



**Molecular characterization of COVID-19 therapeutics:
Luteolin as an allosteric modulator of the spike protein
of SARS-CoV-2**

Journal:	<i>Molecular Systems Design & Engineering</i>
Manuscript ID	ME-ART-08-2021-000119
Article Type:	Paper
Date Submitted by the Author:	18-Aug-2021
Complete List of Authors:	de Pablo, Juan; Liew Family Professor of Molecular Theory and Simulation, Institute for Molecular Engineering Alvarado, Walter; University of Chicago, Biophysical Sciences Bylehn, Fabian; University of Chicago Menendez, Cintia; Universidad Nacional del Sur, Chemistry department Perez, Gustavo; University of Chicago

SCHOLARONE™
Manuscripts

The design of small-molecules which target the spike protein of SARS-CoV-2 has been the focus of recent studies in the search for an effective treatment strategy against COVID-19. In this study, molecular dynamics simulations are used to study how several previously identified therapeutics alter the structure and dynamics of the spike protein. We analyze the strain and molecular stiffness induced by each small molecule upon binding and identify a previously uncharacterized distal binding site that induces significant allosteric conformational changes to the protein–protein interaction that takes place between the subunits of the spike protein and the active site of the angiotensin-converting enzyme 2 (ACE2), a key binding partner required for host cell infectivity. The fundamental understanding obtained in this study will enable the development and engineering of novel small molecules that disrupt viral entry and provide an informed perspective on how allosteric modulators serve as a template for the design of suitable therapeutics.

Cite this: DOI: 00.0000/xxxxxxxxxx

Molecular characterization of COVID-19 therapeutics: Luteolin as an allosteric modulator of the spike protein of SARS-CoV-2[†]

Walter Alvarado,^{a‡} Gustavo R. Perez-Lemus,^{b‡} Cintia A. Menendez,^{b‡} Fabian Bylehn,^{b‡} and Juan J. de Pablo^bReceived Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The interactions between the receptor binding domain (RBD) of SARS-CoV-2 and the angiotensin-converting enzyme 2 (ACE2) are crucial for viral entry and subsequent replication. Given the large and featureless contact surfaces between both proteins, finding a suitable small-molecule drug that can bind and disrupt regulatory pathways has remained a challenge. A promising therapeutic alternative has been the use of small compounds that bind at the protein-protein interface or at distal "hot spots" and induce conformational changes that inhibit or stabilize protein-protein interactions (PPIs). In this work, we conduct large-scale all-atom explicit solvent simulations of the top scoring compounds from a recent large-scale high-throughput docking screening to investigate their interaction with the RBD domain of the spike (S) protein in complex with ACE2. We identify several promising candidates that exhibit a large negative free energy of binding to the RBD/ACE2 complex based on *ab initio* thermodynamic integration calculations. A systematic structural analysis of two glycosylation profiles of the RBD/ACE2 complex reveal the important role glycans play in modulating the binding of small-molecules. Cross correlation, fluctuation and strain analysis identify several of these compounds as effective PPI modulators that inhibit or stabilize the protein-protein interactions of RBD/ACE2 based on glycosylation profile. Among them, Luteolin, a drug currently approved for asthma, exhibits an intense allosteric effect when it binds to a previously unidentified distal binding site of the RBD domain far from the RBD and ACE2 interface which may serve as a potential target for future drug discovery.

Design, System, Application The design of small-molecules which target the spike protein of SARS-CoV-2 has been the focus of recent studies in the search for an effective treatment strategy against COVID-19. In this study, molecular dynamics simulations are used to study how several previously identified therapeutics alter the structure and dynamics of the spike protein. We analyze the strain and molecular stiffness induced by each small molecule upon binding and identify a previously uncharacterized distal binding site that induces significant allosteric conformational changes to the protein-protein interaction that takes place between the subunits of the spike protein and the active site of the angiotensin-converting enzyme 2 (ACE2), a key binding partner required for host cell infectivity. The fundamental understanding

obtained in this study will enable the development and engineering of novel small-molecules that disrupt viral entry and provide an informed perspective on how allosteric modulators serve as a template for the design of suitable therapeutics.

1 Introduction

As of February 2021, the human coronavirus referred to as severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) has infected millions around the world, and is responsible for the COVID-19 disease that has so far led to well over 2.61 million deaths worldwide.¹ As vaccinations begin to roll out, viable treatments for COVID-19 are still necessary to treat those who may contract the disease. One promising treatment strategy is the identification of previously approved drugs that could be repurposed to act on different stages of SARS-CoV-2 infection and host response.

The surface-anchored viral spike (S) glycoprotein mediates coronavirus entry into host cells. The S protein is a type I transmembrane protein comprising two functionally distinct regions, S1 and S2, that mediate receptor binding and membrane fusion,

^a Biophysical Sciences, University of Chicago, 929 E. 57th Street, Chicago, IL 60637, USA. E-mail: walt@uchicago.edu

^b Pritzker School of Molecular Engineering, University of Chicago, 5640 S. Ellis Avenue, Chicago, IL 60637, USA.

[†] Electronic Supplementary Information (ESI) available: See DOI: 00.0000/00000000.

[‡] These authors contributed equally to this work.

respectively. Similar to SARS-CoV-1, responsible for the 2003 SARS outbreak in Asia, the S1 subunit of the SARS-CoV-2 engages the same angiotensin-converting enzyme 2 (ACE2) receptor for host cell entry and utilizes the serine protease TMPRSS2 for S protein priming.² Of the two regions, the S2 subunit is the most conserved among different coronavirus genera and nine residues involved in the interaction between ACE2 and the receptor binding domain (RBD) of the S1 subunit are conserved between SARS-CoV and SARS-CoV2.^{3,4}

Specific RBD-receptor binding determines host cell infection and has been the target of several drug design efforts. Recent flow cytometry studies have shown that SARS-CoV-2 RBD exhibits a higher binding affinity to human ACE2 (hACE2) and bat ACE2 (bACE2) receptors than SARS-CoV-1 RBD.⁵ Deletions in the RBD of closely related S protein sequences have also been shown to inhibit binding to hACE2.⁶ In addition, several studies indicate that the RBD is an attractive antigen for specific antibody detection.^{7,8} Taken together, these results suggest that SARS-CoV-2 RBD is a promising target for the development of novel vaccines and antiviral drugs.

Recent studies have identified key surface contacts that can be leveraged to disrupt the RBD/ACE2 protein-protein interaction (PPI).^{9,10} PPIs are physical contacts of high specificity between proteins that play an important role in cellular function.¹¹ PPIs are characterized by their large, flat, and featureless binding interfaces, as in the case of the RBD/ACE2 complex. One major challenge in PPI inhibition is designing small molecules that can compete with the high binding affinity of a natural protein partner. In addition, the large and flat nature of the protein interaction surface often lacks clear binding pockets or grooves that can act as binding sites for smaller ligands. A promising alternative is the identification or engineering of allosteric modulators that stabilize protein-protein interactions, thereby interfering with the downstream pathways they mediate.¹² The key strategy in this approach involves indirectly affecting PPI interfaces by targeting distal binding sites that are structurally distinct. Allosteric modulators induce a conformational change that either inhibits or stabilizes association with another protein and have shown promise in the design of suitable therapeutics.^{13,14}

In this work, we focus on a set of small-molecule drugs that can be repurposed for the therapeutic treatment of COVID-19. Specifically, we focus on the top ten scoring compounds proposed in a recent high-throughput supercomputer-based docking screen that was performed in vacuum and relied on minimization of the potential energy of the ACE/RBD complex.¹⁵ Here we perform molecular simulations in explicit water and determine the relative and absolute binding free energies of each drug at various binding locations and with different glycosylation profiles. We find the affinity of the ligands to the complex to be different than previously predicted on the basis of energy minimization, with several of the top candidate drugs unbinding from the protein over the course of nanosecond long trajectories. We study the effect glycans have on binding affinity for several ligands by comparing two glycoforms of the RBD/ACE2 complex: a simple and "Abundant" model. Notably, our free energy results show that for an RBD/ACE2 complex glycosylated with small sugar moi-

eties, high-affinity binding ligands stabilize the binding between RBD and ACE2. An RBD/ACE complex with glycans that were determined as the most abundant from a glycoproteomics study conducted by Zhao *et al.* was analyzed for comparison.¹⁶ The presence of these complex-glycans affect the binding affinity of several drug candidates and their mechanism of action. We then analyze the strain and molecular stiffness induced by each small molecule upon binding and identify a "hotspot", comprising only a few amino acid residues, as a potential target for drug discovery. Surprisingly, binding of Luteolin – one of the drugs considered here, to this distal site, induces a profound allosteric conformational change to residues near the RBD/ACE2 binding interface, which disrupts non-native contacts at the protein-protein interface. We also provide additional quantitative insights into the binding mechanism of Luteolin to the RBD/ACE2 complex using cross-correlation analysis and Principal Component Analysis (PCA). We conclude with a discussion of our findings and several suggestions for future experimental studies.

2 Methods

2.1 Molecular Docking

Autodock 4.2 was used for the molecular docking between target proteins and ligands using the Lamarckian genetic algorithm (LGA), and pseudo-Solis and Wets local search method.¹⁷ The initial configuration of receptor-binding domain (RBD) of the spike protein of SARS-CoV-2 bound to the ACE2 receptor was taken from crystal structure (PDB: 6M0J) after an NPT relaxation for 100ns and averaged for the last 20ns (see section below for simulation details). The search space was centered around the S-protein and spanning the binding interface of the ACE2 receptor. The initial configurations of ligands were randomized before each docking calculation. A total of 200 docking runs were performed and each run was set to terminate after 25,000,000 energy calculations. The best pose of each ligand was selected for further analysis in MD simulations.

2.2 Molecular Dynamics

The recently determined crystal structure of the receptor-binding domain (RBD) of the spike protein of SARS-CoV-2 bound to the ACE2 receptor was used as an initial structure (PDB: 6M0J). Crystal waters were removed force field parameters of the S-protein/ACE2 complex were determined with the Antechamber program using the ff14SB and GLYCAM-06j-1 force fields. An octahedron box with 40,000 TIP3P water molecules and 23 Na⁺ ions was added. Energy minimization included 3000 steps which involved 1500 steepest descent steps constrained to heavy atoms, followed by a second minimization of 30000 steps involving 15000 steepest descent steps. Equilibration was performed through a gradual temperature increase from 0 K to 300 K over 400 ps using Langevin thermostat with a temperature coupling constant of 1.0 ps in a constant volume ensemble. Density equilibration and production runs were conducted using a constant pressure ensemble (NPT). All simulations were performed using periodic boundary conditions and 2 fs time step. Long-range electrostatic interactions were modeled using the Particle Mesh Ewald

method with a non-bonded cut-off of 10 Å and the SHAKE algorithm. The ligands included for MD simulations were described by the General Amber Force Field (GAFF). Partial charges for all small-molecules were generated using the AM1-BCC charge model. Their initial position was selected from the best score in docking calculations.

2.3 MM/GBSA Calculations

The relative binding free energy between ACE2 and RBD domains were calculated using the GBSA method implemented in MMPBSA.py with `igb2=2` and `mbondi2` parameter.¹⁸ This relative free energy is defined as: $\Delta\Delta G_{\text{Binding}} = \Delta G^{\text{Lig}} - \Delta G^{\text{N}}$ where the superscript Lig denotes the RBD/ACE2 binding free energy with the presence of the ligand in simulations and N denotes the RBD/ACE2 system with no ligands. Three different replicas were used for every ligand/no ligand system, each one with 5000 snapshots sampled every 20ps from a previously 300ns equilibrated system.

2.4 Thermodynamic Integration Calculations

The absolute binding free energy for ligands is defined as: $\Delta G_{\text{Absolute}} = \Delta G^{\text{L}} - \Delta G^{\text{RL}}$, where ΔG^{RL} is the free energy change of the ligand annihilation in the RBD/ACE2 complex, and ΔG^{L} is the free energy change of ligand annihilation in water. To calculate these free energy changes, we use Thermodynamic Integration (TI) implemented in PMEMD for Amber20. A one step annihilation protocol with soft core potentials was implemented. Runs were conducted from a starting equilibrated ligand position extracted via K-means clustering from previous 300ns MD simulations. In this way, three independent replicas for each ligand were considered, as well as three replicas for solvation in pure water. Twelve windows were selected using Gaussian quadrature with 10ns of simulation time per window. To keep the ligand from wandering in TI calculations, we used a soft restraint of 10 kcal/molÅ².

2.5 Contact Maps and RMSF Calculations

Contacts maps by residue and Root Mean Squared Fluctuations (RMSF) calculations were generated using the native contacts and RMSF functions in CPPTRAJ.¹⁹ The native contacts were defined relative to crystal PDB (6M0J) with a cutoff distance of 7 Å. The maps were averaged over 1000 snapshots taken every 100ps for each ligand. Root Mean Squared Fluctuations calculations were calculated with respect to a 100ns averaged structure using the mass-weighted average over the CA, N and C atoms and reported by residue. A total of 5000 snapshots were used taken every 20ps.

2.6 Strain Analysis

The effects of functional and non-functional fluctuations after ligand binding were performed through the application of the discrete form of the strain formalism from continuum theory.^{20,21} In this form, derivatives are replaced by differential operators. The local neighborhood for a given central atom, i , is determined by

a radius R that contains n number of atoms j . Distances between atoms i and j are related through the deformation matrix F ,

$$x_j - x_i = F (x_{0,j} - x_{0,i}) \quad (1)$$

where x_i and x_j are the instantaneous position of atom i and j , and $x_{0,i}$ and $x_{0,j}$ correspond to their positions at any other given timestep, respectively. An optimized F^* is then sought, that minimizes the difference between actual distances and projected distances to an affine deformation,

$$F^* = \min \sum_{j=1}^n [x_j - x_i - F (x_{0,j} - x_{0,i})]^2 \quad (2)$$

The atomic strain tensor is determined by,

$$\varepsilon = \frac{1}{2} (F^T F - I) \quad (3)$$

whose magnitude is defined as the L2-norm of the shear term,

$$\text{Tr} \left(\left(\varepsilon - \frac{1}{3} \text{Tr} \varepsilon \cdot I \right)^2 \right) \quad (4)$$

given that proteins are generally incompressible.^{20,22}

For this study, a 10Å radius around each Cα atom for our analysis was considered. A simulation of RBD/ACE2 without ligand was used as reference for strain calculations to elucidate the binding of several ligands at different sites.

2.7 Principal Component Analysis

Principal Component Analysis (PCA) calculations were performed using the Bio3D package for R.²³ PCA is a dimensionality reduction technique that is effective in identifying correlated motions in atomic simulations of proteins from experimental structures or MD trajectories. Essential correlated conformational changes between structures can be represented in this low-dimensional subspace spanned by the first few principal components (PCs).

Mathematically, PCs are evaluated by diagonalizing the correlation matrix C_{ij} ,

$$R^T C R = \Lambda \quad (5)$$

for coordinates i and j ,

$$C_{ij} = \langle (x_i - \langle x_i \rangle) (x_j - \langle x_j \rangle) \rangle \quad (6)$$

where x_1, \dots, x_{3N} are the mass-weighted atomic coordinates of the protein, averaged over all sampled structures from simulation trajectories. The diagonal of the matrix Λ contains eigenvalues that correspond to the eigenvectors of R . Eigenvectors with the largest eigenvalues account for the highest proportion of variance within the dataset (first PC) and decrease sequentially while maintaining orthogonality to the first PC. The first several PCs are often considered “essential dynamics”, and the rest are neglected without significant loss of information.²⁴

3 Results and discussion

3.1 Calculating Binding Free Energies

The configuration of the complex was derived from a recently published crystal structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor (PDB: 6MOJ), whose conformation noticeably differed from the homology modeled structure used by Smith *et al.*¹⁵ The glycosylation profile of the RBD/ACE2 protein was based on a simple glycan model and a glycomics-informed glycoproteomics structure ("Abundant" model) generated by Zhao *et al.*¹⁶ which included one glycan group on the RBD domain and six on ACE2 (Figure S2). Molecular dynamics simulations were carried out using the Amber20 simulation package.²⁵ To increase sampling, ensure convergence, and extract an equilibrated representative structure, K-means clustering was applied to three independent 100 ns molecular dynamics replicas *sans* ligand. To identify potential binding sites and validate the results of the previous docking study, candidates were docked using the Autodock4 software.¹⁷ Three binding sites were identified for the simple glycan model: the binding interface between the S protein and the ACE2 receptor, a distal RBD region, and a site inside the ACE2 receptor (Figure 1). For the "Abundant" model, the interface and distal binding sites were identified. The structure with the highest binding affinity, as identified through docking, was used as the starting structure for production runs. Three 100 ns replicas for each ligand in complex with RBD/ACE2 were conducted to determine whether ligands would remain bound. Six ligands remained bound to the simple glycan model, and only three for the "Abundant" form. A representative structure with the ligand bound was extracted and used as a starting structure for three additional MD replicas totaling 300 ns for each ligand bound to the RBD/ACE2 complex and used for analysis.

Multiple binding free energy calculations were conducted to determine the binding affinity between the RBD/ACE2 complex and each small ligand (Table 1). The results were compared to those of the high-throughput screening.¹⁵ We observed standard deviations ranging from 0.60-0.91 and 0.16-0.53 between our docking binding energies to that of the previous screening study for the simple and "Abundant" models, respectively. This could be due to the differences in the structures used (homology modeled vs. crystal structure) and the fact that important disulfide bonds and glycans groups were accounted for in our structure but were omitted in the Smith *et al.* analysis.¹⁵ The absolute binding free energies between ligand and protein reported here were determined using thermodynamic integration (TI) in the presence of explicit water. In our TI protocol, each ligand atom is treated as a softcore atom and is subsequently "removed" in a one-step alchemical cycle both in solution and in complex. This alchemical method, also referred to as free energy perturbation (FEB), allows for the calculation of absolute binding free energies (ABFE). We observed larger binding energies with standard deviations as high as 10.79 for the simple glycan structure and 5.04 for the "Abundant" glycoform in our free energy calculations than the docking binding energies. Of the three ligands studied, Luteolin showed the highest binding affinity for both glycoforms of the RBD/ACE2

complex. Luteolin was also the only ligand to bind at the distal site compared to the other candidates which were found to have a stronger binding affinity for the binding interface of RBD and ACE2 (Figure 1).

To determine how each compound affects S protein and ACE2 binding, we calculated the free energy of binding between the RBD domain of the S-protein and the ACE2 receptor in the presence of each ligand using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method. Tables 1 and 2 show the difference in binding free energy with ligand present versus without ($\Delta\Delta G_{\text{Binding}}$) for the simple and "Abundant" glycan models of the complex. In the case of the simple glycan model, our results show that every ligand enhances the binding free energy between the RBD domain and ACE2, acting as a binding stabilizer for the complex. Sapropterin, which binds inside ACE2 (Figure 1), appears to be the most potent stabilizer of the analyzed group. For the "Abundant" glycan model, our results show that Isoniazid acts as an allosteric inhibitor as indicated by the positive binding energy between the RBD domain and ACE2 ($\Delta\Delta G_{\text{Binding}}$) when the ligand is present. In contrast, Eriodictyol was observed to increase the binding affinity of the RBD domain to ACE2 thereby acting more as an allosteric stabilizer. Luteolin was observed to have negligible effect on the binding energy between RBD and ACE2. We hypothesize that inhibition/stabilization of the complex induced by these small molecules may serve to disrupt downstream events such as S protein priming by TMPRSS2, an essential step for cellular entry by the virus.

3.2 Structural Analysis of Luteolin Binding

Figure 2 provides a closer look at the interactions between ligand and protein for each identified binding region. For both glycoforms, several of the potential binding sites were observed to localize to the same regions; their interactions are shown in Figure S3. Given its higher affinity for the RBD/ACE2 complex, we focus on the interactions of Nitrofurantoin (Figure 2A) within the binding interface. In this region, hydrogen bonds are the dominant interactions responsible for stabilizing Nitrofurantoin. These hydrogen bonds are formed between the carbonyl oxygens of Nitrofurantoin and ACE2 residue Lys353 and RBD residue Gln493. Figure 2B shows a representative configuration for Luteolin and the distal RBD binding site. Luteolin is stabilized by a hydrogen bond between its carbonyl oxygen and Tyr369 and pi-alkyl stacking between its aromatic group and Phe377. For comparison, Sapropterin, which was found to bind to only ACE2, is stabilized by hydrogen bonds formed with residues of the ACE2 protein (Figure 2C).

We study the changes of non-native contacts between the S protein and ACE2 receptor upon ligand binding by comparing the non-native contact maps of each compound (Figure 3). Similar changes in non-native contacts are observed for compounds Eriodictyol and Nitrofurantoin, which bind at the RBD/ACE2 interface (Figure 3A, 3B), and Sapropterin, which binds inside ACE2 (Figure 3C). Interestingly, binding of Luteolin (Figure 3D) to the distal binding region shows a clear difference in the number of contacts between the S protein and ACE2 receptor. These results

Table 1 Ligand Binding Energies for Simple Glycan Model

Molecule	Oak Ridge ΔG (kcal/mol)	Docking ΔG (kcal/mol)	TI ΔG (kcal/mol)	RBD/ACE2 $\Delta\Delta G$ (kcal/mol)
Luteolin	-7.4	-6.19	-11.31 ± 0.62	-13.93 ± 1.63
Protirelin	-7.3	-8.45	-1.22 ± 5.79	-11.22 ± 5.10
Nitrofurantoin	-7.2	-9.22	-28.98 ± 4.83	-13.96 ± 2.84
Sapropterin	-7.1	-5.67	-6.10 ± 1.21	-18.89 ± 3.64
Vidarabine	-7.1	-6.24	-3.39 ± 1.70	-17.77 ± 1.66
Eriodictyol	-7.1	-7.86	-9.36 ± 2.41	-15.99 ± 3.53

Table 2 Ligand Binding Energies for "Abundant" Glycan Model

Molecule	Oak Ridge ΔG (kcal/mol)	Docking ΔG (kcal/mol)	TI ΔG (kcal/mol)	RBD/ACE2 $\Delta\Delta G$ (kcal/mol)
Luteolin	-7.4	-7.71	-15.04 ± 5.04	0.0027 ± 8.94
Isoniazid	-7.3	-6.57	-7.29 ± 4.54	11.87 ± 7.26
Eriodictyol	-7.1	-8.16	-12.65 ± 2.36	-8.06 ± 9.92

suggest that an allosteric effect is induced when Luteolin binds to the distal binding site.

Figure 4 shows the estimated Root Mean Squared Fluctuations and the shear strain on each residue of the RBD/ACE2 complex for several ligands. Strain is a natural quantity to study local protein deformations, which are a good indicator of mechanical allosteric coupling between two binding events on the same protein. This deformation upon ligand binding is calculated as the shear strain measured relative to the average conformation from all frames of the RBD/ACE2 sans ligand. Average Root Mean Square Fluctuations (RMSF), a measure of the displacement of a residue relative to the initial crystal structure (PDB: 6MOJ), suggest high levels of deviations for RBD residues Asn481, Gly482 and Val483 near the binding interface of ACE2, particularly for the Luteolin ligand (Figures 4A,5A). Strain analysis also shows a significant peak at these residues (Figures 4B,5B). Consistent with the results of Figure 3D, this profound allosteric effect is induced when Luteolin is bound to the distal RBD binding site for both the simple and "Abundant" glycan forms of the RBD/ACE2 complex.

The formation of cross-correlation networks allows the transmission of information when the binding of a molecule at one site of the protein induces a change in local structure elsewhere in the protein. From our structural analysis, Luteolin was the only ligand to induce a conformational change to the RBD protein after binding to a distal site. To validate the extent by which the atomic fluctuations of the complex are correlated with one another in the presence of Luteolin, a dynamic cross-correlation network analysis was conducted using the Bio3D software.²³ Anti-correlations between -0.40 and -0.60 are represented as blue lines between several distant regions of the RBD, as seen in Figures 5C and 5D (ACE2 has been omitted for clarity). A comparison of simulations of the RBD/ACE2 with and without Luteolin shows the disruption of the dynamic cross-correlation network by this drug. The disappearance of elements of the network, which include anti-correlated sites between this distal binding site and the RBD/ACE2 interface, may be critical in mediating allosteric transitions and ACE2 binding.

Principal component analysis (PCA) was used to gain quantitative insights into the binding of Luteolin to the RBD/ACE2 complex. The orthogonal eigenvectors of the resulting principal

components (PCs) describe the maximal variance of the distributions of structures. Details of the data processing are described in the Materials and Methods section. The two dominant principal components (PCs) were sufficient to describe the observed conformational changes at the RBD/ACE2 interface. The first principal component (PC1) corresponds to the rotational motion of the RBD/ACE2 complex. The second principal component (PC2) captures the conformational change observed at the RBD/ACE2 interface induced by Luteolin binding (Figure 6). As previously discussed, residues near the binding site were observed to be anti-correlated to residues near the ACE2 binding site. Binding of Luteolin disrupts this cross-correlation network, thereby inducing a conformational shift near the ACE2 binding site. Conformational changes to the RBD protein, as captured by PCA, confirm not only correlated motions of the RBD protein but suggest that the "hotspot" identified here may be a useful target in the discovery and design of new therapeutics that modulate the RBD/ACE2 protein-protein interface. Taken together, our results suggest that Luteolin may serve as an important small-molecule allosteric modulator that affects RBD/ACE2 protein dynamics and may provide an alternative therapeutic approach.

4 Conclusion

The repurposing of approved drug candidates provides an alternative strategy to identify lead candidates for viral infections. The time and expense of bringing a drug to market are significantly reduced when the safety and pharmacokinetic profiles of existing drugs are already known. Drug repositioning is guided by a rational approach that requires detailed knowledge of the target structure, the spatial arrangement, interactions and structural conformation of the compound, and the mechanism of action.²⁶ For protein-protein complexes, the full binding inhibition is hard to achieve using small molecules, and novel approaches based on the concept of interfacial inhibition are needed, where macromolecules are trapped in dead-end complexes that cannot fulfill their biological function.²⁷ In this sense, allosteric modulation of PPIs of the RBD/ACE2 complex may serve as an approach in drug discovery. While current high-throughput virtual screenings have identified several potential therapeutic candidates, the mechanism by which they affect protein-protein interactions remains unclear.

In this study, we have focused on determining the binding mechanism and associated conformational changes in the presence of the top compounds from a docking screen recently conducted by Smith *et al.*¹⁵ After filtering compounds by their residence time in complex through MD simulations, rigorous binding free energy calculations were conducted in explicit water using thermodynamic integration. Our results have revealed a range of free energies and binding positions that were not anticipated in the original docking studies. The majority of such positions lie far from the interface, and the free energy extrema correspond to the shallow interface region of binding (Tables 1-2 and Figure S1). The relative RBD/ACE2 binding free energy changes induced by the top ligands considered here indicate a binding enhancement of the RBD/ACE2 complex, even for drugs bound in the interface region. This enhancement is accompanied by a modification of the non-native contact maps of the protein interface, which lead to changes in the manner in which the proteins interact. Taken together, our results suggest that these ligands could not act as ordinary competitive inhibitors for RBD, since there is no binding disruption, but, given the contact modification and the energy differences with respect to the pure RBD/ACE2 complex, we propose that they may inhibit the system by serving as allosteric or direct PPI stabilizers.^{12,28}

PPIs with large, flat and featureless surfaces, as in the case in the RBD/ACE2 complex, lack good drug-binding pockets for ligands. Recent approaches have shown that allosteric modulators provide an alternative strategy to target PPIs such as the RBD/ACE2 complex. Our results have uncovered pronounced effects at the RBD/ACE2 interface upon ligand binding at a distal site. In particular, upon Luteolin binding, RMSF and strain analysis unveil significant levels of fluctuations and strain in RBD regions away from the binding site. In addition, cross-correlational analysis reveals the disruption of anti-correlated motions between the Luteolin binding site and the binding interface of RBD/ACE2. Future research will require investigating the possible binding motifs for Luteolin and its effects on S protein binding to the ACE2 receptor. Beyond the finding that Luteolin might serve to inhibit viral entry, the discovery of this distal binding site offers potential for the design and engineering of future therapeutics for COVID-19.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

The simulations reported here were carried out on the GPU cluster supported by the NSF through grant DMR1828629. The authors are grateful to the Research Computing Center of the University of Chicago for additional computational resources.

Notes and references

- 1 E. Dong, H. Du and L. Gardner, *The Lancet infectious diseases*, 2020, **20**, 533–534.
- 2 M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche *et al.*, *cell*, 2020, **181**, 271–280.
- 3 P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang *et al.*, *nature*, 2020, **579**, 270–273.
- 4 Y. Huang, C. Yang, X.-f. Xu, W. Xu and S.-w. Liu, *Acta Pharmacologica Sinica*, 2020, **41**, 1141–1149.
- 5 W. Tai, L. He, X. Zhang, J. Pu, D. Voronin, S. Jiang, Y. Zhou and L. Du, *Cellular & molecular immunology*, 2020, **17**, 613–620.
- 6 Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K.-Y. Yuen *et al.*, *Cell*, 2020, **181**, 894–904.
- 7 L. Premkumar, B. Segovia-Chumbez, R. Jadi, D. R. Martinez, R. Raut, A. Markmann, C. Cornaby, L. Bartelt, S. Weiss, Y. Park *et al.*, *Science immunology*, 2020, **5**.
- 8 N. M. Okba, M. A. Müller, W. Li, C. Wang, C. H. GeurtsvanKessel, V. M. Corman, M. M. Lamers, R. S. Sikkema, E. de Bruin, F. D. Chandler *et al.*, *Emerging infectious diseases*, 2020, **26**, 1478–1488.
- 9 D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan, *Science*, 2020, **367**, 1260–1263.
- 10 R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo and Q. Zhou, *Science*, 2020, **367**, 1444–1448.
- 11 J. Westermarck, J. Ivaska and G. L. Corthals, *Molecular & Cellular Proteomics*, 2013, **12**, 1752–1763.
- 12 D. Ni, S. Lu and J. Zhang, *Medicinal research reviews*, 2019, **39**, 2314–2342.
- 13 M. J. Gorzynski, J. Grembecka, Y. Zhou, Y. Kong, L. Roudaia, M. G. Douvas, M. Newman, I. Bielnicka, G. Baber, T. Corpora *et al.*, *Chemistry & biology*, 2007, **14**, 1186–1197.
- 14 J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells and K. M. Shokat, *Nature*, 2013, **503**, 548–551.
- 15 M. D. Smith and J. C. Smith.
- 16 P. Zhao, J. L. Praissman, O. C. Grant, Y. Cai, T. Xiao, K. E. Rosenbalm, K. Aoki, B. P. Kellman, R. Bridger, D. H. Barouch *et al.*, *Cell host & microbe*, 2020, **28**, 586–601.
- 17 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *Journal of computational chemistry*, 2009, **30**, 2785–2791.
- 18 B. R. Miller III, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *Journal of chemical theory and computation*, 2012, **8**, 3314–3321.
- 19 D. R. Roe and T. E. Cheatham III, *Journal of chemical theory and computation*, 2013, **9**, 3084–3095.
- 20 M. R. Mitchell, T. Tlusty and S. Leibler, *Proceedings of the National Academy of Sciences*, 2016, **113**, E5847–E5855.
- 21 P. Gullett, M. Horstemeyer, M. Baskes and H. Fang, *Modelling and Simulation in Materials Science and Engineering*, 2007, **16**, 015001.
- 22 J.-P. Eckmann, J. Rougemont and T. Tlusty, *Reviews of Modern Physics*, 2019, **91**, 031001.
- 23 B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. Caves, *Bioinformatics*, 2006, **22**, 2695–2696.
- 24 A. Amadei, A. B. Linssen and H. J. Berendsen, *Proteins: Structure, Function, and Bioinformatics*, 1993, **17**, 412–425.

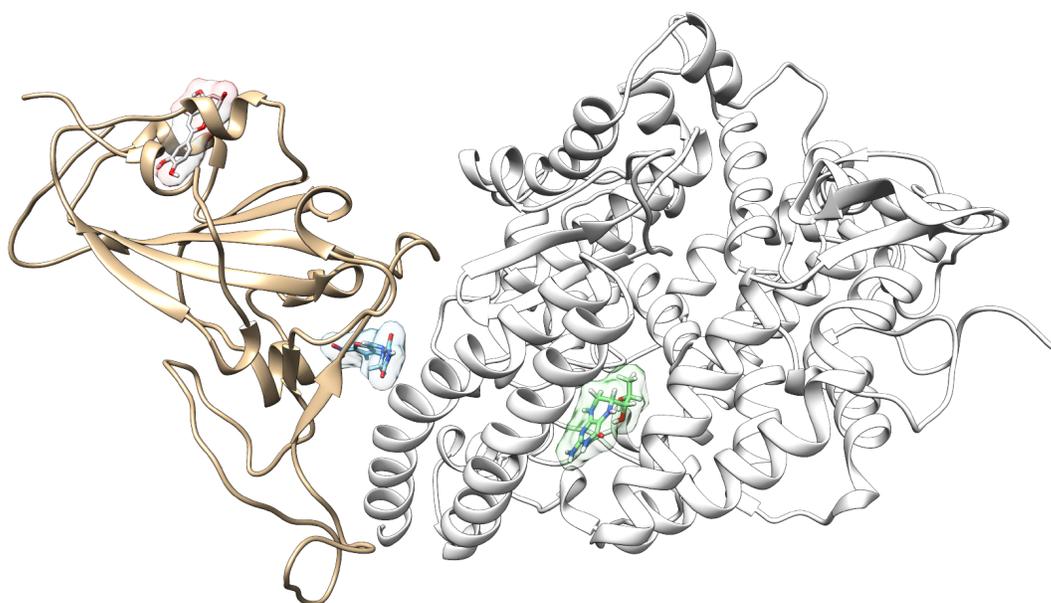


Fig. 1 The RBD domain (tan) of SARS-CoV-2 recognizes ACE2 (white) as its receptor. We identify and characterize the interactions at three potential small-drug binding sites located at the binding interface between the RBD and the ACE2 receptor, inside the ACE2 protein, and a new previously unidentified distal site to which the drug Luteolin has a high binding affinity (red). Nitrofurantoin is shown in blue and Sapropterin is shown in green.

- 25 D. A. Case, K. Belfon, I. Ben-Shalom, S. R. Brozell, D. Cerutti, T. Cheatham, V. W. D. Cruzeiro, T. Darden, R. E. Duke, G. Giambasu *et al.*, 2020.
- 26 W. H. Prusoff, T. Lin, E. M. August, T. G. Wood and M. E. Marongiu, *The Yale journal of biology and medicine*, 1989, **62**, 215.
- 27 Y. Ferrandez, W. Zhang, F. Peurois, L. Akendengué, A. Blangy, M. Zeghouf and J. Cherfils, *Scientific reports*, 2017, **7**, 1–13.
- 28 P. Thiel, M. Kaiser and C. Ottmann, *Angewandte Chemie International Edition*, 2012, **51**, 2012–2018.

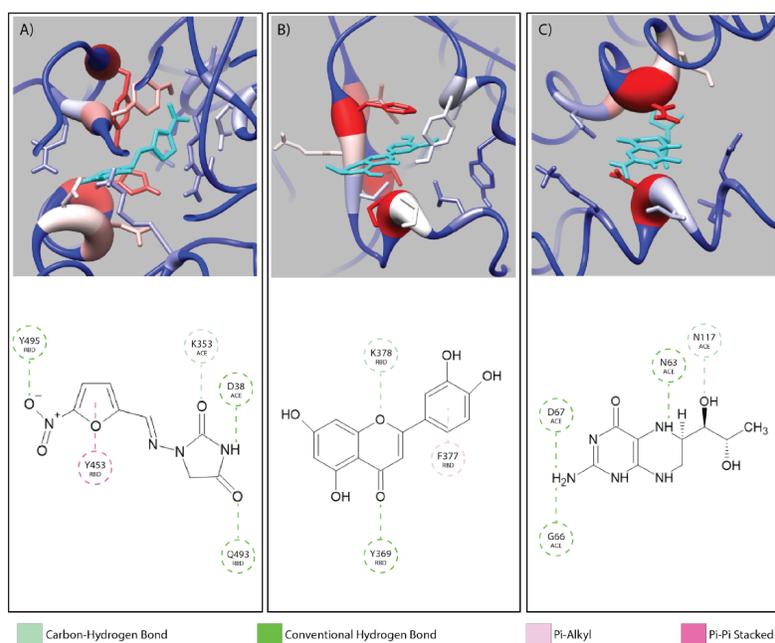


Fig. 2 The interaction diagrams for equilibrated configurations of ligands at the interface (left, Nitrofurantoin), in the RBD domain (middle, Luteolin) and in ACE2 region (right, Sapropterin) are shown. Hydrogen bonds act as the dominating interactions responsible for stabilizing Nitrofurantoin which are formed between the carbonyl oxygens of Nitrofurantoin and the Lys353 residue of ACE2 and RBD residue Gln493. Luteolin is stabilized by a hydrogen bond between its carbonyl oxygen and Tyr369 and pi-alkyl stacking between its aromatic group and Phe377 of the RBD domain. Sapropterin, which is found buried in the ACE2 cavity is stabilized by hydrogen bonds.

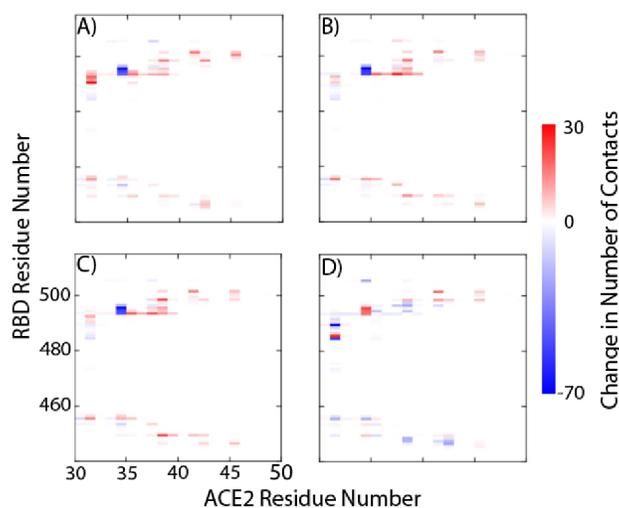


Fig. 3 Relative protein-protein non-native contact maps in the presence of A) Eriodictyol, B) Nitrofurantoin, C) Sapropterin, and D) Luteolin. The relative non-native contact maps measure the change in contacts relative to the complex with no ligands (red more contacts, blue less contacts). From the graphs we see that the first three panels have identical contact profiles compared to D.

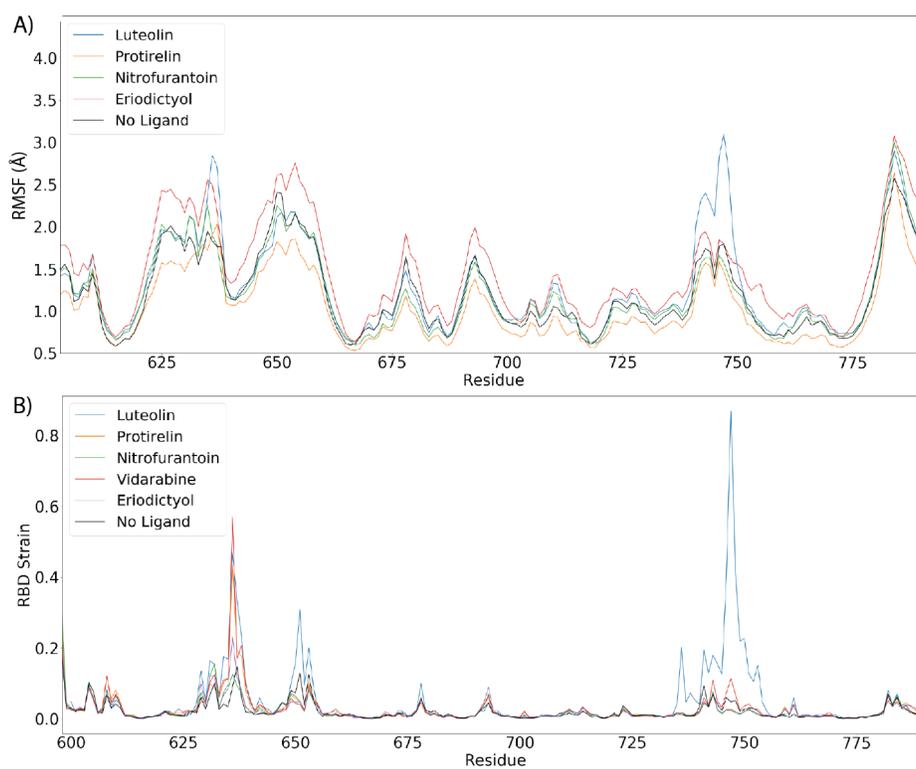


Fig. 4 Luteolin induces large allosteric strain when RBD domain is bound to ACE2. A) Root Mean Squared Fluctuations (RMSF) between the RBD/ACE2 complex with top scoring ligand. B) Shear strains mapped. For shear strain calculations, only $C\alpha$ atoms are included. Strain analysis suggest a strong allosteric strain to the ACE2 binding region of the RBD domain when RBD is in complex with ACE2 and Luteolin.

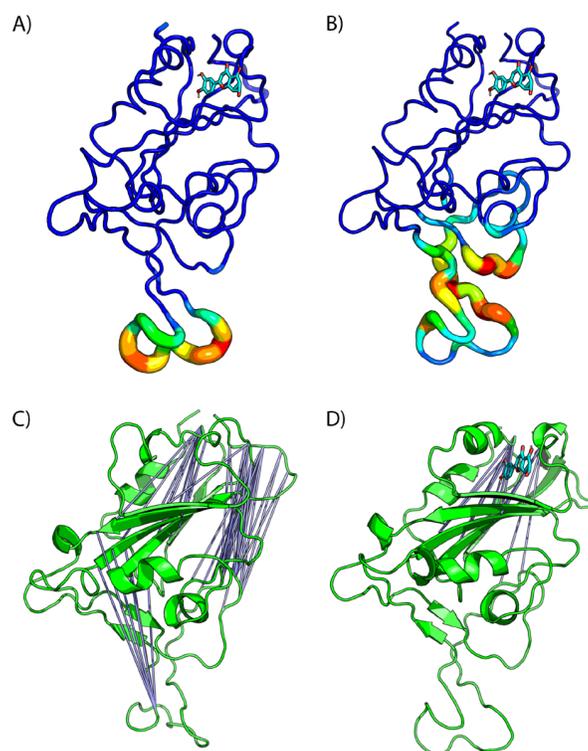


Fig. 5 Bound Luteolin induces large strain at RBD/ACE binding region. A) Difference in estimated Root Mean Squared Fluctuations (RMSF) between the RBD domain and Luteolin in complex are shown. (B) Shear strains mapped onto RBD/LUT complex. Regions flanking disulfide bonds have the highest atomic fluctuations that contribute the deformation energy of the ACE2 binding region. C,D) Distal site binding disrupts intramolecular RBD interactions inducing conformational changes at ACE2 binding interface. Visualization of residue-residue cross correlations. Blue lines indicate anti-correlation motions with values between -0.4 and -0.6. Higher correlations between distal sites *sans* ligand (C) and in the presence of Luteolin (D).

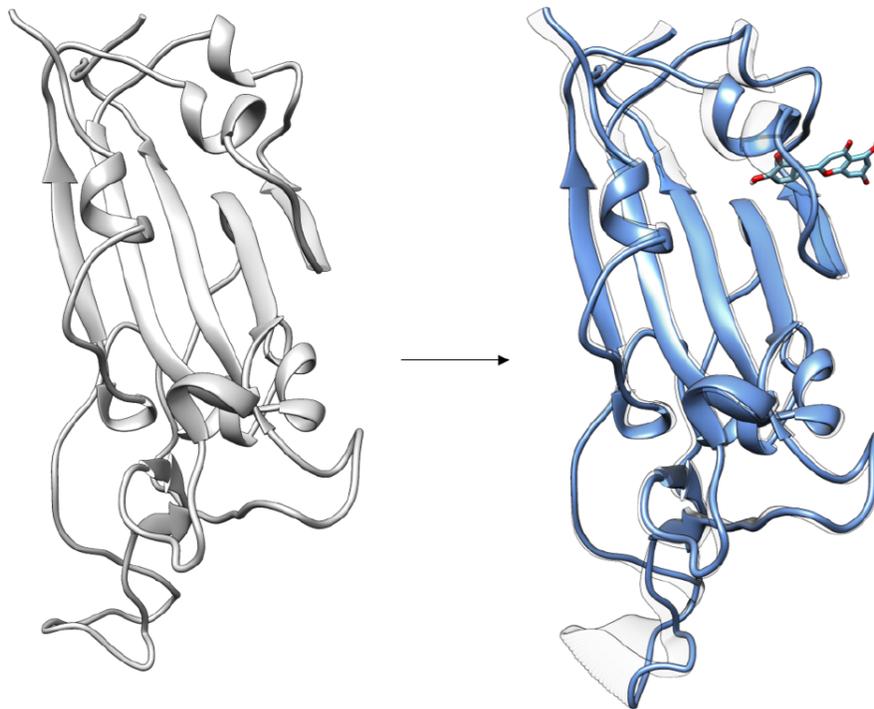


Fig. 6 Distal binding by Luteolin induces conformational changes at ACE2 binding site. Induced conformational changes at loop binding regions are visualized as captured by the second dominant principal component. Images of important conformational changes are superimposed to emphasize conformational changes introduced after Luteolin binding to distal binding site.