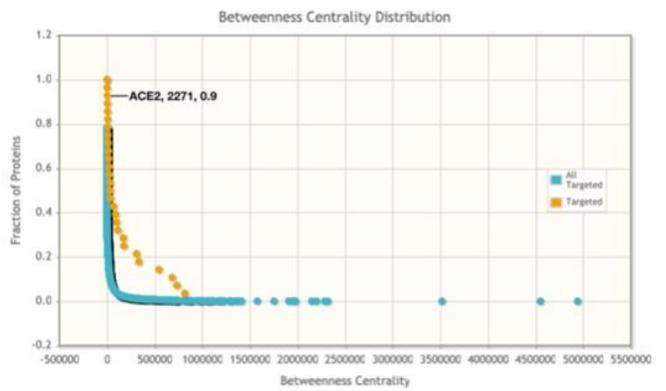


Figure 3

Cumulative degree and betweenness centrality distributions among host proteins involved in SARS-CoV interaction. Host proteins that are targeted SARS-CoV, have an approximate degree and betweenness centrality. A: Degree distribution highlighting ACE2 position and B: Betweenness centrality distribution highlighting ACE2 position. These findings are statistically significant by Fligner-Killeen (median) test.



Figure 4

Multiple sequence alignment of S-protein from SARS-CoV-2, SARS-CoV and RatG13 to understand the conservation of amino acids.

```

AFR58728.1_SARS_Cov      AELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERK 343
QII57328.1_SARS_CoV-2  SETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRK 356
QHR63300_RaTG13        SETKCTLKSFTVEKGIYQTSNFRVQPTDSIVRFPNITNLCPFGEVFNATTFASVYAWNRK 356
                          * * *:*** :***** * :.***** * *****:
AFR58728.1_SARS_Cov      KISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTG 403
QII57328.1_SARS_CoV-2  RISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTG 416
QHR63300_RaTG13        RISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVRQIAPGQTG 416
                          :*****: ***** *****:*****: **:*****
                          Arg426      Tyr436                                Pro462
AFR58728.1_SARS_Cov      VIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPD 463
QII57328.1_SARS_CoV-2  KIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYLYRLFRKSNLKPFERDISTEIQAG 476
QHR63300_RaTG13        KIADYNYKLPDDFTGCVIAWNSKHIDAKEGGNFYLYRLFRKANLKKPFERDISTEIQAG 476
                          ***** **::***:..:.. **::** * :.:*:*****. :.
                          Thr486 Gly488      Tyr491
AFR58728.1_SARS_Cov      GKPCTP-PALNCYWPLNDYGFYTTGIGYQEYRVVVLSFELLNAPATVCGPKLSTDLIKN 522
QII57328.1_SARS_CoV-2  STPCNGVEGFNCYFPLQSYGFQPTNGVGYQEYRVVVLSFELLHAPATVCGPKKSTNLVKN 536
QHR63300_RaTG13        SKPCNGQTGLNCYYPLRYGFYPTGVGHQEYRVVVLSFELLNAPATVCGPKKSTNLVKN 536
                          ..** .:***:* ** * * :*:*****:***** **:*:*
AFR58728.1_SARS_Cov      QCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCXFGGVS 582
QII57328.1_SARS_CoV-2  KCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS 596
QHR63300_RaTG13        KCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS 596
                          :***** *.:* *****: * **::***: * *****:* *****

```

Figure 5

Aligned region of S-protein from SARS_CoV-2, SARS-CoV and RatG13 to understand the conservation of amino acids at RBD domain region.



Figure 6

Amino acids involved in B-Chain of SARS-CoV and ACE2 (D-Chain) interaction. The residues highlighted in blue colour corresponds to B-Chain of SARS-CoV and residues in green colour corresponds to ACE2 (D-Chain).

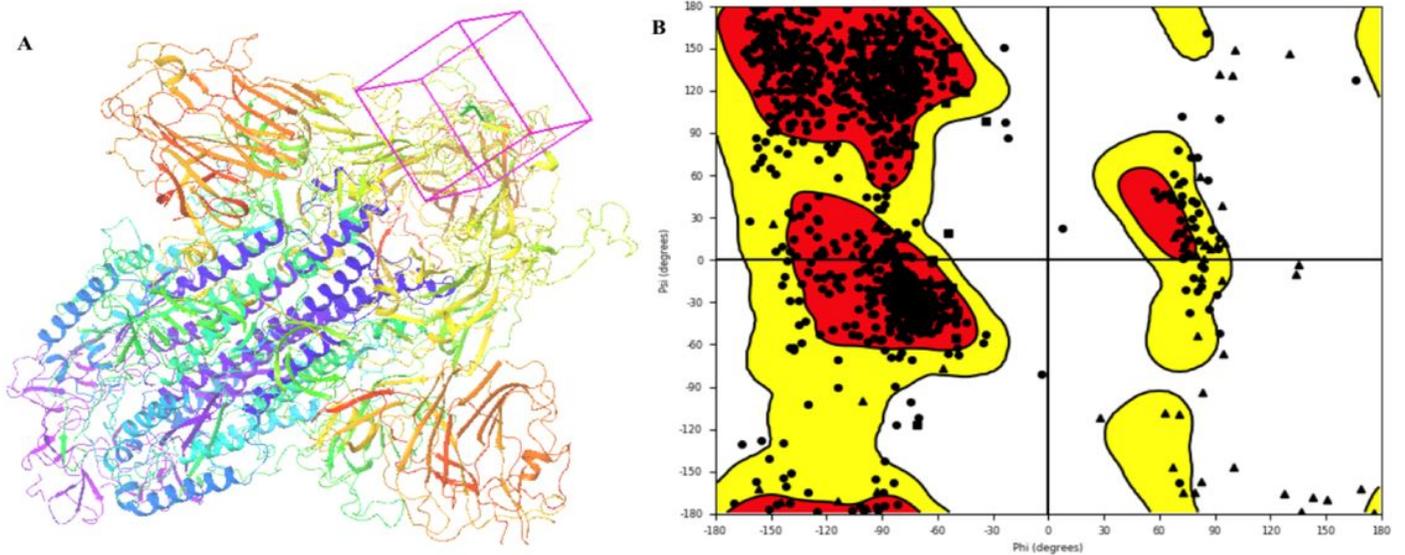


Figure 7

A: The modelled structure of SARS_nCoV-2 with Arg426, Tyr436, Pro462, Thr486, Gly488 and Tyr491 centric grid box generated for molecular docking, B: Ramachandran plot having >95% of amino acids plotted in allowed region (Red and Yellow).

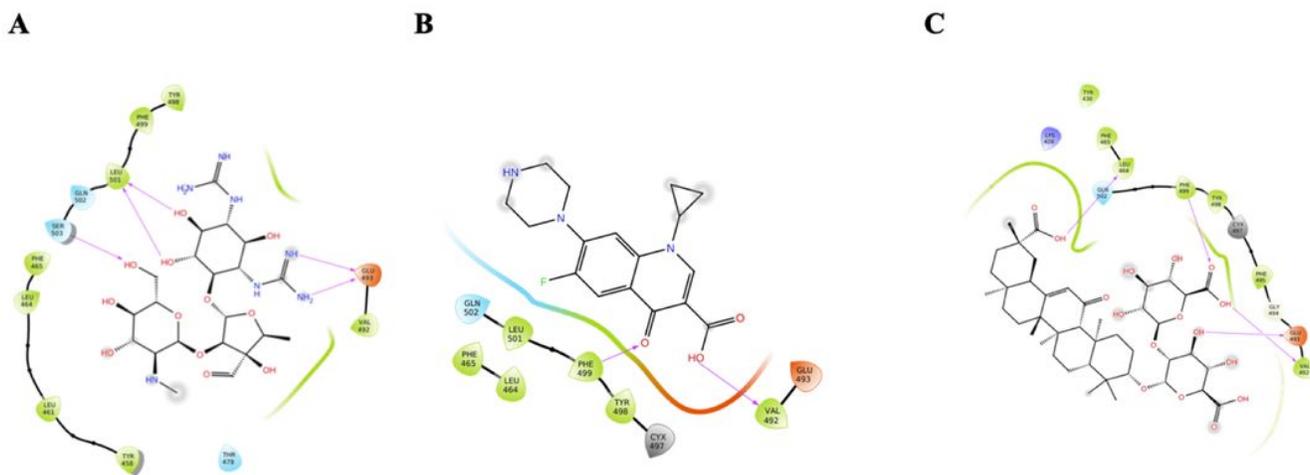


Figure 8

Two dimensional (2D) illustration of small molecule interaction in the ACE2 binding site on RBD of SARS-CoV-2 S. A: Interaction of streptomycin facilitated by 5 H-Bonds shown in pink arrow. B: Interaction of Ciprofloxacin facilitated by 2 H-Bonds shown in pink arrow. C: Interaction of Glycyrrhizic acid, facilitated by 5 H-Bonds shown in pink arrow.

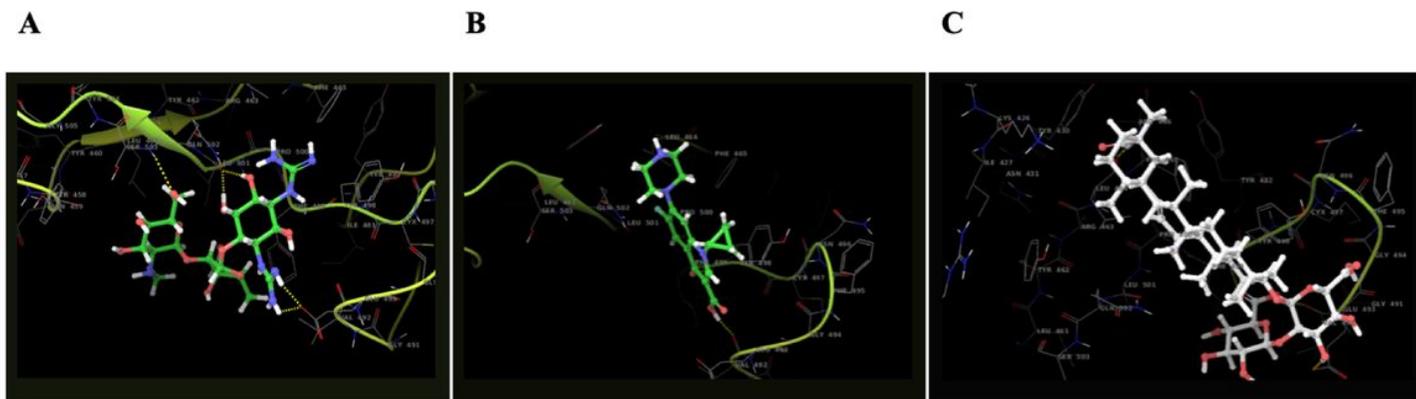


Figure 9

Three dimensional (3D) illustration of small molecule interaction in the ACE2 binding site on RBD of SARS-CoV-2 S. A: Interaction of streptomycin facilitated by 5 H-Bonds. B: Interaction of Ciprofloxacin facilitated by 2 H-Bonds. C: Interaction of GA, facilitated by 5 H-Bonds.

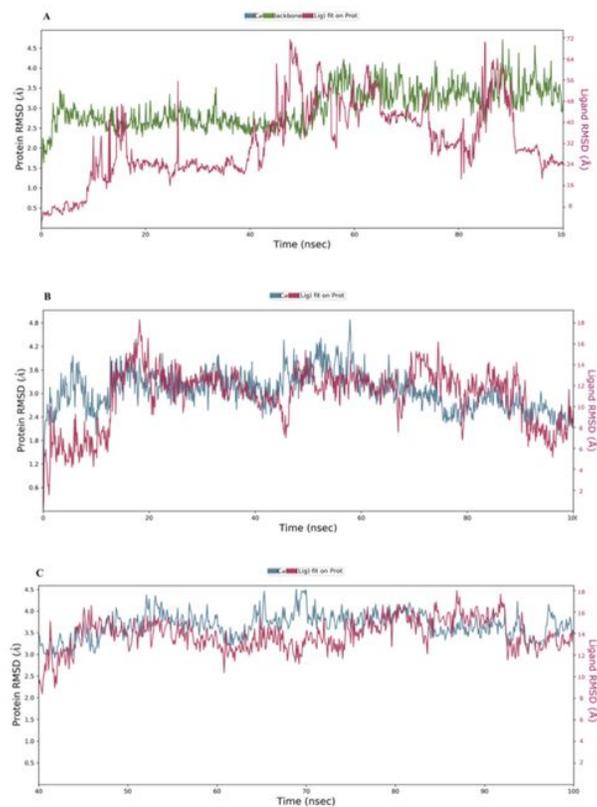


Figure 10

The Root Mean Square Deviation (RMSD) plot for protein ligand complexes generated after 100 ns MD simulation in the NPT ensemble. A: Streptomycin/SARS_nCoV-2 S complex B: Ciprofloxacin SARS_nCoV-2 S complex C: GA/SARS_nCoV-2 S complex.

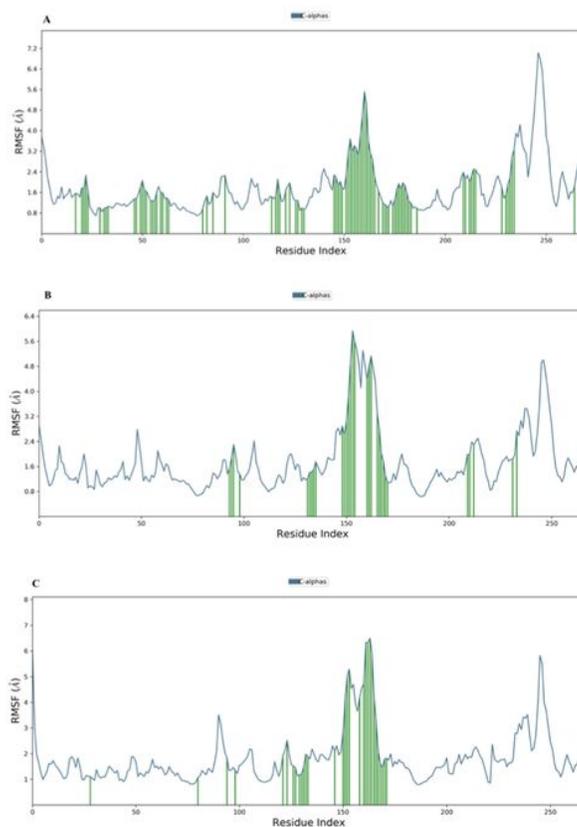


Figure 11

The Root Mean Square Fluctuation (RMSF) plot for protein ligand complexes generated after 100 ns MD simulation in the NPT ensemble. A: Streptomycin/SARS_nCoV-2 S complex B: Ciprofloxacin SARS_nCoV-2 S complex C: GA/SARS_nCoV-2 S complex.

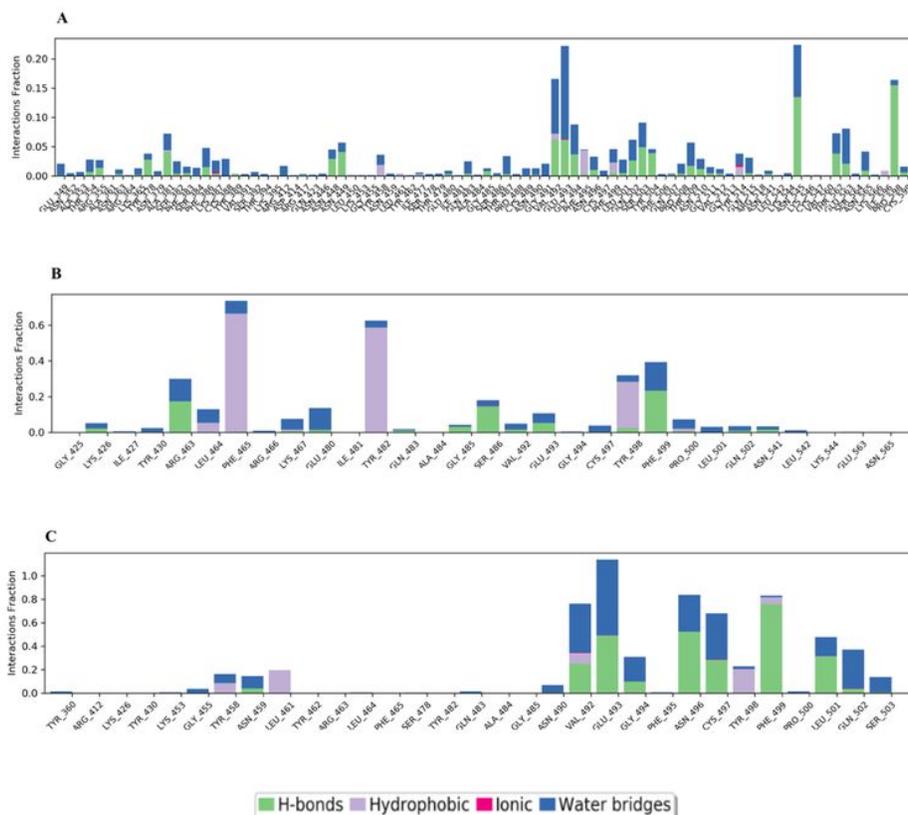


Figure 12

Histogram of protein ligand contacts generated for residues which occur more than 64.0% of the simulation time in the selected trajectory. A: Streptomycin/SARS_nCoV-2 S complex B: Ciprofloxacin SARS_nCoV-2 S complex C: GA/SARS_nCoV-2 S complex.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)