



Pathway enrichment analysis of virus-host interactome and prioritization of novel compounds targeting the spike glycoprotein receptor binding domain–human angiotensin-converting enzyme 2 interface to combat SARS-CoV-2

Pavan Gollapalli^a, Sharath B. S^b, Hrvoje Rimac^{c,d}, Prakash Patil^a, Suchetha Kumari Nalilu^a, Shivanandha Kandagalla^d and Praveenkumar Shetty^a

^aCentral Research Lab, K.S. Hegde Medical Academy, Nitte (Deemed to be University), Deralakatte, Mangalore, India; ^bDepartment of Biotechnology and Bioinformatics, Kuvempu University, Shankaraghatta, Shivamogga, India; ^cDepartment of Medicinal Chemistry, University of Zagreb, Faculty of Pharmacy and Biochemistry, Zagreb, Croatia; ^dLaboratory of Computational Modelling of Drugs, Higher Medical and Biological School, South Ural State University, Chelyabinsk, Russia

Communicated by Ramaswamy H. Sarma

ABSTRACT

SARS-CoV-2 has become a pandemic causing a serious global health concern. The absence of effective drugs for treatment of the disease has caused its rapid spread on a global scale. Similarly to the SARS-CoV, the SARS-CoV-2 is also involved in a complex interplay with the host cells. This infection is characterized by a diffused alveolar damage consistent with the Acute Respiratory Disease Syndrome (ARDS). To explore the complex mechanisms of the disease at the system level, we used a network medicine tools approach. The protein-protein interactions (PPIs) between the SARS-CoV and the associated human cell proteins are crucial for the viral pathogenesis. Since the cellular entry of SARS-CoV-2 is accomplished by binding of the spike glycoprotein binding domain (RBD) to the human angiotensin-converting enzyme 2 (hACE2), a molecule that can bind to the spike RDB-hACE2 interface could block the virus entry. Here, we performed a virtual screening of 55 compounds to identify potential molecules that can bind to the spike glycoprotein and spike-ACE2 complex interface. It was found that the compound ethyl 1-{3-[(2,4-dichlorobenzyl) carbamoyl]-1-ethyl-6-fluoro-4-oxo-1,4-dihydro-7-quinolinyl}-4-piperidine carboxylate (the S54 ligand) and ethyl 1-{3-[(2,4-dichlorobenzyl) carbamoyl]-1-ethyl-6-fluoro-4-oxo-1,4-dihydro-7-quinolinyl}-4 piperazine carboxylate (the S55 ligand) forms hydrophobic interactions with Tyr41A, Tyr505B and Tyr553B, Leu29A, Phe495B, respectively of the spike glycoprotein, the hotspot residues in the spike glycoprotein RBD-hACE2 binding interface. Furthermore, molecular dynamics simulations and free energy calculations using the MM-GBSA method showed that the S54 ligand is a stronger binder than a known SARS-CoV spike inhibitor SSAA09E3 (N-(9,10-dioxo-9, 10-dihydroanthracen-2-yl) benzamide).

ARTICLE HISTORY

Received 26 June 2020
Accepted 20 October 2020

KEYWORDS

SARS-CoV-2; protein–protein interactions (PPIs); spike glycoprotein; virtual screening; molecular docking; molecular dynamics

1. Introduction

COVID-19 is a coronavirus disease caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). The antiviral therapeutics acting on the virus exhibit various modes of action, like disabling the viral RNA synthesis and virus replication or blocking virus attachment to the host cell receptors (angiotensin-converting enzyme 2, ACE2) or to the viral structural proteins in order to inhibit the viral self-assembly process (Canrong et al., 2020). The host-SARS-CoV interaction is established through various strategies and various host cellular mechanisms are utilized by the virus for its successful multiplication during the infection (Zumla et al., 2020). The mechanism of the viral infection can be elucidated systematically by identifying the protein-protein interactions (PPIs) during the virus-host interplay (Yang et al.,

2019; Chuang et al., 2019). Viruses induce malfunction of the host cell by mimicking interacting domains of the host proteins, thus manipulating the signalling networks and cellular responses for their benefit (Pawson & Warner, 2007). In such a way, by interacting with the host proteins, viruses (e.g. SARS-CoV-2) alter host responses at a systems level.

One of the most effective ways to develop a potential drug for SARS-CoV-2 within a short period of time is by repurposing the existing compounds. Recently, several compounds with anti-COVID 19 properties were identified by this approach. Combination therapy was found to be more efficient in combating certain viruses like HIV and coronavirus and hence the synergistic effect of lopinavir, oseltamivir, and ritonavir was used against the SARS-CoV-2 protease (MPro). A better insight into the interaction between these three drugs and MPro was obtained by performing molecular

docking and molecular dynamic simulations (Nisha Muralidharan et al., 2020). A comparative analysis of SARS-CoV-2 MPro with proteases of other viruses from the Coronaviridae family and further virtual screening of phytochemicals and active ingredients of ayurvedic anti-tussive medicines in India, and the synthetic anti-viral drugs revealed several potential SARS-CoV-2 MPro inhibitors, such as delta-d-viniferin, myricitrin, chrysanthemin, myristicin, taiwanhomo-flavone A, lactucopicrin 15-oxalate, nympholide A, afzelin, biorobin, hesperidin and phyllaemblicin B. These molecules showed an equally strong binding to other SARS-CoV-2 targets, e.g. RdRp and hACE-2 (Joshi et al., 2020). Recent studies using molecular dynamic (MD) simulations of isothymol-ACE2 docked complex revealed that isothymol is a functional inhibitor of ACE2 activity and the components of *Ammoides verticillata* essential oils can be used as potential inhibitors of the ACE2 receptor-SARS-CoV-2 interaction (Abdelli et al., 2020). *In silico* studies on the binding affinity of a truncated ACE2 (tACE2) for spike glycoprotein RBD by protein-protein docking and MD simulations demonstrated that the tACE2 has a high binding affinity for the RBD when compared to the intact ACE2 and thus forms a more stable complex (Basit et al., 2020). Drugs that can interfere with the SARS-CoV-2 RBD binding to human ACE2 (hACE2) can potentially prevent SARS-CoV-2 from entering human cells. Nine short peptides that have this potential were designed by Liu et al. (2020) and MD simulations of the free peptides and their SARS-CoV-2 RBD-bound forms showed a high binding affinity of peptides to SARS-CoV-2 spike glycoprotein (Lupala et al., 2020).

In the present work, we employed computational approaches to model protein-protein interactions of the host-virus complex and functional enrichment and pathway analysis of the gene/protein set was performed. As was already said, the virus entry into the host cell is initiated by its binding to human ACE2 via the receptor-binding domain (RBD) of the spike glycoprotein and hence serves as a potential drug target (Lupala et al., 2020). Therefore, the genes/proteins which are the first neighbours of the spike glycoprotein in the interaction network were used to gain mechanistic insights into the virus-host interplay. This information was then used for the virtual screening of a small library of compounds against the spike glycoprotein RBD. The top hit molecules from this screening were then docked to the SARS-CoV-2 spike glycoprotein RBD-ACE2 interface, after which molecular dynamic simulations of the top scored compound and a reference ligand were performed to compare their binding affinities.

2. Materials and methods

2.1. Construction of the PPI network

The Search Tool for the Retrieval of Interacting Genes/Proteins database specific for viral-host interactions (STRINGvirus v11.0) was used to construct the network of the human-SARS coronavirus protein-protein interactions (Cook et al., 2018). Given the set of viral proteins, the STRINGvirus database generates a PPI network between the query proteins and their associated human proteins, with emphasis on

primary interactions. The SARS-CoV-2 shares a high nucleotide sequence identity of 79.7% with the human SARS-CoV (Zhou et al., 2020). Hence, human protein data associated with the SARS-CoV were used here to construct the protein-protein interaction network. First, based on the virus seed proteins, an interaction network was constructed associated with the human proteins. These interactions were derived based on different sources: text mining, experiments, databases, co-expression, neighbourhood, gene fusion, and co-occurrence with a mean confidence level of 0.4. Later, the number of interactions was increased to 200. Cystoscope 3.3.0 (Su et al., 2014) with default settings was used for the network visualization to analyse and calculate the properties of the nodes.

2.2. Topological analysis of the PPI network

Several topological measures, i.e. degree (k), betweenness centrality (BC), eccentricity, closeness centrality (CC), network density, diameter, average number of clusters, average shortest path length, and clustering coefficient were adopted to evaluate nodes of the PPI network (Albert & Barabási, 2002; Barabasi and Oltvai, 2004). These topological parameters were calculated using the NetworkAnalyzer (Fierner et al., 2013). The input and output values of the node are received as mathematical functions (Jeanquartier et al., 2015).

2.3. Functional enrichment analysis

A comprehensive analysis and visualization of a functionally enriched set of genes was performed using ClueGO (Bindea et al., 2009), a Cytoscape plug-in that significantly improves the biological interpretation of large lists of genes. A functionally organized GO/pathway term network was created by integrating gene ontology (GO) terms as well as KEGG pathways. We considered the first neighbours of the hub spike glycoprotein for the functional enrichment analysis. A total of 76 neighbours were found to interact with the spike glycoprotein. Parameters specified for protein/gene list enrichment analysis were set as follows: statistical test-enrichment/depletion (two-side hypergeometric test), correlation test-Bonferroni step down, Min GO level-3, Max GO level-8, Kappa score threshold-0.4, GO fusion-false, GO group-true and $p \leq 0.05$.

2.4. Identification of novel compounds against the spike glycoprotein

2.4.1. Collection of compounds

The drug-like compounds were collected from the ZINC15 database (<http://zinc15.docking.org>) (Sterling & Irwin, 2015), by using the search term 'spike glycoprotein'. The structures of all the compounds were obtained in SMILES format. The three-dimensional (3D) conformation of compounds was protonated at the physiological pH and biologically relevant tautomers were generated for each molecule. A known inhibitor of the SARS-CoV spike-glycoprotein, SSAA09E3 (N-(9,10-dioxo-9,10-dihydroanthracen-2-yl) benzamide), which

prevents the fusion of the viral membrane with the host cellular membrane and blocks the interaction of the SARS-spike glycoprotein with the ACE2 receptor (Adedeji et al., 2013) was taken as the reference molecule.

2.4.2. Virtual screening and molecular docking

The 3D structure of the coronavirus spike glycoprotein receptor-binding domain (RBD) complexed with the ACE2 receptor (PDB entry: 6LZG; resolution: 2.50 Å) was obtained from the RCSB Protein Data Bank (Wrapp et al., 2020; Berman et al., 2000). Both the receptor and the ligand molecules were prepared for docking using the UCSF Chimera 1.14 program (Pettersen et al., 2004). Initial docking calculations were performed using AutoDock Vina (ADT) (Trott & Olson, 2009) with a modified *Python* script on Ubuntu 16.04 LTS platform.

To detect the probable binding sites for all ligands with the spike glycoprotein RBD (chain B) and hACE2 (chain A), we employed a blind docking procedure for both chains separately. Thereafter, we opted for the spike glycoprotein RBD active pocket sites rather than the hACE2 receptor because hACE2 is expressed in various types of human cells and targeting hACE2 might cause more side effects. Before docking, the hACE2 domain (chain A) was deleted from the original PDB complex (6LZG). Additionally, ligand and water molecules were removed from the structure, polar hydrogen atoms and Gasteiger charges were added. All 55 ligand structures collected from the ZINC15 database (Sterling & Irwin, 2015) and the reference SSAA09E3 ligand were imported into UCSF Chimera 1.14 as SMILES strings, after which their structure was optimized using OpenBabel 2.3.2 (O'Boyle et al., 2011). As a final step before docking, receptor and ligand molecules were saved in the PDBQT format using MGL 1.5.6 of AutoDockTools (ADT) (Morris et al., 2009).

Initially, the spike glycoprotein RBD-hACE2 complex was used to obtain the interface residues by using PDBSum (Laskowski et al., 2018). A grid map of size $16 \times 16 \times 16$ Å was generated with a 1.0 Å spacing to cover the interface area (centred at $-34.655, 30.216, 0.971$). For initial compound screening, both exhaustiveness and the number of binding modes were set to 10. Docking calculations were first performed for the spike glycoprotein RBD (chain B). The top five molecules underwent a second round of docking, with the exhaustiveness parameter set to 100 for a better conformation search. Docking was conducted for both the spike glycoprotein RBD alone (chain B) and the spike glycoprotein RBD-ACE2 complex (at the interface). Top hit molecules were then analysed and visualized using Maestro 12.3 (Schrodinger Release 2020) and ChimeraX (Goddard et al., 2018).

2.5. Molecular dynamics (MD) simulations

MD simulations for all ligands (the reference SSAA09E3 ligand and the S54, S5, S21, S43, and S55 ligands) were run in complex with the SARS-CoV-2 spike glycoprotein (6LZG). Docked positions with the highest affinities to the protein were used as starting points for the MD simulations. AMBER

ff14SB force field (Maier et al., 2015) was used to model the enzyme and GAFF2 force field (as implemented in *Antechamber* (Wang et al., 2006), was used in the case of ligands. Such protein-ligand complexes were solvated in a truncated octahedral box of TIP3P water molecules spanning a 12 Å thick buffer, and Na⁺ and Cl⁻ ions were added according to Machado and Pantano (2020) to achieve a neutral environment with a salt concentration of 0.15 M (with the number of water molecules 41647, Na⁺ ions 124, and Cl⁻ ions 99 in the case of the standard SSAA09E3 ligand, number of water molecules 41633, Na⁺ ions 124, and Cl⁻ ions 99 in the case of the S54 ligand, number of water molecules 41650, Na⁺ ions 124, and Cl⁻ ions 100 in the case of the S5 and S43 ligands, number of water molecules 41659, Na⁺ ions 124, and Cl⁻ ions 100 in the case of the S21 ligand, and number of water molecules 41634, Na⁺ ions 124, and Cl⁻ ions 99 in the case of the S55 ligand). Such structures were then submitted for geometry optimization in the AMBER16 program (Case et al., 2014), employing periodic boundary conditions in all directions. For the first 1500 cycles, the complex was restrained and only water molecules were optimized, after which another 2500 cycles of optimization followed where both water molecules and the complex were unrestrained. Optimized systems were gradually heated from 0 to 300 K and equilibrated during 30 ps using NVT conditions, followed by productive and unconstrained MD simulations of 300 ns employing a time step of 2 fs at constant pressure (1 atm) and temperature (300 K), the latter held constant using Langevin thermostat with a collision frequency of 1 ps^{-1} . Bonds involving hydrogen atoms were constrained using the SHAKE algorithm (Ryckaert et al., 1977), while the long-range electrostatic interactions were calculated employing the Particle Mesh Ewald method (Darden et al., 1993). The non-bonded interactions were truncated at 11.0 Å. Analysis of the trajectories was performed using the *cpptraj* module of AmberTools16 (Roe & Cheatham, 2013).

2.6. Binding free energy calculations and decomposition

The binding energy, ΔG_{bind} , of simulated complexes was calculated using the MM-GBSA (Molecular Mechanics – Generalized Born Surface Area) protocol (Genheden & Ryde, 2015; Hou et al., 2011), available as a part of AmberTools16 (Case, 2016). MM-GBSA is a method for the calculation of ΔG_{bind} from snapshots of MD trajectory (Ferenczy, 2015) with an estimated standard error of 1–3 kcal/mol (Genheden & Ryde, 2015). ΔG_{bind} is calculated in the following manner:

$$\Delta G_{\text{bind}} = \langle G_{\text{complex}} \rangle - \langle G_{\text{protein}} \rangle - \langle G_{\text{ligand}} \rangle$$

where the symbol $\langle \rangle$ represents the average value over 100 snapshots collected from a 30 ns part of the corresponding MD trajectories. The whole trajectory was divided into 10 parts of 30 ns length and ΔG_{bind} was calculated for all 10 parts of the simulation and reported as mean \pm standard deviation. The calculated MM-GBSA binding free energies were decomposed into specific residue contribution on a per-residue basis according to established procedures. This protocol calculates the contributions to ΔG_{bind} arising from

Table 1. Extended network topological measures of two protein-protein interaction networks.

Network parameters	Giant network (Core network)	Subnetwork
Number of nodes/edges	374/5828	302/5025
Clustering coefficient	0.250	0.261
Average number of neighbours	31.328	33.278
Network density	9	9
Shortest path (%)	29%	29%
Characteristic path length	2.423	2.227

each amino acid side chains and identifies the nature of the energy change in terms of interaction and solvation energies, or entropic contributions (Gohlke et al., 2003; Rastelli et al., 2010). In this case, the entropy term was not calculated.

2.7. Pharmacokinetics and metabolic site prediction

Identification of potential metabolic sites of a drug can give key information of its pharmacokinetic and pharmacodynamic characteristics. Drugs are commonly metabolized by a special class of enzymes which are known as cytochrome P450 (CYP) enzymes. In this concern, by using the SMARTCyp 3.0 tool (Rydberg et al., 2019), metabolic sites for CYP mediated metabolism were predicted for the top hit molecule.

3. Results and discussion

3.1. Sars-CoV-human protein-protein interaction network

The protein-protein interaction was constructed by assembling the SARS-CoV associated human proteins using the STRINGvirus database. Based on various experimentally collected data, we obtained nearly 360 human proteins associated with SARS-CoV (supplementary material, Table 1). It was also found that these proteins are involved in crucial pathways of the viral infection. The core part (the core network) of the human SARS-CoV-host PPI network (the giant network) generated by using the STRING database consisted of 374 nodes and 5827 edges (Figure 1). The number of edges connected to a designated node is termed a degree, implying the significance of the protein in the biological interactions. The highest degree in the core network was found to be 43, while the average degree was 15.6. The PPI network is characterized by a small number of highly connected nodes, while most of the nodes have only a few connections.

The nodes which degrees or BC are in the top 5% were considered as the key nodes, i.e. the critical points. Out of 374 nodes in the network, the top 10 nodes with the highest BC values were: the spike glycoprotein, ACVR1B, CD44, ALB, MYC, B2M, CREB1, PHB2, STAT3, and IL6. These include both the viral and the human proteins, with the spike glycoprotein identified as the hub node that was further validated as an important target protein. To distinguish these nodes in the network and their roles, they are highlighted in a different colour (Figure 1, supplementary material). The spike glycoprotein was identified as the hub protein with the highest degree and the second highest BC value, while ACVR1B

is the second hub protein with the highest BC value and the second highest degree. The proteins which are directly interacting (first neighbours) with the spike glycoprotein are shown in Figure 2 (supplementary material).

The SARS-CoV spike glycoprotein, identified as the key protein, is involved in binding to the ACE2 receptor, a human cell receptor, through its receptor-binding domain (RBD). RBD-up conformation of the spike glycoprotein is a prerequisite for the formation of the RBD-ACE2 complex (Walls et al., 2020). A drastic conformational change is found to be triggered in the S2 domain of spike glycoprotein due to the specific interaction between the receptor-binding domain and ACE2 receptor, which leads to the viral fusion with the cellular membrane and the nucleocapsid release into the cytoplasm of the host cell (Lan et al., 2020). The second identified hub protein (ACVR1B) in the interactome is a transmembrane serine/threonine kinase activin type-1 receptor.

3.2. Network topology analysis

Network Analyzer v.3.3.1 was employed to evaluate the confidence of the core interactome, using the power law fit of the form $y = ax^b$. Power law uses the least square method to determine the topological parameters and considers the points with positive coordinate values for the fit. The betweenness centrality (BC), closeness centrality (CC), and topology correlation coefficient scores of 0.770, 0.321, and 0.250, respectively, were considered as network topology parameters. Additionally, the neighbourhood connectivity (0.472) and the shortest path length distribution were also considered in the analysis. The hub proteins in the interactome were determined by using network topology parameters like BC and topological correlation coefficient with a cut-off value of 0.7 and 0.6, respectively. These topological parameters considered for the network generation by using the above cut-off values were graphically plotted (Figure 3A–D, supplementary material). The extended global network topological measures of the two protein-protein interaction networks, i.e. the giant or the core network and the backbone or the subnetwork, are presented in Table 1. Therefore, the biological process is essentially regulated by the bottleneck node in the interactome (Table 2, supplementary material).

3.3. Functional enrichment and pathway analysis

Functional enrichment analysis was carried out using ClueGO, a plug-in for Cytoscape. A total of 41 GO terms were collected, out of which 29, 6, and 6 GO terms corresponded to biological processes (Figure 2(a)), molecular function (Figure 2(b)) and pathways (Figure 2(c)), respectively (listed in the supplementary material, Tables 3–5). Genes related to the specific GO terms are presented in Figure 3.

The protein interaction network from STRINGvirus showed a dense network of the spike glycoprotein with 76 first neighbour nodes. We carried out a functional enrichment analysis of these protein interactions, which showed that the proteins in the network play a major role in biological

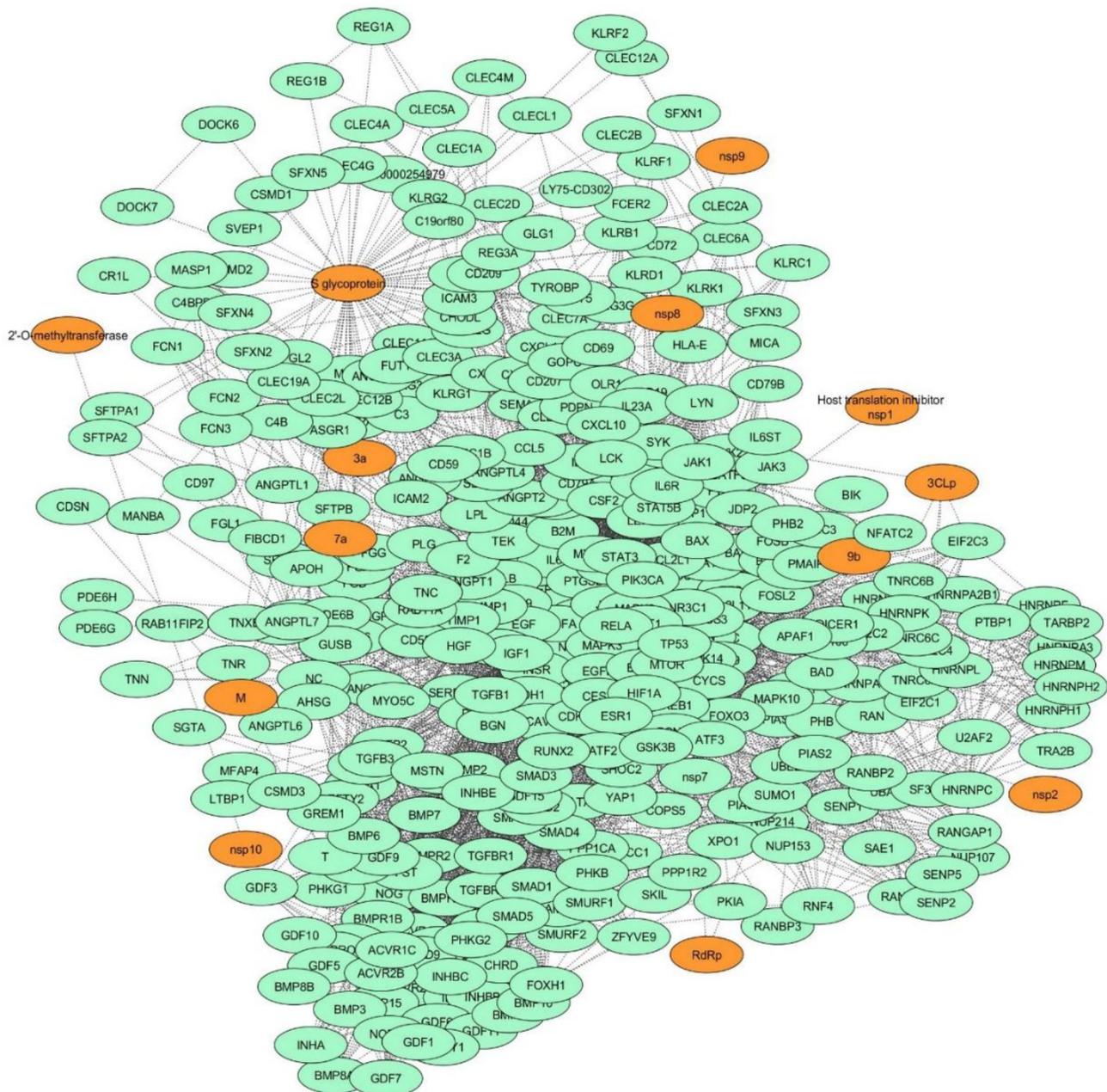


Figure 1. Extended interactome construction of SARS-CoV-human interactions obtained from the viral proteins (orange colour) and associated human proteins showing 374 nodes and 5827 edges.

processes related to the viral entry into the host cell (GO:0046718), the heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules (GO: 0007157), the transition metal ion homeostasis (GO: 0055076), the natural killer cell-mediated immunity (GO:0042267), the glycogen catabolic process (GO: 0005980) and the regulation of humoral immune response mediated by circulating immunoglobulins (GO: 0002923), as well as molecular functions, such as the viral receptor activity (GO: 0001618), the MHC protein binding (GO:0042287), the phosphorylase kinase activity (GO:0004689) and the mannose binding (GO:0048029). Enriched pathways (REACTOME) included immunoregulatory interactions between a lymphoid and a non-lymphoid cell, DAP12 interactions, glycogenolysis, and the complement cascade. The GO term related to the reactions is reported as the

genes involved in exocytosis of tertiary granule membrane (R-HAS-6798747, CLEC15A, CLEC5A, and OLA1). The interaction of SARS-CoV spike glycoprotein with cellular receptors is indispensable for the viral entry into the host cells. From these enriched GO terms, we have identified the genes associated with the viral entry into the host cell and the viral receptor activity: CD209, CD55, CLEC4M, CLEC5A, and CLECAG. The CLEC4M (C-type lectin domain family 4 member M) acts as an attachment receptor for SARS-CoV (Marzi et al., 2004). It was demonstrated that both the ACE2 and CLEC4M (CD2009) are highly expressed in human lung microvascular endothelial cells and lymphatic endothelial cells, respectively (Jing et al., 2007). The ACE2 and other proteins are associated with CLEC4M through a primary interaction and act as a receptor for binding of the viral spike

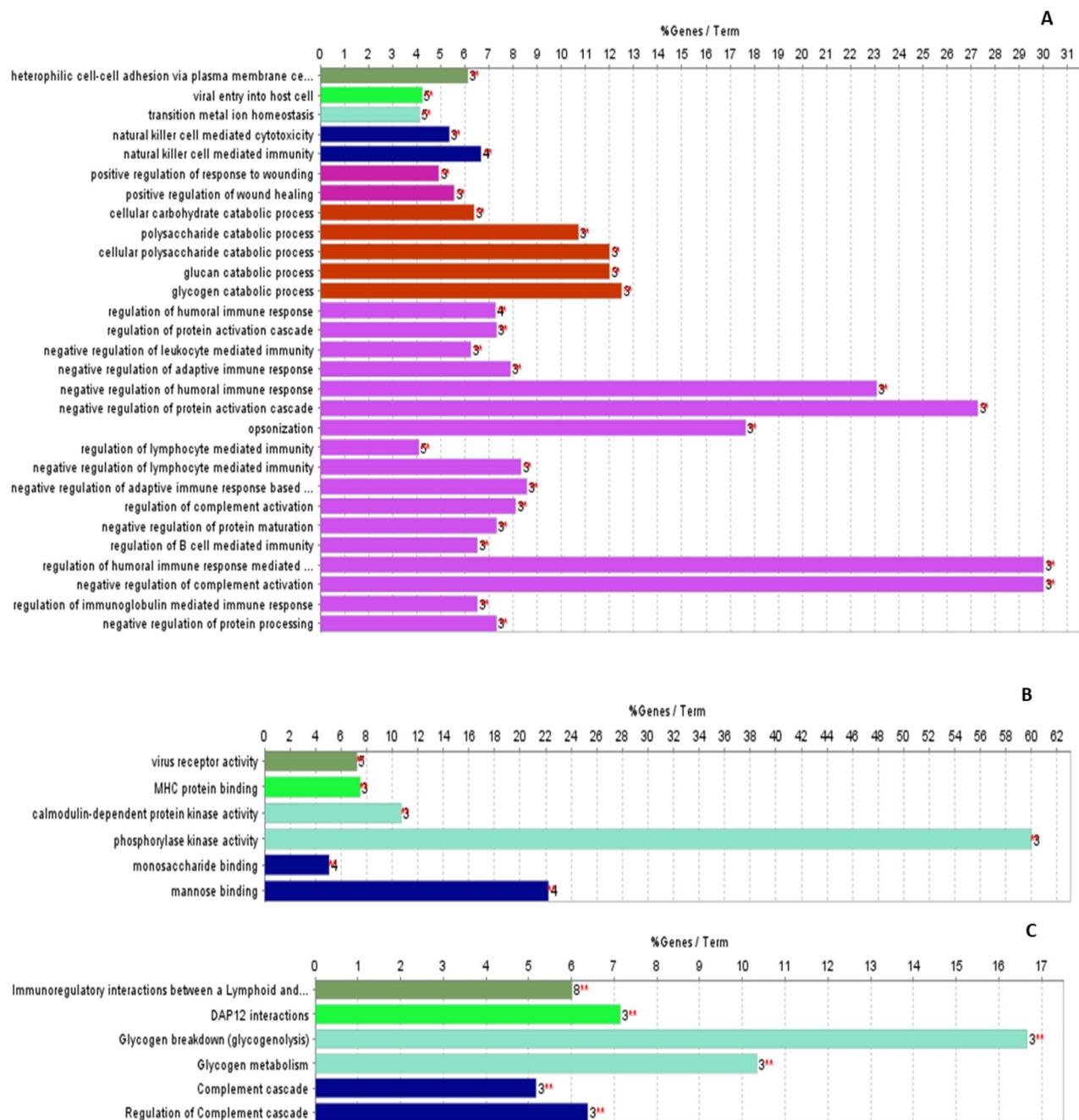


Figure 2. GO/pathway terms specific for upregulated genes from the first neighbour nodes of the spike glycoprotein representing: the biological function (a), molecular function (b), and pathway analysis (c) (60 nodes). The bars represent the number of genes associated with the terms. The percentage of genes per term is shown as a bar label. Terms with up- and down-regulated genes are shown in red and green, respectively. The colour gradient shows the gene proportion of each cluster associated with the term. Equal proportions of the two clusters are represented in white.

glycoprotein (Figures 2 and 3). Several studies have demonstrated that the interaction between the spike glycoprotein and the ACE2 receptor is found to be crucial for the viral entry into the host and thus targeting this mechanism and identifying the inhibitors to interrupt this interaction could result with promising lead compounds (Zumla et al., 2016).

3.4. Selection of ligand binding site in the spike glycoprotein and binding interaction analysis

Recent reports showed that the SARS-CoV-2 spike glycoprotein uses the ACE2 receptor to enter the host cell and that

the RBDs of SARS-CoV-2 and SARS-CoV spike glycoproteins bind with similar affinity to the human ACE2 receptor (Walls et al., 2020). The ligand-binding specificity of the spike glycoprotein RBD was detected by running blind docking and setting both the exhaustiveness and the number of modes to 10. Using blind docking, we analysed the cavity at the spike glycoprotein RBD-hACE2 interface. The results suggest that the majority of ligands' best scored poses were found in the A and C pockets. Additionally, most of the ligands bind at the interface (the C pocket) and have higher binding affinities compared to the A pocket (Figure 4). Therefore, targeting this position may contribute to the interruption of the

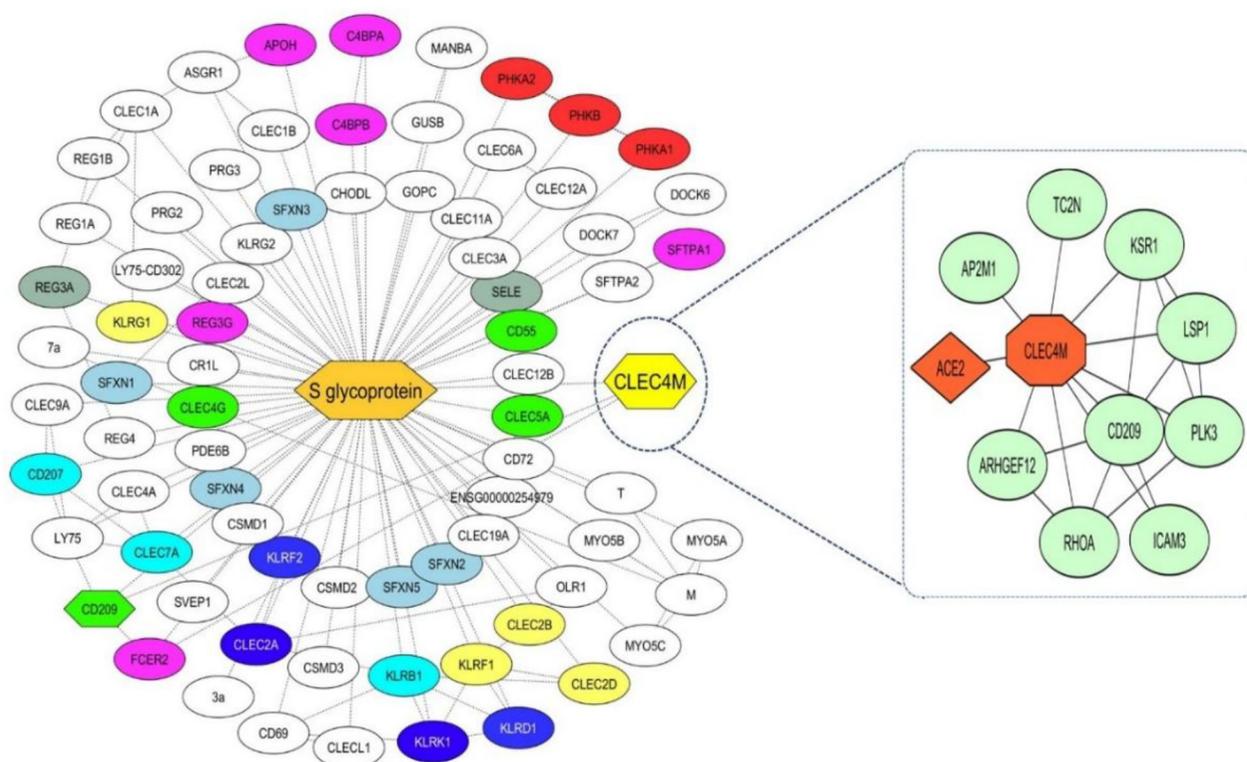


Figure 3. Functional assessment analysis of the spike glycoprotein first neighbour proteins. The genes recognized as close neighbours of the spike glycoprotein are highlighted in different colours based on their functional enrichment: genes for the viral entry into the host cell and the viral receptor activity are shown in green, genes for the heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules in grey, genes for the transition metal ion homeostasis in light blue, genes for the natural killer cell-mediated immunity in dark blue, genes for the glycogen catabolic process in red, genes for the regulation of humoral immune response mediated by circulating immunoglobulin in pink, genes for the MHC protein binding in sky blue, and genes for the immunoregulatory interactions between a lymphoid and a non-lymphoid cell in yellow. The genes shown in the subnetwork (orange) show the interaction of CLEC4M protein with the human ACE2 protein (receptor for the spike glycoprotein), and the genes which are involved in the viral entry into the host.

Table 2. Docking result analysis of the top 5 ligands against the spike glycoprotein RBD alone and the spike glycoprotein-hACE2 complex interface.

Lead Compounds	Binding energy (kcal/mol)	No. of H-bonds	Spike glycoprotein-hACE2 complex (at the interface)		
			H-bond interaction with	Distance (Å)	Hydrophobic interaction
S5	-8.6	1	His 34 A	2.85	Tyr 505B; Tyr495B; Tyr453B
S21	-8.2	1	Arg403B	2.90	Tyr495B; Tyr453B
S43	-8.0	1	Arg403B	3.09	Tyr495B; Tyr505B; Tyr453B; Ala387A; Pro389A
S54	-9.2	3	Asp30A; Asn33A; Gly496B	3.23	Val93A; Leu29A; Tyr453B; Tyr505B; Tyr495B
				3.24	
				3.04	
S55	-8.7	2	Asp30A; Gly496B	3.11	Val93A; Leu29A; Tyr495B; Phe497B; Tyr505B
				3.03	
Standard (SSAA093)	-8.5	1	Arg 403B	2.98	Tyr495B; Ala387A; Tyr505B; Tyr453B

interaction between the spike glycoprotein RBD and the hACE2. Due to its high ranking and position advantages, we used this cavity (the C pocket) for further studies under the assumption that targeting this region may induce conformational changes that could inhibit virus-host interactions and prevent viral entry.

Among the 55 compounds identified from the ZINC15 database, binding energies for most of the ligands were found to be between -6 and -7 kcal/mol (46 ligands). One ligand (S28) showed a lower affinity with the binding energy of -5.8 kcal/mol and eight ligands showed binding energies of -7 kcal/mol or lower (supplementary material, Table 6). Among them, the top 5 ligands were selected based on the interaction pattern with the spike glycoprotein RBD (chain B)

for the second round of the docking by setting the exhaustiveness parameter to 100. Analysis of the top 5 molecules interacting with the spike glycoprotein alone showed a good inhibition potential. The S05 molecule forms two H-bonds, with Asn501B and Gly496B of spike glycoprotein and has a binding energy of -7.9 kcal/mol. Similarly, calculated binding energies for the S54 and S55 ligands were -8.4 and -7.9 kcal/mol, respectively. However, they form only one favourable H-bond interaction, with Gly498B and Asn501B, respectively. On the other hand, the other three molecules form no H-bonds with the spike glycoprotein RBD (Table 2). After this, we merged the hACE2 domain with the spike glycoprotein RBD-S54 complex, and found a hydrophobic interaction between spike glycoprotein RBD and Tyr41A of

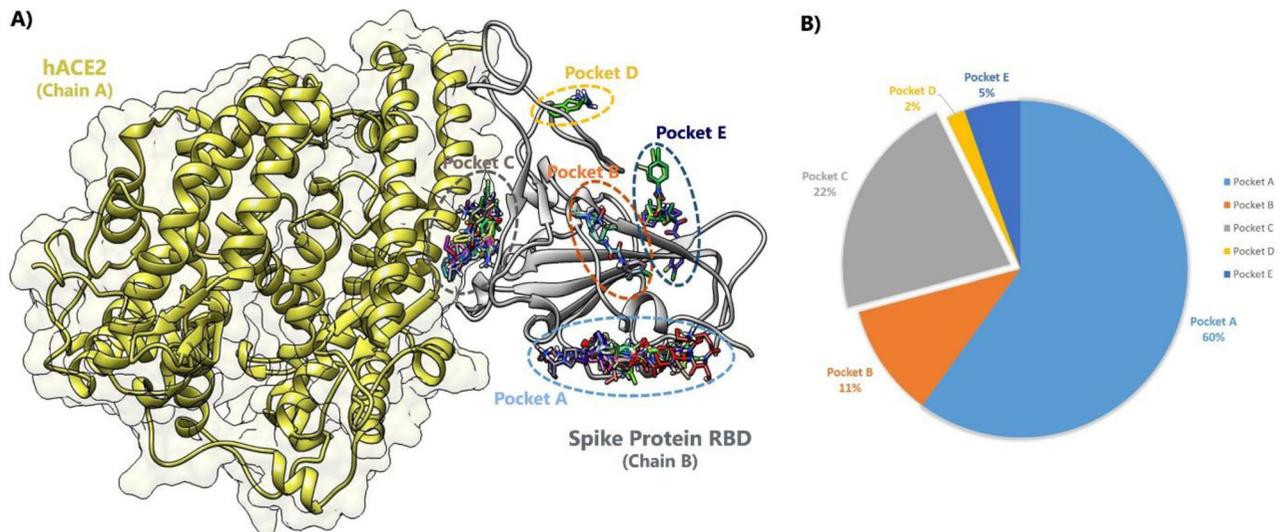


Figure 4. (a) Blind docking of using the spike glycoprotein RBD (grey) for ligand specificity search. Active pocket (pockets A to E) regions show clusters of ligands in different sites. (b) Pie chart showing the percentage of ligands at different active sites bound to the spike glycoprotein RBD.

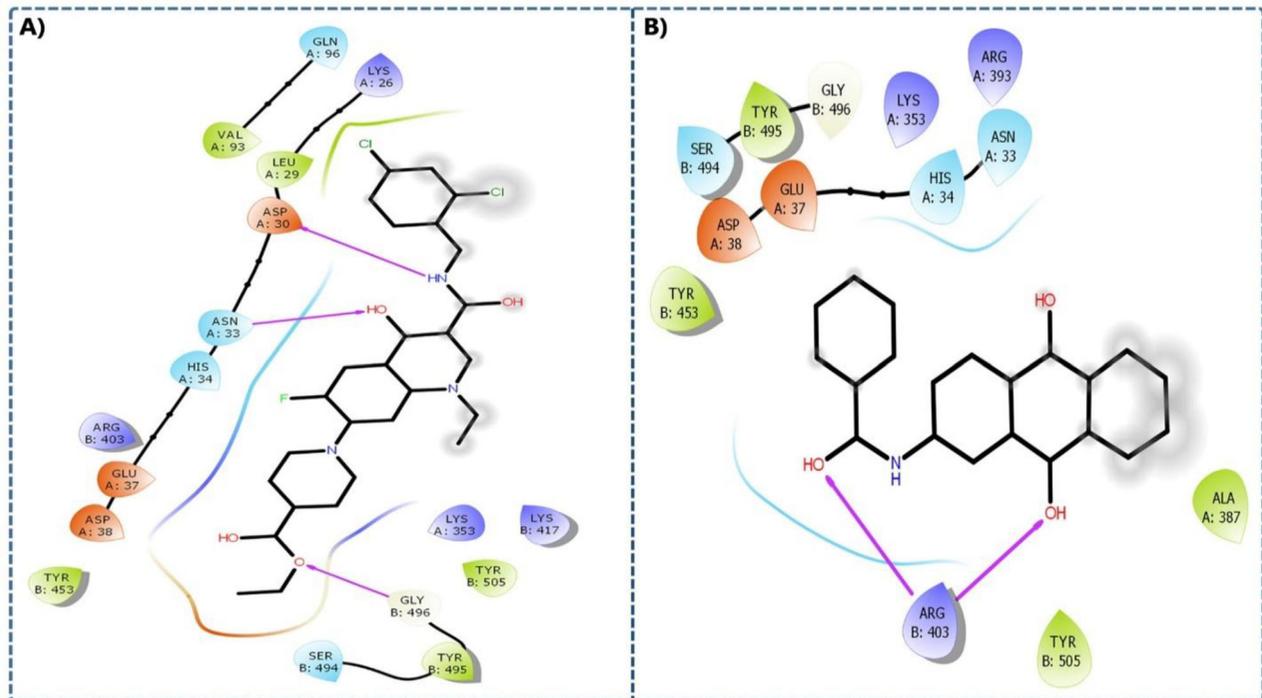


Figure 5. Binding interactions of the S54 and SSAA093 ligands at the interface of the spike glycoprotein RBD-hACE2 complex.

the hACE2 domain (Yan et al., 2020). The binding analysis of the S54 ligand compared to the reference SSAA093 ligand is shown in Figure 5.

The results suggest that the H-bond interaction with the Asn501B residue of the spike glycoprotein may play a key role in destabilizing hACE2 (chain A) by forming hydrophobic interactions with the Tyr41A residue since Tyr41A was found to play a crucial role in the interaction of the spike glycoprotein RBD-hACE2 complex formation (Lan et al., 2020; Yan et al., 2020). These results encouraged us to investigate the complicated role of the spike glycoprotein RBD-hACE2 complex in

virus entry into the cells. To focus on the role of amino acid residues Asn501B and Tyr41A, we performed a docking simulation of the top 5 ligands at the interface of the spike glycoprotein RBD-hACE2 complex (amino acid residues were taken from PDBSum) by setting exhaustiveness to 100 (supplementary material, Table 7 and Figure 4). These calculations showed significant H-bond interactions of all 5 ligands and the reference SSAA093 ligand with amino acids His34A, Arg403B, Asp30A, Asn33A, and Gly496B at the interface. Equally, all 5 ligands form hydrophobic interactions with amino acids Tyr505B, Tyr495B, Tyr453B, Pro389A, Ala387A, Val93A, and

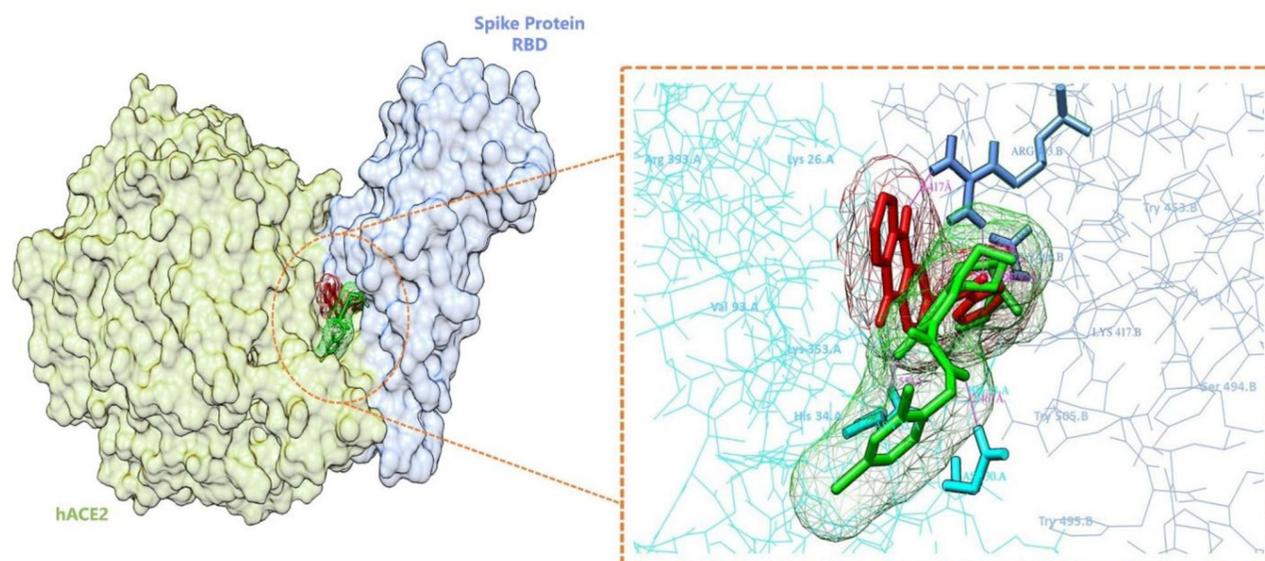


Figure 6. Superimposed conformations of the S54 and SSAA093 ligands at the interface of the spike glycoprotein RBD-hACE2 complex. The S54 ligand is shown in green, while the SSAA093 ligand is shown in red.

Backbone RMSD for spike glycoprotein complexes over time

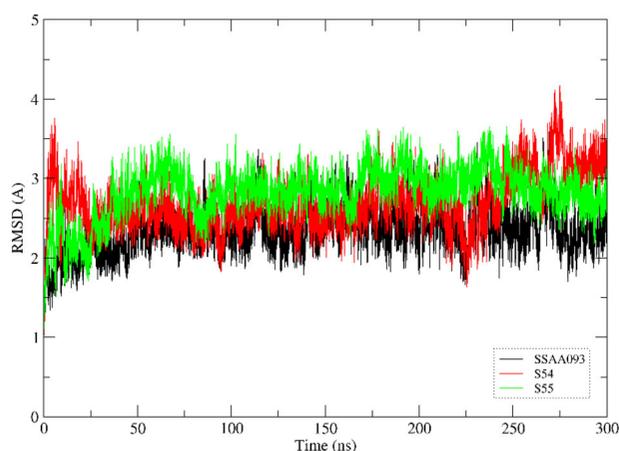


Figure 7. Backbone root-mean-square deviation (RMSD) over time for complexes with the reference SSAA093, the S54, and the S55 ligand.

Leu29A. The binding modes of S54 and the reference SSAA093 ligand at the interface of spike-ACE2 are shown in Figure 6.

S55, which has a similar fingerprint as S54, forms two H-bond interactions with Asp30A and Gly496B and forms favourable hydrophobic interactions with Val93A, Leu29A, Tyr495B, Phe497B, and Tyr505B. Tyr505B, Tyr495B, and Tyr453B were also found to play a key role in forming interaction with the spike glycoprotein alone. Interestingly, Robetta alanine scanning (Kortemme et al., 2004) results also replicated the docking results by confirming the mutagenic amino acid residue Tyr41A (hACE2) with a DDG complex score of 4.88 kcal/mol and Tyr505B (Spike glycoprotein) with DDG complex score of 1.54 kcal/mol. The docking calculations at the spike glycoprotein RBD-hACE2 interface showed an even better binding energy and formation of H-bond interactions, and the detailed interaction analysis for the top 5 ligands is reported in Table 2. Among the top 5 ligands,

S54 and S55 were found to show good binding energies, H-bond interactions, hydrophobic interactions, as well as alanine scanning by forming favourable interaction with Tyr41A and Tyr505B. Therefore, apart from H-bond interacting amino acids Asn501B, Asp30A, Asn33A, Gly496B, the hydrophobic residues Tyr41A and Tyr505B were also found to be important hotspots responsible for the binding of the spike glycoprotein to the hACE2 domain. Interestingly, S54 forms interactions with hotspot residues Tyr41A and Tyr505B along with the H-bond interactions at the interface of the spike glycoprotein RBD (Gly496B) and the hACE2 domain (Asp30A, Asn33A). These observations clearly explain the main reason behind the tighter affinity of the SARS-CoV-2 spike glycoprotein to hACE2 when compared to that of SARS-CoV (Masters, 2006). Our results suggest that S54 is a potent inhibitor of both spike glycoprotein alone, as well as the spike glycoprotein RBD-hACE2 complex. Structure-based drug design showed that the interaction of S54 at the interface of the spike-hACE2 involves forming 3 hydrogen bonds and favourable hydrophobic interactions, which play a major role in destabilizing the spike-hACE2 interaction, thus inhibiting viral entry into human cells.

The top 5 molecules were subjected to molecular fingerprinting analysis (FP) with MACCS and Morgan circular fingerprint method to check their similarity against all the 55 molecules collected from ZINC15 database. The importance of the FP method selection for virtual screening was highlighted and the difference in results obtained by different FP approaches was analysed (Cereto-Massagué et al., 2015; Matsuyama & Ishida, 2018). The Extended Connectivity Fingerprint Diameter (ECFP) offers the highest precision on average, according to database search by compound similarity based on FP (Riniker & Landrum, 2013). In our analysis, the compound S54 and S55 are structurally similar, with both having piperazine, benzene, and 1H-quinolin-4-one in the structure. The only difference between these two compounds is the presence of the nitrogen atom in the

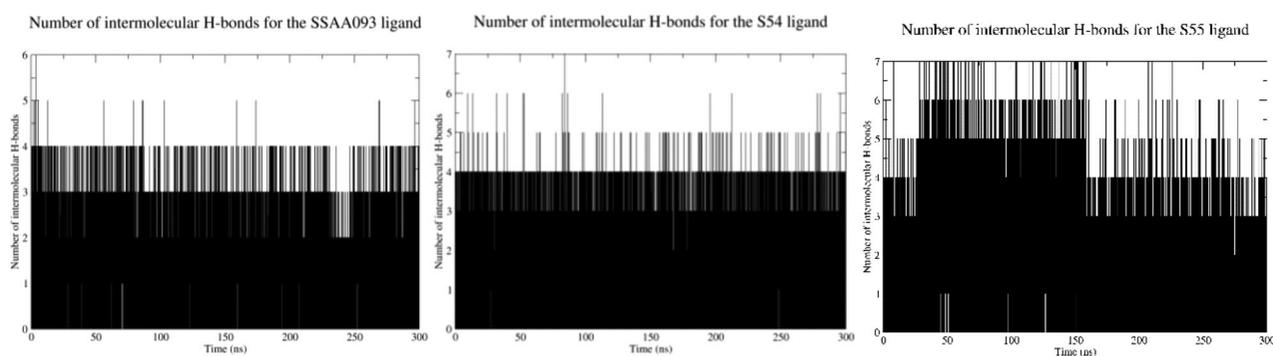


Figure 8. Number of intermolecular H-bonds for complexes with the reference SSAA093 ligand (left), the S54 (middle), and the S55 ligand (right).

Table 3. Contribution of top individual residues to the binding of tested ligands with the MM-GBSA approach. All values are in kcal/mol.

Standard SSAA093		S54		S55	
Residue	Contribution	Residue	Contribution	Residue	Contribution
Pro389A	-1.84	Pro389A	-2.59	Tyr495B	-2.54
Arg403B	-1.72	Lys26A	-1.56	Arg403B	-2.50
His34A	-1.37	Thr92A	-1.15	His34A	-2.05
Asn33A	-1.14	His34A	-0.99	Lys417B	-1.59
Tyr453B	-0.61	Asn33A	-0.98	Gly496B	-1.14
Gln388A	-0.50	Arg403B	-0.86	Tyr453B	-0.97
Tyr495B	-0.46	Tyr505B	-0.86	Gln409B	-0.94
Tyr505B	-0.36	Val93A	-0.76	Tyr505B	-0.86
Ala387A	-0.30	Gln388A	-0.57	Phe497B	-0.83
Arg393A	-0.28	Asn90A	-0.48	Lys353A	-0.82
Total	-23.09	Total	-32.13	Total	-44.55

piperazine ring of the S55 compound, which makes the compound more rotatable. Further, molecular fingerprinting analysis showed a similarity score of 0.79 between S54 and S55 in Morgan circular fingerprint. Other lead compounds such as S5, S21 and S43 also share common traits, such as the chloro-fluorobenzene functional group. The S5 compound along with chloro-fluorobenzene, contains also the 5-azaspiro [3.5] nonane functional group and the S21 compound contains pyrrolidine and cyclohexane rings along with the chloro-fluorobenzene group. Finally, S43 compound contains a piperidine ring along with the chloro-fluorobenzene group. Among these, the S43 compound shows the similarity of score 0.60 and 0.63 with S5 and S21, respectively (supplementary material, Table 8).

3.5. Molecular dynamics (MD) simulations and free energy calculations

MD simulations were carried out for all the top five compounds (S54, S5, S21, S43, and S55) and the reference SSAA093 ligand complexed with the spike glycoprotein. All complexes were found to be stable throughout the entire duration of the simulation (300 ns). However, since only the ligands S54 and S55 had ΔG_{bind} significantly lower than the reference SSAA093 ligand, they will be discussed more thoroughly (the complete MM-GBSA results are shown in Table 9 supplementary material).

Figure 7 shows the backbone mass-weighted root-mean-square deviation (RMSD) for the S54, S55 and the reference SSAA093 ligand complexes through time. It can be seen that

all three complexes achieve the equilibrium state very early and remain stable for the entire duration of the simulation. For the SSAA093 and the S54 ligand, this can also be seen in the intermolecular H-bond graph (Figure 8), where the number of H-bonds for all three ligands remains constant through time (with the reference SSAA093 ligand forming on average 1.39 ± 0.62 , ligand S54 1.33 ± 0.78), while the ligand S55 (1.42 ± 1.25 intermolecular H-bonds) undergoes a slight conformational change around the 150 ns mark, but without any significant influence on the protein structure. This indicates that the docking procedure was successful in finding the correct binding poses for the SSAA093 and the S54 ligands since these ligands did not need to optimize their conformation inside the binding pocket, while the S55 ligand had to go through some conformational changes to obtain the optimal pose. This is in accordance with the docking results, where the binding affinity of the S54 and S55 ligands are almost the same, while the MM-GBSA results differ.

Binding energies for all the complexes were calculated using the MM-GBSA protocol. Since all the complexes are stable throughout the entire simulation, their entire trajectories (300 ns) were divided into ten segments of 30 ns. ΔG_{bind} was calculated for all segments individually and the final ΔG_{bind} was calculated as mean \pm standard deviation. For the reference SSAA093 ligand $\Delta G_{\text{bind}} = -23.09 \pm 6.01$ kcal/mol, for the S54 ligand $\Delta G_{\text{bind}} = -32.13 \pm 4.16$ kcal/mol, and for the S55 ligand $\Delta G_{\text{bind}} = -44.55 \pm 6.65$ kcal/mol (Table 3). ΔG_{bind} for ligands S5, S21, S43 are -20.85 ± 2.74 kcal/mol, -26.89 ± 3.31 kcal/mol, and -26.06 ± 1.33 kcal/mol, respectively (supplementary material, Table 9). It has to be

emphasized that since the entropy term was not calculated, these results are overestimated in their absolute terms (Genheden & Ryde, 2015). However, since all ligands bind to the same binding site of the same protein, the entropic contribution in both cases would also be approximately the same. Therefore, this method can be used in predicting relative binding energies in biomolecular complexes and their comparison (Homeyer & Gohlke, 2012). For this reason, the obtained ΔG_{bind} of the tested compounds should only be analysed relative to each other. That being said, the MM-GBSA results are in accordance with the H-bonds analysis: the S55 ligand forms a slightly higher number of intermolecular H-bonds than the S54 ligand and has a slightly higher binding affinity, similar to the docking results. Additional decomposition of ΔG_{bind} (supplementary material, Table 9

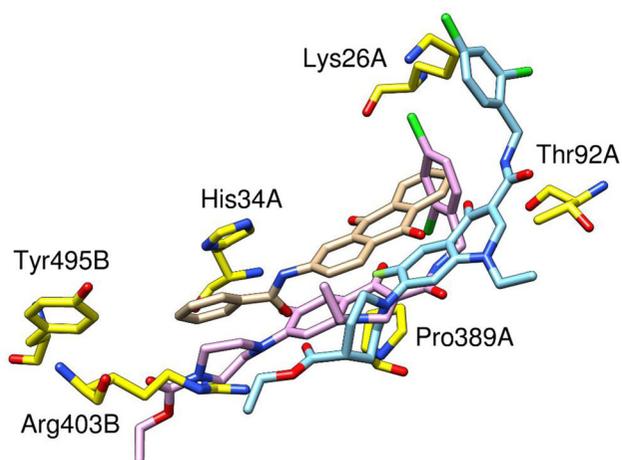


Figure 9. Overlay of the spike glycoprotein complexes with the reference SSAA093 ligand (tan), the S54 ligand (light blue), and the S55 ligand (pink) after 300 ns MD simulation with top contributing amino acid residues (yellow) shown.

(A), 9 (B), and 9 (F)) show that throughout the simulation time, S54 and S55 ligands are more stable in their binding sites, with a lower standard deviation of all types of interactions, excluding the non-polar solvation interaction for the S54 ligand, and with significantly more favourable van der Waals and electrostatic interactions.

From Table 3 it is also visible which amino acid residues contribute the most in the binding of the tested ligands. In the case of the reference ligand and the S54 ligand, the most contributing residue is Pro389A, but for the S54 ligand, its contribution is significantly higher. Additionally, the S54 ligand forms strong bonds with residues Lys26A and Thr92A, which are not present in the list of top 10 contributing residues in the case of the standard SSAA093 ligand. On the other hand, the standard SSAA093 ligand forms a much stronger bond with the His34A residue than the S54 ligand. However, even with these differences, the most important amino acid residues for these two complexes are more similar than compared to the complex with the S55 ligand. As opposed to them, in the case of the S55 ligand, eight of the ten most important amino acid residues come from the B chain, which is a result of its position deeper inside the interaction pocket (Figures 9 and 10). Given that the structural difference between ligands S54 and S55 is only in one nitrogen atom (piperidine and piperazine rings, respectively), it goes to show how small structural changes can have a significant influence on ligand binding. An interesting observation can also be made when comparing the per residue root-mean-square fluctuation (RMSF) for all three complexes (Figure 11). While the A chain has practically the same RMSF in all three cases, RMSF for the chain B shows significant differences in residues 355–395. For all amino acid residues except Ser375, RMSF is the lowest for the S55 ligand, followed by S54 ligand and the reference ligand, which corresponds to their decreasing binding affinities. It also seems

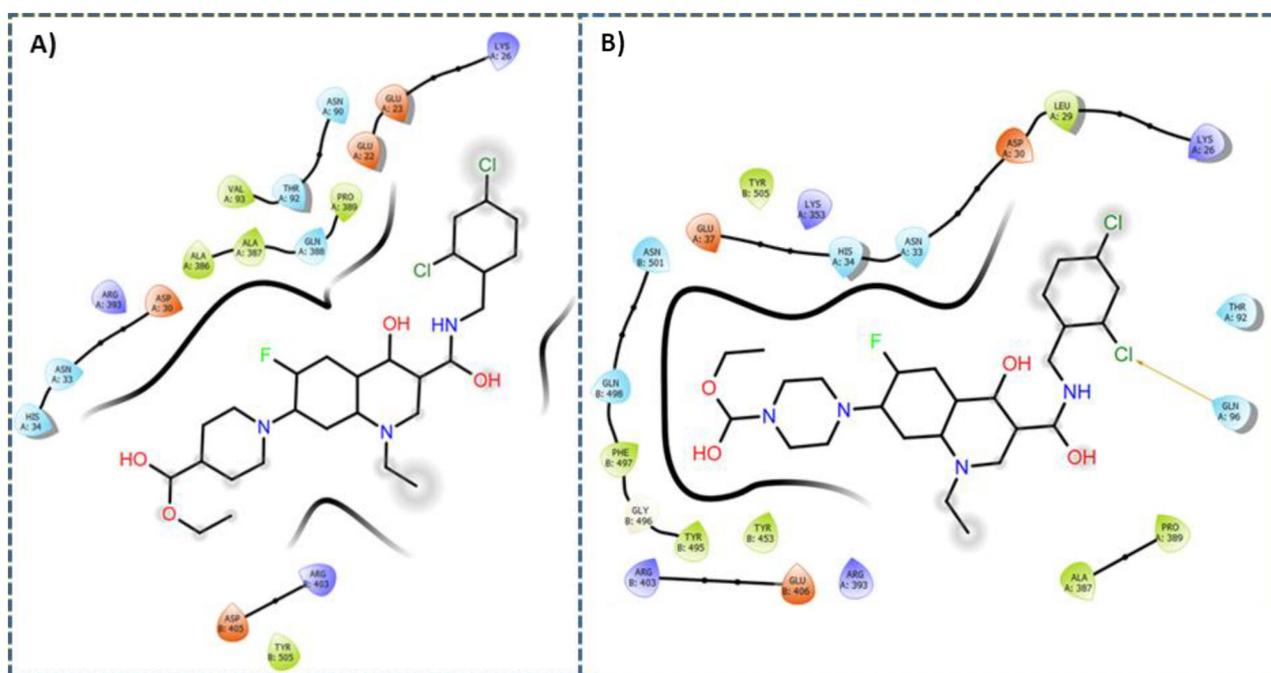


Figure 10. Binding interactions of the S54 and the S55 ligand at the interface of the spike glycoprotein RBD-hACE2 complex after 300 ns MD simulations.

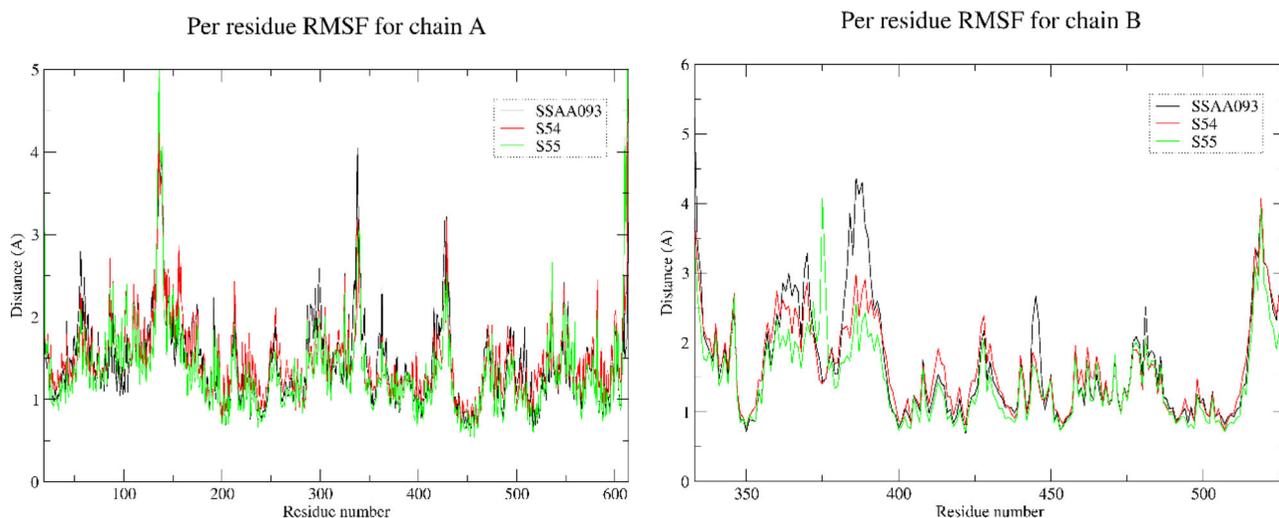


Figure 11. Per residue root-mean-square fluctuation (RMSF) for complexes with the reference SSAA093, the S54 ligand, and the S55 ligand.

that the S55 ligand stabilizes the chain B the most, which is in accordance with the fact that, among the three ligands, it has the strongest interactions with it. However, this part of the chain B is not in vicinity of the binding pocket, so the exact stabilization mechanism remains unknown.

3.6. Pharmacokinetic prediction and metabolic site analysis

All pharmacokinetic properties of S54 and S55 were performed using PreADMET and Molinspiration tool/online server (Jan et al., 2020). The results are in the acceptable range and follow Lipinski's rule of five (supplementary material, Tables 10 and 11). Additionally, the metabolic site analysis of S54 (ZINC000043069833) and S55 (ZINC000095551034) has predicted that it is not an inhibitor of cytochrome enzymes, but is a CYP 3A4 substrate with 4 putative metabolic sites: C17, C15, C12, C27 and C7, C8, C11, respectively which is depicted in supplementary material, Figure 5A and 5B.

The database Smiles Arbitrary Target Specification-based fragment was used by SMARTCyp in combination with an accessibility descriptor to obtain the ranking for the site of metabolism (SOMs) (Hashem & Mahrouse, 2018). Here, the primary site of CYP-mediated metabolism for the S54 ligand was predicted to be C17 and C15 of the amine group and C12 of the aryl fluoride group. Similarly, the S55 ligand belongs to quinoline-3-carboximide class of organic compounds, where the quinoline ring is substituted by one carboxamide group at the 3-position. These compounds were found to inhibit the Nipah virus glycoprotein G/F-mediated cell-cell fusion expressed in African green monkey Vero cells after 24 h relative to untreated control (Niedermeier et al., 2009).

5. Conclusion

In this study, we adopted a systems biology method to construct an extended PPI network of SARS-CoV and associated human proteins. Our findings suggest that the spike glycoprotein has the highest degree and the second-highest BC,

and ACVR1B has the second highest degree and the highest BC. The spike glycoprotein is mainly involved in binding to the human ACE2 receptor, while human ACVR1B is involved in the transmembrane receptor protein serine/threonine kinase signalling pathway. Both proteins are essential in the viral entry and causing infection in humans. Furthermore, studies on the SARS-CoV spike glycoprotein RBD inhibition with the top five ligands were successfully carried out using molecular dynamics approach. Ligands S54 and S55 were found to be selectively interacting with the Tyr41A and Tyr505B hotspots inside the binding pocket via formation of an inclined tape over the binding site with the OH group. These results demonstrate the likelihood of the ethyl 1-{3-[(2,4-dichlorobenzyl) carbamoyl]-1-ethyl-6-fluoro-4-oxo-1,4-dihydro-7-quinolinyl}-4 piperidine carboxylate (the S54 ligand) and ethyl 1-{3-[(2,4-dichlorobenzyl) carbamoyl]-1-ethyl-6-fluoro-4-oxo-1,4-dihydro-7-quinolinyl}-4 piperazine carboxylate (the S55 ligand) activity to block the virus spike glycoprotein RBD from docking to hACE2. The trajectory analysis of the spike RBD-hACE2-S54/S55 complexes also displayed structural stability and lower binding free energy when compared to the complex with the reference SSAA093 ligand. However, these computationally validated results need to be investigated on *in vivo* models before classifying molecules as potential COVID-19 inhibitors.

Acknowledgements

The authors are thankful to Registrar, Nitte (Deemed to be University), Mangalore, India for providing all the facilities to complete this work. The authors also acknowledge the University of Zagreb, University Computing Centre (SRCE) for granting computational time on the ISABELLA cluster.

CRedit authorship contribution statement

P.G., S.B.S., P.K.S. conceived and planned the experiments. P.G., S.B.S. carried out the experiments. H.R. carried out all MD simulations and MM-GBSA calculations. P.G., S.B.S., P. P., S. K., S.K., P.K.S. contributed to the interpretation of the results. P.G., and S.B.S. took the lead in writing the manuscript. All

authors provided critical feedback and helped shape the research, analysis and manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Abdelli, I., Hassani, F., Bekkel Brikci, S., & Ghalem, S. (2020). In silico study the inhibition of angiotensin converting enzyme 2 receptor of COVID-19 by *Ammoides verticillata* components harvested from Western Algeria. *Journal of Biomolecular Structure and Dynamics*, 1–14. <https://doi.org/10.1080/07391102.2020.1763199>
- Adedeji, A. O., Severson, W., Jonsson, C., Singh, K., Weiss, S. R., & Sarafianos, S. G. (2013). Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms. *Journal of Virology*, 87(14), 8017–8028. <https://doi.org/10.1128/JVI.00998-13>
- Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47–97. <https://doi.org/10.1103/RevModPhys.74.47>
- Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews. Genetics*, 5(2), 101–113. <https://doi.org/10.1038/nrg1272>
- Basit, A., Ali, T., & Rehman, S. U. (2020). Truncated human angiotensin converting enzyme 2; a potential inhibitor of SARS-CoV-2 spike glycoprotein and potent COVID-19 therapeutic agent. *Journal of Biomolecular Structure and Dynamics*, 1–10. <https://doi.org/10.1080/07391102.2020.1768150>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. Issue <http://www.rcsb.org/pdb/status.html>. <https://doi.org/10.1093/nar/28.1.235>
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., & Galon, J. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*, 25(8), 1091–1093. <https://doi.org/10.1093/bioinformatics/btp101>
- Case, D. A. (2016). *Amber 2016 Reference Manual*. University of California.
- Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., & Pujadas, G. (2015). Molecular fingerprint similarity search in virtual screening. *Methods (San Diego, Calif.)*, 71, 58–63. <https://doi.org/10.1016/j.ymeth.2014.08.005>
- Cook, H. V., Doncheva, N. T., Szklarczyk, D., von Mering, C., & Jensen, L. J. (2018). Viruses.STRING: A virus-host protein-protein interaction database. *Viruses*, 10(10), 519. <https://doi.org/10.3390/v10100519>
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98(12), 10089–10092. <https://doi.org/10.1063/1.464397>
- Ferency, G. G. (2015). Computation of Drug-Binding Thermodynamics. In G. M. Keserü & D. C. Swinney (Eds.), *Thermodynamics and kinetics of drug binding* (pp. 37–61). Wiley-VCH Verlag GmbH & Co. KGaA.
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5), 449–461. <https://doi.org/10.1517/17460441.2015.1032936>
- Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science: A Publication of the Protein Society*, 27(1), 14–25. <https://doi.org/10.1002/pro.3235>
- Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*, 330(4), 891–913.
- Hashem, H. M., & Mahrouse, M. A. (2018). In vitro metabolism study of a novel P38 kinase inhibitor: In silico predictions, structure elucidation using MS/MS-I. *Future Medicinal Chemistry*, 10(2), 201–220. <https://doi.org/10.4155/fmc-2017-0126>
- Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area method. *Molecular Informatics*, 31(2), 114–122. <https://doi.org/10.1002/minf.201100135>
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 51(1), 69–82. <https://doi.org/10.1021/ci100275a>
- Jeanquartier, F., Jean-Quartier, C., & Holzinger, A. (2015). Integrated web visualizations for protein-protein interaction databases. *BMC Bioinformatics*, 16(1), 195. <https://doi.org/10.1186/s12859-015-0615-z>
- Joshi, R. S., Jagdale, S. S., Bansode, S. B., Shankar, S. S., Tellis, M. B., Pandya, V. K., Chugh, A., Giri, A. P., & Kulkarni, M. J. (2020). Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. *Journal of Biomolecular Structure and Dynamics*, 1–16. <https://doi.org/10.1080/07391102.2020.1760137>
- Kortemme, T., Kim, D. E., & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Sci STKE*, 2004(219), pl2. <https://doi.org/10.1126/stke.2192004pl2>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215–220. <https://doi.org/10.1038/s41586-020-2180-5>
- Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci*, 27(1), 129–134. <https://doi.org/10.1002/pro.3289>
- Liu, C., Ma, Y., Zhao, J., Nussinov, R., Zhang, Y. C., Cheng, F., & Zhang, Z. K. (2020). Computational network biology: Data, models, and applications. In *Physics Reports*, 846, 1–66. Elsevier B.V. <https://doi.org/10.1016/j.physrep.2019.12.004>
- Lupala, C. S., Kumar, V., Li, X., Su, X.-d., and Liu, H. (2020). *Computational analysis on the ACE2-derived peptides for neutralizing the ACE2 binding to the spike protein of SARS-CoV-2*. DOI. <https://doi.org/10.1101/2020.05.03.075473>
- Machado, M. R., & Pantano, S. (2020). Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations. *Journal of Chemical Theory and Computation*, 16(3), 1367–1372. <https://doi.org/10.1021/acs.jctc.9b00953>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Marzi, A., Gramberg, T., Simmons, G., Möller, P., Rennekamp, A. J., Krumbiegel, M., Geier, M., Eisemann, J., Turza, N., Saunier, B., Steinkasserer, A., Becker, S., Bates, P., Hofmann, H., & Pöhlmann, S. (2004). DC-SIGN and DC-SIGNR interact with the glycoprotein of marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *Journal of Virology*, 78(21), 12090–12095. <https://doi.org/10.1128/jvi.78.21.12090-12095.2004>
- Masters, P. S. (2006). The molecular biology of coronaviruses. In *Advances in Virus Research*, (65, 193–292. *Adv Virus Res*, [https://doi.org/10.1016/S0065-3527\(06\)66005-3](https://doi.org/10.1016/S0065-3527(06)66005-3)
- Matsuyama, Y., & Ishida, T. (2018). Stacking multiple molecular fingerprints for improving ligand-based virtual screening. in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* (pp. 279–288). Springer Verlag. https://doi.org/10.1007/978-3-319-95933-7_35
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*, 30(16), 2785–2791. <https://doi.org/10.1002/jcc.21256>
- Nisha Muralidharan, R., Sakthivel, D., Velmurugan, M. M. & Gromiha, (2020). Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. *Journal of Biomolecular Structure and Dynamics*, 1–6. <https://doi.org/10.1080/07391102.2020.1752802>

- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open babel: An Open chemical toolbox. *Journal of Cheminformatics*, 3(10), 33. <https://doi.org/10.1186/1758-2946-3-33>.
- Pawson, T., & Warner, N. (2007). Oncogenic re-wiring of cellular signaling pathways. *Oncogene*, 26(9), 1268–1275. <https://doi.org/10.1038/sj.onc.1210255>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Rastelli, G., Rio, A. D., Degliesposti, G., & Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry*, 31(4), 797–810. <https://doi.org/10.1002/jcc.21372>
- Riniker, S., & Landrum, G. A. (2013). Open-source platform to benchmark fingerprints for ligand-based virtual screening. *Journal of Cheminformatics*, 5(1), 26. <https://doi.org/10.1186/1758-2946-5-26>.
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7), 3084–3095.
- Ryckaert, J. P., Ciccotti, G., & Berendsen, H. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23, 327–341.
- Rydberg, P., Gloriam, D. E., Zaretski, J., Breneman, C., & Olsen, L. (2010). SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Medicinal Chemistry Letters*, 1(3), 96–100. <https://doi.org/10.1021/ml100016x>
- Niedermeier, S., Singethan, K., Rohrer, S. G., Matz, M., Kossner, M., Diederich, S., Maisner, A., Schmitz, J., Hiltensperger, G., Baumann, K., Holzgrabe, U. & Schneider-Schaulies, J. (2009). A small-molecule inhibitor of Nipah virus envelope protein-mediated membrane fusion. *Journal of Medicinal Chemistry*, 52(14), 4257–4265. <https://doi.org/10.1021/jm900411s>
- Schrödinger Release 2020-1: Maestro (2020). Schrödinger, LLC. New York.
- Sterling, T., & Irwin, J. J. (2015). ZINC 15-Ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>
- Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). Biological network exploration with Cytoscape 3. *Current Protocols in Bioinformatics*, 47, 8.13.1–8.13.24.
- Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., Talavera-López, C., Maatz, H., Reichart, D., Sampaziotis, F., Worlock, K. B., Yoshida, M., & Barnes, J. L. HCA Lung Biological Network (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine*, 26(5), 681–687. <https://doi.org/10.1038/s41591-020-0868-6>.
- Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), NA–NA. <https://doi.org/10.1002/jcc.21334>.
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2), 281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
- Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics & Modelling*, 25(2), 247–260. <https://doi.org/10.1016/j.jmkgm.2005.12.005>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)*, 367(6483), 1260–1263. DOI:10.1126/science.abb2507.
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)*, 367(6485), 1444–1448. <https://doi.org/10.1126/science.abb2762>
- Yang, S., Fu, C., Lian, X., Dong, X., & Zhang, Z. (2019). Understanding human-virus protein-protein interactions using a human protein complex-based analysis framework. *MSystems*, 4(2), e00303-18. <https://doi.org/10.1128/msystems.00303-18>.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
- Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C., & Yuen, K. Y. (2016). Coronaviruses - drug discovery and therapeutic options. *Nature Reviews. Drug Discovery*, 15(5), 327–347. <https://doi.org/10.1038/nrd.2015.37>