

Free energy landscapes for RBD opening in SARS-CoV-2 spike glycoprotein simulations suggest key interactions and a potentially druggable allosteric pocket

Lucy Fallon^{1,2#}, Kellon Belfon^{1,2#}, Lauren Raguette^{1,2}, Yuzhang Wang^{1,2}, Christopher Corbo³, Darya Stepanenko^{1,4}, Abbigayle Cuomo², Jose Guerra⁵, Stephanie Budhan², Sarah Varghese⁶, Robert C. Rizzo^{1,4}, and Carlos Simmerling^{1,2*}

¹Laufer Center for Physical and Quantitative Biology, Stony Brook University

²Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

³Graduate Program in Molecular and Cellular Pharmacology, Stony Brook University

⁴Department of Applied Mathematics and Statistics, Stony Brook University

⁵Department of Biochemistry and Cell Biology, Stony Brook University

⁶Undergraduate Program in Biology, Stony Brook University

These authors contributed equally to this work.

Abstract

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is an enveloped, positive-sense single-stranded RNA virus that is responsible for the COVID-19 pandemic. The viral spike is a class I viral fusion glycoprotein that extends from the viral surface and is responsible for viral entry into the host cell, and is the primary target of neutralizing antibodies. However, antibody recognition often involves variable surface epitopes on the spike, and the receptor binding domain (RBD) of the spike hides from immune recognition underneath a glycan shield aside from brief dynamic excursions to search for the host-cell surface receptor ACE2. Using an atomistic model of the glycosylated wild-type spike in the closed and 1-up RBD conformations, we identified specific interactions that stabilize the closed RBD, and mapped the free energy landscape for RBD opening. We characterized a transient pocket associated with a hinge motion during opening of the RBD, suggesting the possibility of allosteric control of the RBD via this region. Substitution of a conserved alanine to bulkier leucine in the pocket shifted the RBD equilibrium to favor the open, exposed state, as did removal of a conserved lysine that forms a critical salt-bridge in the closed, hidden state. Results from our virtual screening, MD simulations and free energy landscape calculations for wild-type spike suggest that small molecules can spontaneously bind to the highly conserved hinge pocket, and that such binding can shift the RBD equilibrium to favor the open state. Stabilizing the open state may facilitate antibody recognition by forcing the spike to expose critical RBD epitopes, and also could increase the likelihood of premature triggering of the spike fusion machinery via S1 shedding, neutralizing the infectious ability of the virus.

Introduction

The emergence of COVID-19 in late 2019 sparked a global pandemic, causing > 1 million deaths as of 2019¹, and crippling the international economy. The disease is caused by the SARS-CoV-2 coronavirus, a positive-sense single-stranded RNA virus that can cause respiratory distress, pneumonia, and death. The severity of the pandemic, coupled with the globe's past history with coronavirus outbreaks, ignited a massive effort to develop effective therapeutics. A particularly promising target in the viral life cycle for

therapeutic design is the spike glycoprotein, a class I membrane fusion protein²⁻⁵ that decorates the surface of the virus.⁶⁻⁷ The goal of COVID-19 vaccines currently under development is to expose the human immune system to the spike glycoprotein prior to viral infection.⁸⁻⁹

The SARS-CoV-2 spike is a homotrimeric glycoprotein consisting of two subunits, S1 and S2 (**Figure 1A**), and is cleaved by host cell proteases at two distinct sites.^{2, 10-13} The N-terminal S1 subunits sit atop the spike and are responsible for recognizing and binding the host cell receptor angiotensin converting enzyme 2 (ACE2) and stabilizing the S2 core.^{11, 13-19} At some point after the S1 region binds to ACE2, the S1 subunits dissociate to expose the S2 core, which undergoes dramatic conformational changes to initiate membrane fusion.^{3, 20-23} Each S1 subunit consists of an N-terminal domain (NTD), a receptor binding domain (RBD), and two C-terminal domains (CTD1 and CTD2); the S1/S2 interface lies at the C-terminal end of CTD2 (**Figure 1B-C**).²⁴⁻²⁶

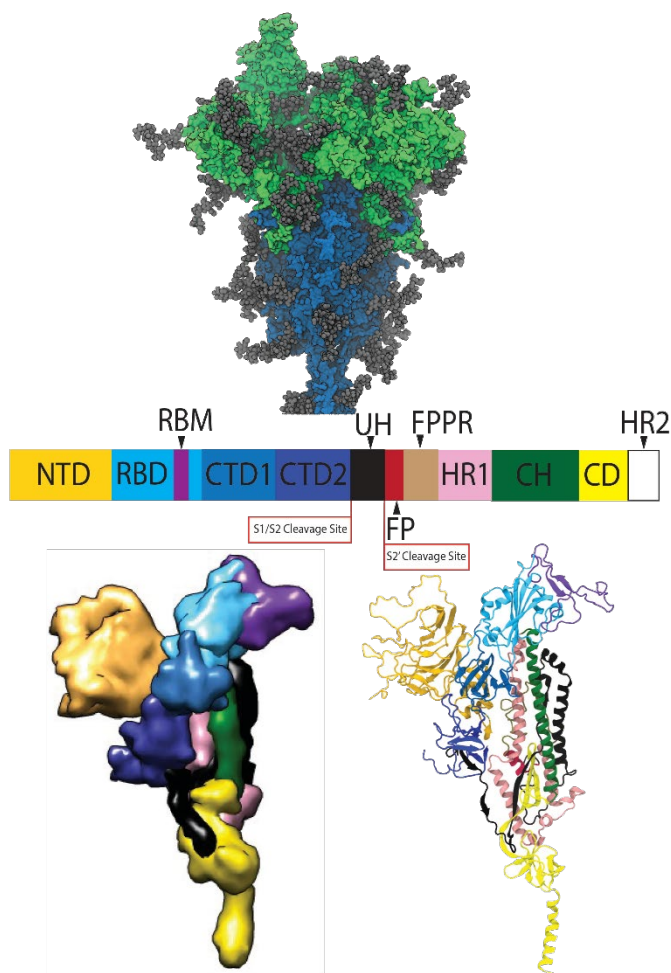


Figure 1. (A, upper) structure of the trimeric SARS-CoV-2 spike glycoprotein ectodomain from 6VXX²⁶, with the S1 subunit shown in green the S2 subunit shown in blue, and the glycans in gray. (B, middle) domain organization of the spike protomer along the protein sequence. (C, bottom) protomer structure, with colors matching the domain organization in the middle image (space-filling on the right, ribbon diagram on left).

In SARS-CoV-2, the RBD in the S1 region is responsible for recognizing and binding ACE2.^{15-19, 27} The RBD alternates between two distinct conformational states relative to the remainder of the spike: 'open' and 'closed' (**Figure 2A,B**).^{6, 25-26, 28} A two-stranded hinge region links the RBD and CTD1, and allows the RBD to shift between the two geometries.²⁹ An open RBD is a prerequisite for ACE2 binding; in the closed state binding of ACE2 is precluded by a steric clash with the RBDs of other protomers.^{26, 28-30} Another key feature of the closed state is that the RBD is shielded by the extensive glycans decorating the surface; only upon opening of the RBD does the receptor binding motif (RBM) protrude out of the glycan shield (**Figure 3**).³¹ This allows it to recognize and contact the ACE2 receptor (**Figure 2C**) -- but also makes the critical RBD residues vulnerable to neutralization by antibody binding.³¹⁻³³

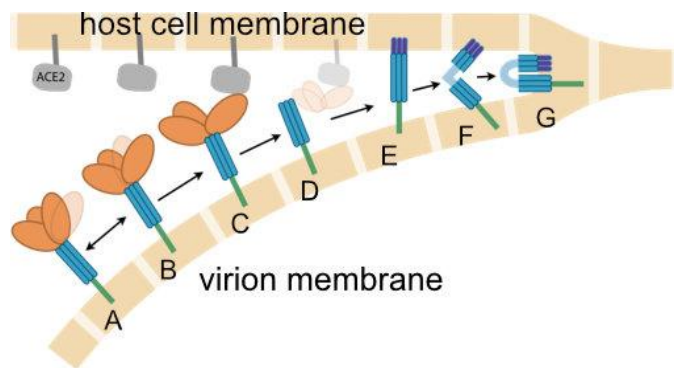


Figure 2. Cartoon illustration of the presumed role of the spike in fusion of the viral (lower beige blocks) and host cell (upper blocks) membranes. The RBD on the S1 subunit (orange) is attached to the S2 subunit (blue), and fluctuates between (A) closed and (B) open states. When the spike approaches the ACE2 receptor (gray), the open RBD is capable of binding to ACE2 (C), leading to shedding of the S1 subunit (D), insertion of fusion peptides into the host membrane (E), additional conformational changes to co-localize the membranes (F) and eventual membrane fusion (G). Double arrows indicate reversible dynamics, while single arrows indicate presumably irreversible events.

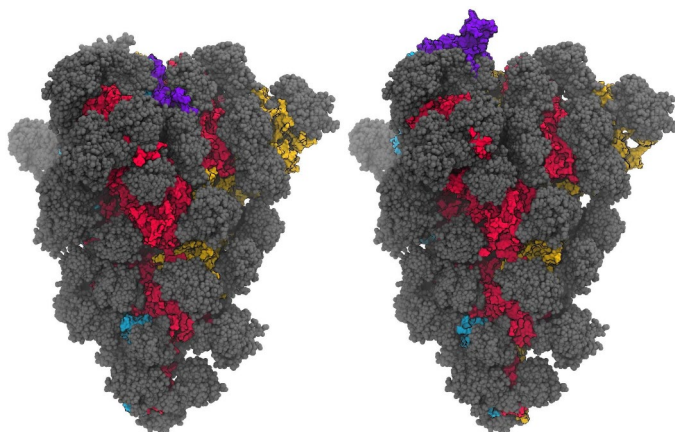


Figure 3. Structure from MD simulations of the spike glycoprotein with all 3 RBD domains closed (left), and with a single RBD in the open position (right, 1-up). A single snapshot of the protein is shown in space-filling model, with red/yellow/blue for the 3 protomers, with the RBM region of the opening RBD

shown in purple. Glycans are shown in gray, with multiple MD snapshots shown to indicate the region of the protein covered by the dynamic glycan shield.

Although the SARS and SARS-CoV-2 S proteins bind to the same host cell receptor using the same RBD region, the residues responsible for ACE2 recognition and binding are substantially different. The isolated RBD of the SARS-CoV-2 S protein binds ACE2 more strongly than that of SARS, but the complete S proteins bind with nearly the same affinity.^{13, 15, 17, 21, 34} SARS-CoV-2 S is thought to occupy the active, open RBD state less often than in the SARS S, but is compensated by the greater affinity of the RBD for ACE2.²¹ Since the RBD is largely obscured by the glycan shield in the closed state³¹, the shifted equilibrium allows the spike to spend less time in the vulnerable open conformation without sacrificing its affinity for ACE2.

While the S1 subunit is responsible for receptor binding, the S2 subunit contains the fusion machinery of the spike.³ The S2 is composed of an upstream helix (UH) at the N-terminus, followed by the fusion peptide (FP), a fusion peptide proximal region (FPPR), a conserved heptad repeat region (HR1), a long central α -helix (CH), and a connector domain (CD) that leads into the C-terminal 'stalk' which contains the second heptad repeat (HR2), and the transmembrane domain (TM) that spans the viral membrane, followed finally by an intracellular region.^{21, 25-26} Both the S1 and S2 subunits are heavily decorated with glycans.³⁵

Although the pre-fusion spike initially is stably folded, following S1 shedding (**Figure 2D**) the S2 core is metastable, and undergoes large-scale refolding that initiate membrane fusion.²⁰ S2 is thought to begin the fusion process by releasing the fusion peptides that insert into the host cell membrane, presumably forming a short-lived pre-hairpin intermediate tethered between the two membranes (**Figure 2E**).^{3, 20, 36} This structure then refolds again to adopt the post-fusion structure, with a six-helical bundle^{5, 20, 37} formed from the HR1 and HR2 helices that are attached to the spike stalk and fusion peptides, respectively (**Figure 2F,G**). This dramatic conformational rearrangement allows one or more spikes to bring the two lipid bilayers into contact, generating a fusion pore and releasing the virion's contents into the host cell.^{21, 33}

The SARS-CoV-2 spike possesses two distinct cleavage sites that must both be processed to foster efficient membrane fusion. The S1/S2 site (₆₈₂RRAR₆₈₅), contains a dibasic furin cleavage site insertion not present in SARS and many other coronaviruses.¹⁰ Cleavage at the S1/S2 site for SARS-CoV-2 is thought to occur during biosynthesis in the secretory pathway, and serves to disconnect the S1 subunit from the S2 and permit S1 dissociation (shedding). The S2' (₈₁₄KRSF₈₁₇) cleavage site^{2, 5} is in the S2 domain and is believed to be cleaved by membrane-bound TMPRSS2^{5, 38-40} or cathepsin^{5, 41} host proteases, allowing for release of the fusion peptides at the end of HR1 that are subsequently inserted into the host membrane. The timing of S2 cleavage relative to ACE2 binding and S1 shedding remains unclear.

Studies of the wild-type (WT) cleaved spike for both SARS-CoV and SARS-CoV-2, either isolated or on viral particles, observe detectable amounts of post-fusion S2 structures even in the absence of ACE2.^{6-7, 20-21, 23, 29, 42-43} These results suggest that the S1 subunit can spontaneously shed and prematurely trigger the irreversible S2 transition to the post-fusion state. Binding of antibodies can also trigger premature spike activation.³³

Our focus here is on the dynamics of the RBD, through which it transitions from the hidden and binding-incompetent closed position, to the open conformation in which it can bind ACE2 but is also susceptible to immune surveillance. Locking the spike in either the open or closed conformation could

potentially interfere with the viral entry into host cells. Sealing it in the closed state would eliminate its ability to bind ACE2, preventing infection. This approach has been demonstrated through introduction of disulfide bonds that lock the RBD closed, preventing ACE2 binding.⁴⁴⁻⁴⁷ Small molecules may also be able to reduce RBD opening for wild-type spike, as suggested by the identification of a linoleic acid binding pocket in the RBD of the SARS-CoV-2 spike.⁴⁸⁻⁴⁹ Conversely, substitutions that increase population of the open RBD⁴⁷ may more frequently expose antibody epitopes, potentially facilitating a neutralizing immune response⁵⁰⁻⁵¹. Shifting the distribution to the open state also could weaken the S1-S2 interface and help to prematurely trigger S1 shedding, irreversibly neutralizing the spike^{6, 16, 21}.

Rational design of small molecules that can penetrate the glycan shield and serve as probes or modulators of RBD positioning would be facilitated by a complete, atomically-detailed model of the RBD dynamics. Here, we use a wide array of computational methods to model a conformation transition pathway and free energy landscape between the all-closed and 1-up RBD states in the fully glycosylated, solvated spike models based on prior simulations³¹, cryo-EM²⁵⁻²⁶ and crystal¹⁹ structures, and mass spectroscopy experiments³⁵. The data provide a detailed description of RBD opening, indicating specific interactions that may stabilize the closed, hidden state, and provide opportunities to control RBD positioning. The results are briefly summarized here.

Analysis of the RBD transition profile identified formation of a transient pocket at the hinge region beneath the RBD (**Figure 4**). This pocket is present only in the open RBD state, and is formed by the bottom of the RBD, the top of the CTD1 domain, and the inner surface of the two protein strands connecting these domains. Importantly, the amino acids composing the hinge-pocket are well-conserved across other betacoronaviruses, despite substantial variations in the RBD and rest of S1. This is consistent with an important mechanistic role for the hinge region.

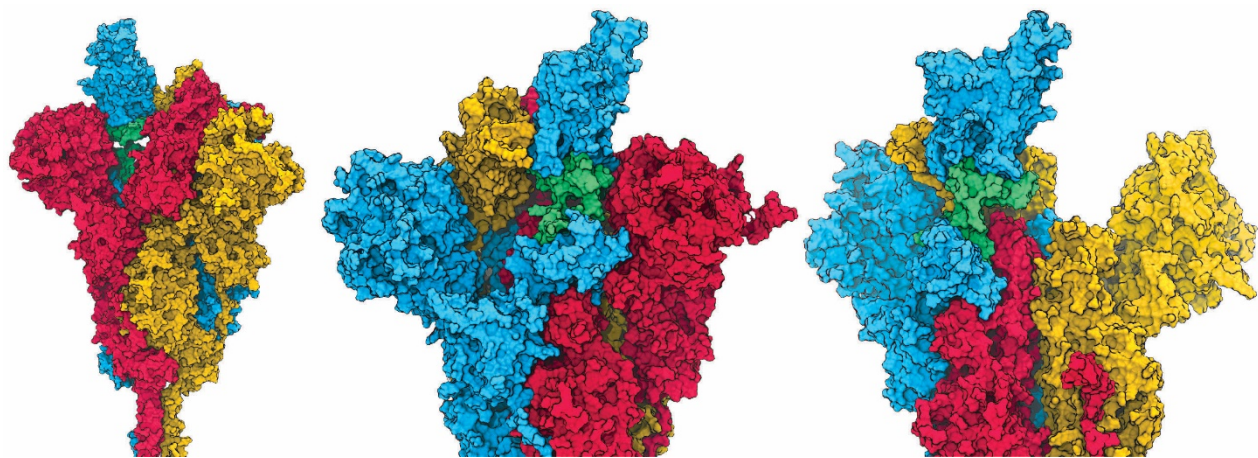


Figure 4. Space-filling model of the 1-up spike trimer, colored blue for protomer with open RBD, and red and yellow for the counterclockwise and clockwise protomers with closed RBD, respectively. Glycans and water are omitted. The RBD hinge region is shown in green. Three different views are shown: (left) pocket accessibility through gap left by opened RBD; (middle) narrow tunnel between hinge connectors; and (right, red S1 subunit not shown) concave pocket between RBD and CTD1 domains.

We examined specific interactions near the hinge region that influence RBD positioning. Consistent with experiments,^{7,21} the closed RBD is favored on free energy landscapes of the wild-type spike. The conserved K528 forms a salt bridge with D389 only when the RBD is closed,²⁶ and simulation of K528A shifts the RBD equilibrium toward the open, exposed state. The pocket is tightly packed when the RBD is closed; substitution of A522 with bulkier Leu also remodels the simulated free energy landscape to favor the open, exposed RBD. These results suggest that binding of a small molecule that makes appropriate contacts in the allosteric hinge pocket could preferentially stabilize the spike in the open conformation, potentially counteracting the immune shielding of the closed RBD and also weakening the interaction between the S1 and S2 subunits.

We conducted a virtual screen using a large library of known compounds to determine if the pocket was large enough to accommodate a small molecule and identified several promising candidates that were studied while bound to the spike during MD simulations. The most promising candidates was used for calculations of the free energy landscape for RBD opening in the full trimeric spike; these support the hypothesis that a bound ligand could dramatically shift the equilibrium of the RBD from strongly favoring the closed position in the apo spike, to a flatter landscape with significant population of the open RBD with a ligand bound at the hinge region. The results suggest that further optimized ligands may be able to provide precise control over RBD position, prevent or reduce shielding of the antigenic epitopes on the RBD, and perhaps amplify the existing tendency of the spike to undergo irreversible inactivation via premature shedding of the S1 subunit.

Results and Discussion

MD simulations of the spike in closed and 1-up states

Following addition of missing loops and glycans, followed by solvation, the complete spike ectodomain system consisted of ~ 1.3 million atoms (see Methods for details on construction of the initial structures for each state). We performed three independent 300 ns MD simulations of the spike with all RBDs closed, as well as three independent runs of the 1-up spike (with only one RBD in the open position; in this work we refer to an individual RBD as being open or closed, while the spike trimer can be in the closed, 1-up, 2-up or 3-up state).

We first confirmed that the spike was reasonably stable in our simulation model. We show in **Figure 5** the best-fit backbone atom RMSD of the S1 and S2 subunits, during our 1-up and closed simulations, against their respective cryo-EM reference structures (see “Reference Models” in Methods). The simulations of the closed spike were largely stable, with each S2 subunit showing low deviations of ~ 2 Å, and the more flexible S1 subunit sampling RMSD values of ~ 3 Å. Although the 1-up simulations are also stable, with similar S2 RMSD values, the additional space from reduced packing of the RBD leads to a more flexible S1 subunit with larger fluctuations and average RMSD values near 4-5 Å. These observations are consistent with prior simulations reports and lower resolution in the cryo-EM datasets for open RBDs.^{25, 31} Overall, the spike behavior appears reasonable in our model.

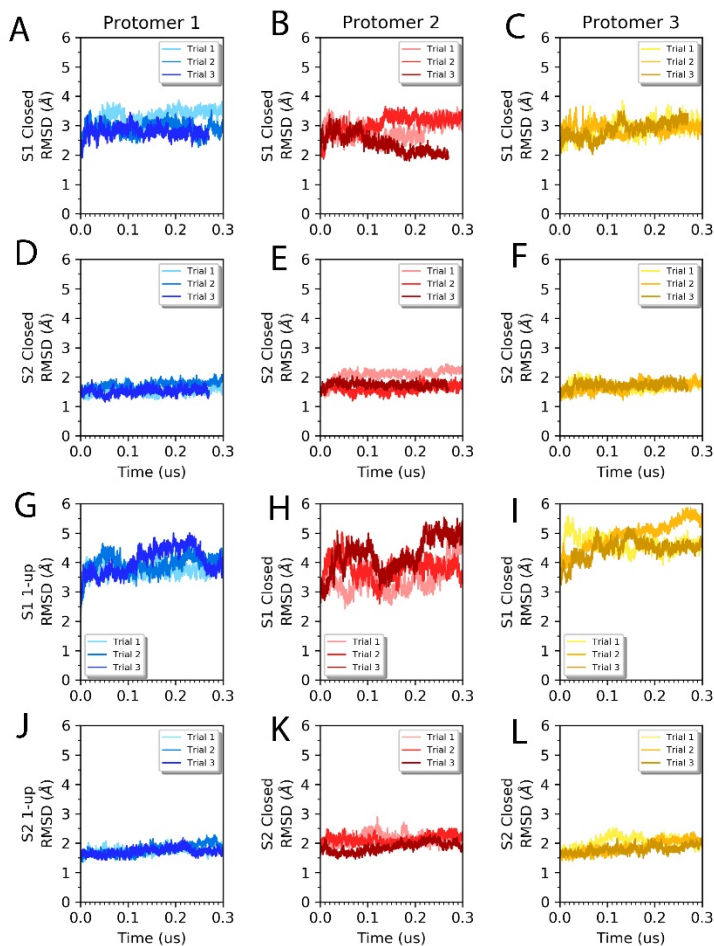


Figure 5. RMSD of backbone atoms of each protomer (blue, red, yellow) over three MD runs of the spike over 300 ns. (A,B,C) S1, closed spike system; (D,E,F) S2, closed; (G,H,I) S1, 1-up; (J,K,L) S2, 1-up. Each shade in a plot represents one of three independent trials. Columns represent protomers 1, 2 and 3 with colors matching **Figure 4**.

We next examined the behavior of individual regions of the spike in the closed and 1-up simulations (**Figures S6, S7**). For the closed system, the domains tended to be quite stable with most sampling RMSD values between 1-2 Å. Similar domain stability is seen in the 1-up simulations, with larger changes in the NTD domains that become more loosely constrained when the packing at the top of the spike is reduced. The CTD2 domain in one protomer shows higher deviation in both systems, likely due to an inaccurately modeled surface loop in the initial structures. The cleaved S1/S2 site is highly mobile in both the closed and 1-up simulations (data not shown), consistent with it typically being unresolved in cryo-EM experiments. In contrast, the uncleaved S2' site is highly stable and samples RMSD values of only ~ 0.5 Å in all simulations, with R815 in the KRSF segment remaining partially buried and presumably inaccessible to proteases. This is consistent with inaccessibility of S2' to proteases in pre-fusion SARS spikes.⁵² The fusion peptide regions proximal to the S2' site are similarly stable, with small changes occurring in some trials, but the results are anecdotal and were not subjected to detailed analysis.

Quantifying RBD positioning

In addition to RMSD values, we characterized the extent of RBD opening using two collective variables (CVs) that describe the opening angle of the RBD and rotation of the RBD in the plane roughly perpendicular to the central helices (see Methods for angle definitions). All RBDs in the 1-up and closed states remained in their respective conformations for the entirety of the 0.3 μ s simulations, with no spontaneous transitions between open and closed RBDs (**Figure 6**, with 2 additional independent runs shown in **Figure S8**). This metastable RBD behavior on the sub- μ sec timescale is consistent with the behavior observed by Casalino et al. in 4 μ sec simulations.³¹ In the 1-up system, the two closed RBDs sample different angle ranges due to their different positions relative to the cavity left by the open RBD.

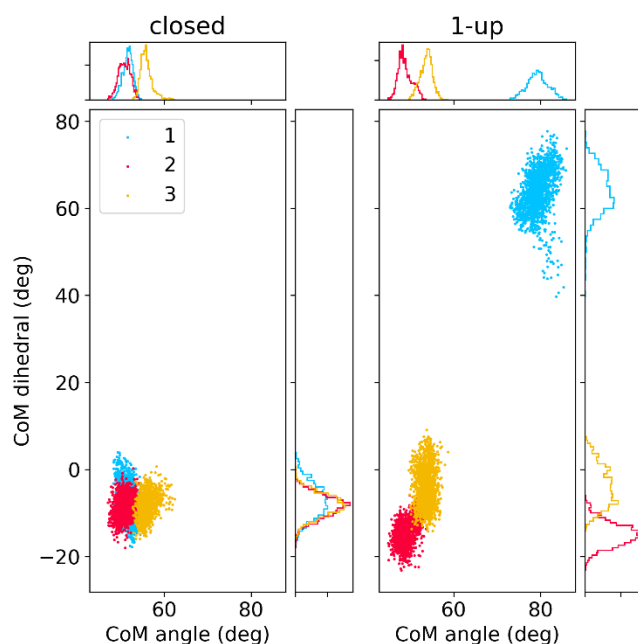


Figure 6. Collective variables quantifying RBD position in MD simulations of closed (left) and 1-up (right) spike systems. Dots represent MD snapshots, and each protomer is shown in a different color corresponding to Figure 4.

Glycans

Behavior of the glycans in the SARS-CoV-2 spike was previously described by Casalino et al., who proposed that the N-glycans at N165 and N234 in the NTD play an essential structural role in RBD conformational transitions.³¹ In our closed simulations, the glycans on N234 point towards the solvent, rather than the core of the protein. In the 1-up simulations, the glycans on N234 of the adjacent NTD direct inward, occupying the vacancy left by the RBD and interacting with the closed, opposing RBD. During the same simulations, the glycans at N165 interacted with various residues on the open RBD and inserted itself into the vacancy with those at N234. These observations are consistent with those reported by Casalino et al. using multiple force fields and MD software packages.³¹

Mapping the RBD opening pathway

Since brute-force MD simulations were unable to capture spontaneous opening or closing of the RBD during the 300 ns runs, we used a nudged elastic band (NEB) approach⁵³ to optimize a low-energy pathway between the spike in the closed and 1-up RBD states (see Methods). Briefly, NEB uses a set of multiple simulations (“beads”) that connect two fixed conformational endpoints. Virtual springs ensure that the beads remain spaced apart along the path, but the intermediate beads are otherwise free to explore changes in the positions of all atoms. An annealing optimization provides a model for a low-energy pathway connecting the defined conformations of the endpoints.

The NEB pathway maps the RBD transition in full Cartesian space, but to facilitate visual analysis we project the resulting dynamics onto the two CVs that quantify the RBD position; these are shown as a function of progress along the closed-to-open RBD transition pathway in **Figure 7**, along with overlaid structure snapshots of a single protomer during the transition shown in **Figure 8**. At roughly 15% progress along the pathway, the CoM angle rises, indicating the start of RBD opening. Changes earlier in the path correspond to modest adjustments in the NTD as it shifts away from the RBD to provide clearance for opening. Next, the RBD lifts up out of the cavity formed by the NTD and RBD of the counterclockwise protomer, giving rise to a sharper increase in the CoM angle value until ~ 60% progress along the path. As the RBD clears the cavity it rotates outward, quantified by the increase in the CoM dihedral starting at roughly 30% progress.

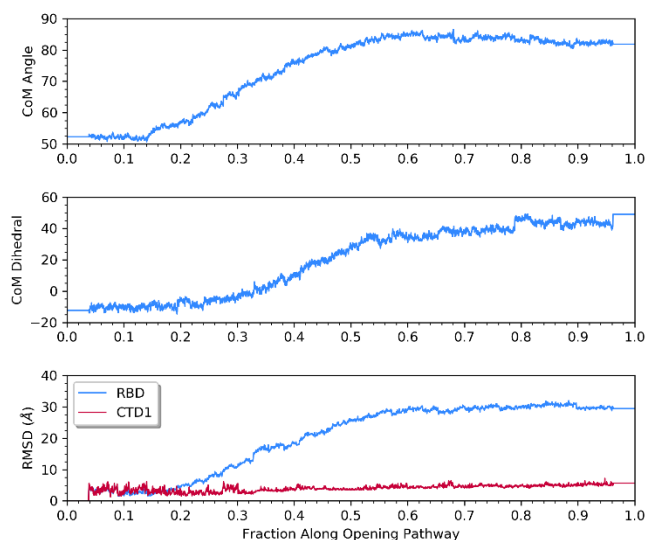


Figure 7. Spike structure properties during the RBD opening pathway, with the closed state on the left and the 1-up spike on the right. (Top) RBD opening angle, (middle) RBD rotation dihedral, (bottom) RMSD values for RBD and CTD1 domains, after best-fit to central helices (CH).

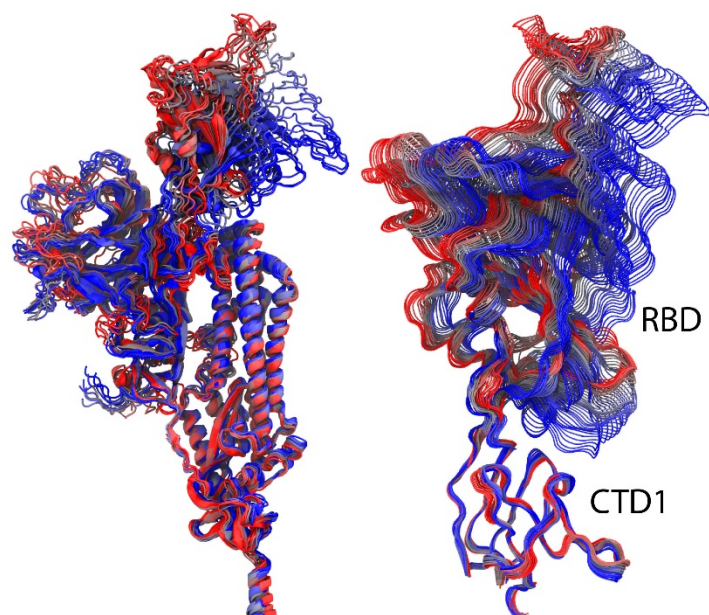


Figure 8. Spike structure snapshots during the RBD opening pathway, colored by opening pathway progress from blue to red. (Left) Structure of a single protomer, best-fit to central helices to emphasize the larger change in RBD as compared to CTD1 domain; (right) close-up of the pocket formed between the RBD and CTD1, best-fit to CTD1 to emphasize the hinge motion between these domains.

RBD opening is achieved through a hinge

As the RBD lifts it moves away from the rest of the spike, leading to an RMSD increase to ~ 30 Å relative to its position in the closed state (**Figures 7 and 8**, best-fit to the S2 core CH helices). The attached CTD1 domain also shifts, but remains much closer to the position in the closed RBD state, deviating less than 5 Å over the pathway. These values largely reflect rigid-body shifts in the individual domains, with much smaller RMSD values of $\sim 1-2$ Å when these domains are self-fit (**Figure S9**), likely due to the presence of multiple disulfide bonds proximal to the hinge pocket (**Table S2**). Thus the most significant change in RBD opening is in the relative position of the RBD and CTD1 domains, with a smaller outward shift of CTD1, consistent with experiments²⁵⁻²⁶.

The hinge connecting the RBD to the CTD1 is composed of two antiparallel strands, one connecting the C-terminal end of the first β -strand in CTD1 to the N-terminal end of the RBD (“N-connector”), and the other connecting the C-terminal of the RBD back to form the remainder of CTD1 (“C-connector”). Both connectors are located on one side of the domain interface, on the exterior of the spike, allowing it to serve as a hinge around which the RBD can rotate up and outward (**Figure 7**). As hinges, both connectors undergo local conformational changes as the spike RBD opens, with backbone dihedral changes present in I332, T333 and P527 (**Figure S10**).

Specific interactions in the hinge region may stabilize the closed RBD

In the all-closed cryo-EM structure 6VXX²⁶, a salt bridge is apparent between K528 on the C-connector and D389 on the RBD α -3 helix. This salt bridge is also present at the beginning of the RBD opening pathway, but the interaction is broken as the RBD moves upward and away from CTD1 (**Figure**

9A, with distance vs. pathway shown in **Figure S11**). The absence of this salt bridge with an open RBD supports the hypothesis that the hinge pocket region modulates RBD opening. Cryo-EM structures also suggest increased hinge flexibility when the RBD is open; many experimental structures are missing density in the hinge region, including K528 and the entire N-connector in the initial model for our 1-up spike system (6VSB).²⁵

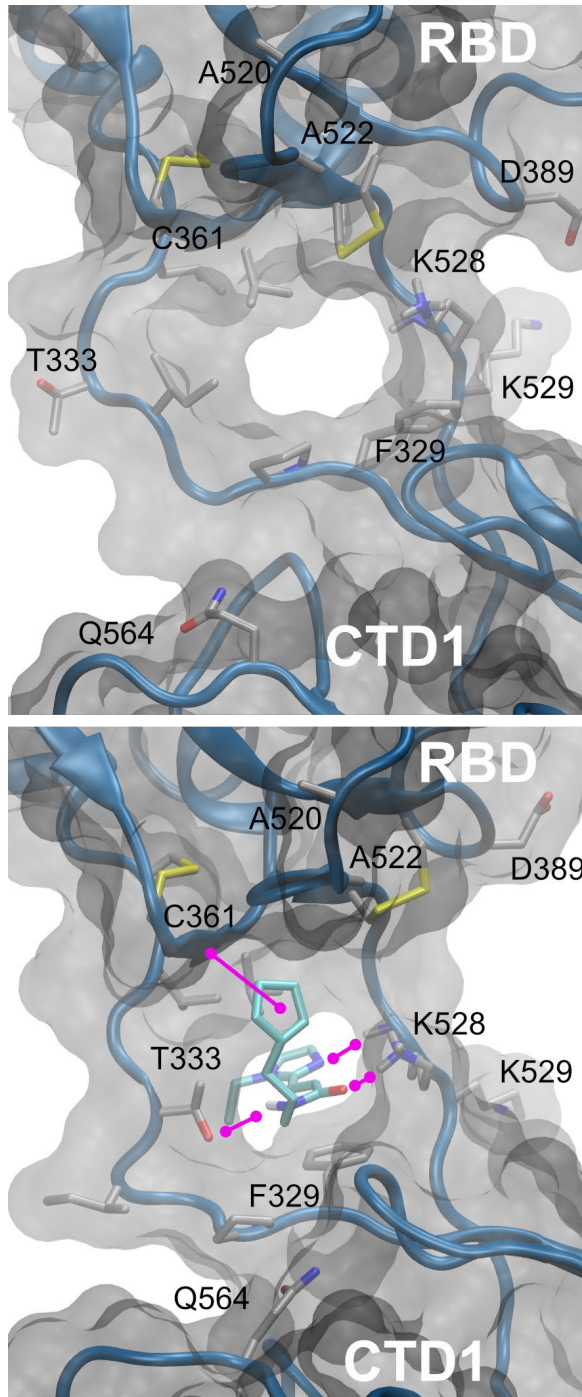


Figure 9. Hinge pocket and tunnel in simulated open RBD structure used for virtual screening (A, upper) and after MD simulation of the upper structure in complex with ligand Z158703301 (B, lower). Pocket is viewed from inside the spike. Sidechains discussed in the text are labeled, and key interactions shown in magenta. For clarity the solvent, most sidechains and hydrogen atoms are not shown.

The hinge region is strongly conserved both within identified SARS-CoV-2 isolates, as well as other betacoronaviruses (**Figure S12, Table S6**), supporting a mechanistic function for this region. In particular, the salt-bridge pair D389 and K528 are both well conserved, while the immediately adjacent K529 is poorly conserved; this difference is consistent with our suggested role for K528 in stabilizing the RBD:CTD1 interface when the RBD is closed and the lack of specific interactions observed in MD for K529 (**Figure 9**). N331 on the C-connector and nearby N343 on the RBD are both glycosylated, with the glycans extending into solution during MD and perhaps shielding the conserved hinge region from antibody recognition (**Figure S13**).

RBD opening via the hinge opens a pocket between RBD and CTD1

In the RBD-closed state, both hinge strands typically are resolved in cryo-EM structures, and the sulfur atoms of RBD disulfide C391-C525 dock into the hydrophobic space between the connector strands. The two domains are tightly packed, with no gaps in the solvent-accessible surface. A520 on the RBD and Q564 on CTD1 are in close contact in the interior of the hinge pocket when the RBD is closed, but move apart to ~ 27 Å as the RBD opens (**Figure S14A**). The disulfide on the RBD base moves upward, leaving a short tunnel between the connector strands and creating a pocket in the expanding space at the former interface between the RBD and CTD1 (**Figures 1, 8, 9**). SASA analysis indicates that the pocket becomes solvent accessible halfway along the opening pathway, reaching a maximum value of 600 Å² (**Figure S15**).

This transient but well-conserved hinge region provides an appealing target for possible intervention in the spike function by modulating the intrinsic open/closed RBD equilibrium. To explore the druggability of the hinge pocket, we used fpocket⁵⁴ to extract descriptors across the opening RBD pathway (see Methods). The volume of the pocket ranged from 0 (closed RBD) to 900 Å³, with a median of 579 Å³ when the RBD is open ($> 45\%$ RBD opening progress, **Figure S15**), comparable to a median of 530 Å³ from a survey⁵⁵ of druggable pocket volumes. Transient decreases in the pocket volume of the pocket near 70% and 80% progress correspond to χ 1 rotation of F329 that temporarily positioned the aromatic ring inside the hinge pocket.

Next, we used fpocket to estimate the potential druggability of the open hinge pocket. The druggability score ranges from 0 (unlikely) to 1 (very likely) and is dependent on overlapping apolar α -spheres, and an amino acid specific hydrophobic and polarity score.⁵⁴ A survey of known small molecule binding pockets in protein crystal structures yielded druggability scores ranging from 0.5 to 1.⁵⁶ Application to the spike hinge pocket showed reasonable druggability scores during opening, with many snapshots scoring above 0.5 (**Figure S15**). Pocket druggability is expected to vary as amino acids rotate in and out of the pocket, and as the connector strand conformations vary. For example, the transient rotation of F329 is associated with reduced druggability (**Figure S15**). Overall, the fpocket analysis suggests that the hinge pocket can adopt conformations with a volume and druggability score consistent with known druggable binding pockets.

Free energy landscapes for RBD opening in wild-type and substituted spike systems

Since it is highly challenging to sample RBD dynamics in a statistically meaningful way for the full spike system, we calculated free energy landscapes in the form of potentials of mean force (PMFs) for the RBD closed/open transition in the complete, glycosylated spike. Uncertainties in our structure (such as model quality for flexible loops) likely prevent us from obtaining accurate quantitative free energy values for the open-to-closed RBD transition, but the qualitative changes resulting from hinge pocket amino acid substitutions or small molecule binding are likely to be qualitatively informative due to cancellation of systematic errors.

The WT spike shows a strong preference for the closed RBD state (**Figure 10A**), consistent with cryo-EM experiments^{6, 21} on the wild-type system, as compared to more frequent observation of open RBD in experiments using the common “2P” stabilizing substitutions^{52, 57} that also may weaken²¹ the interaction between the closed RBD and the S2 subunit. The PMFs suggest that RBD opening involves three states: the closed state, open state, and a more diffuse, weakly populated intermediate state. The intermediate state (CoM angle 65-70°; CoM dihedral 18-26°) is stabilized in part by a transient salt bridge between K386 on the RBD and D985 at the top of CH. In the closed and intermediate state, K528 helps stabilize the RBD by forming a network of electrostatic interactions involving D389-K386-S383 to D985 at the top of CH in the S2 subunit. Further opening results in breaking of the K528-D389 salt bridge. In the open state, a cluster of salt bridges forms between D427/D428 on the RBD and R403/R408 on the closed clockwise RBD.

Focusing on the hinge pocket, we estimated the impact of deleting the salt bridge involving K528 and D389 via simulations of the K528A spike. Consistent with our structure-based hypothesis, the K528A substitution significantly destabilizes the closed RBD, and the free energy surface is shifted toward easier RBD opening (**Figure 10B**). The intermediate state is flatter, likely due to increased flexibility of D389-K386-S383 in the absence of K528.

A shift in the open RBD position is seen in a cryo-EM structure⁴⁷ of the spike with amino acid substitutions at the CTD2:S2 interface. We hypothesized that substitutions in the hinge pocket may also affect RBD dynamics. The small, nonpolar A522 is tightly packed into the hinge pocket when the RBD is closed, but becomes exposed when the RBD opens (**Figure 9A**). Replacement of A522 with a bulkier side chain such as Leu could be readily accommodated with an open RBD, but introduce a steric clash in the tightly packed RBD:CTD1 interface in the closed-RBD structure. The simulated free energy landscape for the A522L substitution supports our hypothesis, and is strongly shifted to favoring the intermediate and open RBD (**Figure 10C**). The minimum for the closed RBD is shifted to higher angles, consistent with a steric clash preventing full closing, and the poorly packed interface destabilizes the closed RBD. These A522L results suggest that introducing new chemical matter into the hinge pocket interface could disrupt the ability of the spike to adopt a closed-RBD state to evade immune surveillance.

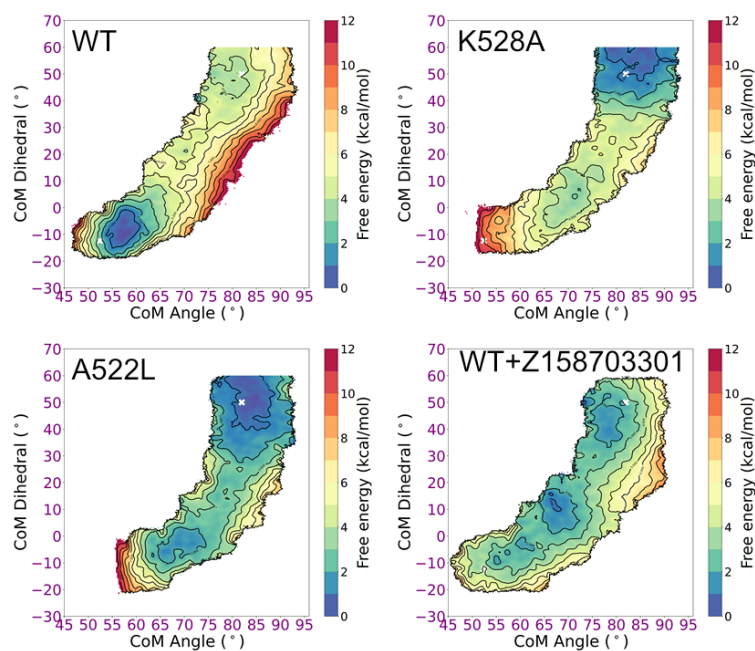


Figure 10: Free energy landscapes for RBD opening in A) WT, B) K528A, C) A522L, D) WT+ Z158703301. Contour lines indicate 1 kcal/mol intervals. WT has a global minimum with closed RBD, K528A prefers an open RBD, and both systems with additional chemical groups in the hinge pocket (WT with bound ligand, A522L) show significant flattening of the free energy landscape.

Virtual screening of small molecules targeting the hinge pocket

Although these proof-of-principle sequence substitutions support the proposed allosteric nature of the hinge region, it may be preferable to alter RBD dynamics without modification of the viral genome. The location of the hinge pocket at the base of the RBD, facing the interior of the spike, makes it unlikely that it could be accessible for antibody or nanobody binding. In particular, the pocket is ~ 20 Å farther down the RBD than a cryptic epitope observed in the crystal structure⁵⁸ of a complex between the CR3022 antibody and isolated RBD (**Figure S16**). However, the pocket appears more readily accessible to small molecules diffusing from the outside of the protein, through a 20+ Å gap between the NTD of the clockwise protomer and CH of the counterclockwise protomer; this space was previously occupied by the closed RBD (**Figure 1**). Small molecules also are more likely to bypass the glycan shield.

We hypothesized that the transition from an open-to-closed state could be blocked by a small molecule docked to an open-like state, making intermolecular interactions similar to those that would normally be present between RBD and CTD1 domains when in the closed-like state. Specifically, we aimed to replace key interactions lost upon RBD opening, including (1) occupying the tunnel between the hinge connectors, perhaps forming hydrogen bonds to the protein backbone, (2) engaging the protonated sidechain of K528 to compensate for the lost salt bridge to D389, and (3) extending a steric “prop” to support the base of the RBD near C361 and keep it from closing onto the CTD1 (**Figure 9A**).

We carried out virtual screening using DOCK6,⁵⁹ with a setup that followed well-tested protocols we previously employed to identify small molecule inhibitors through targeting HIV gp41,⁶⁰⁻⁶² fatty acid binding proteins,⁶³⁻⁶⁴ botulinum neurotoxins,⁶⁵⁻⁶⁶ HER2,⁶⁷ Ebola GP2,⁶⁸ Zika glycoprotein E⁶⁹. We screened

libraries of purchasable molecules downloadable from the ZINC database⁷⁰. Compounds were scored based on footprint similarity scores⁷¹⁻⁷², initially using residues at the base of the RBD as a reference, then supplemented with energy-minimized interactions of the spike with small molecules from an initial screen. The footprint was manually modified to emphasize interactions of interest (see Methods for additional details). Compounds from the first round of screening were not stably bound during MD, but were used to develop a hybrid footprint for additional screening that resulted in six lead compounds (**Table S8 and Figure S21**, with additional details for compound selection and analyses provided in Supplementary Information).

MD using hits from virtual screening show reversible ligand binding and allosteric RBD control

Since the hinge pocket is transient, it is unclear (1) if the presence of a small molecule ligand would be sufficient to shift the equilibrium of the RBD opening, and (2) if a docked ligand could stay bound against the pressure of the closing RBD. Since MD simulations of the entire spike are computationally demanding, we carried out initial testing using a reduced construct of the hinge region, consisting of a single protein chain including only the RBD, CTD1 and connector chains from one protomer (576 amino acids, Q321-S591, see Methods and **Figure S4**). To assess suitability as a model system, MD simulations for this 15x smaller system were initially carried out in the absence of a docked ligand, starting from the open conformation used for virtual screening. Despite the removal of the surrounding protomers, both domains in the smaller construct were stable during MD simulations of 2 μ sec, with domain backbone RMSD values near 2-3 Å (**Figure S9**). The hinge pocket behavior from the NEB trajectory of the full spike also was recapitulated in unrestrained MD for the smaller fragment, sampling a similar distance range between C α atoms of A520 and Q564 with several opening/closing events and a preference for the closed state of the RBD (**Figure S14**).

We next carried out fully unrestrained simulations for this RBD+CTD1 construct with six different docked ligands (**Figure S21**). Three of the simulations resulted in substantial shifts in the ligand pose within 50-200 ns, with two of these showing closing of the RBD as was observed for the unbound construct (**Figure S17**).

The simulation with ligand Z739327691 showed remarkable behavior. In the docked pose, it exhibited good VDW packing in the cleft, but the only protein-ligand hydrogen bond present was between the side chain of K528 and the charged carboxyl group of the ligand (**Figure S21**). After \sim 175 ns of MD, the ligand dissociated and diffused into solution, reaching as far as \sim 60 Å from the hinge pocket before re-approaching and autonomously inserting into the tunnel between the connector strands. Spontaneous protein-ligand binding has been reported in previous MD simulations.⁷³ The ligand then formed a hydrogen bond to the backbone carbonyl of K528, where it remained for \sim 30 ns before dissociating into solution once again. This observation suggests that spontaneous insertion of a small molecule between the strands connecting the RBD and CTD1 domains may be a reasonable event, and that hydrogen bonds between the ligand and the connector backbone may be important for binding.

The ligand Z902245578 remained in the pocket during the simulation, but the docked pose for did not exhibit hydrogen bonds to the tunnel connectors, and it shifted position several times and did not adopt any stable hydrogen bonds to the protein. The ligand Z153687312 also remained bound, but lost the initial hydrogen bond to the backbone of K528 (instead forming one with the backbone of P320 on the N-connector) and did not make a direct contact with the K528 side chain during the simulation.

The remaining ligand, Z158703301, exhibited all of the desired features in the docked pose, including hydrogen bonds to both the K528 backbone and side chain, and VDW contacts with the RBD side of the pocket (**Figure 9B**). It remained stably bound for ~ 0.5 μsec of unrestrained MD (**Figure S18**). In contrast to the unbound RBD+CTD1 simulations (**Figure S14**), the A520 - Q564 distance did not fall below 10 Å at any point, and showed several transitions where the RBD opened even further than in the unbound version, sampling distances over 20Å before attempting unsuccessfully to close, and then opening again (**Figure S18**). The key protein-ligand contacts observed in the docked pose were largely stable. These included the hydrogen bond to the backbone of K528, the packing of the cyclopentane ring against the base of the RBD, and the hydrogen bond to the N ζ of K528 (indicated via magenta lines in **Figure 9B**, with time traces shown in **Figure S18**). New contacts were also formed; the aromatic ring of F329 repositioned to stack underneath the ligand, and an adjustment in the N-connector after $\sim 100\text{ns}$ MD allowed formation of an additional ligand hydrogen bond to the side chain O γ of T333 (**Figure 9B**). Although T333 initially pointed to the outside of the pocket in the structure we had selected for virtual screening (**Figure 9A**), the adjustment appears reasonable since the “T333-in” conformation is observed in other structures sampled along the RBD opening pathway, as well as in experimental models such as PDB ID 7A98¹⁶ in which the open RBD is bound to ACE2. During MD, it was observed that the ligand appeared unable to adopt strong hydrogen bonds simultaneously to both T333 O γ and K528 N ζ ; these interactions alternated on the ~ 100 ns timescale. This suggests possible avenues for future optimization of ligand geometry.

Two independent simulations of ~ 0.5 μsec were performed for the same complex with Z158703301 (**Figure S18**). In the second run, the ligand again stayed bound in the pocket and the RBD remained open, with behavior comparable to that described above, including formation of the additional hydrogen bond to T333. In the third run, the ligand dissociated early in the simulation, which was followed by collapse of the hinge pocket and closing of the RBD. A possible explanation for the difference is that the hydrogen bond to T333 did not form in this run. This dissociation event indicates that the long residence of the ligand in the pocket seen in the other two runs is not simply a result of the ligand being trapped in the docked pose.

Overall, these simulations of the RBD+CTD1 construct suggest that a small molecule may indeed fit into the hinge pocket, with multiple stable interactions on the μsec timescale, and that, while bound, it can prevent the RBD from closing. The allosteric role of the ligand is supported by the ligand-bound simulation in which the RBD closed shortly after ligand dissociation, in contrast to a reproducibly stable open RBD with bound ligand.

The bound ligand remodels the free energy landscape to favor the open RBD

Although the RBD+CTD1 construct appears to be a reasonable model for the hinge pocket, we also tested the docked ligand in a 380 ns simulation of the entire glycosylated spike trimer. Control simulations of a 3-up unbound spike model showed that all 3 RBDs remained open (**Figure 11A**). These RBDs exhibited larger fluctuations in the angle and especially dihedral space as compared to the 1-up simulation (**Figure 6**), and sampled larger opening angles than in the closed to 1-up NEB path (**Figure 7**), likely due to reduced constraints from neighboring RBDs in the 3-up spike.

Following placement of Z158703301 into each hinge pocket of the 3-up spike trimer (see Methods), MD simulation of the same length showed that all 3 RBDs again remained open, but with

reduced fluctuations as compared to the unbound 3-up spike (**Figure 11B**). With the ligands bound to the hinge pocket, the RBDs were restricted closer to the region explored in the 1-up model, perhaps because the ligand was selected from virtual screening against the pocket geometry obtained from 1-up simulations. This suggests that binding of appropriate small molecule probes to the flexible hinge pocket may help to reduce RBD motion and aid in structure determination for open-RBD states, and that rational modification of the ligand size could allow for specific positioning of the RBD.

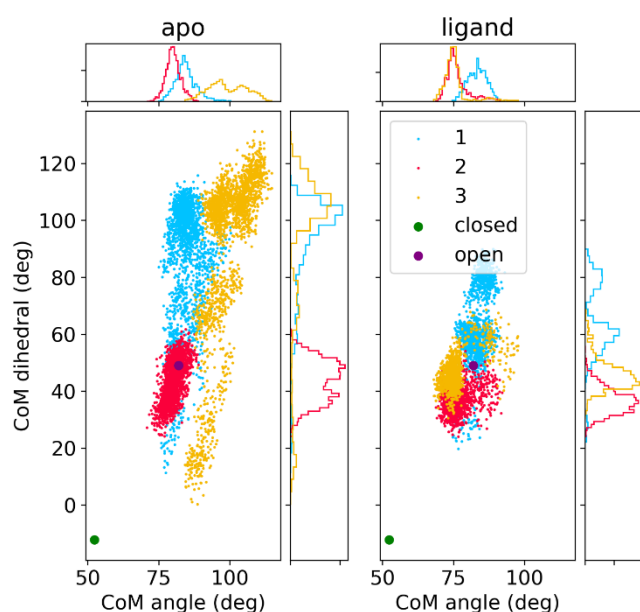


Figure 11. Scatter plot of RBD angle and dihedral measurements of apo 3-up model (left) and ligand bound 3-up model (right) from snapshots during unrestrained MD simulations. Blue, red and yellow dots represent the 3 protomers. Values at the NEB endpoints are shown as larger green and purple circles.

The bound ligand shifts the free energy landscape to favor open RBD

As with the unbound spike simulations, standard MD timescales are too short to probe the potential impact of the bound ligand on RBD equilibrium positioning. We repeated the free energy landscape calculation for RBD closing in the 1-up spike system (see **Figure 10A**), but with the Z158703301 ligand bound to the hinge pocket (**Figure 10D**). Initial calculations were carried out without any restraints on the ligand, however the process of artificially forcing the RBD to sample the completely closed state caused a shift in the ligand pose to relieve the steric clash with the sulfur of C525 on the RBD, producing an artificial minimum. Since our goal was to probe the impact on the RBD of a bound ligand, we used restraints between the ligand and K528 to help retain the ligand in the hinge pocket as the RBD was forcibly closed (see Methods).

As expected, the presence of the ligand in the hinge pocket had a similar impact on the RBD energy landscape as the A522L substitution, significantly flattening the opening profile and stabilizing the open RBD, consistent with the design goal (**Figure 10D**). With the ligand in the hinge pocket, the K528-D389 salt bridge does not form when the RBD closes. Furthermore, the presence of the ligand in the

hinge pocket shifts the opening RBD further toward the clockwise protomer, strengthening interactions seen in the intermediate (K386-D985, K378-D405) and open (D427-R403, D428-R408) RBD states (**Figure S19**).

These simple ligands identified through virtual screening of purchasable compounds almost certainly could be further optimized for improved affinity; for example, the highly conserved K528 forms a stable salt bridge with D389 in the closed-RBD state, but the Z158703301 ligand lacks an acidic group near K528, and instead formed a hydrogen bond with K528 that was only present in ~ 50% of the frames. Refining the ability of the ligand to strongly engage the highly conserved K528, and simultaneously hydrogen bond to T333, could further improve the ability to remain bound to the pocket and stabilize the RBD in the open state.

Conclusions

Experimental structures have provided a wealth of data on most of the spike structure, but less is known about flexible regions, and especially the detailed mechanisms by which the spike transitions between the observed states of the RBD. We performed computational analysis of the fully atomic detail SARS-CoV-2 spike glycoprotein dynamics in explicit solvent, using long standard MD simulations, steered MD, RBD opening pathway mapping, and umbrella sampling calculations of the associated free energy landscape. We identified a conserved allosteric pocket adjacent to a hinge region that is critical for the RBD opening and closing motion. This pocket exists when the RBD is in the open state, and vanishes when the RBD closes and contacts the CTD1 domain.

Reasonable druggability scores for the hinge pocket with an open RBD motivated virtual screening that provided several potential allosteric ligands. We developed a simple screening method to identify potential ligands and explored their stability and allosteric impact over hundreds of nanoseconds of MD, suggesting that the transient pocket can be stably bound and held open by an appropriate ligand. Comparison of free energy landscapes for the complete, glycosylated spike trimer suggested that the presence of a ligand in the hinge pocket is able to shift the RBD conformational equilibrium, significantly stabilizing the open RBD conformation. Optimization of binding affinity was not pursued here, and these simulations are meant only to demonstrate that a small molecule bound to this pocket could have an allosteric effect on the RBD. However, the results did suggest specific interaction motifs that may be worth exploiting via rational ligand design. In addition to changes in RBD dynamics with bound, optimized ligands, our results suggest that appropriate substitutions of A522 and K528 also should alter the RBD equilibrium, and ACE2 binding affinity.

Modulation of RBD dynamics in SARS-CoV-2 could have several consequences. Compared to SARS, SARS-CoV-2 better avoids immune surveillance by reducing the tendency of the RBD to adopt the open state. This less-frequent RBD opening reduces the effective concentration of RBD available to bind the ACE2 receptor, potentially leading to an unfavorable impact on ACE2 binding. The potential negative effect is compensated in SARS-CoV-2 via sequence changes that improve the direct interaction of the RBD and ACE2, with a net overall effect of reducing RBD exposure to immune surveillance while maintaining the net ACE2 affinity, and infectivity. Counteracting the immune shielding by forcing the RBD to remain exposed may be an interesting strategy to increase viral vulnerability. Furthermore, the molecular events that trigger irreversible S1 shedding and spike inactivation, experimentally observed in the absence of

ACE2, remain unclear. Forcing the RBD to adopt the open state for extended periods could weaken the interface between S1 and S2 subunits, potentially leading to increased premature shedding and irreversible disabling of the ability of the virus to infect host cells.

Methods

General Protocols

Simulations described here used the ff14Sbonlysc,⁷⁴ GLYCAM,⁷⁵ and OPC3⁷⁶ force fields for the protein glycans and water, respectively, with salt described by the Joung and Cheatham monovalent ion set⁷⁷⁻⁷⁸. Unless otherwise specified, all simulations used default settings in Amber v20⁷⁹, with a 4 fs timestep with hydrogen mass repartitioning⁸⁰, an 8.0 Å direct space cutoff with particle mesh Ewald⁸¹ for long-range electrostatics, a Langevin thermostat with collision frequency of 1.0 ps⁻¹, a Berendsen barostat with pressure coupling constant of 0.5 ps, and SHAKE on all bonds involving hydrogen atoms with 0.00001 Å tolerance. Simulations were carried out using the *sander* and *pmemd.CUDA* modules of Amber20⁷⁹. Structure visualization and salt bridge identification were performed with VMD⁸². RMSD values, CoM angle and dihedral values, and other structural measurements were calculated using the *cpptraj*⁸³ module of Amber. Backbone RMSD values used C α , C, and N atoms. The *imshow()* and *contour()* functions in the *pyplot* module of *matplotlib* library⁸⁴ were used to visualize the free energy landscapes. All amino acid position numbers refer to those in the full-length spike sequence.

Glycosylated Spike Model Building and Simulation Protocols

The full-length, glycosylated wild-type (WT) spike ectodomain model was described by Casalino et al.³¹, in both the closed and 1-up RBD states (the “initial models” below). These models are based on cryo-EM structures (the “experimental structures” below) of the soluble ectodomain of the spike (6VXX, closed;²⁶ 6VSB, chain A open²⁵), which included amino acid substitutions (e.g. in 6VSB: R682S, R683G, R685S, K986P, and V987P) to stabilize the spike in the pre-fusion state. These variants are easier to image and handle experimentally, but these substitutions also potentially alter²¹ the structure and dynamics of the spike. Thus, our models were converted back to the WT sequence, as well as cleaved at the S1/S2 interface furin site (R685|S686) to better mimic the expected state *in situ*. The models from Casalino et al.³¹ include both N-linked and O-linked glycans on each protomer,^{32, 35, 85} with a glycosylation profile congruent with cryo-EM and MS reports on the spike. Each protomer has 22 N-linked glycans and 1 (chains B and C) or 2 (chain A) O-linked glycans (**Table S1**). The initial models had the full-length stalk embedded into a membrane.³¹ Since our focus was on RBD dynamics, we simplified the initial model by removing the membrane and truncating the stalk at V1164 in each protomer.

The experimental models, particularly the 1-up structure, were missing density for the spike stalk, loops and several sidechains, including a portion of the RBM. Starting from the glycosylated structures from Casalino et al.³¹, we replaced the RBD of each protomer in both initial models with RBD coordinates obtained from the crystal structure of the RBD:ACE2 complex (PDB 6M0J¹⁹). The non-RBM region from 6M0J was first aligned with the non-RBM backbone atoms of each protomer. The atoms between A334 and G526 from the initial model were replaced with those from the crystallographic structure 6M0J. This resolved missing loops, as well as repositioned two cysteine residues (C480 and C488) close enough to introduce a disulfide bond that was absent from the cryo-EM structures and our

initial models. The remaining disulfides not resolved in the cryo-EM structures were included based on their distance and sequence conservation against the consensus sequences of SARS coronavirus; all of the involved cysteines were conserved in the SARS spike protein. A complete list of all the disulfides is provided in **Table S2**. This protocol resulted in the “amended models” used in the remainder of this work.

The amended models were solvated with a 30.0 Å minimum distance to the box edge, yielding 403,743 molecules of explicit water, as well as a 200 mM NaCl buffer. The 1-up and closed systems each consisted of the same 1,298,646 atoms. Each was equilibrated using a 10-step protocol. First, the water molecules were minimized for 1,000 steps using steepest descent, and then for an additional 9,000 steps with conjugate gradient, while the rest of the system was restrained with 1 kcal/(mol·Å²) Cartesian positional restraints. The systems were then heated to 310 K at constant volume over 0.5 ns, again with 100 kcal/(mol·Å²) positional restraints applied to all atoms except hydrogens and waters. The box size and density were then equilibrated over 1 ns with constant pressure of 1 bar, with positional restraints maintained. The restraints were then lowered to 10 kcal/(mol·Å²) for an additional 1 ns of equilibration, before a second minimization of 10,000 steps of conjugate gradient with only backbone atoms restrained using a force constant of 10 kcal/(mol·Å²). The next three steps of equilibration were done for 1 ns each at constant NPT with positional restraints on protein backbone atoms at 10, 1, and 0.1 kcal/(mol·Å²), respectively. This was followed by a final 1 ns of unrestrained MD at constant NPT. These equilibration steps used a 1 fs integration time step. Production simulations of the closed and 1-up spike systems were both conducted in triplicate, each for 0.3 μs at 310 K and constant NPT.

Building models for sequence substitutions

The 1-up structure used for building the WT model before equilibration was used as initial structure for building the sequence variants A522L and K528A. The sidechain atoms of residues to be substituted were deleted, and the new sidechains were introduced using the *tLeap* module of Amber20. The protocols used to build and equilibrate the WT system were repeated for the substituted spike systems.

Reference models

The cryo-EM structures 6VYB and 6VXX²⁶ were used as reference structures for RMSD analysis on the closed and 1-up structures, respectively. Since the experimental structures were missing coordinates for some regions, we built complete models to facilitate comparison by substituting the coordinates of backbone Ca, C, N, and O atoms of each experimental structure into our closed and 1-up models, and retaining coordinates for all other atoms. These are denoted as the “reference models”. RMSD calculations were limited to amino acids that were resolved in the 6VSB 1-up cryo-EM structure; these are listed for each domain in **Table S3**.

Collective variables used to quantify the RBD open/closed transition

We defined center-of-mass (COM) variables to quantify RBD motion during MD, the RBD-opening pathway from NEB, and also as the collective variables (CVs) for the umbrella sampling simulations. Gui et al.²⁸ used the angle between the long axis of RBD and the plane vertical to the symmetry axis of S protein to quantify the state of RBD. Here, to facilitate use as restraints during umbrella sampling, a COM angle was defined to measure the open/closed movement of the RBD. The first point of the angle was chosen

to be the center of the opening RBD (COM groups are provided in **Table S4**). The second point is in CTD1 on the same protomer, and the third is in the center of a short helical region in the upstream helix. The second group was chosen at approximately the fulcrum of RBD opening, and the third group is roughly at the same distance along the CH vector as the second group. These were chosen to make this COM angle definition similar to that used by Gui et al.²⁸

As the RBD opens and becomes less constrained by packing against neighboring RBDs, we also observed RBD rotation along its own long axis. We added a second CV, the COM dihedral, to improve sampling of this rotation and speed convergence in the open state. The first two points of the COM dihedral are in CTD1, while the other 2 are in the RBD. **Figure S1** gives a visual representation of both CV definitions.

Nudged Elastic Band Pathway Optimization

We used a GPU implementation⁸⁶ of a large-system variant⁵³ of the nudged elastic band (NEB) method⁸⁷ to explore a pathway of the spike transition between the closed and 1-up RBD states. We previously used this protocol to obtain pathways for biomolecular recognition dynamics that were subsequently validated through experiments.⁸⁸⁻⁹³ The equilibrated closed and 1-up structures were used as the two pathway endpoints, which in NEB are held fixed. An additional 30 intermediate simulations (beads) were placed to map the transition pathway, giving a total of 32 simulations run simultaneously, with bead 0 corresponding to the equilibrated closed state, and bead 32 corresponding to the equilibrated 1-up state. NEB springs were applied to the backbone heavy atoms of all residues with a 1 kcal/(mol·Å²) force constant. NEB was run in four stages: heating, equilibration, annealing, and production. The set of 32 simulations were first heated from 100 to 300 K over 0.5 ns at constant volume, then the box size was allowed to equilibrate for 1 ns at constant pressure and temperature. Annealing was done over 5 ns at constant volume: 1 ns of heating from 300 to 400 K, 1 ns at 400 K, 1 ns heating from 400 to 500 K, 1 ns at 500 K, then cooling to 300 K over 0.5 ns and a final 0.5 ns at 300 K. The final annealed structure from each bead was used to initiate 15 ns of NEB at 300 K and constant P to locally explore the annealed pathway.

Steered molecular dynamics

The NEB calculations applied springs only to the protein backbone, leading to potential discontinuities in other degrees of freedom along the optimized path that could hinder calculation of free energy landscapes. Therefore, steered molecular dynamics (SMD⁹⁴) using the Amber NFE module was used to generate initial structures for umbrella sampling. The starting point was the frame from NEB used for ligand virtual screening, to make sure the apo SMD path was comparable with the ligand-bound SMD path. The structure first was steered to match the 1-up endpoint of NEB by reducing RMSD from the initial value (1.95 Å) to zero, using a force constant of 50000 kcal/(mol·Å²) during 0.5 ns at 310 K and constant volume. The RMSD region included the C α atoms of the RBD (residues 338-517), CTD1 (residues 324-327, 538-585) and helices in S2 (residues 747-782, 946-966, 987-1034) of only one protomer. The 1-up spike was then closed gradually along the final NEB pathway in a stepwise fashion (31 steps, one for each subsequent NEB bead) until the RBD was fully closed. At each step, the last frame from the neighboring bead was used as the new reference, and the RMSD was reduced to zero as in the previous

step. The simulation conditions and lengths for each step were the same as the first step. This strategy was applied to both apo and ligand-bound wild-type and substituted full spike models.

Umbrella sampling and free energy landscapes

To map the energy landscape for the open-to-closed RBD transition in the apo and bound WT, and the A522L and K528A spike systems, we performed 2-dimensional umbrella sampling (US). A 2-D grid was generated, ranging from 45° to 95° in CoM angle and -30° to 70° in CoM dihedral, spaced evenly by 2° in both dimensions. The range of values for the grid was selected based on the range of values sampled during the NEB simulations with additional space for expanding the grid. The SMD trajectories for apo, bound, and substituted spike systems were mapped onto the 2D grid, and the median potential energy structure occupying each grid point was selected to initiate US simulations. This resulted in a total of 111 occupied grid blocks for the bound S-protein, 102 grid blocks for the apo, 114 grid blocks for A522L, and 102 grid blocks for K528A S-protein. After visual inspection of the initial PMF, additional points were added to expand the sampled grid beyond local minima. Initial structures for expanded grid points were obtained from the final structure sampled in the US run of the grid left or right in CoM angle space; if not available, the grid above or below in CoM dihedral space was selected. The US protocol described above was repeated for the newly added grid points. This brought the total number of grid points to 332 for the bound WT, 328 for apo WT, 308 for A22L, and 280 for K528A (**Figure S2**).

Umbrella sampling was carried out with 1 window (simulation) for each selected grid point, simulated in parallel. For each window, the CoM angle and CoM dihedral of each initial structure were restrained to their respective grid center using a harmonic potential with force constants of 1600 and 1400 kcal/(mol·rad²) respectively. For the ligand-bound US, restraints were used between the distance between the backbone nitrogen of K513 and the nitrogen in the imidazole ring of the ligand and the distance between the sidechain nitrogen of K513 and the carbonyl oxygen of the ligand. A half flat-bottom potential was used, where there was no energy penalty when the distances were below (4 Å, K528 backbone; 5 Å, K528 sidechain), and a harmonic potential energy when the distance was larger, using a force constant of 10 kcal/(mol·Å²). Each grid structure was equilibrated for 2 ns at 310 K under NPT conditions, followed by a production run of 16 ns under the same conditions, saving structures every 20 ps and values of the CVs every 0.4 ps. PMFs for each system were calculated using the weighted histogram analysis method⁹⁵ (WHAM) in WHAM v2.0.10⁹⁶, with a convergence tolerance of 0.00001 kcal/mol and histogram boundaries of 45 ° and 95° for the CoM angle and -30° and 70° for the CoM dihedral reaction coordinates, respectively. Bin widths of 0.25° for CoM angle and 0.35° for CoM dihedral were optimized using the Shimazaki-Shinomoto algorithm⁹⁷

Selection of structure for virtual screening

A snapshot from the full-spike NEB pathway calculation was selected for virtual screening and MD simulations. The NEB pathway was examined visually to select a structure for subsequent virtual screening. A520 and Q563 reside on the RBD and CTD1, respectively, and form a close contact between backbone C α atoms in the closed-RBD state (distance of 4.5 Å). The distance was monitored during RBD opening, and reached a plateau near 27 Å (see Results). A frame was selected at the start of the plateau (75% progress on the opening path), since the later stages did not result in larger opening of the pocket and mainly corresponded to adjustments in the RBD:CTD1 N-connector strand running from F329 to

N334. We had greater confidence in the structural details obtained from the annealed pathway rather than the endpoint structure, since this connector was absent in the 6VSB cryo-EM structure used to build the initial models.

Fpocket analysis

Fpocket 3.0.4⁵⁴ was used to extract descriptors of the hinge pocket. The algorithm requires a selection of grid points to define the pocket. Therefore, we used mdpocket⁹⁸ to obtain the grid points for the hinge pocket, using the trajectory from the NEB bead that contained the structure selected for virtual screening. The structures in the NEB trajectory were aligned along the lower portion of the central helices (S1006-V1036). Since the hinge pocket is contained inside a protomer, the other two protomers were deleted to reduce the calculation time. The analysis yielded the same results whether glycans were retained or removed. Default fpocket settings were used to select and cluster the α -spheres. The grid points that corresponded to the cluster of α -spheres located in the hinge pocket (**Figure S3**) were extracted using Chimera⁹⁹ v1.14. All 395 extracted grid points were used to define the pocket to extract descriptors using a second run of mdpocket, which calculated descriptors using trajectories from the entire RBD opening pathway (all 32 beads). The descriptors used were pocket volume, pocket residues, and SASA score. The druggability score was calculated on each snapshot from the entire RBD opening pathway using the fpocket module.

RBD+CTD1 construct

In order to more rapidly test the interaction of the pocket with docked ligands in MD, and also to provide a smaller system for virtual screening, we excised a subset of the amino acids from the complete spike, including the domains on both sides of the hinge pocket (all of the RBD and CTD1 domains; 271 residues total, corresponding to the portion from Q321 to S591 in the intact spike; see **Figure S4**). Five disulfide bonds present in these domains for the full spike were retained in the fragment (**Table S2**). Glycans were removed since none were within 10 Å of the pocket. Water was removed and the protein was re-solvated with a minimum distance to box edge of 10 Å, resulting in addition of ~ 27,000 water molecules, using the same parameters as for the full spike simulations. The resulting system consisted of 86,270 atoms for the unbound RBD+CTD1 construct.

This smaller construct was equilibrated using a 10-step protocol similar to that described above. First, the water molecules and hydrogen atoms were minimized for 1,000 steps using steepest descent, and then for an additional 9,000 steps with conjugate gradient while the rest of the system was positionally restrained with 10 kcal/(mol·Å²) restraints. Next, the systems were heated to 298 K at constant volume with all atoms except hydrogens and waters being restrained with 100 kcal/(mol·Å²) positional restraints. Next, the box size and density were equilibrated over 1 ns at constant temperature (298K) and pressure (1 bar) with coupling constant of 0.1, with positional restraints retained on all atoms except the water and hydrogens. The restraints were then lowered to 10 kcal/(mol·Å²) for an additional 1 ns of equilibration at 298K and 1 bar, with pressure coupling constant of 0.5. This was followed with a second minimization of 10,000 steps of conjugate gradient, with only backbone atoms (CA, N, C) positionally restrained using a force constant of 10 kcal/(mol·Å²). Next, three steps of equilibration were performed for 1 ns each, at constant NPT, with positional restraints on only protein backbone atoms with restraint force constants of 10, 1, and 0.1 kcal/(mol·Å²), respectively. This was followed by a final 1 ns of

fully unrestrained MD at constant NPT with a pressure coupling constant of 1.0. Unrestrained production MD was carried out for 2 μ sec at 310K and 1 bar, using a time step of 4 fs.

For the ligand-bound RBD+CTD1 constructs, the same protocol was followed as described above for the apo construct. The ligand was not restrained during initial minimization in order to adjust the docked pose to the simulation environment.

Virtual Screening with DOCK

Briefly, docking spheres were generated using the program *sphgen*¹⁰⁰ which defines a negative image space filling model of the protein surface. DOCK spheres are subsequently used to orient ligand scaffolds (rigid subgroups) prior to sampling ligand internal degrees of freedom in the pocket via the Anchor & Grow algorithm.⁵⁹ For the present setup, 52 spheres total were selected based on their proximity to the site being targeted which were used to define a bounding box for the program *grid*¹⁰¹ based on 0.3 Å grid spacing. Docking grids help speed up the calculations by pre-computing van der Waals (VDW) and Electrostatic (ES) contributions from the protein at each grid point.¹⁰¹ As in prior work, the VDW term employed a 6-9 potential and the ES term employed a distance dependent dielectric (4r). The overall docking setup is shown in **Figure S5**.

We used DOCK6.9⁵⁹ to screen two libraries of purchasable molecules downloadable from the ZINC database⁷⁰. Criteria for library inclusion included a ligand formal charge range of -2 to 2 and a logP range between -1 to 5 which helps to ensure candidates have lead-like properties from the outset. Library 1 contained 2.5M molecules with a MW range of 200 to 500 g/mol, mainly composed of smaller molecules. Library 2 contained 4.5M molecules with a somewhat larger MW range of 325 to 500 g/mol. The screening protocol employed the DOCK FLX¹⁰² approach which samples rigid body and internal degrees of freedom for the ligand while the protein is held rigid. The FLX protocol retains the best scoring ligand conformer (pose) for each molecule as determined by the grid energy which is termed single grid energy (SGE) score. Each pose was energy-minimized in Cartesian space which is termed DOCK Cartesian energy (DCE) score. This procedure further optimizes the ligand translational and rotational degrees of freedom and enables Footprint Similarity (FPS) scores to be computed as discussed below. DCE minimizations employed a restraint tether of 10 kcal/(mol·Å²) to help keep poses in their originally docked geometry.

As a logical way to help select compounds for experimental testing, we employed Footprint Similarity Scoring (FPS)⁷¹⁻⁷² which quantifies the degree of overlap (Euclidian distance) in terms of non-bonded per-residue VDW and ES interaction energies with the protein. FPS_{SUM} quantifies the overlap between the VDW and ES patterns while FPS_{VDW} or FPS_{ES} reflect the individual components. References are typically based on the 3D binding geometry for a known molecule. Alternatively, as employed in the present work, a text-based list of residue numbers with their accompanying VDW and ES energies can be used.

We hypothesized that the transition from an open-to-closed state could be blocked by a molecule docked to an open-like state which made intermolecular interactions similar to those that would normally be present between RBD and CTD1 domains when in the closed-like state. Visual examination of a sequential series of snapshots extracted from the NEB pathway trajectory suggested that residues 390-395 plus 513-525 at the base of the RBD docked into the pocket in the closed-like conformation. Construction of the reference used for virtual screening employed the following steps: (1) Residues 390-395 plus 513-525 were extracted and saved as a separate MOL2 file. (2) Per-residue interaction maps between residues

390-395 plus 513-525 were computed with all residues from the frame selected for docking (see above). (3) Inspection of the underlying energies led to the conclusion that the residue subset 390-394 plus 516-524 would be most suitable for the identification of molecules that would occupy the hinge pocket which is only present in the open state. (4) The text-based footprint derived from this smaller subset was then manually adjusted to emphasize key interactions deemed important for binding. Specifically, the ES interaction at residue K528 was set to -7 kcal/mol due to its high conservation and important role in stabilizing the closed state. The VDW interactions at N360 and C361 were increased to -2 kcal/mol as these residues comprise a hydrophobic cleft within the hinge pocket a molecule could occupy and impede the closing motion. Additionally, the VDW interactions at V362 and A363 were increased to -4.5 and -3.0 kcal/mol respectively to help identify molecules that would interact favorably with a tunnel in the open state. Finally, a strong ES interaction at R567 in CTD1 was determined to be of little interest due to the distance from the pocket, so the value was decreased to 0 kcal/mol. The overall objective was to screen molecules into the open-like conformation and use the text-based footprint described above, termed the "peptide reference", to help identify small molecules that mimic closed-state interactions. It should be emphasized that FPS overlap was but one guiding factor used during compound prioritization.

Candidate molecules from docking were parametrized for MD simulations using the AMBER16¹⁰³ suite of programs. Briefly, DOCK6 poses for ligands from the virtual screen (MOL2 format) were used as input for the *antechamber* module which was used to compute AM1-BCC¹⁰⁴⁻¹⁰⁵ partial atomic charges and assign GAFF¹⁰⁶ force field parameters. Full details for virtual screening methods and results are provided in Supplementary Information.

Preparing the ligand-bound full spike, and the 3-up spike model

The NEB frame used for virtual screening was used to build a 1-up complete spike model with a single ligand bound below the open RBD. A structure of the full spike from the NEB pathway with open RBD and T333 pointing into the pocket was selected and aligned with the structure after 300 ns of MD for the ligand-bound RBD+CTD1 fragment by overlapping the backbone of both connectors (F329 – P337, C525 – T531), and the ligand coordinates were transferred from the fragment to the full spike system. To avoid steric clashes that may make equilibration unstable, the pose of the ligand was further optimized by minimization in DOCK6. After minimization, any water molecules within 4 Å of the ligand in the original NEB frame were deleted. The system was equilibrated using the same protocol as described above for the amended spike models, except that in the first minimization step, steepest descent was only run for 10 steps followed by conjugate gradient, and the spring constant was 100 kcal/(mol·Å²). Since we started from an equilibrated structure of the spike, 10 steps of steepest descent were sufficient.

A 3-up apo spike model was built using SMD, for control simulations and for simulations of 3-up spike with bound ligands. The equilibrated closed structure was used as the initial model. Three independent RMSDs were used as CVs to guide the opening of the three RBDs. The RMSD regions were C α atoms of residues 338-517, 747-782, 946-966 and 987-1034, which included one RBD and three α -helices in the S2 subunit including UH, CH and a helix in HR1. A weak Cartesian positional restraint of 1 kcal/(mol·Å²) was applied to the S2 helices to maintain stability during SMD. A reference structure for the target RMSD was generated by aligning the open protomer from the equilibrated 1-up model to each of the three protomers by overlapping the S2 subunit. The newly generated RBD coordinates and the original S2 coordinates were saved, creating a 3-up RBD model that was suitable to use as an RMSD

target for SMD. During subsequent SMD simulation starting from the closed spike, the three RMSDs were gradually decreased from their initial values to zero using a force constant of 10000 kcal/(mol·Å²) during 20 ns of simulation at 310 K and constant volume.

The final structure from this SMD step was used to make the 3-up ligand bound model. To accommodate the ligand, the hinge pockets of all the three protomers were changed via SMD to match the same structure used for placement of the ligand in the 1-up full spike as described above. The RMSD region for SMD included the C α atoms in the N-connector and C-connector. The target RMSD was reduced from the initial values to zero (final value 0.05 Å) during 10 ns of simulation at 310 K and constant volume. The same positional restraints were used as in the previous step. The 3 ligands were placed similarly as in the 1-up model by aligning the connectors to each of the three hinge pockets. The poses of the ligands were optimized similarly by DOCK6.⁵⁹ Water molecules within 4 Å of any of the three ligands were deleted. The same equilibration protocol from the 1-up bound model was used to equilibrate the system. Production MD was in NVT at 310K.

Sequence conservation analysis

A total of 28 sequences belonging to a variety of coronaviruses were utilized to quantify amino acid conservation at various locations of the spike. Lineages were chosen based on previous work done by Cagliani et al.¹⁰⁷ who noted the clustering of SARS-CoV-2 with SARS-CoV in pangolins and bats. The authors used knowledge of previous zoonotic spillover events that resulted in SARS and MERS outbreak to determine the relevance of pangolin and bat coronaviruses. Sequences were obtained using the genbank¹⁰⁸ and GISAID¹⁰⁹ databases. Lineages were chosen based on previous work done by Cagliani et al.¹⁰⁷ who note the clustering of SARS-CoV-2 with SARS-CoV in pangolins and bats. The authors used knowledge of previous zoonotic spillover events that resulted in SARS and MERS outbreak to determine the relevance of pangolin and bat coronaviruses. Differing from Cagliani, the GISAID¹⁰⁹ reference sequence was utilized and two common human coronaviruses were included: 229E and OC43. **Table S5** details the specific accession numbers and hosts of coronaviruses chosen. Alignment of sequences was performed using default alignment parameters on the T-Coffee¹¹⁰ online server. Following the alignment, the sequences were processed to remove gaps that were present in the SARS-CoV-2 sequence. **Table S6** shows the calculated percent identity for the hinge pocket region alongside the sequence alignment. Lastly, sequence logos were generated utilizing the WebLogo server¹¹¹.

Supplementary Information

Additional figures, tables, methods, and detailed description of virtual screening protocol and results.

Author contributions

LF: MD simulations, NEB pathway calculations and structure visualization; KB: free energy calculations and fpocket analysis; LR: glycan modeling and data analysis; YW: Collective variables, SMD simulations and model building; CC: virtual screening; DS: data analysis; AC: glycan modeling, data analysis, and structure visualization; JG: sequence conservation analysis; SB and SV: analysis of experimental structures; RR: virtual screening design and interpretation; CS: ligand-bound simulations and overall project design.

Competing Interests

The authors declare no competing interests.

Acknowledgements

We are grateful to Dan Raleigh, Ronit Freeman, and Arvind Ramanathan for helpful discussions, and to Sarina Bromberg for assistance with Figure 2. We thank Rommie Amaro and her group for helpful discussions and for early access to the spike models.

This work was funded in part by the SUNY Research Seed Grant Program, the Stony Brook University Office of the Vice President for Research, the Research Corporation for Science Advancement (COVID Initiative grant #27350), NIH grants R01GM107104 (C.S.), R35GM126906 (R.C.R) and the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act.

This work used resources services, and support provided via the COVID-19 HPC Consortium (<https://covid19-hpc-consortium.org/>), which is a unique private-public effort to bring together government, industry, and academic leaders who are volunteering free compute time and resources in support of COVID-19 research. The authors thank Stony Brook Research Computing and Cyberinfrastructure, and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance Lired and SeaWulf computing systems, the latter of which was made possible by National Science Foundation grant #1531492. Additional computer time was provided by Brookhaven National Laboratory and AiMOS at Rensselaer Polytechnic Institute.

References

1. World Health Organization Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int/> (accessed 12/27/2020).
2. Belouzard, S.; Chu, V. C.; Whittaker, G. R., Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106* (14), 5871-5876.
3. Li, F., Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Rev Virol* **2016**, *3* (1), 237-261.
4. Bosch, B. J.; van der Zee, R.; de Haan, C. A.; Rottier, P. J., The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* **2003**, *77* (16), 8801-11.
5. Tang, T.; Bidon, M.; Jaimes, J. A.; Whittaker, G. R.; Daniel, S., Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research* **2020**, *178*, 104792.
6. Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Kräusslich, H.-G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G., Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **2020**, *588* (7838), 498-502.

7. Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J. H.; Welsch, S.; Blanc, F. E. C.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; van Zandbergen, G.; Landry, J.; de Azevedo, N. T. D.; Mosalaganti, S.; Schwarz, A.; Covino, R.; Mühlebach, M. D.; Hummer, G.; Krijnse Locker, J.; Beck, M., In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **2020**, *370* (6513), 203-208.
8. Corey, L.; Mascola, J. R.; Fauci, A. S.; Collins, F. S., A strategic approach to COVID-19 vaccine R&D. *Science* **2020**, *368* (6494), 948-950.
9. Callaway, E., The race for coronavirus vaccines: a graphical guide. *Nature* **2020**, *580* (7805), 576-577.
10. Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N. G.; Decroly, E., The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research* **2020**, *176* (February), 104742-104742.
11. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S., SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271-280.e8.
12. Jaimes, J. A.; André, N. M.; Chappie, J. S.; Millet, J. K.; Whittaker, G. R., Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *J Mol Biol* **2020**, *432* (10), 3309-3325.
13. Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F., Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences* **2020**, *117* (21), 11727-11734.
14. Zhang, H.; Penninger, J. M.; Li, Y.; Zhong, N.; Slutsky, A. S., Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Medicine* **2020**, *46* (4), 586-590.
15. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F., Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**.
16. Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J., Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588* (7837), 327-330.
17. Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K. Y.; Wang, Q.; Zhou, H.; Yan, J.; Qi, J., Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181* (4), 894-904.e9.
18. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q., Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)* **2020**, *2* (March), 1444-1448.
19. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**.
20. Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B. J.; Rey, F. A.; Velesler, D., Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc Natl Acad Sci U S A* **2017**, *114* (42), 11157-11162.
21. Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rawson, S.; Rits-Volloch, S.; Chen, B., Distinct conformational states of SARS-CoV-2 spike protein. *Science* **2020**, *369* (6511), 1586-1592.

22. Liu, C.; Yang, Y.; Gao, Y.; Shen, C.; Ju, B.; Liu, C.; Tang, X.; Wei, J.; Ma, X.; Liu, W.; Xu, S.; Liu, Y.; Yuan, J.; Wu, J.; Liu, Z.; Zhang, Z.; Wang, P.; Liu, L., Viral Architecture of SARS-CoV-2 with Post-Fusion Spike Revealed by Cryo-EM. *bioRxiv* **2020**, 2020.03.02.972927.
23. Liu, C.; Mendonça, L.; Yang, Y.; Gao, Y.; Shen, C.; Liu, J.; Ni, T.; Ju, B.; Liu, C.; Tang, X.; Wei, J.; Ma, X.; Zhu, Y.; Liu, W.; Xu, S.; Liu, Y.; Yuan, J.; Wu, J.; Liu, Z.; Zhang, Z.; Liu, L.; Wang, P.; Zhang, P., The Architecture of Inactivated SARS-CoV-2 with Postfusion Spikes Revealed by Cryo-EM and Cryo-ET. *Structure* **2020**, *28* (11), 1218-1224.e4.
24. Huang, Y.; Yang, C.; Xu, X.-f.; Xu, W.; Liu, S.-w., Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica* **2020**, *41* (9), 1141-1149.
25. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)* **2020**, *1263* (March), 1260-1263.
26. Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Velesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**.
27. Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L., Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology* **2020**, *17* (6), 613-620.
28. Gui, M.; Song, W.; Zhou, H.; Xu, J.; Chen, S.; Xiang, Y.; Wang, X., Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Research* **2017**, *27* (1), 119-129.
29. Song, W.; Gui, M.; Wang, X.; Xiang, Y., Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* **2018**, *14* (8), e1007236.
30. Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E., The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. *Biophysical Journal* **2020**.
31. Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E., Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **2020**, *6* (10), 1722-1734.
32. Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J., Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports* **2020**, *10* (1), 14991.
33. Walls, A. C.; Xiong, X.; Park, Y. J.; Tortorici, M. A.; Snijder, J.; Quispe, J.; Cameroni, E.; Gopal, R.; Dai, M.; Lanzavecchia, A.; Zambon, M.; Rey, F. A.; Corti, D.; Velesler, D., Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **2019**, *176* (5), 1026-1039 e15.
34. Zou, J.; Yin, J.; Fang, L.; Yang, M.; Wang, T.; Wu, W.; Bellucci, M. A.; Zhang, P., Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Relative Free Energy Calculations. *J Chem Inf Model* **2020**, *60* (12), 5794-5802.
35. Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M., Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369* (6501), 330-333.

36. Madu, I. G.; Roth, S. L.; Belouzard, S.; Whittaker, G. R., Characterization of a Highly Conserved Domain within the Severe Acute Respiratory Syndrome Coronavirus Spike Protein S2 Domain with Characteristics of a Viral Fusion Peptide. *Journal of Virology* **2009**, *83* (15), 7411-7421.
37. Xia, S.; Zhu, Y.; Liu, M.; Lan, Q.; Xu, W.; Wu, Y.; Ying, T.; Liu, S.; Shi, Z.; Jiang, S.; Lu, L., Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol* **2020**.
38. Johnson, M. C.; Lyddon, T. D.; Suarez, R.; Salcedo, B.; LePique, M.; Graham, M.; Ricana, C.; Robinson, C.; Ritter, D. G., Optimized Pseudotyping Conditions for the SARS-COV-2 Spike Glycoprotein. *Journal of Virology* **2020**, *94* (21), e01062-20.
39. Reinke, L. M.; Spiegel, M.; Plegge, T.; Hartleib, A.; Nehlmeier, I.; Gierer, S.; Hoffmann, M.; Hofmann-Winkler, H.; Winkler, M.; Pöhlmann, S., Different residues in the SARS-CoV spike protein determine cleavage and activation by the host cell protease TMPRSS2. *PLOS ONE* **2017**, *12* (6), e0179177.
40. Hoffmann, M.; Kleine-Weber, H.; Pöhlmann, S., A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell* **2020**, *78* (4), 779-784.e5.
41. Millet, J. K.; Whittaker, G. R., Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology* **2018**, *517*, 3-8.
42. Fan, X.; Cao, D.; Kong, L.; Zhang, X., Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature Communications* **2020**, *11* (1), 3618.
43. Yao, H.; Song, Y.; Chen, Y.; Wu, N.; Xu, J.; Sun, C.; Zhang, J.; Weng, T.; Zhang, Z.; Wu, Z.; Cheng, L.; Shi, D.; Lu, X.; Lei, J.; Crispin, M.; Shi, Y.; Li, L.; Li, S., Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **2020**, *183* (3), 730-738.e13.
44. Xiong, X.; Qu, K.; Ciazynska, K. A.; Hosmillo, M.; Carter, A. P.; Ebrahimi, S.; Ke, Z.; Scheres, S. H. W.; Bergamaschi, L.; Grice, G. L.; Zhang, Y.; Bradley, J.; Lyons, P. A.; Smith, K. G. C.; Toshner, M.; Elmer, A.; Ribeiro, C.; Kourampa, J.; Jose, S.; Kennet, J.; Rowlands, J.; Meadows, A.; O'Brien, C.; Rastall, R.; Crucisio, C.; Hewitt, S.; Price, J.; Calder, J.; Canna, L.; Bucke, A.; Tordesillas, H.; Harris, J.; Ruffolo, V.; Domingo, J.; Graves, B.; Butcher, H.; Caputo, D.; Le Gresley, E.; Dunmore, B. J.; Martin, J.; Legchenko, E.; Treacy, C.; Huang, C.; Wood, J.; Sutcliffe, R.; Hodgson, J.; Shih, J.; Graf, S.; Tong, Z.; Mescia, F.; Tilly, T.; O'Donnell, C.; Hunter, K.; Pointon, L.; Pond, N.; Wylot, M.; Jones, E.; Fawke, S.; Bullman, B.; Bergamaschi, L.; Turner, L.; Jarvis, I.; Omarjee, O.; De Sa, A.; Marsden, J.; Betancourt, A.; Perera, M.; Epping, M.; Richoz, N.; Bower, G.; Sharma, R.; Nice, F.; Huhn, O.; Stark, H.; Walker, N.; Stirrups, K.; Ovington, N.; Dewhurst, E.; Li, E.; Papadia, S.; Nathan, J. A.; Baker, S.; James, L. C.; Baxendale, H. E.; Goodfellow, I.; Doffinger, R.; Briggs, J. A. G.; The, C.-N. C.-B. C., A thermostable, closed SARS-CoV-2 spike protein trimer. *Nature Structural & Molecular Biology* **2020**, *27* (10), 934-941.
45. Hsieh, C.-L.; Goldsmith, J. A.; Schaub, J. M.; DiVenere, A. M.; Kuo, H.-C.; Javanmardi, K.; Le, K. C.; Wrapp, D.; Lee, A. G.; Liu, Y.; Chou, C.-W.; Byrne, P. O.; Hjorth, C. K.; Johnson, N. V.; Ludes-Meyers, J.; Nguyen, A. W.; Park, J.; Wang, N.; Amengor, D.; Lavinder, J. J.; Ippolito, G. C.; Maynard, J. A.; Finkelstein, I. J.; McLellan, J. S., Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **2020**, *369* (6510), 1501-1505.
46. McCallum, M.; Walls, A. C.; Bowen, J. E.; Corti, D.; Veerler, D., Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nature Structural & Molecular Biology* **2020**, *27* (10), 942-949.

47. Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Gobeil, S. M. C.; Kopp, M.; Li, D.; Parks, R.; Hsu, A. L.; Borgnia, M. J.; Haynes, B. F.; Acharya, P., Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature Structural & Molecular Biology* **2020**, *27* (10), 925-933.
48. Toelzer, C.; Gupta, K.; Yadav, S. K. N.; Borucu, U.; Davidson, A. D.; Kavanagh Williamson, M.; Shoemark, D. K.; Garzoni, F.; Staufer, O.; Milligan, R.; Capin, J.; Mulholland, A. J.; Spatz, J.; Fitzgerald, D.; Berger, I.; Schaffitzel, C., Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **2020**, *370* (6517), 725-730.
49. Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J.-H.; Portnoff, A. D.; Patel, N.; Massare, M. J.; Yates, J. R.; Nemazee, D.; Paulson, J. C.; Glenn, G.; Smith, G.; Ward, A. B., Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* **2020**, *370* (6520), 1089-1094.
50. Yang, J.; Wang, W.; Chen, Z.; Lu, S.; Yang, F.; Bi, Z.; Bao, L.; Mo, F.; Li, X.; Huang, Y.; Hong, W.; Yang, Y.; Zhao, Y.; Ye, F.; Lin, S.; Deng, W.; Chen, H.; Lei, H.; Zhang, Z.; Luo, M.; Gao, H.; Zheng, Y.; Gong, Y.; Jiang, X.; Xu, Y.; Lv, Q.; Li, D.; Wang, M.; Li, F.; Wang, S.; Wang, G.; Yu, P.; Qu, Y.; Yang, L.; Deng, H.; Tong, A.; Li, J.; Wang, Z.; Yang, J.; Shen, G.; Zhao, Z.; Li, Y.; Luo, J.; Liu, H.; Yu, W.; Yang, M.; Xu, J.; Wang, J.; Li, H.; Wang, H.; Kuang, D.; Lin, P.; Hu, Z.; Guo, W.; Cheng, W.; He, Y.; Song, X.; Chen, C.; Xue, Z.; Yao, S.; Chen, L.; Ma, X.; Chen, S.; Gou, M.; Huang, W.; Wang, Y.; Fan, C.; Tian, Z.; Shi, M.; Wang, F.-S.; Dai, L.; Wu, M.; Li, G.; Wang, G.; Peng, Y.; Qian, Z.; Huang, C.; Lau, J. Y.-N.; Yang, Z.; Wei, Y.; Cen, X.; Peng, X.; Qin, C.; Zhang, K.; Lu, G.; Wei, X., A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* **2020**, *586* (7830), 572-577.
51. Hansen, J.; Baum, A.; Pascal, K. E.; Russo, V.; Giordano, S.; Wloga, E.; Fulton, B. O.; Yan, Y.; Koon, K.; Patel, K.; Chung, K. M.; Hermann, A.; Ullman, E.; Cruz, J.; Rafique, A.; Huang, T.; Fairhurst, J.; Libertiny, C.; Malbec, M.; Lee, W.-y.; Welsh, R.; Farr, G.; Pennington, S.; Deshpande, D.; Cheng, J.; Watty, A.; Bouffard, P.; Babb, R.; Levenkova, N.; Chen, C.; Zhang, B.; Romero Hernandez, A.; Saotome, K.; Zhou, Y.; Franklin, M.; Sivapalasingam, S.; Lye, D. C.; Weston, S.; Logue, J.; Haupt, R.; Frieman, M.; Chen, G.; Olson, W.; Murphy, A. J.; Stahl, N.; Yancopoulos, G. D.; Kyratsous, C. A., Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **2020**, *369* (6506), 1010.
52. Kirchdoerfer, R. N.; Wang, N.; Pallesen, J.; Wrapp, D.; Turner, H. L.; Cottrell, C. A.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B., Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific Reports* **2018**, *8* (1), 1-11.
53. Bergonzo, C.; Campbell, A. J.; Walker, R. C.; Simmerling, C., A partial nudged elastic band implementation for use with large or explicitly solvated systems. *International Journal of Quantum Chemistry* **2009**, *109* (15), 3781-3790.
54. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **2009**, *10* (1), 168.
55. Coleman, R. G.; Sharp, K. A., Protein pockets: inventory, shape, and comparison. *J Chem Inf Model* **2010**, *50* (4), 589-603.
56. Schmidtke, P.; Barril, X., Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *Journal of Medicinal Chemistry* **2010**, *53* (15), 5858-5867.

57. Pallesen, J.; Wang, N.; Corbett, K. S.; Wrapp, D.; Kirchdoerfer, R. N.; Turner, H. L.; Cottrell, C. A.; Becker, M. M.; Wang, L.; Shi, W.; Kong, W.-P.; Andres, E. L.; Kettenbach, A. N.; Denison, M. R.; Chappell, J. D.; Graham, B. S.; Ward, A. B.; McLellan, J. S., Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proceedings of the National Academy of Sciences* **2017**, *114* (35), E7348-E7357.
58. Liu, H.; Wu, N. C.; Yuan, M.; Bangaru, S.; Torres, J. L.; Caniels, T. G.; van Schooten, J.; Zhu, X.; Lee, C.-C. D.; Brouwer, P. J. M.; van Gils, M. J.; Sanders, R. W.; Ward, A. B.; Wilson, I. A., Cross-Neutralization of a SARS-CoV-2 Antibody to a Functionally Conserved Site Is Mediated by Avidity. *Immunity* **2020**, *53* (6), 1272-1280.e5.
59. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C., DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry* **2015**, *36* (15), 1132-1156.
60. Holden, P. M.; Kaur, H.; Goyal, R.; Gochin, M.; Rizzo, R. C., Footprint-based identification of viral entry inhibitors targeting HIVgp41. *Bioorg. Med. Chem. Lett.* **2012**, *22* (Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved.), 3011-3016.
61. Allen, W. J.; Yi, H. A.; Gochin, M.; Jacobs, A.; Rizzo, R. C., Small molecule inhibitors of HIVgp41 N-heptad repeat trimer formation. *Bioorg Med Chem Lett* **2015**, *25* (14), 2853-9.
62. McGee, T. D., Jr.; Yi, H. A.; Allen, W. J.; Jacobs, A.; Rizzo, R. C., Structure-based identification of inhibitors targeting obstruction of the HIVgp41 N-heptad repeat trimer. *Bioorg Med Chem Lett* **2017**, *27* (14), 3177-3184.
63. Berger, W. T.; Ralph, B. P.; Kaczocha, M.; Sun, J.; Balius, T. E.; Rizzo, R. C.; Haj-Dahmane, S.; Ojima, I.; Deutsch, D. G., Targeting Fatty Acid Binding Protein (FABP) Anandamide Transporters – A Novel Strategy for Development of Anti-Inflammatory and Anti-Nociceptive Drugs. *PLoS ONE* **2012**, *7* (12), e50968.
64. Zhou, Y.; Elmes, M. W.; Sweeney, J. M.; Joseph, O. M.; Che, J.; Hsu, H. C.; Li, H.; Deutsch, D. G.; Ojima, I.; Kaczocha, M.; Rizzo, R. C., Identification of Fatty Acid Binding Protein 5 Inhibitors Through Similarity-Based Screening. *Biochemistry* **2019**, *58* (42), 4304-4316.
65. Teng, Y. H.; Berger, W. T.; Nesbitt, N. M.; Kumar, K.; Balius, T. E.; Rizzo, R. C.; Tonge, P. J.; Ojima, I.; Swaminathan, S., Computer-aided identification, synthesis, and biological evaluation of novel inhibitors for botulinum neurotoxin serotype A. *Bioorganic & Medicinal Chemistry* **2015**, *23* (17), 5489-95.
66. Zhou, Y.; McGillick, B. E.; Teng, Y. G.; Haranahalli, K.; Ojima, I.; Swaminathan, S.; Rizzo, R. C., Identification of small molecule inhibitors of botulinum neurotoxin serotype E via footprint similarity. *Bioorg Med Chem* **2016**, *24* (20), 4875-4889.
67. Guo, J.; Collins, S.; Miller, W. T.; Rizzo, R. C., Identification of a Water-Coordinating HER2 Inhibitor by Virtual Screening Using Similarity-Based Scoring. *Biochemistry* **2018**, *57* (32), 4934-4951.
68. Singleton, C. D.; Humby, M. S.; Yi, H. A.; Rizzo, R. C.; Jacobs, A., Identification of Ebola Virus Inhibitors Targeting GP2 Using Principles of Molecular Mimicry. *J Virol* **2019**, *93* (15), e00676-19.
69. Telehany, S. M.; Humby, M. S.; McGee, T. D., Jr.; Riley, S. P.; Jacobs, A.; Rizzo, R. C., Identification of Zika Virus Inhibitors Using Homology Modeling and Similarity-based Screening to Target Glycoprotein E. *Biochemistry* **2020**, *in revision (bi-2020-00458x)*.
70. Sterling, T.; Irwin, J., ZINC 15 - ligand discovery for everyone. *J Chem Inf Model* **2015**, *55*.

71. Balias, T. E.; Allen, W. J.; Mukherjee, S.; Rizzo, R. C., Grid-based Molecular Footprint Comparison Method for Docking and De Novo Design: Application to HIVgp41. *J. Comput. Chem.* **2013**, *34* (14), 1226-1240.
72. Balias, T. E.; Mukherjee, S.; Rizzo, R. C., Implementation and Evaluation of a Docking-rescoring Method Using Molecular Footprint Comparisons. *J. Comput. Chem.* **2011**, *32* (10), 2273-2289.
73. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., How Does a Drug Molecule Find Its Target Binding Site? *Journal of the American Chemical Society* **2011**, *133* (24), 9181-9183.
74. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696-3713.
75. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **2008**, *29* (4), 622-55.
76. Izadi, S.; Onufriev, A. V., Accuracy limit of rigid 3-point water models. *The Journal of Chemical Physics* **2016**, *145* (7), 074501.
77. Joung, I. S.; Cheatham, T. E., Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *The Journal of Physical Chemistry B* **2009**, *113* (40), 13279-13290.
78. Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B* **2008**, *112* (30), 9020-9041.
79. Case, D. A.; Belfon, K. A. A.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D.; Cheatham, T.; Cruzeiro, V. W. D. D., T.; ; Duke, R. E.; Giambasu, G.; Gilson, M.; Gohlke, H.; Götz, A.; Harris, R.; Izadi, S.; Измайлов, С. А.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Скрынников, Н. П.; Smith, J.; Swails, J.; Walker, R.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D.; Kollman, P. A., AMBER 2020, University of California, San Francisco. **2020**.
80. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2015**, *11* (4), 1864-1874.
81. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089-10092.
82. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14* (1), 33-8, 27-8.
83. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9* (7), 3084-3095.
84. Hunter, J. D., Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9* (3), 90-95.

85. Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P., Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **2020**.
86. Ghoreishi, D.; Cerutti, D. S.; Fallon, Z.; Simmerling, C.; Roitberg, A. E., Fast Implementation of the Nudged Elastic Band Method in AMBER. *Journal of Chemical Theory and Computation* **2019**, *15* (8), 4699-4707.
87. Jónsson, H.; Mills, G.; Jacobsen, K. W., Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, pp 385-404.
88. Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A. P.; Simmerling, C., Energetic Preference of 8-oxoG Eversion Pathways in a DNA Glycosylase. *Journal of the American Chemical Society* **2011**, *133* (37), 14504-14506.
89. Kuznetsov, N. A.; Bergonzo, C.; Campbell, A. J.; Li, H. Q.; Mechetin, G. V.; de los Santos, C.; Grollman, A. P.; Fedorova, O. S.; Zharkov, D. O.; Simmerling, C., Active destabilization of base pairs by a DNA glycosylase wedge initiates damage recognition. *Nucleic Acids Research* **2015**, *43* (1), 272-281.
90. Lai, C. T.; Li, H. J.; Yu, W. X.; Shah, S.; Bommineni, G. R.; Perrone, V.; Garcia-Diaz, M.; Tonge, P. J.; Simmerling, C., Rational Modulation of the Induced-Fit Conformational Change for Slow-Onset Inhibition in Mycobacterium tuberculosis InhA. *Biochemistry* **2015**, *54* (30), 4683-4691.
91. Li, H. J.; Lai, C. T.; Pan, P.; Yu, W. X.; Liu, N. N.; Bommineni, G. R.; Garcia-Diaz, M.; Simmerling, C.; Tonge, P. J., A Structural and Energetic Model for the Slow-Onset Inhibition of the Mycobacterium tuberculosis Enoyl-ACP Reductase InhA. *Acs Chem Biol* **2014**, *9* (4), 986-993.
92. Li, H. Q.; Endutkin, A. V.; Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A.; Zharkov, D. O.; Simmerling, C., A dynamic checkpoint in oxidative lesion discrimination by formamidopyrimidine-DNA glycosylase. *Nucleic Acids Research* **2016**, *44* (2), 683-694.
93. Li, H. Q.; Endutkin, A. V.; Bergonzo, C.; Fu, L.; Grollman, A.; Zharkov, D. O.; Simmerling, C., DNA Deformation-Coupled Recognition of 8-Oxoguanine: Conformational Kinetic Gating in Human DNA Glycosylase. *Journal of the American Chemical Society* **2017**, *139* (7), 2682-2692.
94. Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. In *Steered Molecular Dynamics*, Berlin, Heidelberg, Springer Berlin Heidelberg: Berlin, Heidelberg, 1999; pp 39-65.
95. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., Multidimensional free-energy calculations using the weighted histogram analysis method. **1995**, *16* (11), 1339-1350.
96. Grossfield, A. WHAM: an implementation of the weighted histogram analysis method. Version 2.0.10. <http://membrane.urmc.rochester.edu/content/wham/>.
97. Shimazaki, H.; Shinomoto, S., A method for selecting the bin size of a time histogram. *Neural computation* **2007**, *19* (6), 1503-27.
98. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X., MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics (Oxford, England)* **2011**, *27* (23), 3276-85.

99. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25* (13), 1605-12.
100. DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R., Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem* **1988**, *31*, 722-729.
101. Meng, E.; Shoichet, B.; Kuntz, I., Automated Docking with Grid-Based Energy Evaluation. *Journal of Computational Chemistry* **1992**, *13*, 505-524.
102. Mukherjee, S.; Balius, T. E.; Rizzo, R. C., Docking validation resources: Protein family and ligand flexibility experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986-2000.
103. Case, D. A.; Cerutti, D. S.; Cheatham, T. E., 3rd; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., AMBER 16. *University of California, San Francisco* **2016**.
104. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **2000**, *21* (2), 132-146.
105. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **2002**, *23* (16), 1623-41.
106. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004**, *25* (9), 1157-74.
107. Cagliani, R.; Forni, D.; Clerici, M.; Sironi, M., Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2. *Journal of Virology* **2020**, *94* (12), e00411-20.
108. Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W., GenBank. *Nucleic Acids Res* **2013**, *41* (Database issue), D36-42.
109. Elbe, S.; Buckland-Merrett, G., Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **2017**, *1* (1), 33-46.
110. Notredame, C.; Higgins, D. G.; Heringa, J., T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **2000**, *302* (1), 205-17.
111. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res* **2004**, *14* (6), 1188-90.