Leaving no stone unturned: Allosteric targeting of SARS-CoV-2 Spike protein at putative druggable sites disrupts human angiotensin-converting enzyme interactions at the receptor binding domain.

Fisayo A. Olotu, Kehinde F. Omolabi, Mahmoud E.S. Soliman

PII: S2352-9148(20)30601-8
DOI: https://doi.org/10.1016/j.imu.2020.100451
Reference: IMU 100451

To appear in: Informatics in Medicine Unlocked

Received Date: 22 July 2020
Revised Date: 11 October 2020
Accepted Date: 11 October 2020

Please cite this article as: Olotu FA, Omolabi KF, Soliman MES, Leaving no stone unturned: Allosteric targeting of SARS-CoV-2 Spike protein at putative druggable sites disrupts human angiotensin-converting enzyme interactions at the receptor binding domain., Informatics in Medicine Unlocked, https://doi.org/10.1016/j.imu.2020.100451.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 The Author(s). Published by Elsevier Ltd.
Leaving no stone unturned: Allosteric targeting of SARS-CoV-2 Spike protein at putative druggable sites disrupts human angiotensin-converting enzyme interactions at the receptor binding domain.

Fisayo A. Olotu, Kehinde F. Omolabi and Mahmoud E. S Soliman

Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa
*Corresponding Author: Mahmoud E.S. Soliman
Email: soliman@ukzn.ac.za
Website: http://soliman.ukzn.ac.za
Telephone: +27 (0) 31 260 8048
Fax: +27 (0) 31 260 7872
Abstract

The systematic entry of SARS-CoV-2 into host cells, as mediated by its Spike (S) protein, is highly essential for pathogenicity in humans. Hence, targeting the viral entry mechanisms remains a major strategy for COVID-19 treatment. Although recent efforts have focused on the direct inhibition of S-protein receptor-binding domain (RBD) interactions with human angiotensin-converting enzyme 2 (hACE2), allosteric targeting remains an unexplored possibility. Therefore, in this study, for the first time, we employed an integrative meta-analytical approach to investigate the allosteric inhibitory mechanisms of SARS-CoV-2 S-protein and its association with hACE2. Findings revealed two druggable sites (Sites 1 and 2) located at the N-terminal domain (NTD) and S2 regions of the protein. Two high-affinity binders; ZINC3939013 (Fosaprepitant – Site 1) and ZINC27990463 (Lomitapide – Site 2) were discovered via site-directed high-throughput screening against a library of ~1500 FDA approved drugs. Interestingly, we observed that allosteric binding of both compounds perturbed the prefusion S-protein conformations, which in turn, resulted in unprecedented hACE2 displacement from the RBD. Estimated $\Delta G_{\text{bind}}$ for both compounds were highly favorable due to high-affinity interactions at the target sites. In addition, Site 1 residues; R190, H207, K206 and K187, I101, R102, I119, F192, L226, V126 and W104 were identified for their crucial involvement in the binding and stability of ZINC3939013. Likewise, energy contributions of Q957, N953, Q954, L303, Y313, Q314, L858, V952, N953, and A956 corroborated their importance to ZINC27990463 binding at the predicted Site 2. We believe these findings would pave way for the structure-based discovery of allosteric SARS-CoV-2 S-protein inhibitors for COVID-19 treatment.
Keywords: SARS-CoV-2, Spike protein, allosteric targeting, virtual high-throughput screening, receptor binding domain, high-affinity binding.

1. Introduction

The novel coronavirus disease also referred to as COVID-19 is caused by the SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2), with incidences first reported in Wuhan China in December 2019. This disease has, however, persisted till mid-2020, spreading across 212 countries with over 3,513,507 cases reported coupled with increasingly high casualties numbering over 245,544 globally. SARS-CoV-2 belongs to a large group of coronaviruses which are known to cause respiratory infections and related complications. These RNA viruses are spherical, pleomorphic, positive-sensed, single-stranded and polyadenylated. Of all known viruses, coronaviruses (CoVs) have the largest RNA genome, with diverse pathogenic effects in animals and humans. This virus class is divided into four genera namely: alpha-CoV, beta-CoV, gamma-CoV and delta CoV, with the beta-CoV class prominent for their disease-causing effects in humans (HCoVs). Seven HCoVs have been characterized to date; among which four (HCoV-HKU1, HCoV-OC43, HCoV-NL63 and HCoV-229E) cause very mild respiratory symptoms. On the other hand, MERS-CoV, SARS-CoV, and SARS-CoV-2 cause severe respiratory and gastrointestinal infections which, in most cases, can be fatal. Although SARS-CoV-related infections were zoonotically transmitted into human populations, human to human transmissions has further contributed towards viral super-spread via respiratory aerosols.

The entry of SARS-CoV-2 coupled with its replication process in target human cells is achieved by the functionalities of a cohort of components, majorly non-structural and structural proteins, that make up the virus. Generally, about 16 non-structural proteins (NSPs) mediate diverse pro-
pathogenic functions such as replication, processing and proof-reading of genomic frames, host immune evasion among many others, as previously reported. More so, CoVs comprises of four major structural proteins that are integral to their pathogenesis. These are the nucleocapsid (N), envelope (E), membrane (M) and spike (S) proteins. The N protein makes up the nucleocapsid and other viral genome-related processes while the M protein is the most abundant of the four, playing major roles in maintaining viral structural integrity as well as coordinating other structural proteins. E protein, on the other hand, is crucial to the maturation of the virus while the trimeric S protein mediates viral entry into the host cell via the endosomal or non-endosomal route.

Two domains make up the S protein namely the N-terminal S1 domain and the C-terminal S2-membrane-anchored domain. The S2 region is extensively conserved in CoVs while constituent S1 region residues are highly diverge across the CoV strains. These domains have been further characterized into subdomains due to specific functionalities with respect to host receptor recognition and binding (S1), coupled with membrane fusion and entry (S2) (Figure 1). Similar to SARS-CoV architecture, some recent reports have sub-categorized the SARS-CoV-2 S1 ectodomain into the N-terminal domain (NTD), a conserved receptor-binding domain (RBD) which recognizes the human angiotensin-converting enzyme 2 (hACE2), and subdomains 1 and 2 (SD1 and SD2). During infection, proteolytic cleavage or priming of the S protein is crucial for viral fusion and entry into host cells, a process mediated by host cell proteases such as the transmembrane serine protease 2 (TMPRSS2) and Cathepsin L, at the S1/S2 (boundary between S1 and S2 subunits) and S2’ (immediately upstream S2 fusion peptide - FP) cleavage sites. The S protein primarily exists in a metastable prefusion complex prior to cleavage, after which notable conformational arrangements occur in order to fuse the viral membrane into
the host cell. In addition, the RBD adopts disparate conformational motions to engage the host cell receptor.

Figure 1.

These conformational states are transient, designated as up (open) and down (close) conformations. The up conformation corresponds to the hACE2 accessible state while the down state cannot engage the host cell receptor. The S2 domain, on the other hand, consists of the functionally important fusion peptide (FP), which is critical for viral fusion and formation of the post-fusion complex; heptad repeats 1 and 2 (HR1 and HR2); transmembrane domain (TM) and cytoplasmic tail (CT). The HRs of the S-protein trimer interact to form a fusion core of six-helical bundle which helps bring the membranes of the virus and host cell in close proximity for fusion and entry. Therefore, the roles of SARS-CoV-2 S-protein present it as an important therapeutic target, which would enable the prevention of viral entry and fusion in host cells.
Numerous studies have been reported over the past months with regards to the possibility of blocking direct interactions between SARS-CoV-2 S-protein and hACE2. Most of these studies were aimed at targeting the S protein RBD domain with antibodies, peptide-based or small molecule compounds that binds with a much higher affinity to block S-protein-hACE2 interactions. Also, targeting host proteases such as TMPRSS2 was explored in a recent study, with consequential impediments on SARS-CoV-2 entry.

Identification of other functional (allosteric) sites on the prefusion S protein could present another dynamic and effective approach of preventing SARS-CoV-2 infectivity relative to its interaction with the host cell ACE2 and proteases. This alternative target approach for SARS-CoV-2 S protein is important because its RBD (similar to other CoVs) has been associated with a high mutational propensity which may in turn alter the affinity of small molecule inhibitors or peptide designed to bind therein. Allosteric targeting was explored in a recent study wherein the CoV-conserved S2 HR1 region was identified as an important target site for the development of broad-spectrum inhibitors of human CoVs. The resulting peptide inhibitor (EK1) was evaluated in vivo and exhibited desirable safety and efficacy. More so, the Protein Contact Network (PCN) paradigm was used to map functional allosteric loci on SARS-CoV S protein.

Relatively, this study was implemented to (i) identify potential druggable sites across the S1 and S2 domains of the SARS-CoV-2 S protein other than the RBD-hACE2 interface (ii) perform high-throughput (virtual) screening of ~1500 FDA approved drugs against the most druggable site(s) (iii) investigate the binding dynamics and interaction mechanisms of the compounds and their consequential effects on the S-protein RBD-ACE2 complex.

We believe this systematic study will be able to provide structural and molecular insights into possible allosteric sites on SARS-CoV-2 S protein suitable for selective targeting and structure-
based design. This would open up avenues for the development of allosteric inhibitors with the potential of disrupting hACE2 interactions at the S-protein RBD.

2. **Computational methodologies**

2.1 Protein retrieval and preparation

The three-dimensional structure of SARS-CoV-2 S-protein (prefusion) was retrieved from PDB with entry 6VSB. This, as previously reported, represents the S-protein RBD conformation in its up (open) state, which is most suitable for hACE2 binding. Also, to model binding interactions between the prefusion SARS-CoV-2 S-protein (S1/S2) and the hACE2, a crystalized structure with PDB entry 6M0J was separately retrieved. This complex depicts binding between the RBD domain (truncated) of SARS-CoV-2 S-protein and the protease domain (PD) of hACE2. Co-crystallized molecules not relevant to this study were removed while missing residues (gaps) in the structures were filled using the MODELLER algorithm. This preparation was performed on the UCSF Chimera Graphic User Interface (GUI). Subsequently, using the structural superposition method, we were able to model a complex between prefusion S-protein (S1/S2) monomer (RBD - up conformation) and the hACE2 protein (Figure 2).
2.2 Predictive-mapping, cross-validation and characterization of allosteric sites

Possible druggable sites other than the SARS-CoV-2 RBD interface were predicted using approaches previously reported.\textsuperscript{57–60} Herein, we employed multiple tools for site identification and validation, which include SiteMap\textsuperscript{61}, Fpocket\textsuperscript{62}, Discovery studio 2016 Client\textsuperscript{63} and Prankweb.\textsuperscript{64} SiteMap is an exhaustive tool which ranks protein pockets based on properties such as druggability, surface exposure, hydrophobicity and hydrophilicity among others\textsuperscript{65–67}. These details were then used to characterize the predicted pockets after which other predictive algorithms were used complementarily for cross-validation. Two highly ranked sites were then selected for further analyses.

2.3 High-throughput screening against predicted SARS-CoV-2 S-protein allosteric sites

Furthering on the rationale of the study, we mapped out the two most druggable sites on the target protein and virtually screened against them a large chemical library of FDA approved drugs (\~1500 compounds) derived from the ZINC repository (http://zinc.docking.org/substances/subsets/fda/). This screening was performed using high-performance computing-integrated Autodock vina\textsuperscript{68} prior to which coordinates of the predicted sites were mapped using gridboxes. Corresponding binding scores were retrieved from the resulting .pdbqt files and were used to filter down to the topmost 20 compounds for each predicted Sites 1 and 2. Subsequently, two compounds with the highest binding scores (most negative) were selected for the two predicted sites yielding complexes that were subjected to further simulation studies. As explained in 2.1, the prefusion S-proteins (ligand-bound and unbound) were superimposed with the RBD-hACE2 complex (6M0J) after which the single
truncated RBD was removed. By so doing, we obtained models of allosterically-bound and unbound pre-fusion S-protein-ACE2 complex. This, as aimed in this study, would provide structural and dynamical insights into the mechanistic effects of allosteric targeting on SARS-CoV-2 host entry machinery.

2.4 Molecular dynamics (MD) simulations

Although computationally expensive (1673 residues), we proceeded with long-timescale MD simulation runs for the systems on AMBER18 Graphical Processing Unit (GPU) using its embedded modules. Protein parameters were defined using the FF14SB forcefield while ligand parameters were generated with the antechamber and parmchk modules. Likewise, the LEAP program was used to define coordinate and topology files for the ligand-bound and unbound protein complexes. This program, also, was used to neutralize (addition of counter-ions; Na$^+$ and Cl$^-$) and solvate the systems in a TIP3P water box of size 10Å. Structural minimization was first carried out partially for 5000 steps with a restraint potential of 500kcal mol$^{-1}$. Å$^2$ followed by another 100000 steps of full minimization with no restraints. A canonical (NVT) ensemble with a 5kcal mol$^{-1}$ Å$^2$ harmonic restraints was used to heat the systems gradually from 0 – 300k for 50ps, after which the systems were equilibrated for 10000ps at a constant 300k temperature without restraints in an NPT ensemble. Atmospheric pressure was maintained at 1bar with a Berendsen barostat while each protein system was subjected to a production run of 350ns. Studied systems include ZINC3939013-S-protein-hACE2 (allosteric Site 1), ZINC27990463-S-protein-hACE2 (allosteric Site 2), and unbound S-protein-hACE2. Corresponding trajectories were saved at every 1ps time-frame until the end of the simulation followed by data plot analyses using Microcal Origin software. Snapshots were also taken and analyzed to monitor structural events and ligand interaction dynamics across the trajectories on the UCSF Chimera user
interface (GUI) and Discovery Studio Client.

2.5 Calculations of binding free energies and per-residue decomposition

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method was used to evaluate binding affinities of the predicted allosteric S-protein binders at their target sites. Binding energy profiles for both compounds, inclusive of their energy components, were estimated using 1000 snapshots from the terminal 30ns of MD trajectories where conformational stabilities were visible. This approach was important in order to minimize the effects of conformational disorder or entropy on ligand interactions. The equations below mathematically express binding energy calculations:

\[
\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{inhibitor}}) \quad (1)
\]

\[
\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S = \Delta H - T\Delta S \quad (2)
\]

\[
\Delta G_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdW}} \quad (3)
\]

\[
\Delta G_{\text{sol}} = \Delta G_{\text{ele,sol}(GB)} - \Delta G_{\text{np,sol}} \quad (4)
\]

\[
\Delta G_{\text{np,sol}} = \gamma S\text{ASA} + \beta \quad (5)
\]

As shown, internal (\(\Delta E_{\text{int}}\)), electrostatic (\(\Delta E_{\text{ele}}\)) and van der Waals (\(\Delta E_{\text{vdW}}\)) energies sum up the gas-phase energy (\(\Delta G_{\text{gas}}\)) while the solvation free energy (\(\Delta G_{\text{sol}}\)) is defined by the polar solvation (\(\Delta G_{\text{ele,sol}}\)) and non-polar contribution to solvation (\(\Delta G_{\text{np,sol}}\)) terms. The MM/GBSA method was used to estimate the Generalized Born (GB) for \(\Delta G_{\text{ele,sol}}\) while the linear relationship between the surface tension proportionality constant (\(\gamma = 0.0072 \text{ mol}^{-1} \text{ Å}^{-2}\)), solvent accessible surface area (SASA, Å²), and \(\beta\) constant was used to solve \(\Delta G_{\text{np,sol}}\). Furthermore, estimated \(\Delta G_{\text{bind}}\) was decomposed into individual residue energies, most especially those that constitute the predicted allosteric pockets where the ligands were bound. This method was essential to identify specific residues that contribute crucially to the stability and inhibitory activities of potential allosteric inhibitors.

3. Results and Discussion
3.1 Potential allosteric sites on the prefusion SARS-CoV-2 S-protein – Identification, Cross-Validation and Characterization

Based on the study rationale, we set out to identify possible sites for drugging the target protein other than the RDB domain where interaction with hACE2 reportedly occurs. Prediction and cross-validation were done using aforementioned tools with results showing corresponding residues (domain). This represents a combinatorial method for identifying, validating and cross-validating possible allosteric sites that are druggable in SARS-CoV-2 S-protein. From the multiple predictions, **Site 1** entails residues that make up the NTD (7-307) of SARS-CoV-2 S-protein while **Site 2** residues overlap across the FP (816-831), HR1 (898-985) and CR (838-897) (Table 1). The architectures of these pockets are shown in Figure 3. Furthermore, defining the druggability of a site on target proteins depends on the size (volume) and hydrophobicity (with minimal hydrophilicity) while, on the other hand, high hydrophilicity, reduced hydrophobicity, small pocket size and shallowness characterize “difficult-to-drug” and undruggable pockets. While large hydrophilicity could have repulsive effects on ligand mobility at the binding site, a small or shallow cavity would impede ligand access, fitness, optimal binding and stability.
Table 1: Identification and cross validation of potential allosteric sites on SARS-CoV-2 prefusion S-protein using multiple predictive algorithms.

<table>
<thead>
<tr>
<th>Predicted sites</th>
<th>Binding site prediction and cross-validation</th>
<th>Corresponding S-protein domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiteMap</td>
<td>Fpocket</td>
<td>DS 2016 Client</td>
</tr>
<tr>
<td>Site 6</td>
<td>1068, 1069, 1077, 1078, 1079, 1080, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1105, 1106, 1107</td>
<td>1010, 1012, 1015, 1019, 1108, 1119, 1208, 1210, 1211, 1212, 1214, 1215, 1218, 909, 910, 911, 1089, 1090, 1091, 1092, 1093, 1094, 1104, 1105, 1106, 1107, 1108, 1109, 1207-1216</td>
</tr>
</tbody>
</table>

NTD, N-terminal domain; FP, fusion peptide; HR1, heptad repeat 1; CD, connector domain; RBD, receptor binding domain; SD2, subdomain 2; SD1, subdomain 1; CH, central helix; βH – β hairpin; TM, transmembrane region.
More so, a >0.83 (Halgren Dscore) threshold applies for druggable sites \(^6\) while a 0.8 cut-off point was set for the SiteScore, which, altogether are parameters used to distinguish between sites that are druggable and non-druggable.

From our results, four prediction algorithms employed herein commonly identified and ranked specific regions on the target protein, indicating their possibilities as allosteric sites. As earlier stated, constituent residues of the top two ranked sites correspond to the fusion peptide (FP), Heptad repeat 1 (HR1), connecting region/domain (CR/CD) and N-terminal domain (NTD).

From Table 2, Sites 1 → 3 ranks above the 0.83 Halgren Dscore threshold making them suitable for therapeutic targeting. Relatively, Site 1 appears to be highly surface-exposed with a score of 0.933 while a large pocket size and volume for Site 2 could favor the use of large-molecule compounds.

**Table 2:** Characterization of the predicted allosteric sites on SARS-CoV-2 prefusion S-protein based on site attributes

<table>
<thead>
<tr>
<th>Predicted sites</th>
<th>Druggability score (Dscore)</th>
<th>SiteScore</th>
<th>Surface-exposure</th>
<th>Pocket Size</th>
<th>Pocket Volume (Å³)</th>
<th>Hydrophobicity</th>
<th>Hydrophilicity</th>
<th>Hydrogen donor/acceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.893</td>
<td>0.935</td>
<td>0.933</td>
<td>482</td>
<td>245.82</td>
<td>0.168</td>
<td>0.336</td>
<td>0.875</td>
</tr>
<tr>
<td>Site 2</td>
<td>1.046</td>
<td>0.899</td>
<td>0.794</td>
<td>503</td>
<td>294.31</td>
<td>0.243</td>
<td>0.203</td>
<td>1.922</td>
</tr>
<tr>
<td>Site 3</td>
<td>0.842</td>
<td>0.823</td>
<td>0.862</td>
<td>319</td>
<td>208.6</td>
<td>0.142</td>
<td>0.822</td>
<td>1.886</td>
</tr>
<tr>
<td>Site 4</td>
<td>0.803</td>
<td>0.817</td>
<td>0.852</td>
<td>263</td>
<td>214.30</td>
<td>0.111</td>
<td>0.613</td>
<td>2.539</td>
</tr>
<tr>
<td>Site 5</td>
<td>0.738</td>
<td>0.801</td>
<td>0.800</td>
<td>218</td>
<td>148.8</td>
<td>0.353</td>
<td>0.705</td>
<td>1.294</td>
</tr>
<tr>
<td>Site 6</td>
<td>0.711</td>
<td>0.786</td>
<td>0.696</td>
<td>267</td>
<td>237.64</td>
<td>0.392</td>
<td>0.884</td>
<td>0.634</td>
</tr>
</tbody>
</table>

Taken together, high surface-exposure coupled with relatively large volumes, hydrophobicity and favorable donor/acceptor properties for Sites 1 and 2 could account for their suitability as targetable allosteric regions on the S-protein other than the RBD (Figure 3). These presumptions are also reflected by the estimated Dscore and SiteScore values. In addition, since these predicted sites are highly functional, particularly the overlapping FP, HR1 and CR, targeting them could
disrupt structural events associated with SARS-CoV-2 fusion into the host-cell as explicitly reported in recent literatures.\textsuperscript{39,41,44,54,72}

3.2 High-throughput screening and identification of potential allosteric binders to the predicted Sites 1 and 2

High-throughput screening using a library of \textasciitilde1500 FDA approved drug compounds (http://zinc.docking.org/substances/subsets/fda/) were performed against the two predicted allosteric sites.

![Autodock vina grid boxes were centered on coordinates (center, size, x = 107.1, 14.45; y = 99.65,](image)

**Figure 3.**

Autodock vina grid boxes were centered on coordinates (center, size, x = 107.1, 14.45; y = 99.65,
16.33, z = 114.74, 15.82) for Site 1 and (center, size, x = 129.2, 17.84; y = 85.23, 24.15, z = 139.11, 18.47) for Site 2.

Figure 4.

Results for the top 20 compounds with the highest binding scores are presented in Supplementary Table S1 and Supplementary Table S2 for Sites 1 and 2 respectively.

From the screening results, overall highest scores were estimated for ZINC3939013 (-10
kcal/mol) at Site 1 and ZINC27990463 (-9.3 kcal/mol) at Site 2.

### 3.3 Allosteric targeting disrupts S-protein conformation and hACE2 association at the RBD

As highlighted in our methods, MD simulations were performed for the prefusion S-protein-hACE2 complexes bound distinctly at two potential allosteric sites.

This approach was essential to investigate the likely effects of allosteric targeting on the entry/fusion mechanisms of SARS-CoV-2 via host hACE2.

**Figure 5.**

Firstly, we monitored conformational events associated with the allosteric binding of Compds A and B at the predicted Sites 1 and 2 of SARS-CoV-2 S-protein. This was done by obtaining snapshots from the MD trajectories from selected time-frames which were then visually compared with their starting structures.

Our findings reveal that the unbound S-hACE2 complex remained intact over the simulation time while the hACE2, after about 50ns, was gradually displaced from the RBD interfaces in S-
proteins bound by ZINC3939013 (Fosaprepitant) at **Site 1** and ZINC27990463 (Lomitapide) at **Site 2** (Figure 4 and 5). These variations, as induced by the allosteric binders, proceeded more prominently until the end of the simulation. As shown, RBD for the unbound S-protein exhibited relative ‘elongated’ or ‘out-stretched’ conformations which most likely favored hACE2 binding. However, this conformation appeared distorted the allosterically-bound S-proteins and could account for displacement motions of the interacting hACE2 from the RBD interface. Therefore, the allosteric-mediated disruption of SARS-CoV-2 S-protein RBD and its interaction with hACE2, as reported herein, is a major finding that could indicate the viability of allosteric targeting in SARS-CoV-2 therapy.

Furthermore, we measured structural stabilities across the ligand-protein complexes relative to the unbound system using the RMSD metrics. As shown in Figure 6, structural instability was highest in the unbound S-protein while its associated hACE2 was relatively stable compared to the allosterically-bound S-proteins which exhibited reductions in \( \text{Ca} \) deviations across the simulation time (RMSD <2Å). Interestingly, hACE2s for the bound proteins showed higher structural motions and could correlate with their displacement motions away from the RBD as earlier revealed. Comparatively, allosteric ligand binding at both sites distinctly lowered RMSD in the S-protein and induced high RMSDs in the hACE2.
This could indicate the structural effects of allosteric targeting on the S-protein and its interaction with hACE2. Estimated mean RMSDs, as presented in Table 2, corroborates conformational variations among the unbound and bound protein complexes.

To minimize the effects of structural disorderliness (entropy) in our calculations, we selected, from the MD trajectories, terminal time-frames (270-300ns) from which the systems appeared to relatively stabilize. These were defined as the finally equilibrated (FE) time-frames and were used for subsequent structural analyses (Table 2).
Table 2: Estimations of relative conformational variations across the studied systems

<table>
<thead>
<tr>
<th>Structural analysis (Å)</th>
<th>Unbound Spike (hACE2)</th>
<th>ZINC393013 Spike (hACE2)</th>
<th>ZINC27990463 Spike (hACE2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole RMSD</td>
<td>17.09±1.1 (1.77±0.23)</td>
<td>13.31±0.9 (1.93±0.32)</td>
<td>11.00±2.0 (2.2±0.4)</td>
</tr>
<tr>
<td>FE-RMSD</td>
<td>11.05±0.3 (1.66±0.26)</td>
<td>9.57±0.4 (1.87±0.31)</td>
<td>7.61±1.4 (2.6±0.1)</td>
</tr>
<tr>
<td>FE-RMSF</td>
<td>5.40±1.9 (0.73±0.32)</td>
<td>4.87±1.7 (1.39±0.45)</td>
<td>3.79±1.2 (3.0±0.82)</td>
</tr>
<tr>
<td>FE-RoG</td>
<td>41.08±0.41 (24.49±0.1)</td>
<td>38.4±0.36 (25.01±0.16)</td>
<td>36.9±0.61 (26.0±0.22)</td>
</tr>
</tbody>
</table>

Receptor binding domain (RBD)

| FE-RMSD | 4.1±0.1 | 5.6±0.3 | 6.1±0.2 |
| FE-RoG  | 15.3±0.2 | 16.1±0.3 | 16.2±0.2 |

From the resulting FE-RMSD plots, unbound S-protein was highly unstable while its associated hACE2 exhibited low structural motion in line with the RMSD calculations, which could also imply that the binding of S-protein stabilized hACE2. In contrast, the allosterically-bound S-proteins (Sites 1 and 2) were notably stable while their corresponding hACE2 showed high structural instability that could correlate with their systemic motions at the S-protein RBD as earlier mentioned.

Furthermore, we measured the Cα mobilities of the proteins in their bound and unbound forms by calculating their relative FE-RoGs. From the plots in Supplementary Figure S1, we observed that atomistic mobilities were distinctly high in the unbound S-protein while its corresponding hACE2 maintained stable motions which on the contrary was prominently high in the allosterically-bound systems. This also agrees with the FE-RMSD analysis which, altogether indicate that high Cα mobility in hACE2s associated with the allosterically-bound S-protein could be as a result of their motions at the S-protein RBD.
We then investigated Cα motions at the S-protein RBDs since this region is reportedly critical for hACE2 binding \(^{54,72}\). FE-RMSD calculation revealed distinct stability (RMSD < 2Å) at the RBD of unbound S-protein while high structural motions or instability characterized RBDs of the allosterically-bound S-proteins. This corroborated high distortions earlier reported at the RBDs of Compd A and B-bound S-proteins.

![Graphs and diagrams showing FE-RMSD and FE-RoG over time](image)

Figure 7.

Presumably, allosteric targeting or inhibition of SARS-CoV-2 S-protein perturbs the RBD
conformation essential for interactions with the host ACE2, leading to eventual dissociation. More so, we determined the degree of Ca mobility at the respective RBDs using the FE-RoG metrics. From the resulting plot, the unbound S-protein RBD exhibited low Ca motions, which were increased in the allosterically-bound S-proteins (Figure 7). This observation agrees with the estimated FE-RMSD which could altogether indicate that allosteric binding of the compounds induced conformational variations that may have altered hACE2 binding and stability at the RBD.

Figure 8.
Systemic fluctuations of constituent residues across the protein units were then monitored by calculating the FE-RMSF values using the stable time-frames (Figure 8). Results showed that the S-protein was, on the average, less flexible in the bound systems relative to unbound while high conformational flexibility characterized the corresponding hACE2 proteins in the bound systems. In line with earlier results, high structural fluctuations manifested by hACE2 in the bound S-proteins could indicate the degree of dynamical variations relative to their starting positions at the RBD. We also mapped out equilibrated fluctuations of the receptor-binding domain (RBD), a crucial component of the S-protein that interacts directly with the hACE2. Likewise, residual
fluctuations across the CTD2 and CTD3 regions were investigated in addition to the S2 subunit (Table 3). As shown in Supplementary Figure S2, the RBD region of unbound S-protein exhibited the least fluctuation while allosteric ZINC27990463 binding induced the highest fluctuation at this region. The posterior S2 domain of the S-protein was highly fluctuated in the unbound form while allosteric binding lowered residual fluctuation in this region, particularly by ZINC27990463 at Site 2. Residual fluctuations were also lowered at the SD1 and SD2 domains of ZINC3939013 and ZINC27990463-bound S-proteins which further indicates the allosteric inhibitory effects were extended into crucial parts of the protein. This could present allosteric SARS-CoV-2 S-protein targeting as a viable strategy for disrupting viral entry mechanisms into human hosts.

Table 3: FE-RMSF calculations of the structural components of SARS-CoV-2 S-protein in their unbound and allosterically-bound forms

<table>
<thead>
<tr>
<th>Structural domains</th>
<th>Unbound Spike</th>
<th>ZINC3939013-Spike</th>
<th>ZINC27990463-Spike</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTD</td>
<td>1.35±1.25</td>
<td>1.64±0.52</td>
<td>1.26±0.6</td>
</tr>
<tr>
<td>RBD</td>
<td>1.75±0.7</td>
<td>1.9±1.24</td>
<td>2.14±0.44</td>
</tr>
<tr>
<td>SD1</td>
<td>1.67±0.39</td>
<td>1.53±0.44</td>
<td>1.35±0.3</td>
</tr>
<tr>
<td>SD2</td>
<td>1.85±0.49</td>
<td>1.72±0.67</td>
<td>1.19±0.31</td>
</tr>
<tr>
<td>S2</td>
<td>1.39±0.89</td>
<td>1.30±0.66</td>
<td>1.15±0.43</td>
</tr>
</tbody>
</table>

Also from the table above, we could deduce possible allostery communication mechanisms wherein ligand binding at Site 1 altered the corresponding Site 2 region (FP, HRI and CR) which altogether constitute the S2. This could provide an avenue to identify and characterize allosteric communication networks within the target protein.

Accordingly, while structural fluctuation was slightly difference among the NTD and S2 domain of unbound S-protein (~0.04Å),
systematic fluctuation increased by 0.34Å and 0.11Å respectively in the Site 1 and Site 2-bound S-protein. In other words, ligand binding at the NTD reduced structural perturbations at the S2 which opposes structural occurrence in the unbound protein.

3.4. Interaction dynamics of Compds A and B at the predicted SARS-CoV-2 S-protein allosteric sites

Structural analyses of ligand orientations at the respective allosteric sites of SARS-CoV-2 S-protein were performed using averaged structures from the MD trajectories (Figure 9). Findings reveal that the allosteric binding of **ZINC3939013** (Fosaprepitant) was stabilized at the NTD. Fosaprepitant contains a terminal triphosphate group that orients towards residues such as N99, K187, N188, R190 and H207. Likewise, its trifluoromethyl group oriented towards D102 while constituent -O and -NH groups mediate interactions with Q173 and N121, among others. These altogether could facilitate high-affinity interactions accountable for its stability and allosteric inhibitory effects against the SARS-CoV-2 and associated hACE2.
On the other hand, systematic orientations of ZINC27990463 show that it transverses the predicted Site 2 and adjacent NTD-RBD linker (L) at the initial time-frames. Chemically, ZINC27990463 (Lomitapide) consists of two trifluoromethyl groups at its two terminal ends, and as observed, while a terminal trifluoromethyl moiety (together with amino and hydroxyl substituents) was oriented into the hydrophobic pocket formed by T732, T734, D830, F833, Y837, Q853, K854, N856, L858, T859, V860, the second trifluoromethyl moiety was extended towards T302, L303, K310, I312 and Y313 of the NTD-RBD linker subdomain.

Also, HR1 residues (N955, A956, L959, N960, L962, V963 and K964) were involved in ligand binding. This trans-domain binding activity of ZINC27990463 could account for its prominent disruptive activities on the S-protein structure, particularly at the RBD.
3.5 Estimations of binding free energies and per-residue contributions

Binding affinities of the compounds were determined using the MM/PBSA technique, which also allowed us to measure the energy contributions of interactive residues at the predicted allosteric sites. Energy calculations, as presented in Table 4 were performed using relatively stable time-frames (270-300ns) to minimize entropic effects that may interfere with ligand binding activities.

Table 4: Binding free energy profiles of the ligand-protein complexes

<table>
<thead>
<tr>
<th>Energy Components</th>
<th>ZINC3939013-Site 1</th>
<th>ZINC27990463-Site 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(kcal mol(^{-1}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta E_{vdW})</td>
<td>-49.93 ± 0.21</td>
<td>-43.19 ± 0.15</td>
</tr>
<tr>
<td>(\Delta E_{ele})</td>
<td>-213.0 ± 1.54</td>
<td>34.68 ± 0.54</td>
</tr>
<tr>
<td>(\Delta G_{gas})</td>
<td>-262.9 ± 1.49</td>
<td>-8.51 ± 0.56</td>
</tr>
<tr>
<td>(\Delta G_{ele,sol(GB)})</td>
<td>214.2 ± 1.38</td>
<td>-21.65 ± 0.51</td>
</tr>
<tr>
<td>(\Delta G_{ap,sol})</td>
<td>-7.5 ± 0.01</td>
<td>-6.4 ± 0.02</td>
</tr>
<tr>
<td>(\Delta G_{sol})</td>
<td>206.7 ± 1.38</td>
<td>-28.05 ± 0.51</td>
</tr>
<tr>
<td>(\Delta H)</td>
<td>-56.2 ± 0.19</td>
<td>-36.56 ± 0.26</td>
</tr>
<tr>
<td>(T\Delta S)</td>
<td>-3.5 ± 0.02</td>
<td>3.99 ± 0.1</td>
</tr>
<tr>
<td>(\Delta G_{bind})</td>
<td>-52.7 ± 0.26</td>
<td>-32.57 ± 0.17</td>
</tr>
</tbody>
</table>

Findings revealed that both targets sites are highly favorable for ligand binding as both compounds exhibited favorable \(\Delta G_{binds}\) of \(-52.7\) kcal/mol (ZINC3939013 at Site 1) and \(-32.57\) kcal/mol (ZINC27990463 at Site 2). These estimations further suggest both predicted sites as high-affinity regions for achieving the allosteric inhibition of SARS-CoV-2 S-protein. Also from the estimations, we could deduce that entropical effects on the \(\Delta G_{binds}\) were minimal.
In addition, we observed that electrostatic effects contributed most notably to the allosteric binding of ZINC3939013 at the NTD region while van der Waals contributions had the highest effect on the binding of ZINC27990463 at the predicted Site 2 pocket. Electrostatic contributions at Site 1 could be due to the high number of electropositive residues that constitute the pocket, as shown in Figure 10, which may form high-affinity interactions with electronegative moieties of the compound. Calculations further revealed that $\Delta E_{vdW}$ and $\Delta E_{ele}$ were more favorable in the gas phase for ZINC3939013 while polar solvation energies were more favorable for ZINC27990463.
at the S2 region of the S-protein. This could imply that while the former was buried in the deep hydrophobic pocket of the NTD, the latter was surface exposed due to its trans-domain binding activity as earlier reported. To understand the mechanistic binding of the compounds at both predicted sites, we decomposed the binding free energies into individual contributions of the interacting residues. These were juxtaposed with structural analysis that showed the type and nature of interactions involved. From the per-residue plot, we observed high $\Delta E_{ele}$ contributions by R190, H207, K206 and K187 with values of -15.5 kcal/mol, -11.90 kcal/mol, -8.35 kcal/mol and -10.55 kcal/mol indicative of their crucial roles in stabilizing ZINC3939013 at the predicted Site 1. R190 and H207 had notably high total energy contributions of -14.42 kcal/mol and -5.3 kcal/mol respectively while vdW contributions were also high in H207, F175 and F192 with values of -3.5 kcal/mol, -2.58 kcal/mol and -2.06 kcal/mol. Unfavorable $\Delta E_{ele}$ interactions, which could be due to steric clashes, were mediated by L176, D178 and E191 but were however compensated by the high negative energies. Coherent with our calculations, 2D analysis of the ligand-residue interactions revealed that K187, R190, K206, H207 formed high-affinity salt bridges, ionic and hydrogen interactions with the phosphoryl group of ZINC3939013. The salt bridge interaction was unique to R190 which could account for its high $\Delta E_{tot}$ contribution as estimated. Also, H-F bonds mediated by I101, R102, I119 coupled with aromatic ($\pi$) interactions by F192, L226, V126, and W104 could further contribute to the stability of the inhibitor at the predicted Site 1. It is also important to mention that these bonds have short distances ranging from 1.2Å - 6Å indicative of their strengths and stabilities. At Site 2, $\Delta E_{ele}$ contribution was highest in Q314 (-11.9 kcal/mol) followed by Q957, N953, and Q954 which had estimated energy values of -2.83 kcal/mol, -2.73 kcal/mol and -1.00 kcal/mol respectively. These energies were however insufficient to compensate unfavorable $\Delta E_{ele}$ contributions by K304 (3.24 kcal/mol).
kcal/mol), K733 (2.54 kcal/mol), K854 (3.48 kcal/mol) and N955 (1.3 kcal/mol). However, vdW contributions by L303 (-1.28 kcal/mol), Y313 (-1.1 kcal/mol), Q314 (-1.37 kcal/mol), L858 (-1.47 kcal/mol), V952 (-1.46 kcal/mol), N953 (-2.0 kcal/mol) and A956 (-1.96 kcal/mol) complemented ligand binding and stability. Overall, total energies above the -1 kcal/mol threshold were contributed by A956 (-2.48 kcal/mol), L858 (-1.65 kcal/mol), L303 (-1.24 kcal/mol) and V952 (-1.16 kcal/mol) indicative of their importance to ligand binding at Site 2. Structural-interaction analysis revealed that halogen (H-F) bond predominated ZINC27990463 binding as mediated by Q314, A956 and N953. In addition, Q957 and N953 engaged two F atoms via hydrogen bonds while L303, L858 and A956 interacted with the constituent rings via aromatic (π-alkyl) interactions. More so, π-π stacked interaction between Y313 and a benzene ring (of the 4-tri-fluoromethyl-1,1’-biphenyl group) could be highly crucial for the stability of the compound.

Taken together, electrostatic energies favored the binding of ZINC3939013 at Site 1 while vdW energies favored ZINC27990463 binding at Site 2, which consequentially, were able to perturb the S-protein RBD and allosterically disrupt hACE2 interactions.

**Conclusion**

The systemic entry of SARS-CoV-2 into the human host cell is a crucial process that underlies its virulence and pathogenicity in humans and other animals it infects. This mechanism is mediated by its interaction with the host ACE2 (hACE2) via attachment and fusion. Potential intervention approaches in SARS-CoV-2 treatment include therapeutic strategies that could prevent SARS-CoV-2 S-protein binding to hACE2. In this study, we implemented an exhaustive approach to identify drug molecules that could potentially bind to SARS-CoV-2 S-protein at other sites other than the RBD. Pertinent to the allosteric targeting approach implemented herein
was the identification of highly druggable sites inherent in the S-protein (S1/S2), which was carried out using multiple pocket prediction algorithms for identification and validation of possible allosteric sites. Predicted pockets were then characterized based on their attributes after which two highly probable pockets were selected. These were then screened distinctly against a library of ~1500 FDA approved drugs retrieved from the ZINC database. Amongst all, **ZINC3939013** (Fosaprepitant) and **ZINC27990463** (Lomitapide) were obtained for both Sites 1 and 2. All atom-MD simulation methods were then used to investigate their interaction dynamics and possible inhibitory effects on the S-protein (S1/S2)-hACE2 complex. Findings revealed that the allosteric binding of the compounds induced varying degrees of conformational alterations across the protein structures, including the RBD where hACE2 interactions reportedly occur. Time-based analysis revealed a systematic displacement of the hACE2 away from the allosterically-bound S-protein RBD domain, which on the other hand exhibited reduced Cα motions. Conformational variations across the different components of the S1/S2 protein provided further structural insights into the inhibitory mechanisms of the allosteric binders; **ZINC3939013** and **ZINC27990463**.

Analyses of interaction dynamics revealed that while **Compd A** (**ZINC3939013**) was optimally buried within the deep hydrophobic pocket of **Site 1**, **Compd B** (**ZINC27990463**) exhibited trans-domain binding across **Site 2** and its adjacent NTD-RBD linker (L) region. This binding activity demonstrated by **Compd B** could account for its prominent disruptive effects on the S-protein-hACE2 complexes. $\Delta G_{\text{binds}}$ of -52.7 kcal/mol and -32.57 kcal/mol were estimated for **ZINC3939013** and **ZINC27990463** respectively. Per-residue energy decomposition revealed the crucial contributions of R190, H207, K206 and K187, I101, R102, I119, F192, L226, V126, and W104 to the binding of **ZINC3939013**. Also, energy contributions of Q957, N953, Q954, L303,
Y313, Q314, L858, V952, N953, and A956 to **ZINC27990463** binding and stability were reported. While *de novo* binding site characterization performed herein were based on computational modeling methods, experimental methods such as Surface Plasmon Resonance (SPR), Isothermal Titration Calorimetry (ITC), Biolayer Interferometry (BLI), and Microscale Thermophoresis (MST) can be employed for further validation. These implementations will provide additional insights into the targetability and suitability of these pockets for novel COVID-19 therapeutics. Findings from this study paves way for novelty in the structure-based design of high-affinity allosteric inhibitors or disruptors of SARS-CoV-2 association with host hACE2 thereby preventing viral entry.

**Acknowledgement**

Authors thank the college of health sciences, University of KwaZulu-Natal, South Africa for providing infrastructural support and we also acknowledge the Center for High Performance Computing (CHPC), Capetown, South Africa, for providing computational resources.

**Declaration of Interest**

Authors declare no conflict of interest.

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**References**


(3) González, J. M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A. E., and Enjuanes, L.


Crit. Care 21.


(53) Paola, L. Di, and Giuliani, A. Mapping active allosteric loci SARS-CoV Spike Proteins by means of Protein Contact Networks.


Authorship Form

The following information is required for submission. Please note that failure to respond to these questions/statements will mean your submission will be returned. If you have nothing to declare in any of these categories then this should be stated.

Please state any sources of funding for your research

All sources of funding should be declared as an acknowledgement at the end of the text. Authors should declare the role of study sponsors, if any, in the collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. If the study sponsors had no such involvement, the authors should so state.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Consent

Studies on patients or volunteers require ethics committee approval and fully informed written consent which should be documented in the paper.

Authors must obtain written and signed consent to publish the case report from the patient (or, where applicable, the patient's guardian or next of kin) prior to submission. We ask Authors to confirm as part of the submission process that such consent has been obtained, and the manuscript must include a statement to this effect in a consent section at the end of the manuscript, as follows: "Written informed consent was obtained from the patient for publication of this case report and accompanying images. A copy of the written consent is available for review by the Editor-in-Chief of this journal on request".

Patients have a right to privacy. Patients’ and volunteers’ names, initials, or hospital numbers should not be used. Images of patients or volunteers should not be used unless the information is essential for scientific purposes and explicit permission has been given as part of the consent. If such consent is made subject to any conditions, the Editor in Chief must be made aware of all such conditions.

Even where consent has been given, identifying details should be omitted if they are not essential. If identifying characteristics are altered to protect anonymity, such as in genetic pedigrees, authors should provide assurance that alterations do not distort scientific meaning and editors should so note.

Not applicable to this study.

Author contribution

Please specify the contribution of each author to the paper, e.g. study design, data collections, data analysis, writing, others, who have contributed in other ways should be listed as contributors.

FAO conceptualized, implemented, analyzed, interpreted and wrote the manuscript, KFO performed molecular dynamics simulation, while MES revised and approved the manuscript for submission.
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: