

# The inherent flexibility of receptor binding domains in SARS-CoV-2 spike protein

Hisham M Dokainish<sup>1</sup>, Suyong Re<sup>2,3</sup>, Takaharu Mori<sup>1</sup>, Chigusa Kobayashi<sup>4</sup>, Jaewoon Jung<sup>1,4</sup>, Yuji Sugita<sup>1,3,4\*</sup>

<sup>1</sup>Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako, Japan; <sup>2</sup>Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan; <sup>3</sup>Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Kobe, Japan; <sup>4</sup>Computational Biophysics Research Team, RIKEN Center for Computational Science, Kobe, Japan

**Abstract** Spike (S) protein is the primary antigenic target for neutralization and vaccine development for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It decorates the virus surface and undergoes large motions of its receptor binding domains (RBDs) to enter the host cell. Here, we observe Down, one-Up, one-Open, and two-Up-like structures in enhanced molecular dynamics simulations, and characterize the transition pathways via inter-domain interactions. Transient salt-bridges between RBD<sub>A</sub> and RBD<sub>C</sub> and the interaction with glycan at N343<sub>B</sub> support RBD<sub>A</sub> motions from Down to one-Up. Reduced interactions between RBD<sub>A</sub> and RBD<sub>B</sub> in one-Up induce RBD<sub>B</sub> motions toward two-Up. The simulations overall agree with cryo-electron microscopy structure distributions and FRET experiments and provide hidden functional structures, namely, intermediates along Down-to-one-Up transition with druggable cryptic pockets as well as one-Open with a maximum exposed RBD. The inherent flexibility of S-protein thus provides essential information for antiviral drug rational design or vaccine development.

\*For correspondence: sugita@riken.jp

Competing interest: See page 19

Funding: See page 19

Preprinted: 06 August 2021

Received: 20 November 2021

Accepted: 15 March 2022

Published: 24 March 2022

Reviewing Editor: Qiang Cui, Boston University, United States

© Copyright Dokainish et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Editor's evaluation

Using extensive molecular dynamics simulations with a novel enhanced sampling technique, the authors are able to characterize the structural flexibility of the SARS-CoV2 spike protein and identify new conformational states. These insights will be valuable to the design of novel strategies that modulate the interactions of the spike protein during the infection process.

## Introduction

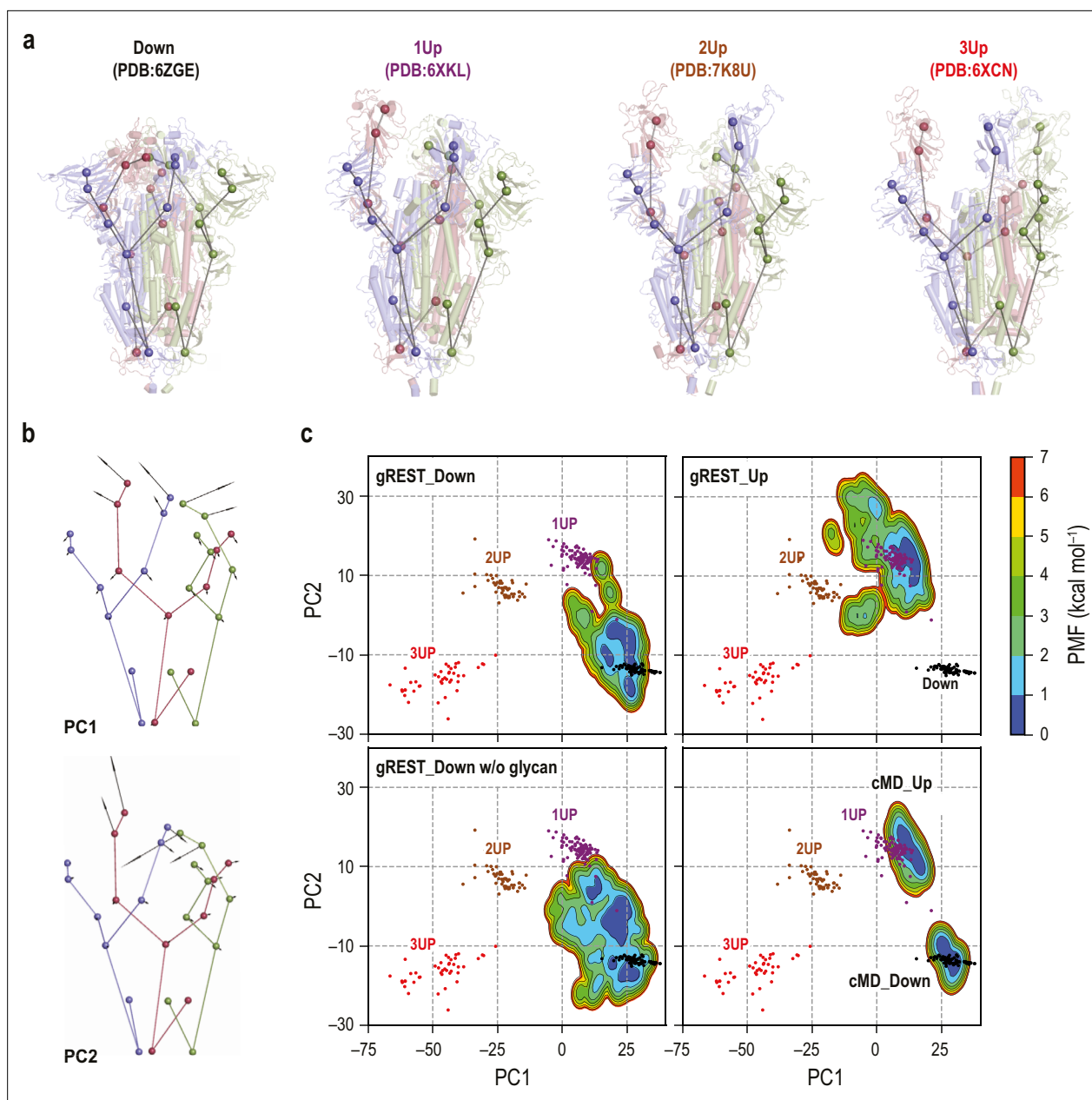
The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused over 430 million infections and 5.9 million deaths, as of February 2022 (<https://coronavirus.jhu.edu/map.html>). It represents an urgent need for an effective medical intervention strategy to avoid further social and economic consequences (Hu et al., 2021). Different types of vaccines, for example, those from Pfizer-BioNTech or Moderna using the mRNA of spike (S) protein, are currently available, and there are several FDA approved drug candidates under consideration (Li et al., 2021; Venkadapathi et al., 2021; Alvarado et al., 2022). At the same time, more infectious mutants such as B.1.617.2 (Delta), B.1.427/B.1.429 (Epsilon), and more recently Omicron have appeared (McCallum et al., 2021; Cao et al., 2022), and some evade from the immune system (Yurkovetskiy et al., 2020; Harvey et al., 2021; Wang et al.,

2021; Gobeil et al., 2021). Furthermore, the virus's ability to infect a wide range of vertebrates, which could act as a reservoir, points out the future risk despite the vaccination progress (Prince et al., 2021). A deeper understanding of the virus molecular structure and infection mechanism is crucial to stop the virus transmission including mutant strains (V'kovski et al., 2021).

SARS-CoV-2, an enveloped positive single-stranded RNA virus, has a large genome of approximately 30 kb encoding 29 proteins (V'kovski et al., 2021; Finkel et al., 2021). A transmembrane homotrimeric class I fusion glycoprotein decorating the virus, known as S-protein, plays a critical role in the viral cell entry (Ke et al., 2020; Shang et al., 2020). In an immediate response to the pandemic, more than 300 S-protein structures determined with cryo-electron microscopy (EM) and X-ray crystallography have been reported and rapidly advanced in our understanding of the S-protein/receptor binding mechanisms (Wrobel et al., 2020; Hsieh et al., 2020; Barnes et al., 2020a; Wrapp et al., 2020; Cai et al., 2020; Wang et al., 2020; Yan et al., 2020). The N-terminal subunit (S1), which is composed of the N-terminal domain (NTD), the receptor binding domain (RBD), and two other subdomains (SD1 and SD2) (Figure 1—figure supplement 1; Wrapp et al., 2020; Walls et al., 2020), initially binds to the host cell receptor angiotensin-converting enzyme 2 (ACE2) (Yan et al., 2020). This binding is followed by priming of the C-terminal subunit (S2) and its large conformational changes leading to the membrane fusion for the cell entry (Shang et al., 2020; Fan et al., 2020). S-protein is covered by 66 N-glycans, 22 per protomer, to evade from the host cell immune system (Grant et al., 2020; Watanabe et al., 2020). To block the initial binding with ACE2, vaccines for stimulating the immune system or antibodies/small-molecule drugs for neutralizing the virus is the primary target for medical interventions. Numerous cryo-EM structures have revealed that RBDs can take Down or several Up forms including one-Up, two-Up, and three-Up conformations (Figure 1a; Wrobel et al., 2020; Hsieh et al., 2020; Barnes et al., 2020a; Wrapp et al., 2020; Cai et al., 2020; Fan et al., 2020; Barnes et al., 2020b; Juraszek et al., 2021). Up forms are accessible to the ACE2 receptor in the cell entry, while the Down is inaccessible (Wrapp et al., 2020; Yan et al., 2020; Xu et al., 2021). Note that most of the cryo-EM structures representing Up forms were determined together with other proteins, such as antibodies or a fragment of ACE2 (Supplementary file 1A, Table 1). A recent single molecule fluorescence resonance energy transfer (smFRET) experiment suggested that the RBD Down-to-Up transition occurs even without any ligand of S-protein, and at least one transient intermediate exists during the transition (Lu et al., 2020). These experiments suggest the potential flexibility of S-protein conformation and the possible involvement of multiple intermediates along transition pathways. The exploration of a wide conformational space of S-protein is necessary for characterizing the transition pathways, which give us better chances to find cryptic drug/antibody binding sites otherwise elusive.

Molecular dynamics (MD) simulations at the atomic level have been conducted to explore conformational dynamics of S-protein including the glycans at its surface, which are largely missed in the cryo-EM structures. Several microsecond-scale simulations showed the conformational flexibility of the stalk region in S-protein, different levels of glycan shielding between Down and Up, and their relevance in the ACE2 binding (Grant et al., 2020; Turoňová et al., 2020; Choi et al., 2021; Pavlova et al., 2021). They also identified specific interactions between side-chain residues or those with glycans to stabilize either Down or Up structure (Casalino et al., 2020; Mori et al., 2021). However, the timescale of the Down-to-Up transition in S-protein is far slower than that is attainable in the current MD simulations, leaving the inherent flexibility and transition pathways largely unknown. Therefore, several enhanced sampling methods including targeted MD (Mori et al., 2021), steering MD (Gur et al., 2020), nudged elastic band/umbrella sampling (Simmerling et al., 2020), two-dimensional umbrella sampling (Pang et al., 2021), adaptive sampling (Zimmerman et al., 2021), and weighted ensemble methods have been applied to investigate the RBD transition from Down to one-Up (Sztain et al., 2021). In these simulations, pre-defined reaction coordinates and/or bias potentials along the coordinates were used to enhance the motions of a single RBD, which limit the number of RBDs to be investigated simultaneously.

Here, we apply the generalized replica exchange with solute tempering of selected surface charged residues (gREST\_SSCR) (Dokainish and Sugita, 2021), which can enhance domain motions of a protein by exchanging the solute temperature of selected surface residues between neighboring replicas like the original gREST method (Kamiya and Sugita, 2018). gREST\_SSCR is distinct from the other simulations applied to S-protein in that no reaction coordinates as well as bias potentials are used, allowing us to examine the inherent flexibility of S-protein involving more than one RBD.



**Figure 1.** Comparisons of spike (S) protein structures in molecular dynamics (MD) simulations with cryo-electron microscopy (cryo-EM) structures. **(a)** Twenty-seven coarse-grained beads representations of four representative cryo-EM structures: Down (PDB ID: 6ZGE), 1Up (6XKL), 2Up (7K8U), and 3Up (6XC�N). Chains A, B, and C in S-protein are shown in red, blue, and green, respectively. **(b)** The lowest two modes (PC1 and PC2) in principal component analysis of the 27-beads model of 289 cryo-EM structures. PC1 and PC2 respectively represent a symmetric and an anti-symmetric Down-to-Up motion of the receptor binding domains. The vector direction of PC1 is reversed so that it points from Down to Up for better visualization. The vectors are magnified 100 times for clarity. **(c)** Free-energy landscapes at 310 K along the PC1 and PC2 obtained from the simulations: (top) gREST\_SSCR simulations with glycans starting from Down (500 ns) and Up (300 ns), (bottom left) gREST\_SSCR simulation without glycans from Down (150 ns), (bottom right) conventional MD simulations with glycans starting from either Down or Up. The positions of cryo-EM structures are also shown for comparison. Wherein, Down, 1Up, 2Up, and 3Up conformations are shown in black, purple, brown, and red dots, respectively.

The online version of this article includes the following video and figure supplement(s) for figure 1:

**Figure supplement 1.** Structural model of the spike protein used in this study.

**Figure supplement 2.** Glycans in our spike protein models.

**Figure supplement 3.** Performance of the gREST\_SSCR simulations.

**Figure supplement 4.** Characterization of the receptor binding domain (RBD) conformational change in the three gREST\_SSCR simulations.

Figure 1 continued on next page

Figure 1 continued

**Figure supplement 5.** Analysis of the intra-domain stability of receptor binding domain (RBD) and N-terminal domain (NTD) in the gREST\_SSCR simulations.

**Figure supplement 6.** Scheme of the protomer rotation and analysis of the rotated trajectories of the gREST\_Down simulation.

**Figure supplement 7.** Free-energy landscape (FEL) along the hinge/twist angles in the gREST\_Down w/o glycan and gREST\_Up simulations.

**Figure 1—video 1.** Down-to-1Up transition of replica 16 in gREST\_Down simulation.

<https://elifesciences.org/articles/75720/figures#fig1video1>

**Figure 1—video 2.** 1Up-to-2Up transition of replica 4 in gREST\_Up simulation.

<https://elifesciences.org/articles/75720/figures#fig1video2>

**Figure 1—video 3.** Down-to-1Up transition of replica 6 in gREST\_Down w/o glycan simulation.

<https://elifesciences.org/articles/75720/figures#fig1video3>

As listed in **Table 1**, the RBD motion in the RBD/SD1 monomer as well as the trimeric S-protein with and without glycans is systematically investigated. The intrinsic flexibility of S-protein observed in the enhanced sampling MD simulations suggests new mechanisms underlying the Down-to-Up transitions, glycan shielding for the binding with antibodies, and unprecedented cryptic pockets in the intermediates between Down and one-Up.

## Results

### Enhanced RBD motions are observed in gREST\_SSCR simulations

**Table 1** lists all simulations performed in this study (for details, SI methods, **Figure 1—figure supplements 1 and 2**). Eight pairs of positively and negatively charged residues at RBD interface in each chain were selected as solute in gREST\_SSCR and their Coulomb and Lennard-Jones interactions were scaled using solute temperatures in 16 copies of the original systems (**Figure 1—figure supplement 3**). After modeling missing residues, two simulations were started from a Down cryo-EM structure (PDBID: 6VXX; **Walls et al., 2020**) in the presence (gREST\_Down) and absence of glycan (gREST\_Down w/o glycan), and one from an one-Up structure (PDBID: 6VYB; **Walls et al., 2020**) (gREST\_Up). MD simulations for longer than 15  $\mu$ s, in total, were carried out for the solvated S-protein including about 655,000 atoms. Analyses of SD1/RBD hinge angle and the C $\alpha$  root mean square deviation (RMSD) of each RBD upon fitting S2 provide simple measures of the RBD Down-to-Up transitions: the RMSD and hinge angle in Up take the values of about 20 Å and 150°, respectively, while the hinge angle in Down is about 116°. **Figure 1—figure supplement 4** indicates that RBD motions are drastically enhanced in all three simulations. Since gREST\_SSCR is free from any pre-defined reaction coordinates, the Down-to-Up transitions happened not only in RBD<sub>A</sub> but also in RBD<sub>B</sub> or RBD<sub>C</sub>. Even two-Up-like conformations were observed in some replicas, for instance, replica 8 in gREST\_Down, replica 4 in gREST\_Up, and replica 16 in gREST\_Down w/o glycan (**Figure 1—figure supplement 4**).

**Table 1.** Molecular dynamics (MD) simulations of spike (S) protein performed in this study.

Name	Model	Method	Simulations length
gREST_Down	Spike Down w/ glycans	gREST_SSCR	500 ns $\times$ 16 replicas
gREST_Up	Spike Up w/ glycans	gREST_SSCR	300 ns $\times$ 16 replicas
gREST_Down w/o glycan	Spike Down w/o glycans	gREST_SSCR	150 ns $\times$ 16 replicas
Monomer_Down	RBD/SD1 monomer Down	cMD	300 ns $\times$ 1 run
Monomer_Up	RBD/SD1 monomer Up	cMD	300 ns $\times$ 2 runs
cMD_Down*	Spike Down w/ glycans	cMD	1000 ns $\times$ 1 run
cMD_Up*	Spike Up w/ glycans	cMD	1000 ns $\times$ 1 run

\*cMD\_Down and cMD\_Up are the same simulations as shown in our previous study **Mori et al., 2021**.

Despite the large-scale motions, intra-domain structures of three RBDs and NTDs were kept stable (*Figure 1—figure supplement 5*).

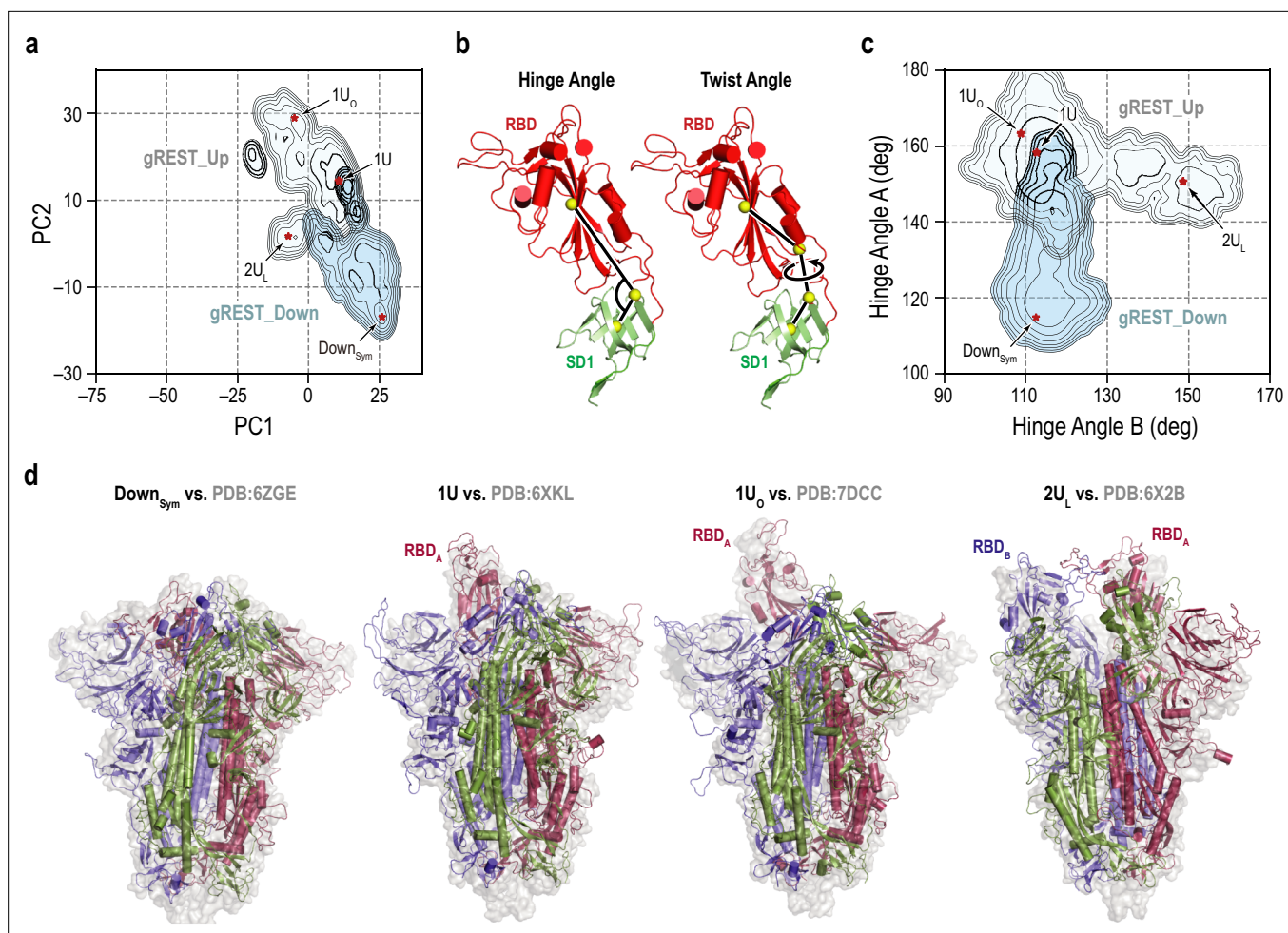
## Simulated structures connect cryo-EM ensembles

As of September 2021, over 300 cryo-EM structures of S-protein, including all Down, one-Up, two-Up, and three-Up conformations, were deposited in Protein Data Bank (*Supplementary file 1A*). Considering the trimeric nature of S-protein and its complex motions, the comparison with experimental structures is not trivial. For this purpose, we introduce (1) the nine-beads representation per protomer as used in the studies by Henderson et al. (*Gobeil et al., 2021; Henderson et al., 2020a*) (the 27-beads model of S-protein) (*Figure 1a* and *Supplementary file 1B*), (2) the rotation scheme to make the most significant RBD motion always happen in Chain<sub>A</sub> (*Figure 1—figure supplements 6 and 7*), and (3) the principal component analysis (PCA) on the 27-beads model upon fitting all the beads to reduce the essential dimensions in the conformational space. In *Figure 1b*, the first principal component (PC1) represents a symmetric Down-to-Up motion involving all three RBDs, while the second component (PC2) reveals an asymmetric motion of RBDs where only RBD<sub>A</sub> undergoes the Down-to-Up motion. The two lowest PCs cover about 85% of the conformational variations observed in the cryo-EM structures. In *Figure 1c*, the cryo-EM structures with distinct RBD conformations, Down, one-Up, two-Up, and three-Up, are found in different ensembles on the PC1-PC2 space (*Figure 1c*).

Next, we project the results of our previous 1  $\mu$ s conventional MD (cMD) simulations (*Mori et al., 2021*) (cMD\_Down and cMD\_Up in *Table 1*) and the three gREST\_SSCR simulations at 310 K on the same space (*Figure 1c*). The gREST\_SSCR distributions at the temperature were obtained using multi-state Bennett acceptance ratio (MBAR) (*Shirts and Chodera, 2008*) by utilizing all the replicas' trajectories. The structure distributions in cMD\_Down and cMD\_Up overlap with the corresponding Down and one-Up cryo-EM structures, while there is a big gap between the distributions on the PC1-PC2 space. Instead, gREST\_SSCR could fill the gap effectively: gREST\_Down samples from Down to one-Up, while gREST\_Up covers one-Up and two-Up ensembles. Using the simulated distributions, we can investigate the inherent flexibility of S-protein, the transition pathways, and the intermediate structures. gREST\_Down w/o glycan gives a wider conformational space than gREST\_Down, suggesting that glycans on the surface of S-protein play key roles in the conformational stability.

To focus on the Down-to-Up motions in S-protein, we define the hinge and twist angles using the C $\alpha$  atoms in RBD and SD1 (*Figure 2b*). The hinge angle directly describes the Down-to-Up transitions, while the twist angle explains side motions of RBD. Larger values of the hinge and twist angles signify the transition toward Up forms. The gREST\_SSCR distributions at 310 K on the hinge-twist angle space (*Figure 2—figure supplement 1*) well overlap with those of the 891 protomers in the 300 cryo-EM structures, as we found on the PC1-PC2 space. Interestingly, the cMD simulation consisting of only a single monomer of SD1 and RBD in the Up form (Monomer\_Up in *Table 1*) sampled Down, one-Up, and one-Open states within 300 ns and gave the structure ensemble close to the cryo-EM structure distributions and the gREST simulations containing a whole S-protein.

On the Hinge<sub>A</sub>-Hinge<sub>B</sub> space, a second Down-to-Up motion is also identified for Chain<sub>B</sub> in gREST\_Up (*Figure 2c*). We did not observe similar motions on the Hinge<sub>A</sub>-Hinge<sub>C</sub> space, suggesting the order of Down-to-Up transitions toward two-Up. Taking together the ensemble distributions on the PC1-PC2 and the Hinge<sub>A</sub>-Hinge<sub>B</sub> spaces, four major structures were identified: Down symmetric (Down<sub>Sym</sub>: (Hinge<sub>A</sub>, Hinge<sub>B</sub>) = (114.9°, 112.5°)), one-Up (1U: (158.3°, 112.7°)), one-Up-open (1U<sub>O</sub>: (163.3°, 108.9°)), and two-Up-like (2U<sub>L</sub>: (150.6°, 148.6°)). They are superimposed to the cryo-EM structures having corresponding S-protein conformations in *Figure 2d*. Down<sub>Sym</sub> and 1U from our simulations are well aligned to the high-resolution cryo-EM structures of Down (PDB:6ZGE; *Wrobel et al., 2020* [RMSD: 3.7 Å]) and one-Up (PDB:6XKL (*Hsieh et al., 2020*) (RMSD: 3.6 Å)), respectively. Intriguingly, RBD<sub>A</sub> in 1U<sub>O</sub> is aligned to one of the RBDs in one of the three-Up structures (PDB:7DCC; *Zhang et al., 2021b* [RMSD: 5.4 Å]). Although 2U<sub>L</sub> from our simulations remains some interactions between RBD<sub>A</sub> and RBD<sub>B</sub>, which is completely lost in two-Up cryo-EM structures, 2U<sub>L</sub> is aligned with those in one of the two-Up structures (PDB:6X2B; *Henderson et al., 2020a* [RMSD: 5.8 Å]). The overall comparison of the simulations' structures to cryo-EM points out the quality of the predicted structures.



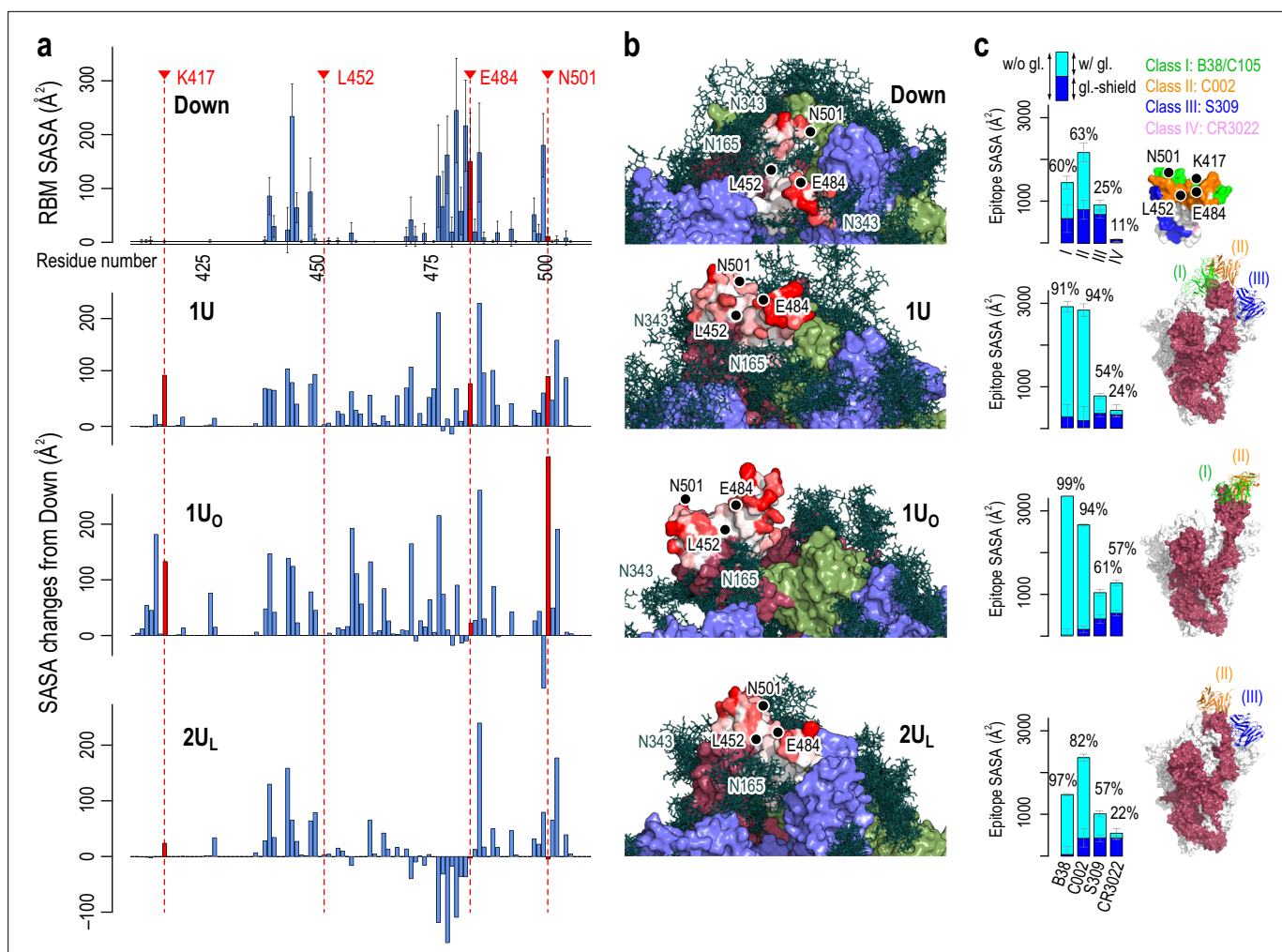
**Figure 2.** Representative receptor binding domain (RBD) conformations from molecular dynamics (MD) simulations vs. cryo-electron microscopy (cryo-EM) PDB structures. **(a)** An overlay of the two free-energy landscapes at 310 K along the PC1 and PC2 obtained from gREST\_Down (light blue) and gREST\_Up (light cyan) simulations. The red dots represent the positions of four representative RBD conformations: symmetric Down (Down<sub>Sym</sub>), 1RBD Up (1U and 1U<sub>O</sub>) and 2RBDs Up-like (2U<sub>L</sub>) conformations. **(b)** Definitions of the hinge and twist angles representing the RBD conformations. The hinge angle is determined by three centers of masses (COMs, yellow spheres) of the core and top residues of SD1 (green, C $\alpha$  atoms only) and the core residues of RBD (red, C $\alpha$  atoms only). The twist angle is determined by the aforementioned COMs with an extra COMs of the bottom residues of RBD. **(c)** An overlay of two free-energy landscapes at 310 K along the hinge angles in RBD<sub>A</sub> and RBD<sub>B</sub> obtained from gREST\_Down (light blue) and gREST\_Up (light cyan) simulations. **(d)** Four representative conformations from MD simulations (cartoon representation) in comparison with cryo-EM structures (light gray surface). Chains A, B, and C in spike (S) protein are shown in red, blue, and green, respectively.

The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** Free-energy landscape (FEL) along the hinge/twist angles in the gREST\_SSCR simulations in comparisons with the conventional molecular dynamics (cMD) simulations for receptor binding domain (RBD)/SD1 monomer and cryo-electron microscopy (cryo-EM) structures.

## The accessibility of RBD in different conformations

To get insights into the contribution of each conformation to ACE2 and neutralizing antibodies (nAbs) binding, the accessibility of RBD is examined in terms of the solvent accessible surface area (SASA) (*Casalino et al., 2020*). **Figure 3a and b** and **Figure 3—figure supplements 1–5** show the per-residue SASA around the receptor binding motif (residues I410 to V510, referred to as RBM hereafter) calculated for Down<sub>Sym</sub>, 1U, 1U<sub>O</sub> and 2U<sub>L</sub>. RBM is rarely exposed in Down<sub>Sym</sub>, where glycans at N165 and N343 largely shield RBM. In contrast, RBM SASA increases in 1U and 1U<sub>O</sub> compared to Down<sub>Sym</sub>, wherein 1U<sub>O</sub> exhibits the utmost increase, suggesting its potential contribution to the receptor and antibody binding. The SASA increases only slightly, even decreases locally around residue C480, in 2U<sub>L</sub> due to the interactions between RBD<sub>A</sub> and RBD<sub>B</sub>. This finding suggests that one-Up is the primary target of the receptor and antibody binding. Focusing on mutational residues of concern (K417, L452,



**Figure 3.** Accessibility of receptor binding motif (RBM). (a) Per-residue solvent accessible surface area (SASA) values of the RBM (residues 410–510) in Down conformation (top) and their changes in Up conformations (bottom three). SASA values were calculated using the probe radius of 7.2 Å. Four mutational residues, K417, L452, E484, and N501, are highlighted in red. (b) The surface representation of RBM SASA (white to red for 0–260 Å<sup>2</sup>, the maximum value in Down, the values higher than this are truncated for consistent color scheme). Chains A, B, and C in the protein are shown in red, blue, and green surfaces, respectively, while a collection of glycans from 10 snapshots are shown in stick representation. (c) Epitope SASA and glycan shielding of four types of neutralized antibodies, B38 (Class I), C002 (Class II), S309 (Class III), and CR3022 (Class IV). Sum of SASA with glycans (cyan) and the glycan shield (blue) gives SASA without glycans. The ratio of the SASA with glycan over that without glycans is shown. The right-most column shows the putative interaction models with three classes of antibodies: Class III (S309, PDBID: 6WPT, blue), Class II (C002, PDBID: 7K8T, orange), and Class I (C105, PDBID: 6XCM, green).

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Glycan shield of spike (S) protein.

**Figure supplement 2.** Accessibility of receptor binding motif (RBM).

**Figure supplement 3.** Glycan effect on the accessibility of receptor binding motif (RBM).

**Figure supplement 4.** Putative interaction models with antibodies.

**Figure supplement 5.** Comparison of Up structures from molecular dynamics (MD) simulations with cryo-electron microscopy (cryo-EM) structures.

E484, and N501 *Harvey et al., 2021*), K417 and N501 are accessible only in one-Up conformations. The mutations of these residues could directly affect the binding with the receptor and antibody. On the other hand, E484 is accessible in both Down and Up, inferring a widespread effect of E484 mutations. L452 is not accessible in any conformations and hence is expected to rarely affect the binding with either receptor or antibody in wild type (*Valdes-Balbin et al., 2021*).

To examine antibody preferences in Up conformations, we also calculate SASA for epitope regions that are recognized by four classes of nAbs (Class I: binds to Up for blocking ACE2, Class II: binds to Up and Down for blocking ACE2, Class III: binds outside RBM but recognize Up and Down, Class IV: binds Up without blocking ACE2) (Harvey *et al.*, 2021). The epitope regions of Up conformations are generally less shielded by glycans compared to Down (at most 99% de-shielded, Figure 3c). Notably, the epitope SASAs of Class I and II antibodies in 1U and 1U<sub>O</sub> are significantly large. The C105 (Class I) and C002 (Class II) antibodies in cryo-EM structures are well aligned to 1U and 1U<sub>O</sub> without steric clash (Figure 3c and see Figure 3—figure supplement 4 for detail). These results suggest that Class I and II antibodies preferentially bind to these two conformations. The corresponding SASA values are small in 2U<sub>L</sub> due to the interaction between RBD<sub>A</sub> and RBD<sub>B</sub>, leading the Class I antibody difficult to align without steric clash. Note that there are differences between 2U<sub>L</sub> and the full two-Up cryo-EM structures. The observation on SASA values in 2U<sub>L</sub> is applicable to the state just before reaching the full two-Up structures. In any Up conformations, the epitope SASAs of Class III antibody is comparable to that of Down, while the region is less shielded by glycans. The S309 (Class III) antibody in cryo-EM structures is still well aligned to 1U and 2U<sub>L</sub>, rationalizing that the Class III antibodies recognize ‘glyco-epitopes’ (Pinto *et al.*, 2020). Note that the epitope region of the Class IV antibody is slightly exposed only in 1U<sub>O</sub>. Further opening of RBD may allow the binding of the Class IV antibody, such as CR3022, as suggested in the previous computational study (Zimmerman *et al.*, 2021). All these results suggest the preferential binding of different antibodies depending on the RBD conformation. Notably, the Up structures from MD simulations are reasonably aligned to those of cryo-EM structures with antibodies (Figure 3—figure supplement 5). This observation supports an idea of conformer selection regime for antibody bindings. As the S-protein RBD is inherently flexible, potential conformations for antibody bindings are likely programmed.

### Simulated structure ensembles align with smFRET experimental results

Using smFRET, Lu *et al.* examined the conformational dynamics of S-protein in the presence or absence of its receptor, hACE2 (Lu *et al.*, 2020). They characterized four structural ensembles including two types of Down (major and minor), at least one intermediate and one-Up conformations, suggesting the inherent flexibility of the RBD region regardless of the receptor binding. However, to date, there is no structural information that describes the observed intermediate states in the smFRET experiment. In addition, despite the abundance of cryo-EM structures, they are unable to solely explain smFRET results as shown in Figure 4—figure supplement 4d. To unravel such inaccessible structural information of the intermediate states, we first classify each trajectory at 310 K using k-means clustering and re-clustering guided by the hinge and twist angles distributions (Figure 4—figure supplements 1–3, Supplementary file 1C). The 13 micro-clusters were identified from gREST\_Down (Figure 4—figure supplement 1) including two clusters representing the symmetric Down (Down<sub>sym</sub>: D1<sub>sym</sub> and D2<sub>sym</sub>), two clusters of asymmetric Down with differences in RBDs twist angle distribution (Down<sub>Asym</sub>: D1<sub>Asym</sub> and D2<sub>Asym</sub>), three clusters with a slight increase in the RBD hinge angle around 120° (Int1: I1a, I1b, and I1c), three clusters with one RBD hinge angle around 130° (Int2: I2a, I2b, and I2c), two clusters with hinge angle around 140° (Int3: I3a and I3b), and one cluster with an Up-like conformation (1U<sub>J</sub>). Similarly, from gREST\_Up, the 13 clusters were identified including 10 clusters representing one-Up conformation (1U: 1Ua–1Uj), two clusters with two-Up-like conformations (2U<sub>L</sub>: 2Ua<sub>L</sub> and 2Ub<sub>L</sub>) and one-Up open cluster (1U<sub>O</sub>). Then we compute the distance between residues 425–431 in RBD<sub>A</sub> and 554–561 in SD1<sub>A</sub> to correlate the simulated structure ensembles with the smFRET intensity reported by Lu *et al.*, 2020. By combining the distance distributions from gREST\_Down and gREST\_Up, we obtain five main conformational ensembles (Figure 4—figure supplement 4): Down symmetric (Down<sub>sym</sub>), Down like (Down<sub>Like</sub>: Down<sub>Asym</sub> and Int1), two intermediates (Int2 and Int3), and Up (1Up). The smFRET distance alone cannot distinguish various Up conformations including 1U, 1U<sub>O</sub>, and 2U<sub>L</sub> and thus we refer to them as 1Up. Down<sub>sym</sub> and Down<sub>Like</sub> give distributions in the range of 30–35 Å, while 1Up shows around a median distance of 47 Å. The distributions of Int2 and Int3 have median distances of around 38 and 40 Å, respectively. Note that Int3 has a large distance distribution that overlaps with Int2 and both intermediates might be indistinguishable in the smFRET experiment. Collectively our simulations show the formation of four main conformations Down<sub>sym</sub>, Down<sub>Like</sub>, Intermediate (Int2 and Int3), and 1Up that correspond to the smFRET intensity of 0.8, 0.5, 0.3, and 0.1 in the absence of hACE2 (Lu *et al.*, 2020). Each ensemble is also characterized with hinge angle distributions that align



in order of smFRET distances: Down<sub>Sym</sub>: Hinge < 120°, Down<sub>Like</sub>: Hinge < 130°, Int2: 120° < Hinge < 140°, Int3: Hinge ~140°, 1Up: 140° < Hinge < 160°.

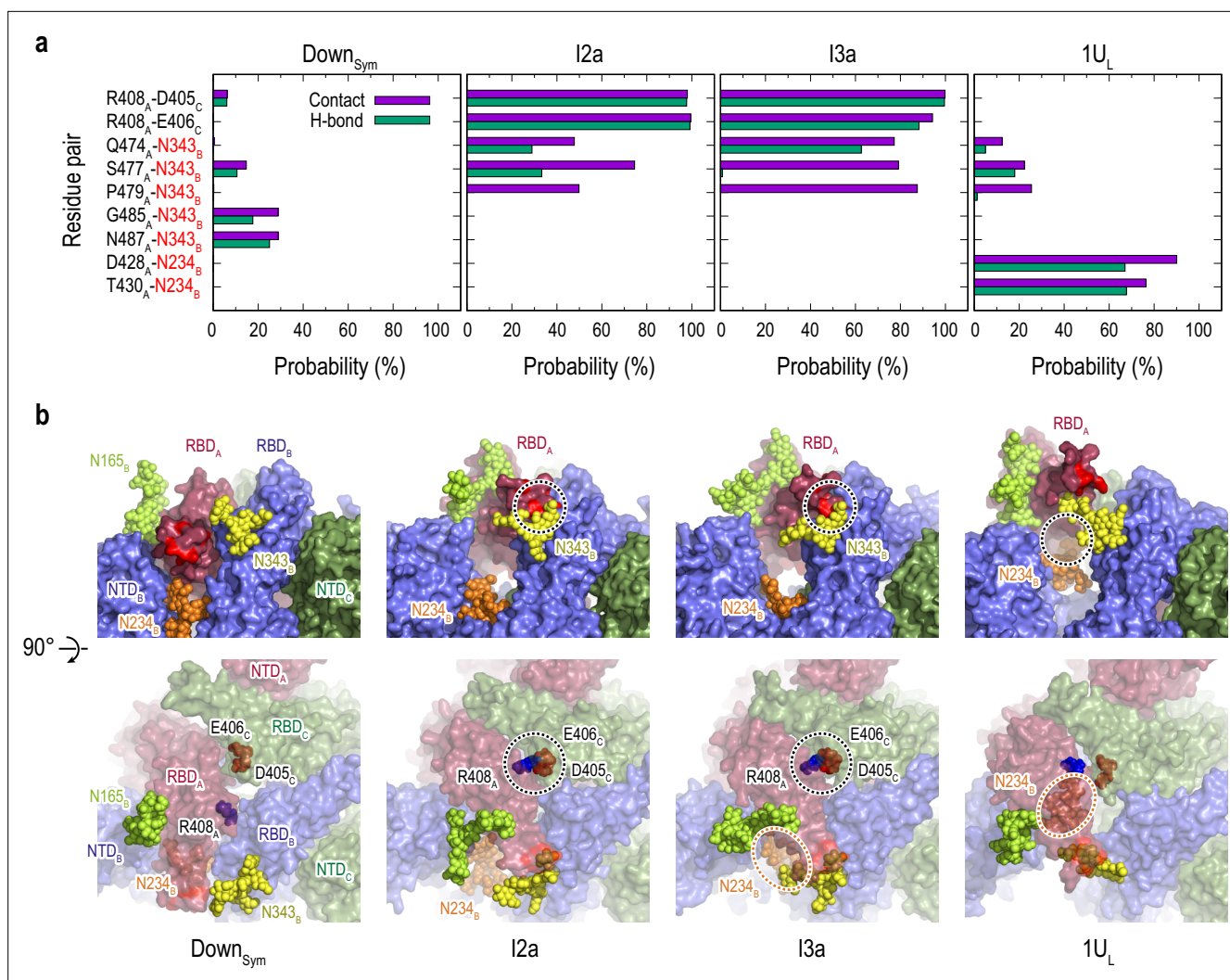
## Transition pathways and transient interactions stabilizing the intermediate structures

We next focus on molecular mechanisms underlying the Down-to-1Up transitions in terms of the correlated motions of the hinge and twist angles in the three RBDs (**Figure 4—figure supplement 5**). Note that such correlated motions are hardly obtained using targeted MD simulations (**Figure 4—figure supplement 6**) or the MD simulations enhanced with pre-defined reaction coordinates (**Sztain et al., 2021; Brotzakis et al., 2021**). **Figure 4—figure supplement 5** indicates that the transition pathway from Down-to-1Up conformation occurs via four main states including two intermediates. First, the transition is initiated from flexible Down structures (Down<sub>Sym</sub> and Down<sub>Like</sub>), where three RBDs show high flexibilities reflected by a wide range of twist angle between 10° and 90° while RBD<sub>A</sub> maintains a hinge angle <125°. The increase in RBD<sub>A</sub> hinge angle to 130° in I2a is accompanied by the reduction in other RBDs' flexibility. It is reflected by smaller changes in twist and hinge angles for the chains B and C. Later, two main pathways are identified through I3a or I3b (hinge 140°), however, comparisons of all five free-energy maps as well as the projection on PC1-PC2 free-energy surface suggest that I3b is an off-target intermediate. In I3a, both hinge and twist angles are increased in RBD<sub>A</sub>, accompanied by a slight reduction in RBD<sub>C</sub> hinge angle. Both I2a and I3a represent stable basins along the free-energy landscapes (FELs) along the PC1-PC2 and hinge-twist angles, emphasizing their role in mediating the transition (**Figure 4—figure supplements 5 and 11**). Lastly, 1U<sub>L</sub> conformation is formed with hinge angle >150° with a concomitant increase in its twist angle. In general, Hinge<sub>A</sub> and Twist<sub>A</sub> are highly correlated (cc: 0.94) with each other (**Figure 4—figure supplement 5c**), while Hinge<sub>A</sub> and Hinge<sub>B</sub> show no correlations (cc: 0.02). The correlations between Hinge<sub>A</sub> and Hinge<sub>C</sub>/Twist<sub>C</sub> exist to some extent (cc: -0.68 and 0.30).

To further analyze the changes throughout the transition pathway, the analysis of contact (**Figure 4—figure supplement 7**) and hydrogen bond (HB) (**Figure 4—figure supplement 8**) were carried out for each cluster observed in gREST\_Down and gREST\_Up as we did in the previous study (**Mori et al., 2021**). In **Figure 4**, several key contacts and HBs are highlighted to see drastic changes of the interactions in the Down-to-1Up transition. The glycans at N343<sub>B</sub> and N234<sub>B</sub> switch the interactions with RBD<sub>A</sub> along the transition from Down to one-Up-like. In the intermediates, the glycan at N343<sub>B</sub> changes its contact partners by inserting underneath RBD<sub>A</sub>. Concurrently, the formation of salt-bridge interactions between RBD<sub>A</sub> and RBD<sub>C</sub>, for instance, R408<sub>A</sub>-D405<sub>C</sub> and R408<sub>A</sub>-D406<sub>C</sub>, lift up RBD<sub>A</sub> from its Down to intermediate structures. Finally, N234<sub>B</sub> inserts into the newly formed cavity in 1U<sub>L</sub> between RBD<sub>A</sub> and RBD<sub>B</sub>, forming new contacts with D428<sub>A</sub> and T430<sub>A</sub>. The glycan at N165<sub>B</sub> is in contact with RBD<sub>A</sub> throughout the transition, likely acting as a main barrier.

The formation of two Up-like conformation is characterized by the increase of hinge/twist angle in RBD<sub>B</sub>, as shown in **Figure 1—figure supplement 7c**. Free energy landscape (FEL) on Hinge<sub>A</sub>-Twist<sub>A</sub> angles shows a similar transition as observed in RBD<sub>A</sub> in gREST\_Down (**Figure 1—figure supplements 7c and 6d**). The relative RBD motions during the 1Up-to-2Up transition is characterized using the hinge angles (**Figure 5—figure supplement 1a**). The increase of hinge angle in RBD<sub>B</sub> is accompanied with a slight decrease of hinge angle in RBD<sub>A</sub> in 2U<sub>A</sub><sub>L</sub> and subsequently in 2U<sub>B</sub><sub>L</sub>. Note that both clusters might represent intermediate states toward the full 2Up structure, in which the increase of RBD<sub>A</sub> hinge angle is regained. The latter decrease likely relates with the increases of the contacts and HBs between RBD<sub>A</sub> and RBD<sub>C</sub>, for instance, K378<sub>A</sub>-E484<sub>C</sub>, Y369<sub>A</sub>-N487<sub>C</sub>, and D427<sub>A</sub>-Y505<sub>C</sub>, from 1U<sub>A</sub>, the top populated cluster in 1U, to 2U<sub>L</sub> structures (**Figure 5** and **Figure 4—figure supplements 7 and 8**). The numbers of contacts and HBs in two-Up-like structures are much less than those in 1U<sub>A</sub>, suggesting that a drastic reduction of the inter-domain interactions between RBD<sub>A</sub> and RBD<sub>B</sub> is required to form a full two-Up conformations. No specific protein-glycan interactions are found to support the 1Up-to-2Up transition, while more contacts between RBD<sub>B</sub> and glycan at N165<sub>C</sub> are observed in the two-Up-like conformations, emphasizing its role as a bottleneck of the transition.

We also examine the effect of glycans on the Down-to-1Up transitions. **Figure 5—figure supplement 1b** shows that the distributions of gREST\_Down w/o glycan on the Hinge<sub>A</sub>-Hinge<sub>B</sub> space become much wider, suggesting that S-protein is more flexible without glycans. In addition, Down conformations are more diverse and asymmetric as one RBD has larger hinge angle distributions than the



**Figure 4.** Protein-protein and protein-glycan interactions critical for Down-to-Up transition. **(a)** Probability of finding the hydrogen bond (green) and contact (purple) pairs between protein residues or protein-glycans that markedly change along the transition pathway (Down<sub>Sym</sub>, I2a, I3a, and 1U<sub>L</sub>). All hydrogen bond (probability of finding of >50%) and contact pairs (probability of finding of >70%) are shown in **Figure 4—figure supplements 7 and 8**, respectively. **(b)** Typical snapshots of the protein-protein and protein-glycan interactions along the transition pathway. Chains A, B, and C in the protein are shown in red, blue, and green surfaces, respectively. Glycans at N165, N234, and N343 are shown with spheres in lime, orange, and yellow color, respectively. The transient N343<sub>B</sub>-RBD<sub>A</sub> contact is highlighted in red surface (top). The salt-bridges formed by R408<sub>A</sub> (blue), E406<sub>C</sub> (brown), and D405<sub>C</sub> (red) are also highlighted with black dashed circles (top and bottom), while the location of N234 glycan is highlighted with orange dashed circles.

The online version of this article includes the following video and figure supplement(s) for figure 4:

**Figure supplement 1.** Clustering for the conformations obtained from the gREST\_Down simulation.

**Figure supplement 2.** Clustering for the conformations obtained from the gREST\_Up simulation.

**Figure supplement 3.** Clustering for the conformations obtained from the gREST\_Down w/o glycan simulation.

**Figure supplement 4.** Simulated single molecule fluorescence resonance energy transfer (smFRET) distance using the gREST\_SSCR trajectory data.

**Figure supplement 5.** Transition pathway from Down to 1Up in the gREST\_Down simulation.

**Figure supplement 6.** Comparison between targeted molecular dynamics (TMD) and gREST\_SSCR simulations.

**Figure supplement 7.** Contact analysis for the main clusters in gREST\_Down and gREST\_Up simulations.

**Figure supplement 8.** Hydrogen bond analysis for the main clusters in gREST\_Down and gREST\_Up simulations.

**Figure supplement 9.** Relationship between the sideways motion of RBD<sub>B</sub> and intrusion of the glycan at N343<sub>B</sub>.

**Figure supplement 10.** Glycan interaction sites in various Up conformations.

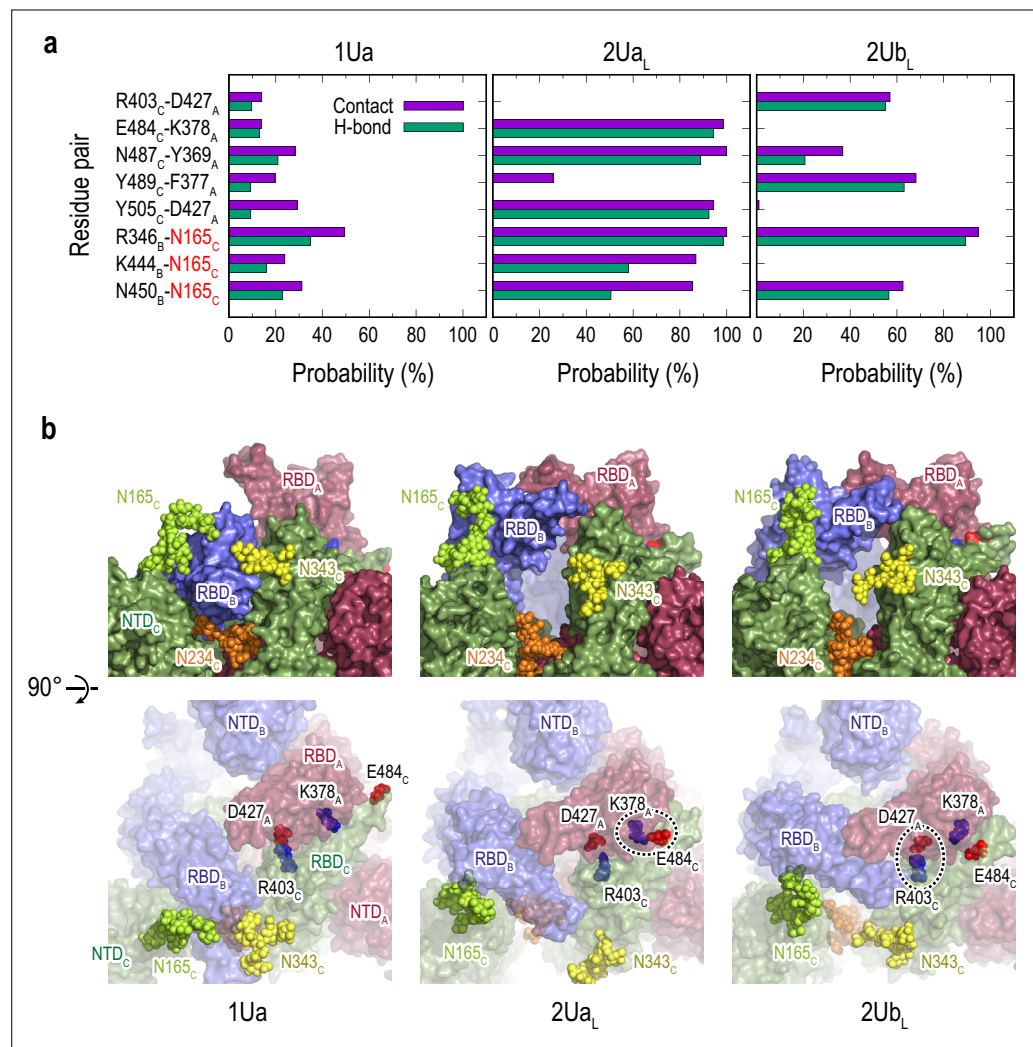
**Figure supplement 11.** Stability of I2a and I3a along transition pathway.

Figure 4 continued on next page

Figure 4 continued

**Figure supplement 12.** Intermediate structures align with cryo-electron microscopy (cryo-EM) structures.**Figure 4—video 1.** Protein-glycans interactions along Down-to-1Up transition pathway.<https://elifesciences.org/articles/75720/figures#fig4video1>

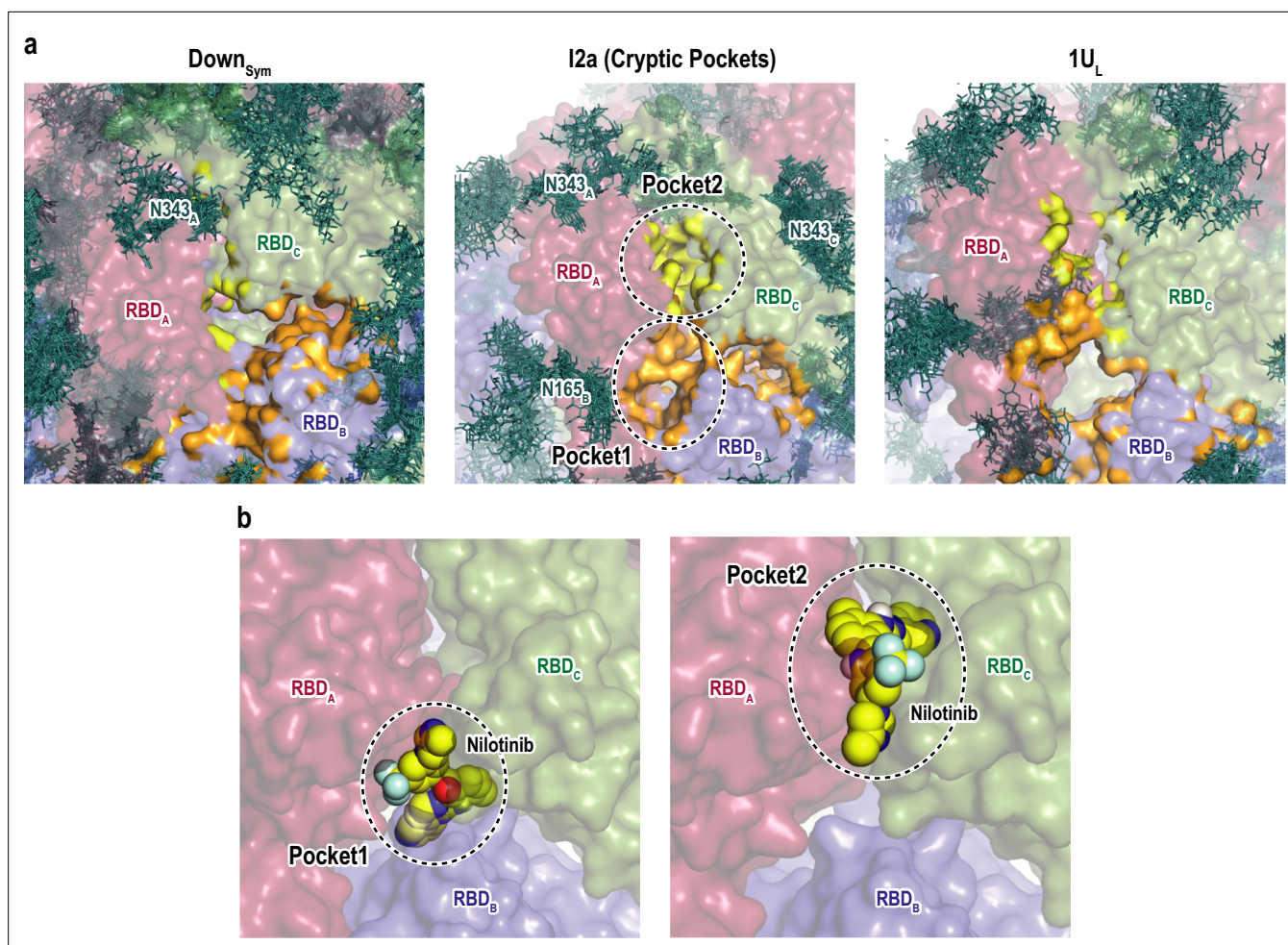
others. The transition mechanisms seem to be the same as those with glycans: RBD<sub>A</sub> and RBD<sub>C</sub> form the transient interactions to support the Down-to-1Up transition. However, the two-Up-like structure distribution is directly connected with Down, suggesting that structural integrity of S-protein is lost without glycans.



**Figure 5.** Selected protein-protein and protein-glycan interactions along the 1Up-to-2Up transition. **(a)** Probability of finding the hydrogen bond (green) and contact (purple) pairs throughout the transition pathway (1Ua, 2Ua<sub>L</sub>, and 2Ub<sub>L</sub>). Where the conformational transition of RBD<sub>B</sub> induces RBD<sub>A</sub>/RBD<sub>C</sub> interactions. All hydrogen bonds and contact pairs are shown in **Figure 4—figure supplements 7 and 8** respectively. **(b)** Typical snapshots of the protein-protein and protein-glycan interactions along RBD<sub>B</sub> transition pathway. Chains A, B, and C in the protein are shown in red, blue, and green surfaces, respectively. Glycans at N165, N234 and N343 are shown with spheres in lime, orange, and yellow colors, respectively. Positive and negatively charged residues are shown as blue and red spheres, respectively. The transient salt-bridges formed by K378<sub>A</sub>/E484<sub>C</sub> and D427<sub>A</sub>/R403<sub>C</sub> are also highlighted with black dashed circles.

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** Free-energy landscape (FEL) in the gREST<sub>Up</sub> and gREST<sub>Down</sub> w/o glycan simulations.



**Figure 6.** Druggable cryptic pockets in the transition intermediates. (a) Snapshots of receptor binding domain (RBD) interface in Down symmetric ( $Down_{Sym}$ ), Intermediate 2a (I2a), and 1Up-like ( $1U_L$ ) conformations. Chains A, B, and C in the protein are shown in red, blue, and green surfaces, respectively, while a collection of glycans from 10 snapshots are shown in stick representation. The cryptic pockets predicted for I2a using P2Rank software are shown in orange (Pocket1) and yellow (Pocket2), respectively. These pockets disappear in both  $Down_{Sym}$  and  $1U_L$ . (b) Nilotinib docked poses (top and third ranked) to two cryptic pockets in I2a by Autodock Vina. The pockets are highlighted with black dashed circles and the nilotinib is shown in sphere representation in yellow.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Cryptic pockets and ligand binding in receptor binding domain (RBD).

### Searching for cryptic binding pockets in the intermediate structures

We applied a machine learning-based algorithm (P2Rank) (Krivák and Hoksza, 2018) to search for the formation of druggable pockets in the intermediate structures (Figure 6—figure supplement 1a). The same search was also carried out for  $Down_{Sym}$  and  $1U_L$  for comparison. Figure 6a shows the formation of two cryptic pockets (pocket1 and pocket2) at the interface of RBDs in one of the intermediates, which is not observed in Down or one-Up. Supplementary file 1D lists these predicted pockets showing relatively high scores in all the intermediates. To test the druggability of these two pockets, we performed virtual screening of FDA approved drugs from ZINC database (Irwin and Shoichet, 2005), where we docked 2115 molecules to the RBD interfaces in the intermediates (following the procedure sketched in Figure 6—figure supplement 1c). Supplementary file 1E shows a list of the index (ID) and binding energies of top-ranked molecules. This includes few anticancer drugs (irinotecan, nilotinib, and ponatinib), antimigraine drugs (ergot alkaloids), and antiviral-drug (glecaprevir). Intriguingly, the previous experimental study has shown that nilotinib affects the SARS-CoV-2 infectivity, although the mechanism remains unknown (Cagno et al., 2021). Figure 6c and Figure 6—figure supplement 1d show the representative binding poses of nilotinib in the three intermediates.

The top-ranked drugs tend to bind to either pocket1 or pocket2 in all first nine binding modes with high binding affinity (**Supplementary file 1F**). These results serve the two cryptic pockets as potential targets to stabilize the intermediate structures and prevent the formation of one-Up conformation responsible for the viral entry.

## Discussion

### Sampling the conformational space of S-protein

Due to the importance of the conformational changes of S-protein in the infection mechanisms and the rational design of antiviral drugs or antibodies, many extensive simulation studies on S-protein have been carried out since the start of the pandemic. The latest challenges involve the use of a million of distributed computer resources, Folding@home, to realize MD simulation of the SARS-CoV-2 proteome for 0.1 s, in total. An alternative approach is to incorporate experimental observation to effectively sample different states. Brotzakis and co-workers recently applied the cryo-EM metainference method to determine the opening pathway and intermediates based on the experimental density map (**Broetzakis et al., 2021**). Yet another approach is to use enhanced sampling methods. Sztain and co-workers performed the weighted ensemble simulations, collecting 130  $\mu$ s trajectories, to characterize the opening pathway (**Sztain et al., 2021**). Pang et al. employed two-dimensional umbrella sampling to characterize the transition pathway and the role of glycans (**Pang et al., 2021**). The gREST\_SSCR simulations are free from reaction coordinates and bias potentials that were used in many previous enhanced sampling simulations of S-protein. By using Fugaku supercomputer as well as GENESIS program (ver2.0) designed for achieving high scalability on that system (**Kobayashi et al., 2017; Jung et al., 2021**), the simulations explored a wide conformational space covering Down, one-Up, one-Open, and two-Up-like structures. The obtained structure ensembles overall agree with the cryo-EM structure distributions and smFRET data, suggesting the existence of the inherent flexibility of S-protein structures without its ligand, such as hACE2 or nAbs. Comparison to cryo-EM and smFRET experiment also indicates the quality of the predicted structures including intermediate states. In fact, comparison of the intermediate structures I2a and I3a with the D614G spike variant cryo-EM structure (PDB:7KRS) (**Zhang et al., 2021a**) shows a good agreement, with C $\alpha$  RMSD of 4.9 and 5.1 Å, respectively (**Figure 4—figure supplement 12**).

Besides the enhancement of conformational sampling by gREST\_SSCR, there are several remaining issues in the simulations. Within our computational time, one single simulation cannot cover a whole FEL containing all the important structures of S-protein. Three-Up structures, which were found with cryo-EM, for instance, PDB:7DCC (**Zhang et al., 2021b**), were not observed in the current simulations. As we discuss in the next section, glycans on the surface of S-protein seem to have dual roles in the transitions, either helping or hindering the Down-to-Up transitions, which makes conformational sampling via MD simulations more difficult. However, conformational sampling using gREST\_SSCR is much better than cMD without any pre-defined reaction coordinates and/or bias potentials, providing sufficient structural information to investigate the inherent flexibility of S-protein and the transition mechanisms.

### Role of flexible Down structures in the Down-to-Up transitions

In this study, we observed the inherent flexibility of S-protein in absence of its ligand, exploring Down, one-Up, one-Open, and two-Up-like structures at 310 K. The flexible nature of RBD forming multiple Down and Up conformations was also demonstrated in the cryo-EM study at physiological pH by **Pramanick et al., 2021**. The trimeric RBDs in S-protein are intrinsically flexible even in Down, as their interfaces are electrostatically repulsive (**Mori et al., 2021**). Not only symmetric structures but also anti-symmetric structures are also observed in Down, suggesting the existence of 'flexible Down structure' (**Figure 4—figure supplement 9**). This flexibility can initiate the Down-to-1Up transition, by allowing the transient reduction of inter-domain interactions in Down<sub>Sym</sub> and allowing the insertion of a glycan at N343<sub>B</sub> underneath RBD<sub>A</sub> (**Figure 4—figure supplement 10**). This is consistent with the mobile RBD conformations in Down observed experimentally by **Ke et al., 2020** and **Wrobel et al., 2020**.

Following the flexible Down structures, our simulations elucidate molecular mechanisms underlying the Down-to-1Up transition involving RBD<sub>A</sub> and the 1Up-to-2Up transitions involving RBD<sub>A</sub> and

RBD<sub>B</sub>. This sequence of events can be explained via the changes of inter-domain interactions as well as protein-glycan interactions. As RBD<sub>A</sub>-RBD<sub>C</sub> interaction becomes stronger in the Down-to-1Up transition (Figure 4, Figure 4—figure supplements 7 and 8), RBD<sub>B</sub> becomes more mobile for the next 1Up-to-2Up transition. Among the inter-domain interactions, we pointed out the importance of transient salt-bridges and HBs between R408<sub>A</sub>-D405<sub>C</sub> and R408<sub>A</sub>-D406<sub>C</sub> for stabilizing the intermediates (Figure 4 and Figure 4—figure supplement 8). The salt-bridge between R408<sub>A</sub> and D405<sub>C</sub> was also observed in the weighted ensemble simulations reported by *Sztain et al., 2021*. This scenario is mediated by protein-glycan interactions, in particular, those involving three glycans at N165<sub>B</sub>, N234<sub>B</sub>, and N343<sub>B</sub> as pointed out in previous studies (*Choi et al., 2021; Casalino et al., 2020; Mori et al., 2021*). In the current study, we highlighted the dual roles of these three glycans: both to stabilize one of the states and to serve the driving force toward the other state. The glycan at N165 stabilizes Down, representing a barrier for the Down-to-1Up transition, while the glycan at N343<sub>B</sub> drives the transition. Finally, the glycan at N234<sub>B</sub> stabilizes one-Up-like conformation. This picture coincides with the experimental results by *Henderson et al., 2020b*, where the population of one-Up is drastically reduced by the glycan deletions at N234 but increases by those at N165. As suggested previously (*Casalino et al., 2020*), the glycans at N165 may also stabilize one-Up conformation but as a minor role because it can adapt various orientations to that conformation (Figure 4—figure supplement 10). The role of glycan at N343 is noteworthy in that it helps the position of RBD lift up from Down and supports it throughout the intermediates. This well explains the 20-fold reduction of the infectivity upon N343Q mutation observed by *Li et al., 2020*. Recently, a similar role of the glycan at N343 has been proposed by *Sztain et al., 2021*. Mutational studies of charged residues at the RBD interface (e.g. R408, D405, and D427) could validate the proposed role of the salt-bridge formation in the transition.

## Implication for vaccine and drug developments

As a consequence of its central role, S-protein has been early identified as potential target for drug repurposing in order to block viral entry. Since the drug repurposing studies (*Han et al., 2021; Bakowski et al., 2021; Deganutti et al., 2020*) mostly target Down or Up from cryo-EM structures, druggable cryptic pockets identified in the intermediates would introduce unprecedented drug targets (*Brotzakis et al., 2021*). Two cryptic pockets at the RBD interface are identified in the highly populated intermediates along the Down-to-1Up transition. Both I2a and I3a show high stability on the FELs on PC1-PC2 and hinge-hinge angles as shown in Figure 4—figure supplement 11. Notably, allosteric sites in intermediate structures were previously used for drug design (*Lu et al., 2021; Shukla et al., 2014*). From our virtual screening of FDA approved molecules, these pockets are druggable and accommodates several small molecules including irinotecan, ergotamine, nilotinib, and ponatinib. For example, nilotinib was shown to reduce SARS-CoV-2 infection by ~50% (*Cagno et al., 2021*), likely via blocking spike interactions with ACE2 (*Tsegay et al., 2021*). Notably, top-ranked molecules (Supplementary file 1E) were also predicted to bind RBD in the previous virtual screening studies (*Deganutti et al., 2020; Murugan et al., 2020*). If these bindings happen, the bound molecules at RBD interface could potentially stabilize the intermediate states and reduces the population of Up conformation, to block ACE2 binding. To shift the conformational equilibrium of S-protein toward the inaccessible Down state for blocking ACE2 and subsequent membrane fusion has been a focus of the recent challenges. Indeed, there have been some reports altering the conformational dynamics of S-protein either via site-specific mutations, disulfide bonds, and binding to small molecule (*Hsieh et al., 2020; Juraszek et al., 2021; Henderson et al., 2020a; Edwards et al., 2021; McCallum et al., 2020; Toelzer et al., 2020*). Despite the promising outcome from shifting S-protein equilibrium, the high rate of mutations in spike might hinder its use as drug and antibodies target or even vaccine development. Indeed, continuous assessment of cryptic and epitope regions in spike variants is necessary.

The understanding of antibody responses is critical for vaccines and nAb developments. The large exposure of RBD<sub>A</sub> in Up forms is shown in the current study, which is consistent with several previous studies (*Grant et al., 2020; Casalino et al., 2020; Sikora et al., 2021*). In consistent with the conformation-dependent SASA variation reported in the previous cryo-EM study (*Pramanick et al., 2021*), we found that the accessibilities of RBM and the antibody epitopes, including mutational sites, depend sensitively on the RBD conformation. The result suggests that each of the Up conformations (1U, 1U<sub>O</sub>, and 2U<sub>J</sub>) tends to bind with distinct classes of antibodies. From our

inspection, 1U<sub>O</sub>, which largely exposes RBM but the antibody epitopes of only Class I and II, is the potentially active conformation for both the ACE2 binding and evading from the antibody attack. Class I and II epitope regions include K417, E484, and N501, and their mutations could effectively enhance the infection either by enhancing the ACE2 binding (N501Y) (Harvey *et al.*, 2021) or reducing the antibody binding (K417N or E484K) (Harvey *et al.*, 2021). Intriguingly, HB analysis shows that K417 and E484 respectively contribute to stabilize the intermediates likely enhancing the population of one-Up conformations. Note that L452 shows a little accessibility (Figure 3) in contrast with the reported severe mutational effect to an antibody recognition (Harvey *et al.*, 2021). The similar observation was also reported previously (Valdes-Balbin *et al.*, 2021), while the recent simulation shows the possible interaction between L452 and an antibody (Cao *et al.*, 2021). A recent comparison of L452R cryo-EM structure with wild type suggested that R452 sterically hinders the nAbs bindings without affecting ACE2 binding (McCallum *et al.*, 2021). Finally, the Up conformations are reasonably aligned with the antibody-bound S-protein cryo-EM structures (Figure 3—figure supplement 5). Hence, the antibody bindings of S-protein can be explained based on the conformer selection mechanism (Csermely *et al.*, 2010). The exploration of conformational diversity of S-protein together with binding free-energy calculations followed by docking simulation would provide valuable structural insights with possible antibody bindings (Ren *et al.*, 2021).

Finally, we highlight the key differences between our simulations and other computational works on S-protein. Current work includes at least two unique features based on the simulation and analysis methods. First, our conformational sampling method, gREST, could avoid the use of pre-defined reaction coordinates and/or bias potentials along the coordinates. Therefore, the gREST simulation is able to generate unexpected structures or structural changes, which are free from the starting cryo-EM structures and the reaction coordinates in biased MD simulations. Two-Up-like structures, Down<sub>Asym1</sub> or 1U<sub>O</sub> were not expected even by us before the extensive gREST simulations and analysis. Since these unexpected structures involve novel molecular interactions, it is necessary to detect which interactions can stabilize the intermediate structures and to examine how the interactions are changed along the transition pathways. Therefore, we analyzed the HBs and contacts in each intermediate, systematically, and quantitatively (the second key feature). The changes of the atomistic interactions involving side chains and glycans could characterize the conformational transition pathways in atomic detail without using simplified reaction coordinates. To understand the effect of mutations in the conformational changes toward the active Up form, the difference of intrinsic flexibility between wild type and mutant S-protein is one of the most important features to be understood. The current simulation and analysis methods are applicable to new variants for giving atomistic pictures of conformational flexibility and transitions of spike proteins, which are both important for understanding the essential characters of new S-protein variants.

## Materials and methods

### gREST\_SSCR simulations

The initial structures for S-protein head regions (residues 28–1135) were prepared based on the cryo-EM structures PDB:6VXX and PDB:6VYB for the Down and Up states, respectively (Walls *et al.*, 2020). Eighteen N-glycans and one O-glycan were added per protomer as suggested in the mass spectrometry experiments and previous computational models (Watanabe *et al.*, 2020; Woo *et al.*, 2020). A full list of included glycans is shown in Figure 1—figure supplement 2. CHARMM-GUI (Lee *et al.*, 2016) was used to prepare the final model including the glycans, ions (0.15 M NaCl), and water molecules. Three gREST\_SSCR simulations were performed in which two started from Down in the presence (500 ns) and absence of glycans (150 ns), and one from Up (300 ns). Eight pairs of charged residues per protomer at the RBD interface were selected as the solute region for gREST. The total number of atoms in solute was 870. All simulations were performed using 16 replicas covering the solute temperature range from 310 to 545 K while maintaining solvent temperature at 310.15 K. All simulations were performed using the new version of GENESIS MD software that was optimized on Fugaku (Kobayashi *et al.*, 2017; Jung *et al.*, 2021). Further detailed information is given in Supporting Information.

## Comparison between cryo-EM structures and MD simulations using PCA

We used cryo-EM structures provided by ‘spike protein and spike receptors’ in Protein Data Bank (<http://www.rcsb.org/covid19>, deposited date 2020/02/04 to 2021/09/09, released by 09/22) in this study. Only structures where the number of residues is greater than 700 and the last residue  $\geq 1122$  for each protomer were selected in the analysis; 891 protomers of 300 structures and 289 trimeric structures from PDB meet the criteria (**Supplementary file 1A, Table 1**). We adopt a method representing the structure with nine beads per protomer as used in the previous work of **Gobeil et al., 2021; Henderson et al., 2020a**. Unlike their study, 27 beads consisting of three chains are used here. This coarse-grained model consists of two beads for RBD, three beads for NTD, one bead for SD1 and SD2, and two beads for the S2 region (CD and S2-b) (see **Supplementary file 1B**). We first executed PCA (**Yang et al., 2009**) of each protomer using the selected cryo-EM structures after converting the nine-beads model. Up/Down for each protomer was determined by the value of projection against the first PC vector. By using the information, the rotation scheme was applied to the trimeric S-protein. PCA was performed using the rotated cryo-EM structures, after converting to the 27-beads model. The PC vectors were calculated upon fitting all the beads. All simulation trajectories were projected onto the PC1 and PC2 vectors and the potential mean forces (PMFs) were calculated for each simulation. The PMF at 310 K in each gREST\_SSCR was obtained using all the trajectories using the MBAR method (**Shirts and Chodera, 2008**). The 891 protomers of the cryo-EM structures were also used to compare hinge/twist angles.

### Pocket search and virtual screening

The P2Rank software (**Krivák and Hoksza, 2018**) was used to identify potential druggable pockets in intermediate structures. The cluster centers of 12a, 13a, and 13b as well as D1<sub>sym</sub> and 1U<sub>L</sub> were used for pocket search. All top-ranked pockets were investigated. Pockets at the RBD interface were selected for further analysis as they exist in all three intermediates but vanish in Down and Up, representing potential cryptic pockets. To check the druggability of these pockets, all FDA approved drugs were downloaded from ZINC database (**Irwin and Shoichet, 2005**). This includes 2115 molecules representing 1379 drug candidates. Open Babel was used to convert PDB to PDBQT. AutoDockTools-1.5.6 was used to prepare RBD receptor (**Morris et al., 2009**). AutoDock Vina was used to dock all 2115 molecules and perform virtual screening (**Trott and Olson, 2010**).

### Modeling of S-protein structure for MD simulations

The full-length S-protein is formed of three highly glycosylated protomers, with a 1273 residue each. The trimeric structure is divided into three regions, the head region which consists of S1 and part of S2 subunits (residues 1–1140), the heptad repeat 2/the transmembrane domain region (residues 1141–1234), and the cytoplasmic tail region (residues 1235–1273) (**Wrobel et al., 2020; Wrapp et al., 2020**). In this study, truncated structures of S-protein including S1 and part of the S2 subunits (residues 28–1135) were used in the simulations. Wherein the starting structure of the Down and Up conformations was based on the cryo-EM structures, PDB:6VXX and the PDB:6VYB, respectively (**Walls et al., 2020**). The 6VXX PDB structure includes multiple missing regions at the NTD (residues 70–79, 144–164, 173–185, and 246–262), the RBD (residues 445–446, 455–461, 469–488, and 502), and the S2 subunit (residues 621–640, 677–688, and 828–853). The 6VYB PDB structure has even more missing regions in the RBD with Up form and the adjacent NTD. Although higher resolution cryo-EM structures were deposited to the PDB later, only these two structures and PDB:6VSB (**Wrapp et al., 2020**) were available when we started this study. Due to the large size and the presence of multiple missing regions, several modeling strategies were used to complete the structures. Wherein, residues 28–292 of NTD were modeled based on the SARS-CoV crystal structure (PDB:5 × 4S at 2.2 Å) (**Yuan et al., 2017**) using Modeller9.19 software (**Sali and Blundell, 1993**). Then part of the modeled region (residues 28–288) was inserted in the cryo-EM structure upon fitting the backbone of residues 263–290. Similarly, the crystal structure of the RBD domain (PDB:6LZG at 2.45 Å) (**Wang et al., 2020**) (residues 336–515) was inserted in the cryo-EM structures upon fitting the C $\alpha$  atoms of residues 336–400. Both modeled NTD and RBD regions show a good alignment with the resolved regions in the 6VXX and 6VYB cryo-EM structures, see **Figure 1—figure supplement 1**. The VMD program (**Humphrey et al., 1996**) was used to superimpose the modeled regions into the cryo-EM structure. Finally, the missing



regions in the S2 subunit were modeled as loop conformations using the top-ranked structure from Modeller9.19 (McCallum *et al.*, 2020). A total of 13 disulfide bonds were included in each protomer including the original 12 disulfide bonds in the cryo-EM structure and one more in the RBD crystal structure. A comparison of our modeled structure and the more recent high-resolution cryo-EM structure (PDB:6ZGE at 2.6 Å) (Wrobel *et al.*, 2020) shows a very good agreement, see **Figure 1—figure supplement 1**. Eighteen N-glycans and one O-glycan were added per protomer as suggested in the previous mass spectrometry experiments and a computational model (Watanabe *et al.*, 2020; Woo *et al.*, 2020). A full list of included glycans is shown in **Figure 1—figure supplement 2**. CHARMM-GUI (Lee *et al.*, 2016) was used to make the final model including the addition of glycans, ions (0.15 M NaCl), and water molecules. In total, three S-protein models were built including the Down conformation in the absence of glycan, glycosylated S-protein in Down, and the glycosylated S-protein in Up conformation (**Figure 1—figure supplement 1d**). The total numbers of atoms in each model are 657,411, 654,427, and 654,494, respectively, with the average box lengths of 186.947, 186.452, and 186.475 Å after equilibration, respectively. Finally, the RBD/SD1 monomer models were made by truncating one protomer from the abovementioned Down and Up models, including residues 315–595 (**Figure 2—figure supplement 1b**).

### Further details of gREST\_SSCR simulations

We recently proposed an enhanced sampling method, the generalized replica exchange with solute tempering of selected surface charged residue (gREST\_SSCR) (Dokainish and Sugita, 2021) to enhance large domain motions in multi-domain proteins. In this method the Coulomb and Lennard Jones parameters of surface charged residues at the domain interfaces are selected as a solute region in gREST (Kamiya and Sugita, 2018). In this study, to enhance conformational dynamics of S-protein, we performed gREST\_SSCR simulations, wherein charged residues at the interfaces between two RBD domains, between RBD and NTD, and between RBD and S2, were selected as the solute region (**Figure 1—figure supplement 3a**). In total, 16 residues in each protomer, consisting of 8 positive and 8 negative charged residues (870 atoms), were selected as solute in gREST: K113, K378, K386, R408, K417, K462, R466, R983, E132, E169, D198, D405, E406, D420, D428, and E471. We performed two sets of 100 ns cMD simulations from Down and Up cryo-EM conformations in the presence and absence of positional restraints on backbone atoms, respectively, and carried out a preliminary gREST simulation where all charged residues in one of RBDs were included in the solute region. H-bonding and contact analysis were performed to identify the abovementioned residues in the solute region. Some pairs (e.g. D428\_R983) represented the native contacts observed in the restrained cMD simulations, Down<sub>Asym</sub> in the cMD simulation in the absence of restraints included the salt-bridge pair, K462\_D198, and newly formed salt-bridge pairs (e.g. K378\_D405) were found in the preliminary gREST simulation. In summary, we carefully selected essential charged residues from the cryo-EM structures, preliminary cMD and gREST simulations to accelerate large conformational changes of RBDs in S-protein. All simulations were performed using 16 replicas covering a solute temperature parameter range from 310.00 to 545.00 K while maintaining solvent temperature at 310.15 K in NVT ensemble. We carried out three gREST\_SSCR simulations: two from Down in the presence (500 ns) and absence of glycans (150 ns), and one from Up (300 ns). The total simulation times correspond to 8, 4.8, and 2.4 μs in gREST\_Down, gREST\_Up, and gREST\_Down w/o glycan, respectively.

All simulations were performed using the new version of GENESIS MD software that was optimized on Fugaku (Kobayashi *et al.*, 2017; Jung *et al.*, 2021). The overall performance of gREST\_SSCR simulations using 16 replicas is 52 ns/day using 2048 nodes on Fugaku. CHARMM 36m force field was used for protein (C36m), carbohydrate, and ions, while CHARMM TIP3P was used as a water model (Huang *et al.*, 2017; Guvench *et al.*, 2011). gREST\_SSCR simulations were performed after a series of equilibration steps. First modeled systems were minimized for 10,000 steps, while applying positional restraint on the backbone atoms. Second, using leap-frog integrator and the Langevin thermostat, we heated the simulation systems to 310.15 K in a stepwise manner for 100 ps. Third, a series of equilibration steps were performed: (1) MD simulations in the NVT ensemble using the velocity Verlet integrator with stochastic velocity rescaling thermostat (Bussi *et al.*, 2007). (2) Those in the NPT ensemble with stochastic velocity rescaling thermostat and MTK barostat (Bussi *et al.*, 2007; Bussi *et al.*, 2009; Jung *et al.*, 2018) (note that all previous steps also included a weak restraints on side chain and glycan dihedral angles). (3) After removing all restraints, another MD simulation in the NPT ensemble

was performed as equilibration using the same protocol. (4) MD simulation in the NVT ensemble was followed as the second equilibration using the same thermostat and the multiple time step integrator with a fast motion time step of 2.5 fs, and slow motion every 5 fs (Tuckerman *et al.*, 1992). (5) Prior to production run, a 2 ns equilibration was performed for 16 replicas. Production runs were then performed for 150, 500, and 300 ns per replica in gREST\_Down w/o glycan, gREST\_Down, and gREST\_Up simulations, respectively. At every 20 ps, replica exchanges were attempted, and trajectories were saved. Electrostatic interactions were computed by smooth particle mesh Ewald (Essmann *et al.*, 1995) method with  $128 \times 128 \times 128$  grids and the sixth-order B-spline function. Temperature is evaluated using the group-based approach with an optimal temperature evaluation, and thermostat is applied at every 10 steps (Jung and Sugita, 2020). Classical MD simulation of RBD/SD1 monomer structures was performed for 300 ns. Two independent simulations were performed starting from Up and one from Down. In all simulations, water molecules were constrained with SETTLE, while bonds involving hydrogens were constrained with SHAKE/RATTLE algorithm (Miyamoto and Kollman, 1992; Ryckaert *et al.*, 1977; Andersen, 1983).

## Simulation trajectory analysis

To characterize the RBD motions, two main criteria are considered: the C $\alpha$  atoms RMSD of RBD upon fitting the S2 C $\alpha$  atoms of cryo-EM structure (residues 689–827 and 854–1134) and the RBD hinge and twist angles. Hinge and twist angles represent relative domain motions of RBD, wherein the hinge angle describes the Down/Up transition while the twist angle describes RBD side motion. The hinge angle is defined with three points, the center of mass of the C $\alpha$  atoms in the SD1 core (residues 324–329, 531–590), the top residues of SD1 (residues 328, 329, 530, 531, 543, and 544), and the center of mass in the RBD core (residues 335–466 and 491–526). To define the twist angle, one more point was added at the lower part of RBD (residues 335, 336, 361, 362, 524, and 525). To examine the intra-domain stability, we computed the C $\alpha$  atom RMSD of RBD (residues 333–528) and NTD (residues 28–306). The C $\alpha$  RMSD of RBD and NTD upon fitting to their own structures at 310 K in gREST\_Down reveal about 1.4 and 2.2 Å, respectively (Figure 1—figure supplement 5a and b). They are comparable to those in the cMD (Mori *et al.*, 2021) (1.4 and 1.6 Å for RBD and NTD, respectively). Slightly larger RMSD values of NTD in gREST\_SSCR are attributed to the loop regions abundant in NTD, as indicated in root mean square fluctuations (RMSF) (Figure 1—figure supplement 5c and d). The C $\alpha$  RMSD of the simulations' representative structures (Down<sub>Sym</sub>, 1U<sub>O</sub>, and 2U<sub>I</sub>) were calculated with respect to cryo-EM structures using all three chains including 2766, 2884, 2450, and 2719 C $\alpha$  atoms, respectively. Since 1U<sub>O</sub> is compared to three-Up cryo-EM structure, two RBDs with a Down form were excluded in the RMSD calculation.

The k-means algorithm in GENESIS software package was used to classify the conformations of S-protein in MD simulations at 310 K. Hereafter, all the analysis is carried out to obtain the canonical ensembles at 310 K. The number of clusters in k-means clustering was set to eight in all cases. The cluster analysis was performed with the same fittings used in the RMSD analysis of RBD. Only the C $\alpha$  atoms included in the original cryo-EM structures (PDB:6VXX and PDB:6VYB) were included to avoid flexible regions in our analysis. Furthermore, the distributions of hinge and twist angles for all the eight clusters were calculated for each protomer (in total, six angles) and subsequently the number of clusters was increased until the hinge/twist distribution showed the minimal overlaps (Figure 4—figure supplements 1–3). In this procedure, 12, 13, and 13 clusters were obtained in gREST\_Down w/o glycan, gREST\_Down, and gREST\_Up simulations, respectively (Supplementary file 1C).

Due to the homo-trimeric nature of S-protein, protomers are indistinguishable in the structure. gREST\_SSCR enhanced motions of RBD regions so that we don't know which protomer reveals large-scale conformational motions in any replicas. For instance, Figure 1—figure supplement 6b shows that RBD<sub>A</sub> undergoes large transition in replica 1, while RBD<sub>C</sub> shows large motion in replica 16. To clarify the discussion in this paper, we applied a rotational scheme that makes RBD<sub>A</sub> undergo the largest conformational transitions in the following ways: (1) We identify all replicas that show significant RBD motions with a hinge angle  $>130^\circ$  in RBD<sub>B</sub> or RBD<sub>C</sub>. (2) We rotate the conformations of those selected replicas where RBD<sub>B</sub> or RBD<sub>C</sub> becomes RBD<sub>A</sub> while rotating the rest of the molecule including glycans (Figure 1—figure supplement 6a). (3) We confirm the rotation scheme by comparing hinge/twist angle free-energy maps before and after rotation (Figure 1—figure supplement 6c and d, 7a, and 7b). We also compared PCA before and after rotations. (4) In cases of two

RBDs showing large hinge angles in the same replica, the protomer with the highest RBD hinge angle becomes Chain<sub>A</sub>.

HB and contact analysis were also performed for major clusters, wherein a 75% and 50% probability threshold were used for the heavy atoms' contacts and HB residue pairs selection in **Figure 4—figure supplements 7 and 8**, respectively. The correspondence analysis to the previous smFRET experiment (Lu *et al.*, 2020) was performed upon calculating the COM distance from residues 425–431 to residues 554–561 using the C $\alpha$  atoms (**Figure 4—figure supplement 4a**). Experimental statistical ratio (Lu *et al.*, 2020) of 77% and 23% was used to combine gREST\_Down and gREST\_Up simulation results, respectively, as shown in **Figure 4—figure supplement 4d**. The VMD and PyMOL programs were used for trajectory and structure visualization (Humphrey *et al.*, 1996; Schrodinger LLC, 2021).

SASA values were calculated using the measure SASA function in VMD (Humphrey *et al.*, 1996). The restrict option, which considers only solvent accessible points near the user specified region, was used for per-domain and per-residue SASA calculations. A range of probe radius, including 1.4 Å (a sphere of water) and 7.2 Å (approximating the hypervariable loops of anti-gp120 antibodies) (Grant *et al.*, 2020), was used for SASA calculations. SASA values were calculated for different RBD conformations, Down, 1U, 1U<sub>O</sub>, and 2U<sub>L</sub>. Down represent the sum of Down<sub>sym</sub> and Down<sub>Asym</sub>, the rest is defined in **Supplementary file 1**. For each conformation, 30 snapshots close to the cluster center were extracted and used for analysis. The calculated SASA values were mapped on the structure using PyMOL software (Schrodinger LLC, 2021). For comparison, we confirmed that our SASA calculations give the results in consistent with the previous work by Amaro and co-workers (**Figure 3—figure supplement 1**; Casalino *et al.*, 2020).

## Acknowledgements

This research used computational resources of the supercomputer Fugaku (the evaluation environment in the trial phase) provided by the RIKEN Center for Computational Science. The results obtained on the evaluation environment in the trial phase do not guarantee the performance, power, and other attributes of the supercomputer Fugaku at the start of its public use operation. The computer resources of Oakforest-PACS were also provided through HPCI System Research project (Project ID: hp200135, hp200153, hp200028, hp210107, hp210177).

## Additional information

### Competing interests

### Funding

Funder	Grant reference number	Author
Ministry of Education, Culture, Sports, Science and Technology	FLAGSHIP 2020 project	Yuji Sugita
Ministry of Education, Culture, Sports, Science and Technology	Priority Issue on Post-K computer	Yuji Sugita
Ministry of Education, Culture, Sports, Science and Technology	Program for Promoting Researches on the Supercomputer Fugaku	Yuji Sugita
Ministry of Education, Culture, Sports, Science and Technology	JPMXP1020200101	Yuji Sugita
Ministry of Education, Culture, Sports, Science and Technology	JPMXP1020200201	Yuji Sugita

Funder	Grant reference number	Author
Ministry of Education, Culture, Sports, Science and Technology	19H05645	Yuji Sugita
Ministry of Education, Culture, Sports, Science and Technology	21H05249	Yuji Sugita
Ministry of Education, Culture, Sports, Science and Technology	20K15737	Hisham M Dokainish
Ministry of Education, Culture, Sports, Science and Technology	19K12229	Suyong Re
Ministry of Education, Culture, Sports, Science and Technology	21H05157	Takaharu Mori
Ministry of Education, Culture, Sports, Science and Technology	19K06532	Takaharu Mori
RIKEN	Dynamic Structural Biology/Glycolipidologue Initiative/Biology of Intracellular Environments	Yuji Sugita
HPCI System Research project	hp200135	Takaharu Mori
HPCI System Research project	hp200153	Takaharu Mori
HPCI System Research project	hp200028	Takaharu Mori
HPCI System Research project	hp210107	Takaharu Mori
HPCI System Research project	hp210177	Takaharu Mori

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Hisham M Dokainish, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing - original draft, Writing - review and editing; Suyong Re, Formal analysis, Methodology, Visualization, Writing - original draft, Writing - review and editing; Takaharu Mori, Formal analysis, Methodology, Software, Validation, Visualization; Chigusa Kobayashi, Formal analysis, Methodology, Software, Visualization; Jaewoon Jung, Methodology, Software; Yuji Sugita, Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing - review and editing

### Author ORCIDs

Hisham M Dokainish  <http://orcid.org/0000-0002-4387-4790>

Suyong Re  <http://orcid.org/0000-0002-3752-6554>

Takaharu Mori  <http://orcid.org/0000-0002-8717-2926>

Chigusa Kobayashi  <http://orcid.org/0000-0002-5603-4619>

Jaewoon Jung  <http://orcid.org/0000-0002-2285-4432>

Yuji Sugita  <http://orcid.org/0000-0001-9738-9216>

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.75720.sa1>

Author response <https://doi.org/10.7554/eLife.75720.sa2>

## Additional files

### Supplementary files

- Supplementary file 1. List of PDBs, clusters and lignads. (A) Cryo-electron microscopy (cryo-EM) structures used in the principal component analysis (PCA). (B) Definition of protomer coarse-grained particles representing rigid domains for PCA. (C) List of clusters for gREST\_Down, gREST\_Up, and gREST\_Down w/o glycan simulations. (D) The receptor binding domain (RBD) interface cryptic pockets predicted by P2Rank. (E) List of the top-ranked molecules from the virtual screening of 2115 FDA approved drugs to RBD interface in I2a, I3a, and I3b intermediate structures. (F) Nilotinib binding energy to I2a, I3a, and I3b intermediates.
- Transparent reporting form

### Data availability

The trajectories were computed with GENESIS 2.0 beta, open source program <https://www.r-ccs.riken.jp/labs/cbrt/> and analyzed using GENESIS 1.6.0 analysis tools <https://www.r-ccs.riken.jp/labs/cbrt/download/genesis-version-1-6/>. Simulation data were deposited at <https://covid.molssi.org/>. Data of gREST simulations from Down and data of gREST\_Up simulations, including model and simulation structures, are available.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Suyong R, Hisham MD, Takaharu M, Chigusa K, Jaewoon J, Yuji S	2021	gREST_SSCR Simulation of Trimeric SARS-CoV-2 Spike Protein Starting From Down Conformation	<a href="https://doi.org/10.34974/wtbx-0r84">https://doi.org/10.34974/wtbx-0r84</a>	MolSSI, 10.34974/wtbx-0r84
Suyong R, Hisham MD, Takaharu M	2021	gREST_SSCR Simulation of Trimeric SARS-CoV-2 Spike Protein Starting From 1Up Conformation	<a href="https://doi.org/10.34974/xn67-xk26">https://doi.org/10.34974/xn67-xk26</a>	MolSSI, 10.34974/xn67-xk26

## References

- Alvarado W, Perez-Lemus GR, Menéndez CA, Byléhn F, de Pablo JJ. 2022. Molecular characterization of COVID-19 therapeutics: luteolin as an allosteric modulator of the spike protein of SARS-CoV-2. *Molecular Systems Design & Engineering* 7:58–66. DOI: <https://doi.org/10.1039/D1ME00119A>
- Andersen HC. 1983. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* 52:24–34. DOI: [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1)
- Bakowski MA, Beutler N, Wolff KC, Kirkpatrick MG, Chen E, Nguyen TTH, Riva L, Shaabani N, Parren M, Ricketts J, Gupta AK, Pan K, Kuo P, Fuller M, Garcia E, Teijaro JR, Yang L, Sahoo D, Chi V, Huang E, et al. 2021. Drug repurposing screens identify chemical entities for the development of COVID-19 interventions. *Nature Communications* 12:3309. DOI: <https://doi.org/10.1038/s41467-021-23328-0>, PMID: 34083527
- Barnes CO, Jette CA, Abernathy ME, Dam KMA, Esswein SR, Gristick HB, Malutin AG, Sharaf NG, Huey-Tubman KE, Lee YE, Robbani DF, Nussenzweig MC, West AP, Bjorkman PJ. 2020a. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588:682–687. DOI: <https://doi.org/10.1038/s41586-020-2852-1>, PMID: 33045718
- Barnes CO, West AP, Huey-Tubman KE, Hoffmann MAG, Sharaf NG, Hoffman PR, Koranda N, Gristick HB, Gaebler C, Muecksch F, Lorenzi JCC, Fink S, Häggglöf T, Hurley A, Millard KG, Weisblum Y, Schmidt F, Hatzioannou T, Bieniasz PD, Caskey M, et al. 2020b. Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* 182:828–842. DOI: <https://doi.org/10.1016/j.cell.2020.06.025>, PMID: 32645326
- Brotzakis ZF, Löhr T, Vendruscolo M. 2021. Determination of intermediate state structures in the opening pathway of SARS-CoV-2 spike using cryo-electron microscopy. *Chemical Science* 12:9168–9175. DOI: <https://doi.org/10.1039/d1sc00244a>, PMID: 34276947
- Bussi G, Donadio D, Parrinello M. 2007. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* 126:014101. DOI: <https://doi.org/10.1063/1.2408420>, PMID: 17212484
- Bussi G, Zykova-Timan T, Parrinello M. 2009. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics* 130:074101. DOI: <https://doi.org/10.1063/1.3073889>, PMID: 19239278
- Cagno V, Magliocco G, Tapparel C, Daali Y. 2021. The tyrosine kinase inhibitor nilotinib inhibits SARS-CoV-2 in vitro. *Basic & Clinical Pharmacology & Toxicology* 128:621–624. DOI: <https://doi.org/10.1111/bcpt.13537>, PMID: 33232578

- Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, Rawson S, Rits-Volloch S, Chen B. 2020. Distinct conformational states of SARS-CoV-2 spike protein. *Science (New York, N.Y.)* **369**:1586–1592. DOI: <https://doi.org/10.1126/science.abd4251>, PMID: 32694201
- Cao Y, Choi YK, Frank M, Woo H, Park SJ, Yeom MS, Seok C, Im W. 2021. Dynamic Interactions of Fully Glycosylated SARS-CoV-2 Spike Protein with Various Antibodies. *Journal of Chemical Theory and Computation* **17**:6559–6569. DOI: <https://doi.org/10.1021/acs.jctc.1c00552>, PMID: 34529436
- Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, Huang W, Li Q, Wang P, An R, Wang J, Wang Y, Niu X, Yang S, Liang H, Sun H, Li T, Yu Y, Cui Q, Liu S, et al. 2022. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**:657–663. DOI: <https://doi.org/10.1038/s41586-021-04385-3>, PMID: 35016194
- Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, Fogarty CA, Barros EP, Taylor BC, McLellan JS, Fadda E, Amaro RE. 2020. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **6**:1722–1734. DOI: <https://doi.org/10.1021/acscentsci.0c01056>, PMID: 33140034
- Choi YK, Cao Y, Frank M, Woo H, Park SJ, Yeom MS, Croll TI, Seok C, Im W. 2021. Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. *Journal of Chemical Theory and Computation* **17**:2479–2487. DOI: <https://doi.org/10.1021/acs.jctc.0c01144>, PMID: 33689337
- Csermely P, Palotai R, Nussinov R. 2010. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **35**:539–546. DOI: <https://doi.org/10.1016/j.tibs.2010.04.009>, PMID: 20541943
- Deganutti G, Prischi F, Reynolds CA. 2020. Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. *Journal of Computer-Aided Molecular Design* **35**:195–207. DOI: <https://doi.org/10.1007/s10822-020-00356-4>, PMID: 33103220
- Dokainish HM, Sugita Y. 2021. Exploring Large Domain Motions in Proteins Using Atomistic Molecular Dynamics with Enhanced Conformational Sampling. *International Journal of Molecular Sciences* **22**:E270. DOI: <https://doi.org/10.3390/ijms22010270>, PMID: 33383937
- Edwards RJ, Mansouri K, Stalls V, Manne K, Watts B, Parks R, Janowska K, Gobeil SMC, Kopp M, Li D, Lu X, Mu Z, Deyton M, Oguin TH, Spreng J, Williams W, Saunders KO, Montefiori D, Sempowski GD, Henderson R, et al. 2021. Cold sensitivity of the SARS-CoV-2 spike ectodomain. *Nature Structural & Molecular Biology* **28**:128–131. DOI: <https://doi.org/10.1038/s41594-020-00547-5>, PMID: 33402708
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**:8577–8593. DOI: <https://doi.org/10.1063/1.470117>
- Fan X, Cao D, Kong L, Zhang X. 2020. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature Communications* **11**:3618. DOI: <https://doi.org/10.1038/s41467-020-17371-6>, PMID: 32681106
- Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, Tamir H, Achdout H, Stein D, Israeli O, Beth-Din A, Melamed S, Weiss S, Israely T, Paran N, Schwartz M, Stern-Ginossar N. 2021. The coding capacity of SARS-CoV-2. *Nature* **589**:125–130. DOI: <https://doi.org/10.1038/s41586-020-2739-1>, PMID: 32906143
- Gobeil SM-C, Janowska K, McDowell S, Mansouri K, Parks R, Stalls V, Kopp MF, Manne K, Li D, Wiehe K, Saunders KO, Edwards RJ, Korber B, Haynes BF, Henderson R, Acharya P. 2021. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science (New York, N.Y.)* **373**:eabi6226. DOI: <https://doi.org/10.1126/science.abi6226>, PMID: 34168071
- Grant OC, Montgomery D, Ito K, Woods RJ. 2020. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports* **10**:14991. DOI: <https://doi.org/10.1038/s41598-020-71748-7>, PMID: 32929138
- Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Golcuk M. 2020. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *The Journal of Chemical Physics* **153**:075101. DOI: <https://doi.org/10.1063/5.0011141>, PMID: 32828084
- Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW, Mackerell AD. 2011. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *Journal of Chemical Theory and Computation* **7**:3162–3180. DOI: <https://doi.org/10.1021/ct200328p>, PMID: 22125473
- Han N, Hwang W, Tzelepis K, Schmerer P, Yankova E, MacMahon M, Lei W, M. Katritsis N, Liu A, Felgenhauer U, Schuldt A, Harris R, Chapman K, McCaughan F, Weber F, Kouzarides T. 2021. Identification of SARS-CoV-2-induced pathways reveals drug repurposing strategies. *Science Advances* **7**:27. DOI: <https://doi.org/10.1126/sciadv.abh3032>, PMID: 34193418
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Peacock SJ, Robertson DL, COVID-19 Genomics UK COG-UK Consortium. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews. Microbiology* **19**:409–424. DOI: <https://doi.org/10.1038/s41579-021-00573-0>, PMID: 34075212
- Henderson R, Edwards RJ, Mansouri K, Janowska K, Stalls V, Gobeil SMC, Kopp M, Li D, Parks R, Hsu AL, Borgnia MJ, Haynes BF, Acharya P. 2020a. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature Structural & Molecular Biology* **27**:925–933. DOI: <https://doi.org/10.1038/s41594-020-0479-4>, PMID: 32699321

- Henderson R**, Edwards RJ, Mansouri K, Janowska K, Stalls V, Kopp M, Haynes BF, Acharya P. 2020b. Glycans on the SARS-CoV-2 Spike Control the Receptor Binding Domain Conformation. *Biophysics* 1:73765. DOI: <https://doi.org/10.1101/2020.06.26.173765>, PMID: 32637959
- Hsieh CL**, Goldsmith JA, Schaub JM, DiVenere AM, Kuo HC, Javanmardi K, Le KC, Wrapp D, Lee AG, Liu Y, Chou CW, Byrne PO, Hjorth CK, Johnson NV, Ludes-Meyers J, Nguyen AW, Park J, Wang N, Amengor D, Lavinder JJ, et al. 2020. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science (New York, N.Y.)* 369:1501–1505. DOI: <https://doi.org/10.1126/science.abd0826>, PMID: 32703906
- Hu B**, Guo H, Zhou P, Shi ZL. 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews. Microbiology* 19:141–154. DOI: <https://doi.org/10.1038/s41579-020-00459-7>, PMID: 33024307
- Huang J**, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14:71–73. DOI: <https://doi.org/10.1038/nmeth.4067>, PMID: 27819658
- Humphrey W**, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14:33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5), PMID: 8744570
- Irwin JJ**, Shoichet BK. 2005. ZINC—a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling* 45:177–182. DOI: <https://doi.org/10.1021/ci049714+>, PMID: 15667143
- Jung J**, Kobayashi C, Sugita Y. 2018. Kinetic energy definition in velocity Verlet integration for accurate pressure evaluation. *The Journal of Chemical Physics* 148:164109. DOI: <https://doi.org/10.1063/1.5008438>, PMID: 29716226
- Jung J**, Sugita Y. 2020. Group-based evaluation of temperature and pressure for molecular dynamics simulation with a large time step. *The Journal of Chemical Physics* 153:234115. DOI: <https://doi.org/10.1063/5.0027873>, PMID: 33353318
- Jung J**, Kobayashi C, Kasahara K, Tan C, Kuroda A, Minami K, Ishiduki S, Nishiki T, Inoue H, Ishikawa Y, Feig M, Sugita Y. 2021. New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems. *Journal of Computational Chemistry* 42:231–241. DOI: <https://doi.org/10.1002/jcc.26450>, PMID: 33200457
- Juraszek J**, Rutten L, Blokland S, Bouchier P, Voorzaat R, Ritschel T, Bakkers MJG, Renault LLR, Langedijk JPM. 2021. Stabilizing the closed SARS-CoV-2 spike trimer. *Nature Communications* 12:244. DOI: <https://doi.org/10.1038/s41467-020-20321-x>, PMID: 33431842
- Kamiya M**, Sugita Y. 2018. Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. *The Journal of Chemical Physics* 149:072304. DOI: <https://doi.org/10.1063/1.5016222>, PMID: 30134668
- Ke Z**, Oton J, Qu K, Cortese M, Zila V, McKeane L, Nakane T, Zivanov J, Neufeldt CJ, Cerikan B, Lu JM, Peukes J, Xiong X, Kräusslich H-G, Scheres SHW, Bartenschlager R, Briggs JAG. 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* 588:498–502. DOI: <https://doi.org/10.1038/s41586-020-2665-2>, PMID: 32805734
- Kobayashi C**, Jung J, Matsunaga Y, Mori T, Ando T, Tamura K, Kamiya M, Sugita Y. 2017. GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *Journal of Computational Chemistry* 38:2193–2206. DOI: <https://doi.org/10.1002/jcc.24874>, PMID: 28718930
- Krivák R**, Hoksza D. 2018. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics* 10:39. DOI: <https://doi.org/10.1186/s13321-018-0285-8>, PMID: 30109435
- Lee J**, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL, MacKerell AD, Klauda JB, Im W. 2016. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* 12:405–413. DOI: <https://doi.org/10.1021/acs.jctc.5b00935>, PMID: 26631602
- Li Q**, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. 2020. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* 182:1284–1294. DOI: <https://doi.org/10.1016/j.cell.2020.07.012>, PMID: 32730807
- Li Y**, Tenchov R, Smoot J, Liu C, Watkins S, Zhou Q. 2021. A Comprehensive Review of the Global Efforts on COVID-19 Vaccine Development. *ACS Central Science* 7:512–533. DOI: <https://doi.org/10.1021/acscentsci.1c00120>, PMID: 34056083
- Lu M**, Uchil PD, Li W, Zheng D, Terry DS, Gorman J, Shi W, Zhang B, Zhou T, Ding S, Gasser R, Prévost J, Beaudoin-Bussièrès G, Anand SP, Laumaea A, Grover JR, Liu L, Ho DD, Mascola JR, Finzi A, et al. 2020. Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host & Microbe* 28:880–891. DOI: <https://doi.org/10.1016/j.chom.2020.11.001>, PMID: 33242391
- Lu S**, He X, Yang Z, Chai Z, Zhou S, Wang J, Rehman AU, Ni D, Pu J, Sun J, Zhang J. 2021. Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nature Communications* 12:4721. DOI: <https://doi.org/10.1038/s41467-021-25020-9>, PMID: 34354057
- McCallum M**, Walls AC, Bowen JE, Corti D, Velesler D. 2020. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nature Structural & Molecular Biology* 27:942–949. DOI: <https://doi.org/10.1038/s41594-020-0483-8>, PMID: 32753755
- McCallum M**, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, Tortorici MA, Navarro MJ, Silacci-Fregni C, Saliba C, Sprouse KR, Agostini M, Pinto D, Culp K, Bianchi S, Jaconi S, Camerini E, Bowen JE, Tilles SW,

- Pizzuto MS, et al. 2021. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science (New York, N.Y.)* **373**:648–654. DOI: <https://doi.org/10.1126/science.abi7994>, PMID: 34210893
- Miyamoto S, Kollman PA. 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**:952–962. DOI: <https://doi.org/10.1002/jcc.540130805>
- Mori T, Jung J, Kobayashi C, Dokainish HM, Re S, Sugita Y. 2021. Elucidation of interactions regulating conformational stability and dynamics of SARS-CoV-2 S-protein. *Biophysical Journal* **120**:1060–1071. DOI: <https://doi.org/10.1016/j.bpj.2021.01.012>, PMID: 33484712
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**:2785–2791. DOI: <https://doi.org/10.1002/jcc.21256>, PMID: 19399780
- Murugan NA, Kumar S, Jeyakanthan J, Srivastava V. 2020. Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach. *Scientific Reports* **10**:19125. DOI: <https://doi.org/10.1038/s41598-020-75762-7>, PMID: 33154404
- Pang YT, Acharya A, Lynch DL, Pavlova A, Gumbart JC. 2021. Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to ACE2. *Biophysics* **1**:e456168. DOI: <https://doi.org/10.1101/2021.08.12.456168>
- Pavlova A, Zhang Z, Acharya A, Lynch DL, Pang YT, Mou Z, Parks JM, Chipot C, Gumbart JC. 2021. Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to ACE2. *The Journal of Physical Chemistry Letters* **12**:5494–5502. DOI: <https://doi.org/10.1021/acs.jpclett.1c01494>, PMID: 34086459
- Pinto D, Park YJ, Beltramello M, Walls AC, Tortorici MA, Bianchi S, Jaconi S, Culap K, Zatta F, De Marco A, Peter A, Guarino B, Spreafico R, Camerini E, Case JB, Chen RE, Havenar-Daughton C, Snell G, Telenti A, Virgin HW, et al. 2020. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**:290–295. DOI: <https://doi.org/10.1038/s41586-020-2349-y>, PMID: 32422645
- Pramanick I, Sengupta N, Mishra S, Pandey S, Girish N, Das A, Dutta S. 2021. Conformational flexibility and structural variability of SARS-CoV2 S protein. *Structure (London, England)* **29**:834–845. DOI: <https://doi.org/10.1016/j.str.2021.04.006>, PMID: 33932324
- Prince T, Smith SL, Radford AD, Solomon T, Hughes GL, Patterson EI. 2021. SARS-CoV-2 Infections in Animals: Reservoirs for Reverse Zoonosis and Models for Study. *Viruses* **13**:494. DOI: <https://doi.org/10.3390/v13030494>, PMID: 33802857
- Ren W, Dokainish HM, Shinobu A, Oshima H, Sugita Y. 2021. Unraveling the Coupling between Conformational Changes and Ligand Binding in Ribose Binding Protein Using Multiscale Molecular Dynamics and Free-Energy Calculations. *The Journal of Physical Chemistry. B* **125**:2898–2909. DOI: <https://doi.org/10.1021/acs.jpcc.0c11600>, PMID: 33728914
- Ryckaert J-P, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**:327–341. DOI: [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
- Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**:779–815. DOI: <https://doi.org/10.1006/jmbi.1993.1626>, PMID: 8254673
- Schrodinger LLC. 2021. The PyMOL Molecular Graphics System. Version 2.4. PyMOL. <https://pymol.org/2/>
- Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. 2020. Cell entry mechanisms of SARS-CoV-2. *PNAS* **117**:11727–11734. DOI: <https://doi.org/10.1073/pnas.2003138117>, PMID: 32376634
- Shirts MR, Chodera JD. 2008. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **129**:124105. DOI: <https://doi.org/10.1063/1.2978177>, PMID: 19045004
- Shukla D, Meng Y, Roux B, Pande VS. 2014. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nature Communications* **5**:3397. DOI: <https://doi.org/10.1038/ncomms4397>, PMID: 24584478
- Sikora M, von Bülow S, Blanc FEC, Gecht M, Covino R, Hummer G. 2021. Computational epitope map of SARS-CoV-2 spike protein. *PLOS Computational Biology* **17**:e1008790. DOI: <https://doi.org/10.1371/journal.pcbi.1008790>, PMID: 33793546
- Simmerling C, Rizzo R, Varghese S, Budhan S, Guerra J, Cuomo A, Stepanenko D, Corbo C, Wang Y, Raguette L, Belfon K, Fallon L. 2020. Free Energy Landscapes for RBD Opening in SARS-CoV-2 Spike Glycoprotein Simulations Suggest Key Interactions and a Potentially Druggable Allosteric Pocket. *Biological and Medicinal Chemistry* **1**:e13502646. DOI: <https://doi.org/10.26434/chemrxiv.13502646>
- Sztain T, Ahn SH, Bogetti AT, Casalino L, Goldsmith JA, Seitz E, McCool RS, Kearns FL, Acosta-Reyes F, Maji S, Mashayekhi G, McCammon JA, Ourmazd A, Frank J, McLellan JS, Chong LT, Amaro RE. 2021. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chemistry* **13**:963–968. DOI: <https://doi.org/10.1038/s41557-021-00758-3>, PMID: 34413500
- Toelzer C, Gupta K, Yadav SKN, Borucu U, Davidson AD, Kavanagh Williamson M, Shoemark DK, Garzoni F, Stauffer O, Milligan R, Capin J, Mulholland AJ, Spatz J, Fitzgerald D, Berger I, Schaffitzel C. 2020. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science (New York, N.Y.)* **370**:725–730. DOI: <https://doi.org/10.1126/science.abd3255>, PMID: 32958580
- Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**:455–461. DOI: <https://doi.org/10.1002/jcc.21334>, PMID: 19499576
- Tsegay KB, Adeyemi CM, Gniffke EP, Sather DN, Walker JK, Smith SEP. 2021. A Repurposed Drug Screen Identifies Compounds That Inhibit the Binding of the COVID-19 Spike Protein to ACE2. *Frontiers in Pharmacology* **12**:685308. DOI: <https://doi.org/10.3389/fphar.2021.685308>, PMID: 34194331



- Tuckerman M**, Berne BJ, Martyna GJ. 1992. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* **97**:1990–2001. DOI: <https://doi.org/10.1063/1.463137>
- Turoňová B**, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, von Bülow S, Gecht M, Bagola K, Hörner C, van Zandbergen G, Landry J, de Azevedo NTD, Mosalaganti S, Schwarz A, Covino R, Mühlebach MD, Hummer G, Krijnse Locker J, Beck M. 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science (New York, N.Y.)* **370**:203–208. DOI: <https://doi.org/10.1126/science.abd5223>, PMID: 32817270
- Valdes-Balbin Y**, Santana-Mederos D, Paquet F, Fernandez S, Climent Y, Chiodo F, Rodríguez L, Sanchez Ramirez B, Leon K, Hernandez T, Castellanos-Serra L, Garrido R, Chen G-W, Garcia-Rivera D, Rivera DG, Verez-Bencomo V. 2021. Molecular Aspects Concerning the Use of the SARS-CoV-2 Receptor Binding Domain as a Target for Preventive Vaccines. *ACS Central Science* **7**:757–767. DOI: <https://doi.org/10.1021/acscentsci.1c00216>, PMID: 34075345
- Venkadapathi J**, Govindarajan VK, Sekaran S, Venkatapathy S. 2021. A Minireview of the Promising Drugs and Vaccines in Pipeline for the Treatment of COVID-19 and Current Update on Clinical Trials. *Frontiers in Molecular Biosciences* **8**:637378. DOI: <https://doi.org/10.3389/fmolb.2021.637378>, PMID: 34179072
- V'kovski P**, Kratzel A, Steiner S, Stalder H, Thiel V. 2021. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews. Microbiology* **19**:155–170. DOI: <https://doi.org/10.1038/s41579-020-00468-6>, PMID: 33116300
- Walls AC**, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**:281–292. DOI: <https://doi.org/10.1016/j.cell.2020.02.058>, PMID: 32155444
- Wang Q**, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q, Zhou H, Yan J, Qi J. 2020. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**:894–904. DOI: <https://doi.org/10.1016/j.cell.2020.03.045>, PMID: 32275855
- Wang P**, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**:130–135. DOI: <https://doi.org/10.1038/s41586-021-03398-2>, PMID: 33684923
- Watanabe Y**, Allen JD, Wrapp D, McLellan JS, Crispin M. 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (New York, N.Y.)* **369**:330–333. DOI: <https://doi.org/10.1126/science.abb9983>, PMID: 32366695
- Woo H**, Park SJ, Choi YK, Park T, Tanveer M, Cao Y, Kern NR, Lee J, Yeom MS, Croll TI, Seok C, Im W. 2020. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *The Journal of Physical Chemistry. B* **124**:7128–7137. DOI: <https://doi.org/10.1021/acs.jpcc.0c04553>, PMID: 32559081
- Wrapp D**, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)* **367**:1260–1263. DOI: <https://doi.org/10.1126/science.abb2507>, PMID: 32075877
- Wrobel AG**, Benton DJ, Xu P, Roustan C, Martin SR, Rosenthal PB, Skehel JJ, Gamblin SJ. 2020. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nature Structural & Molecular Biology* **27**:763–767. DOI: <https://doi.org/10.1038/s41594-020-0468-7>, PMID: 32647346
- Xu C**, Wang Y, Liu C, Zhang C, Han W, Hong X, Wang Y, Hong Q, Wang S, Zhao Q, Wang Y, Yang Y, Chen K, Zheng W, Kong L, Wang F, Zuo Q, Huang Z, Cong Y. 2021. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances* **7**:1. DOI: <https://doi.org/10.1126/sciadv.abe5575>, PMID: 33277323
- Yan R**, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)* **367**:1444–1448. DOI: <https://doi.org/10.1126/science.abb2762>, PMID: 32132184
- Yang LW**, Eyal E, Bahar I, Kitao A. 2009. Principal component analysis of native ensembles of biomolecular structures (PCA\_NEST): insights into functional dynamics. *Bioinformatics (Oxford, England)* **25**:606–614. DOI: <https://doi.org/10.1093/bioinformatics/btp023>, PMID: 19147661
- Yuan Y**, Cao D, Zhang Y, Ma J, Qi J, Wang Q, Lu G, Wu Y, Yan J, Shi Y, Zhang X, Gao GF. 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications* **8**:15092. DOI: <https://doi.org/10.1038/ncomms15092>, PMID: 28393837
- Yurkovetskiy L**, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A, Diehl WE, Dauphin A, Carbone C, Veinotte K, Egri SB, Schaffner SF, Lemieux JE, Munro JB, Rafique A, Barve A, Sabeti PC, Kyratsous CA, Dudkina NV, et al. 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**:739–751. DOI: <https://doi.org/10.1016/j.cell.2020.09.032>, PMID: 32991842
- Zhang J**, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, Walsh RM, Rits-Volloch S, Zhu H, Woosley AN, Yang W, Sliz P, Chen B. 2021a. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science (New York, N.Y.)* **372**:525–530. DOI: <https://doi.org/10.1126/science.abf2303>, PMID: 33727252
- Zhang C**, Wang Y, Zhu Y, Liu C, Gu C, Xu S, Wang Y, Zhou Y, Wang Y, Han W, Hong X, Yang Y, Zhang X, Wang T, Xu C, Hong Q, Wang S, Zhao Q, Qiao W, Zang J, et al. 2021b. Development and structural basis of a two-MAb cocktail for treating SARS-CoV-2 infections. *Nature Communications* **12**:264. DOI: <https://doi.org/10.1038/s41467-020-20465-w>, PMID: 33431876
- Zimmerman MI**, Porter JR, Ward MD, Singh S, Vithani N, Meller A, Mallimadugula UL, Kuhn CE, Borowsky JH, Wiewiora RP, Hurley MFD, Harbison AM, Fogarty CA, Coffland JE, Fadda E, Voelz VA, Chodera JD, Bowman GR. 2021. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets

across the proteome. *Nature Chemistry* **13**:651–659. DOI: <https://doi.org/10.1038/s41557-021-00707-0>, PMID: [34031561](https://pubmed.ncbi.nlm.nih.gov/34031561/)