The Inherent Flexibility of Receptor Binding Domains in SARS-CoV-2 Spike Protein

Author Information

Hisham M. Dokainish[a], Suyong Re[b,c], Takaharu Mori[a], Chigusa Kobayashi[d], Jaewoon Jung[d], and Yuji Sugita[a,b,d]

[a]Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Hirosawa 2-1, Wako, Saitama 351-0198, Japan.

[b]Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, 6-7-1 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

[c]Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health, and Nutrition, 7-6-8, Saito-Asagi, Ibaraki, Osaka 567-0085, Japan.

[d]Computational Biophysics Research Team, RIKEN Center for Computational Science, 6-7-1 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

Contributions

H.M.D. performed simulations. H.M.D., S.R., C.K., T.M. analyzed data and made figures. H.M.D. modelled simulation systems. J.J. and C.K. made software. H..M.D., S.R. and Y.S. wrote manuscript. H.M.D. and Y.S. designed and initiated the research.

Corresponding author

Corresponding to: Yuji Sugita (sugita@riken.jp).

ORCID

Hisham M. Dokainish (https://orcid.org/0000-0002-4387-4790)

Suyong Re (https://orcid.org/0000-0002-3752-6554)

Takaharu Mori (https://orcid.org/0000-0002-8717-2926)

Chigusa Kobayashi (https://orcid.org/0000-0002-5603-4619)

Jaewoon Jung (https://orcid.org/0000-0002-2285-4432)

Yuji Sugita (https://orcid.org/0000-0001-9738-9216)

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 conformational changes of its receptor binding domain (RBD) to enter the host cell, as the abundant structural studies suggest. Here, we observe Down, one-Up, one-Open, and two-Up-like structures in enhanced molecular dynamics simulations without pre-defined reaction coordinates. The RBD_A transition from Down to one-Up is supported by transient salt-bridges between RBD_A and RBD_C and by the glycan at $N343_B$. Reduced interactions between RBD_A and RBD_B induce the RBD_B motions toward two-Up. Glycan shielding for neutralizing antibodies is the weakest in one-Open. Cryptic pockets are revealed at the RBD interfaces in intermediate structures between Down and one-Up. The inherent flexibility in S-protein is, thus, essential for the structure transition and shall be considered for antiviral drug rational design or vaccine development.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused over a 190 million infections and 4 million deaths, as of July 2021 (https://coronavirus.jhu.edu/map.html). It represents an urgent need for an effective medical intervention strategy, to avoid further social and economic consequences[1]. 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[2,3]. At the same time, more infectious mutants such as B.1.617.2 (Delta) and B.1.427 / B.1.429 (Epsilon) have appeared, and some evade from the immune system[4,5,6,7]. Furthermore, the virus ability to infect a wide range of vertebrates, which could act as a reservoir, points out the future risk despite the vaccination progress[8]. A deeper understanding of the virus molecular structure and infection mechanism is crucial to stop the virus transmission including mutant strains[9].

SARS-CoV-2, an enveloped positive single stranded RNA (+ssRNA) virus, has a large genome of approximately 30 kb encoding 29 proteins[9,10]. A transmembrane homotrimeric class I fusion glycoprotein decorating the virus, known as S-protein, plays a critical role in the viral cell entry[11,12]. In an immediate response to the pandemic, over a hundred Cryo-electron microscopy (EM) and X-ray crystal structures of S-protein have been reported and rapidly advanced in our understanding of the S-protein/receptor binding mechanism[13,14,15,16,17,18,19]. 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 S1)[16,20], initially binds to the host cell receptor angiotensin-converting enzyme 2 (ACE2)[19]. This binding is followed by priming of the C-terminal subunit (S2) and its large conformational change leading to the membrane fusion for the cell entry[12,21]. S-protein is covered by 66 N-glycans, 22 per protomer, to evade from the host cell immune system[22,23]. To block the initial binding with ACE2 either by stimulating the immune system using vaccines or by neutralized antibodies or small-molecule drugs is the primary target for medical interventions. Numerous Cryo-EM structures reveal that the RBDs can take inactive Down and active Up forms including one-Up, two-Up and three-Up conformations (Fig. 1a)[13,14,15,16,17,21,24,25]. The RBDs are considered to undergo large conformational changes from the receptor inaccessible Down to the receptor accessible Up states in the ACE2 binding[16,19,26]. Note that most of the Cryo-EM structures representing the active Up forms were determined together with other proteins, such as antibodies or a fragment of ACE2 (Table S1). A recent single molecule

fluorescence resonance energy transfer (smFRET) experiment suggested that the Down-to-Up transition occurs in S-protein even without its ligand, taking at least one transient intermediate[27]. This suggests the necessity to explore a wide conformational space of S-protein for describing transition pathways and intermediate structures. It would give us better understandings of the virus entry mechanisms and contribute to the development of antiviral drugs or antibodies.

Molecular dynamics (MD) simulations at the atomic level have been conducted to explore the conformational stability of S-protein by including the surface glycans, which are largely missed in the experimental structures. Several microsecond-scale MD simulations showed conformational flexibility of the stalk in S-protein, different levels of glycan-shielding between Down and Up, and their relevance in the ACE2 binding[22,28,29]. They also identified specific interactions between side-chain residues or those with glycans to stabilize either Down or Up[30,31]. The timescale of the Down-to-Up transitions in S-protein is far slower than that attainable one in the current MD simulations, leaving the inherent flexibility and transition pathways largely unknown. To overcome such difficulties, several enhanced sampling methods have been applied to investigate the Down-to-1Up-transition, including targeted MD[31], steering MD[32], nudged elastic band/umbrella sampling[33], adaptive sampling,[34] and weighted ensemble methods[35]. In most of these simulations, pre-defined reaction coordinates and/or bias potentials along the coordinates are used to enhance the motions of a single RBD in S-protein.

Here we apply the generalized replica exchange with solute tempering of selected surface charged residues (gREST_SSCR)[36], which enhances domain motions of a protein by exchanging the solute temperature of selected surface residues. 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. As listed in Table 1, the RBD motion in the RBD/SD1 monomer, the trimeric S-protein in the presence and absence of 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 intermediate structures during the transitions.

Results

Enhanced RBD motions in gREST_SSCR simulations

Table 1 lists all simulations performed in this study, wherein three gREST_SSCR simulations are involved. Two starting with a Down Cryo-EM structure (PDBID: 6VXX[20]) in the presence (gREST_Down) and absence (gREST_Down w/o glycan) of glycans and one from one-Up structure (PDBID: 6VYB[20]) (gREST_Up). All were solvated and ionized after modeling missing residues (SI Methods, Figures. S1 and S2). The simulations were performed using GENESIS software (2.0 beta)[37], which was designed to achieve high performance and scalability on the supercomputer Fukagu[38]. In the solvated S-protein system including about 655,000 atoms, 2,048 nodes on Fugaku were used to achieve the performance of around 52 ns/day. The total simulation times including all the replicas are 8 μs, 4.8 μs and 2.4 μs in gREST_Down, gREST_Up and gREST_Down w/o glycan, respectively.

Table 1. MD simulations of S-protein performed in this study.

Name	Model	Method	Simulations Length
gREST_Down	Spike Down w/ glycans	gREST_SSCR	500 ns × 16 replicas
gREST_Up	Spike Up w/ glycans	gREST_SSCR	300 ns × 16 replicas
gREST_Down w/o glycan	Spike Down w/o glycans	gREST_SSCR	150 ns × 16 replicas
Monomer_Down	RBD/SD1 monomer Down	cMD	300 ns × 1 run
Monomer_Up	RBD/SD1 monomer Up	cMD	300 ns × 2 runs
cMD_Down*	Spike Down w/ glycans	cMD	1,000 ns × 1 run
cMD_Up*	Spike Up w/ glycans	cMD	1,000 ns × 1 run

*Classical MD simulations of S-protein performed in our previous study[31].

gREST_SSCR facilitates conformational sampling of a multi-domain protein by scaling the Coulomb and Lennard-Jones interactions of selected charged residues at domain interfaces through the solute temperature exchanges[36]. Here, 8 pairs of positively and negatively charged residues (870 atoms in total) in the RBD region

were selected as the solute region of gREST_SSCR (Figure S3a). 16 copies of each simulation system were used to cover the solute temperature range from 310 to 545 K. For this method to work, the solute temperatures between neighboring replicas must be exchanged frequently. This, indeed, happened in all three simulations, showing sufficient overlaps of the potential energy distributions between neighboring replicas (Figure S3). All simulations show reasonable solute temperature exchange with the average acceptance ratio of 21.2%, 20.3% and 20.8% for gREST_Down, gREST_Up, and gREST_Down w/o glycan, respectively. The SD1/RBD hinge angle and the Cα root mean square deviation (RMSD) of each RBD upon fitting S2 provide simple measures of the 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°. Since gREST_SSCR is free from any pre-defined reaction coordinates, the Down-to-Up transitions happened not only in RBD_A but also in RBD_B or RBD_C. Even two-Up-like conformations were observed in some replicas within 500 ns/replica (for instance, replica 8 in gREST_Down, replica 4 in gREST_Up, and replica 16 in gREST_Down w/o glycan) (Figure S4). Despite the large-scale motions, intra-domain structures of three RBDs and NTDs were kept stable. The Cα 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 (Figures S5a and S5b). They are comparable to those in the conventional MD[31] (1.4 and 1.6 Å for RBD and NTD). 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 S5c and S5d).

Comparison between Cryo-EM and simulated structural ensembles

As of December 2020, 126 Cryo-EM structures of S-protein, including all Down, one-Up, two-Up and three-Up conformations, were deposited in the Protein Data Bank (Table S1). To describe the Down-to-Up transitions found in the Cryo-EM structures, we introduced three computational techniques: (i) the 9-beads representation per protomer as used in the studies by Henderson et al.[7,39] to build up a 27-beads model of S-protein (Figure 1a and Table S2), (ii) the rotation scheme to make the most significant RBD motion always happen in $Chain_A$ (Figures S6 and S7), and (iii) the principal component analysis (PCA) on the 27-beads model of S-protein 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_A undergoes the

Down-to-Up motion. The two lowest PCs cover about 85 % of the conformational fluctuations observed in the Cryo-EM structures. In Figure 1c, S-protein Cryo-EM structures with distinct RBD conformations, Down, one-Up, two-Up, and three-Up, are found in different clusters on the PC1-PC2 space (Figure 1c).

Next, we project the results of our previous $1\mu s$ conventional MD simulations (cMD_Down and cMD_Up) and the three gREST_SSCR simulations on the same space (Figure 1c). To obtain the distributions at 310 K in gREST_SSCR, we applied Multistate Bennett Acceptance Ratio (MBAR)[40] for utilizing all the trajectories at different solute temperatures. The distributions of cMD_Down and cMD_Up overlap with the corresponding Cryo-EM structures (Down and one-Up, respectively), while there is a big gap between the two simulations. Instead, the distributions of gREST_SSCR simulations reach from the starting state to another: gREST_Down: from Down to one-Up, gREST_Up: from one-Up to two-Up, gREST_Down w/o glycan: from Down to one-Up. Due to the insufficient computational times, the initial structure dependences on distributions are found in all the three gREST_SSCR simulations. This issue would be resolved if we continue the simulations much longer timescales. More importantly, gREST_SSCR could cover more than one state, opening the possibility to investigate inherent flexibility of S-protein, transition pathways, and the intermediate structures. The distribution of gREST_Down w/o glycan is wider than that in gREST_Down, suggesting that glycans on the surface of S-protein play key roles in regulating RBD motions.

To focus on the Down-to-Up motions in S-protein, we define the hinge and twist angles using the Cα 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 reveal the transition toward Up conformations. The gREST_SSCR distributions at 310 K on the hinge-twist angle space are compared to protomers of the 126 Cryo-EM structures in Figure S8, showing a good overlap between the Cryo-EM structures and structure ensembles at 310 K in gREST_SSCR as we see on the PC1-PC2 space. Interestingly, Monomer_Up, which consists of only a monomer containing SD1 and RBD (Up) sampled Down, one-Up, and one-Open states within 300 ns cMD simulations. The simulated ensembles in Monomer_Up is close to the Cryo-EM structures and the gREST simulations containing a whole S-protein.

On the $Hinge_A$-$Hinge_B$ space at 310 K, the Down-to-Up motion is also identified in $Chain_B$ in gREST_Up simulations starting from the one-Up structure in RBD_A (Figure 2c). We did not observe similar motions in the $Hinge_A$-$Hinge_C$ space, suggesting that the Up motions of RBD_B is followed from RBD_A to form two-Up

structures. Taking together the information on the PC1-PC2 and the $Hinge_A$-$Hinge_B$ spaces, four major structures were identified: the Down symmetric (D_{Sym}: ($Hinge_A$, $Hinge_B$) = (114.9˚, 112.5˚)), one-Up (1U: (158.3˚, 112.7˚)), one-Up-open ($1U_O$: (163.3˚, 108.9˚)), and two-Up-like ($2U_L$: (150.6˚, 148.6˚)). They are superimposed to the Cryo-EM structures having corresponding RBD conformations in Figure 2d. The simulation-derived D_{Sym} and 1U conformations are better aligned to the recent high-resolution structures of Down (PDB:6ZGE)[13] and one-Up (PDB:6XKL)[14], respectively. Intriguingly, RBD_A in $1U_O$ is well aligned to that in one of the RBDs in three-UP Cryo-EM structure (PDB:7DCC)[41]. RBD_A and RBD_B in $2U_L$ also agree well with those in two-Up Cryo-EM structure (PDB:6X2B)[39], although $2U_L$ in gREST_SSCR remains inter-domain interactions between RBD_A and RBD_B, which is completely lost in the two-Up Cryo-EM structures.

The accessibility of RBD in different conformations

To get insights into the contribution of each conformation to the ACE2 and neutralizing antibodies (nAbs) bindings, the accessibility of RBD is examined in terms of the Solvent Accessible Surface Area (SASA)[30]. Figures 3a and 3b (and Figures S9-S11) show the per-residue SASA around the receptor binding motif (residues I410 to V510, referred to as RBM hereafter) calculated for D_{Sym}, 1U, $1U_O$ and $2U_L$ conformations. Four mutational residues of concern, K417, L452, E484, and N501[5], are highlighted in red. In Down conformation (D_{Sym}), glycans at N165 and N343 largely shield RBM, leaving only small part accessible. RBM SASA increases in one-Up conformations (1U and $1U_O$) compared to Down, wherein $1U_O$ exhibits the utmost increase suggesting its potential contribution to the receptor and/or antibody binding. In contrast, the SASA increases only slightly, even decreases locally around residue C480, in two-Up ($2U_L$) conformation due to the mutual interactions between two RBDs. This finding suggests that one-Up conformations are the primary target of nAbs and the bindings of multiple nAbs likely proceed in a stepwise manner through one-Up conformations. Notably, K417 and N501 are accessible only in one-Up conformations as expected, while E484 is accessible in both Down and Up, inferring a widespread effect of E484 mutations. On the other hand, L452 is not accessible in any conformations and hence is expected to rarely affect the binding with either receptor or antibody, in wild type.[42] In contrast L452R mutation has shown to reduce neutralization[5], suggesting its indirect contribution (e.g. the alternation of the S-protein conformational equilibria) to the immune escape.

Figure 3c shows the calculated SASA for each epitope region of the 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)[5]. The epitope regions of Class I and II antibodies are largely exposed and almost completely de-shieled from glycans in Up conformations. On the other hand, the epitope region for Class III antibody is less exposed and largely shielded by glycans. This agrees with that Class III antibodies recognize "glycoepitopes"[43]. Note that the epitope region of Class IV antibody is slightly exposed only in $1U_O$ conformation. Further opening of RBD may allow the binding of Class IV antibody, such as CR3022, as suggested in a previous computational study[34]. Figure 3c shows the putative binding models with antibodies obtained by aligning the antibodies in Cryo-EM structures with each of Up conformations (see Figure S12 and S13 for detail). The 1U conformations can be targeted by any of Class I-III antibodies. On the other hand, $1U_O$ and $2U_L$ conformations are likely bound with Class I/II and Class II/III, respectively, suggesting the possible evade from some antibody attacks.

Comparison between smFRET and simulated structural ensembles

Using smFRET, Lu et al. showed the conformational dynamics of S-protein in the presence or absence of its receptor, hACE2.[27] They characterized four structural ensembles including two types of Down (major and minor), one intermediate and one-Up, suggesting the inherent flexibility of the RBD region regardless of the receptor binding. To examine whether gREST_SSCR could reproduce the experimental data, we first classified each gREST_SSCR trajectory at 310 K by means of *k*-means clustering and re-clustering guided by the hinge and twist angles distributions (Figures S14-S16, Table S3) and then computed the distance between residues 425-431 in RBD_A and residues 554-561 in $SD1_A$ to correlate the smFRET intensity by Lu et al.[27] By combining the distance distributions from gREST_Down and gREST_Up, we obtained the five conformational ensembles (Figure S17): Down symmetric ($Down_{Sym}$), Down like ($Down_{Like}$), two intermediates (Int2 and Int3), and Up (1Up). The smFRET distance alone cannot distinguish various Up conformations including 1U, $1U_O$, and $2U_L$ and thus we refer them to as 1Up. $Down_{Sym}$ and $Down_{Like}$ 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 are median distances of around 38 Å and 40 Å, respectively. Since Int3 has a large distance distribution that overlaps with Int2 and both intermediates might be indistinguishable in the smFRET experiment, the four conformations corresponding to

the smFRET intensity of 0.8, 0.5, 0.3 and 0.1 in the absence of hACE2[27] might be $Down_{Sym}$, $Down_{Like}$, Intermediate (Int2 and Int3) and 1Up. Each ensemble is also characterized with hinge angle distributions: $Down_{Sym}$:Hinge < 120°, $Down_{Like}$: Hinge < 130°, Int2:120° < Hinge < 140°, Int3:Hinge ~ 140°, 1Up:140° < Hinge < 160°.

Transition pathways and transient interactions stabilizing the intermediate structures

Based on the free-energy landscapes at 310 K and the comparison between the gREST simulations and smFRET experiments, we now have evidence on the inherent flexibility of S-protein in the absence of its ligand. We next focus on molecular mechanisms underlying the Down-to-1Up transitions in S-protein in terms of the correlated motions of the hinge and twist angles in three RBDs (Figure S18). Note that such correlated motions are hardly obtained using targeted MD simulations (Figure S19) or the MD simulations enhanced with pre-defined reaction coordinates[35,44]. As expected, $Hinge_A$ and $Twist_A$ are highly correlated with each other (Figure S18c), while $Hinge_A$ and $Hinge_B$ show almost no correlations (Figure S18a). Some correlations exist between $Hinge_A$ and $Hing_C$/$Twist_B$, which appear to relate with the preference of two intermediates.

To examine the origin of such correlated motions in three RBDs, contact analysis (Figure S20) and hydrogen bond (HB) analysis (Figure S21) were carried out for each cluster observed in gREST_Down and gREST_Up simulations as we did in the previous study[31]. In Figure 4, several key contacts and hydrogen bonds are highlighted to see drastic changes of the interactions in the Down-to-1Up transition. The glycans at $N343_B$ and $N234_B$ switch the interactions with RBD_A along the transition from Down to one-Up-like. In the intermediates, I2a and I3a, the glycan at $N343_B$ changes its contact partners by inserting underneath RBD_A. Concurrently, the formation of salt-bridge interactions between RBD_A and RBD_C, for instance, $R408_A$-$D405_C$ and $R408_A$-$D406_C$, lift up RBD_A from its Down to intermediate structures. Finally, in $1U_L$, $N234_B$ inserts into the newly formed cavity between RBD_A and RBD_B, forming new contacts with $D428_A$ and $T430_A$. The glycan at $N165_B$ is in contact with RBD_A throughout the transition, likely acting as a main barrier for the transition.

We can also characterize the 1Up-to-2Up transition (Figure S22a), where RBD_B undergoes the transitions. The increase in RBD_B hinge angle is accompanied with a slight decrease in RBD_A hinge angle. The latter decrease

likely relates with the increases of the contacts and HBs between RBD_A and RBD_C from 1Ua, the top populated cluster in 1Up, to two-Up-like structures, for instance, $K378_A$-$E484_C$, $Y369_A$-$N487_C$ and $D427_A$-$Y505_C$ (Figures S20 and S21). The numbers of contacts and HBs in two-Up-like structures are much less than those in one-Up, suggesting that a drastic reduction of inter-domain interactions is required toward a full two-Up conformations. No specific protein-glycan interactions are found to support the 1Up-to-2Up transition.

Finally, we examine the effect of glycans on the Down-to-Up transitions. Figure S22b shows the distributions of gREST_Down w/o glycan simulation on the $Hinge_A$-$Hinge_B$ space. Without glycans on the surface of S-protein, the structure distribution becomes 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 distribution than the others. The transition mechanisms seem to be the same as those with glycans: RBD_A and RBD_C form transient interactions to support the Down-to-1Up transition. However, the two-Up-like structure ($2Ub_L$) is more connected with Down distribution, suggesting that structural integrity of S-protein is lost without glycans on the surface.

Searching for cryptic binding pockets in the intermediate structures

We applied a machine learning based algorithm (P2Rank)[45] to search for the formation of druggable pockets in the intermediates structures, I2a, I3a and I3b (Figure S23a). The same search was also carried out for Down and one-Up structures for comparison. Figure 5a shows the formation of two cryptic pockets (pocket1 and pocket2) at the interface of RBD in I2a, which are not observed in Down or one-Up. Table S4 lists these predicted pockets showing relatively high scores in all the intermediates. To test the druggability of the two pockets, we performed virtual screening of FDA approved drugs from ZINC database[46], where we docked 2,115 molecules to the RBD interfaces in I2a, I3a and I3b (following the procedure sketched in Figure 23c). Table S5 shows a list of the index (ID) and binding energies of top ranked molecules. Notably for Nilotinib (Figure 5b), which is ranked at the top 5 and 6, the previous study has shown that it affects the SARS-Cov-2 infectivity[47]. The authors attributed the effect to the reduction of cell entry, but the mechanism remains unknown[47]. Figures 5c and S23d show the top-ranked binding poses of Nilotinib in the three intermediates. Nilotinib binds to either pocket1 or pocket2 in all first 9 binding modes with high binding affinity (Table S6).

These results serve the two cryptic pockets as potential targets to stabilize intermediate structures and to 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 second, 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[44]. Yet another approach is to use enhanced sampling methods. Sztain and co-workers performed the weighted ensemble simulations, collecting 130 microseconds trajectories, to characterize the opening pathway[35]. Fallon et al. employed the umbrella sampling, nudged elastic band, and steered MD to characterize the transition pathway and associated cryptic pockets[33]. Our gREST_SSCR simulations aimed to enhance sampling of multi-domain proteins like S-protein by exchanging the solute temperatures of selected surface charged residues. It is free from reaction coordinates and bias potentials, which were used in many of the 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[37,38], the simulations explored a wide conformational space covering Down, one-Up, one-Open, and two-Up-like structures. The structure ensembles that we obtained in the two gREST_SSCR simulations agree with the Cryo-EM structure distributions and smFRET data, suggesting inherent flexibility of S-protein structures even without its ligand, such as hACE2 or nAbs.

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 free-energy landscape containing all the important structures of S-protein. Three-Up structures, which were found in Cryo-EM, for instance, PDB:7DCC[41], were not observed in our 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 via gREST_SSCR is much better than cMD without any pre-defined reaction coordinates and/or bias potentials, providing sufficient structural information to investigate 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 inherent flexibility of S-protein structures in absence of its ligand, exploring Down, one-Up, one-Open, and two-Up-like structures at 310 K. The trimeric RBDs in S-protein is intrinsically flexible even in Down form, as their interfaces are electrostatically repulsive as shown in our previous study[31]. Not only symmetric structures but also anti-symmetric structures are also observed in Down, suggesting the existence of "flexible Down structure" (Figure S24). This flexibility can initiate the Down-to-1Up transition, by reducing inter-domain interactions found in $Down_{Sym}$ and allowing the insertion of a glycan at $N343_B$ underneath of RBD_A (Figure S25). This is consistent with the experimentally mobile RBD conformations in Down observed by Ke et al.[11] and Wrobel et al.[13].

Following the flexible Down structure, our simulations elucidate molecular mechanisms underlying the Down-to-1Up transition involving RBD_A and the 1Up-to-2Up transitions involving RBD_A and RBD_B. This sequence of events can be explained via the changes of inter-domain interactions as well as protein-glycan interactions. Since RBD_A-RBD_C interaction becomes stronger in the Down-to-1Up transition as we see in Figures 4, S20, and S21, RBD_B is 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 $R408_A$-$D405_C$ and $R408_A$-$D406_C$ for stabilizing the intermediates (Figures 4 and S21). The salt-bridge between $R408_A$ and $D405_C$ was also observed in the weighted ensemble simulations reported by Sztain et al.[35] This scenario is perturbed via protein-glycan interactions, in particular, those involving three glycans at $N165_B$, $N234_B$, and $N343_B$ as pointed out in previous studies[29,30,31]. In the current study, we highlighted the dual roles of these three glycans: the stabilization of one of the states and the driving force toward the other state. The glycans at N165 stabilize Down, representing a barrier for the Down-to-Up transition, while the glycan at $N343_B$ drives the transition. Finally, the glycan at $N234_B$ stabilizes one-Up-like conformation. This picture coincides with the experimental results by Henderson

et al.[48], where the population of one-Up drastically is reduced by the glycan deletions at N234 but increases by those at N165. As suggested previously[30], 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 S25). The role of glycans 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.[49]. Recently, a similar role of the glycans at N343 has been proposed by Sztain et al.[35]

Implication for vaccine and drug developments

Since the drug repurposing studies[50,51,52] so far mostly target Down and Up conformations from Cryo-EM structures, the identification of druggable cryptic pockets in transient intermediates would introduce unprecedented drug targets[44]. Two cryptic pockets at the RBD interface are identified in the highly populated intermediates along the Down-to-1Up transition. From our virtual screening of FDA approved molecules, these pockets are druggable and accommodates several small molecules including Nilotinib. Intriguingly, Cagno et al.[47] reported that Nilotinib reduces SARS-CoV-2 infection by ~50% via interfering with viral cell fusion and replication. In addition, Nilotinib as well as other top ranked molecules (Table S5) were also predicted to bind RBD in the previous virtual screening studies[52,53]. It is yet unclear if the binding of molecules at RBD interface stabilizes 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[14,25,39,54,55,56]. In this study Nilotinib is only shown as an example, while further studies to identify better molecules targeting the intermediate states would pave the way to hinder viral entry.

The understanding of antibody responses is critical for vaccines and neutralizing antibody developments. The large exposure of RBD_A in Up is shown in the current study, which is consistent with several previous studies[22,30,57]. We found that the accessibilities of RBM and the antibody epitopes, including mutational sites, depend sensitively on the RBD conformation, suggesting that each of the Up conformations (1U, $1U_O$ and $2U_L$) tends to bind with distinct classes of antibodies. From our inspection, $1U_O$, 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)[5] or reducing the antibody binding (K417N or E484K)[5]. Intriguingly, HB analysis shows that K417 and E484 respectively contribute to stabilize the intermediates I2a/I3b and $2Ua_L$, likely enhancing the population of one-Up conformations. Note that L452 shows a little accessibility (Fig. 3) in contrast with the reported severe mutational effect to an antibody recognition[5]. The similar observation was also reported previously[42]. A possibility is that L452R mutation affects the structural flexibility of S-protein, which consequently gives rise to the enhanced infection. The last point to mention is that the predicted Up conformations are reasonably aligned with the antibody-bound S-protein Cryo-EM structures (Fig. S13). Hence, the antibody bindings of S-protein can be explained based on the conformer selection mechanism[58]. 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[59].

Methods

gREST_SSCR simulations

The initial structures for S-protein head regions (residue 28-1135) were prepared based on the Cryo-EM structures PDB:6VXX and PDB:6VYB for the Down and Up states, respectively[20]. 18 N-glycans and 1 O-glycan were added per protomer as suggested in previous mass-spectrometry experiments and computational models[23,60]. A full list of included glycans is shown in Figure S2. CHARMM-GUI[61] were used to prepare the final model including the addition of glycans, ions (0.15 M NaCl) and water molecules. Three gREST_SSCR simulations were performed two 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 in the solute region of gREST. The total of 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[37,38]. 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 2020/02/04-2020/12/11, released by 12/30) in this study. Only structures where the number of residues is greater than 700 and the last residue >= 1122 for each protomer were selected in the analysis. 373 protomers of 126 structures and 121 trimeric structures from PDB were satisfied with the criteria and used in this study (Table S1). We adopt a method representing the structure with a 9-beads per protomer as used in the previous work of Henderson et al.[7,39] Unlike their study, 27 beads consisting of three chains are used here. This coarse-grained model consists of 2 beads for RBD, 3 beads for NTD, 1 bead for SD1 and SD2 each, and 2 beads for the S2 region (CD and S2-b) (see Table S2). Principal component analysis (PCA)[62] was performed using the selected Cryo-EM structures, after converting to the 27 beads model. The PC vectors are 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 via the Multistate Bennett Acceptance Ratio (MBAR) method[40]. The 373 protomers of the Cryo-EM structures were also used to compare hinge/twist angles.

Pocket Search and Virtual Screening

The P2Rank software[45] were used to identify potential druggable pockets in intermediate structures. The cluster centres of 12a, 13a and 13b as well as $D1_{Sym}$ and $1U_L$ 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[46]. This includes 2,115 molecules representing 1,379 drug candidates. Open Babel was used to convert PDB to PDBQT. AutoDockTools-1.5.6 was used to prepare RBD receptor[63]. AutoDock Vina were used to dock all 2115 molecules and perform virtual screening[64].

Data Availability

The trajectories are available from the corresponding authors upon reasonable request.

Code Availability

The source code of GENESIS (https://www.r-ccs.riken.jp/labs/cbrt/) is distributed under the GNU Lesser General Public License version 3.

References

1. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **19**, 141-154 (2021).

2. Li Y, Tenchov R, Smoot J, Liu C, Watkins S, Zhou Q. A Comprehensive Review of the Global Efforts on COVID-19 Vaccine Development. *ACS Cent Sci* **7**, 512-533 (2021).

3. Venkadapathi J, Govindarajan VK, Sekaran S, Venkatapathy S. A Minireview of the Promising Drugs and Vaccines in Pipeline for the Treatment of COVID-19 and Current Update on Clinical Trials. *Front Mol Biosci* **8**, 637378 (2021).

4. Yurkovetskiy L*, et al.* Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**, 739-751.e8 (2020).

5. Harvey WT*, et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **19**, 409-424 (2021).

6. Wang P*, et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130-135 (2021).

7. Gobeil SM*, et al.* Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science*, 10.1126/science.abi6226, (2021).

8. Prince T, Smith SL, Radford AD, Solomon T, Hughes GL, Patterson EI. SARS-CoV-2 Infections in Animals: Reservoirs for Reverse Zoonosis and Models for Study. *Viruses* **13**, 494 (2021).

9. V'Kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155-170 (2021).

10. Finkel Y*, et al.* The coding capacity of SARS-CoV-2. *Nature* **589**, 125-130 (2021).

11. Ke Z*, et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498-502 (2020).

12. Shang J*, et al.* Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **117**, 11727-11734 (2020).

13. Wrobel AG*, et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* **27**, 763-767 (2020).

14. Hsieh CL*, et al.* Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501-1505 (2020).

15. Barnes CO*, et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682-687 (2020).

16. Wrapp D*, et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).

17. Cai Y*, et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592 (2020).

18. Wang Q*, et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894-904.e9 (2020).

19. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444-1448 (2020).

20. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).

21. Fan X, Cao D, Kong L, Zhang X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nat Commun* **11**, 3618 (2020).

22. Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci Rep* **10**, 14991 (2020).

23. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330-333 (2020).

24. Barnes CO*, et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842.e16 (2020).

25. Juraszek J*, et al.* Stabilizing the closed SARS-CoV-2 spike trimer. *Nat Commun* **12**, 244 (2021).

26. Xu C*, et al.* Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci Adv* **7**, eabe5575 (2021).

27. Lu M*, et al.* Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host Microbe* **28**, 880-891.e8 (2020).

28. Turonova B*, et al.* In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203-208 (2020).

29. Choi YK*, et al.* Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. *J Chem Theory Comput* **17**, 2479-2487 (2021).

30. Casalino L*, et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent Sci* **6**, 1722-1734 (2020).

31. Mori T, Jung J, Kobayashi C, Dokainish HM, Re S, Sugita Y. Elucidation of interactions regulating conformational stability and dynamics of SARS-CoV-2 S-protein. *Biophys J* **120**, 1060-1071 (2021).

32. Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Golcuk M. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *J Chem Phys* **153**, 075101 (2020).

33. Fallon L*, et al.* Free Energy Landscapes from SARS-CoV-2 Spike Glycoprotein Simulations Suggest that RBD Opening Can Be Modulated via Interactions in an Allosteric Pocket. *J Am Chem Soc* **143**, 11349-11360 (2021).

34. Zimmerman MI, *et al.* SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat Chem* **13**, 651-659 (2021).

35. Sztain T, *et al.* A glycan gate controls opening of the SARS-CoV-2 spike protein. *bioRxiv*, 10.1101/2021.02.15.431212, (2021).

36. Dokainish HM, Sugita Y. Exploring Large Domain Motions in Proteins Using Atomistic Molecular Dynamics with Enhanced Conformational Sampling. *Int J Mol Sci* **22**, 270 (2020).

37. Kobayashi C, *et al.* GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J Comput Chem* **38**, 2193-2206 (2017).

38. Jung J, *et al.* New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems. *J Comput Chem* **42**, 231-241 (2021).

39. Henderson R, *et al.* Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat Struct Mol Biol* **27**, 925-933 (2020).

40. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* **129**, 124105 (2008).

41. Zhang C, *et al.* Development and structural basis of a two-MAb cocktail for treating SARS-CoV-2 infections. *Nat Commun* **12**, 264 (2021).

42. Valdes-Balbin Y, *et al.* Molecular Aspects Concerning the Use of the SARS-CoV-2 Receptor Binding Domain as a Target for Preventive Vaccines. *ACS Cent Sci* **7**, 757-767 (2021).

43. Pinto D, *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290-295 (2020).

44. Brotzakis ZF, Lohr T, Vendruscolo M. Determination of intermediate state structures in the opening pathway of SARS-CoV-2 spike using cryo-electron microscopy. *Chem Sci* **12**, 9168-9175 (2021).

45. Krivak R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J Cheminform* **10**, 39 (2018).

46. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **45**, 177-182 (2005).

47. Cagno V, Magliocco G, Tapparel C, Daali Y. The tyrosine kinase inhibitor nilotinib inhibits SARS-CoV-2 in vitro. *Basic Clin Pharmacol Toxicol* **128**, 621-624 (2021).

48. Henderson R, *et al.* Glycans on the SARS-CoV-2 Spike Control the Receptor Binding Domain Conformation. *bioRxiv*, 10.1101/2020.06.26.173765, (2020).

49. Li Q, *et al.* The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **182**, 1284-1294.e9 (2020).

50. Han N, *et al.* Identification of SARS-CoV-2-induced pathways reveals drug repurposing strategies. *Sci Adv* **7**, eabh3032 (2021).

51. Bakowski MA, *et al.* Drug repurposing screens identify chemical entities for the development of COVID-19 interventions. *Nat Commun* **12**, 3309 (2021).

52. Deganutti G, Prischi F, Reynolds CA. Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. *J Comput Aided Mol Des* **35**, 195-207 (2021).

53. Murugan NA, Kumar S, Jeyakanthan J, Srivastava V. Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach. *Sci Rep* **10**, 19125 (2020).

54. Edwards RJ*, et al.* Cold sensitivity of the SARS-CoV-2 spike ectodomain. *Nat Struct Mol Biol* **28**, 128-131 (2021).

55. McCallum M, Walls AC, Bowen JE, Corti D, Veesler D. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat Struct Mol Biol* **27**, 942-949 (2020).

56. Toelzer C*, et al.* Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **370**, 725-730 (2020).

57. Sikora M, von Bulow S, Blanc FEC, Gecht M, Covino R, Hummer G. Computational epitope map of SARS-CoV-2 spike protein. *PLoS Comput Biol* **17**, e1008790 (2021).

58. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **35**, 539-546 (2010).

59. Ren W, Dokainish HM, Shinobu A, Oshima H, Sugita Y. Unraveling the Coupling between Conformational Changes and Ligand Binding in Ribose Binding Protein Using Multiscale Molecular Dynamics and Free-Energy Calculations. *J Phys Chem B* **125**, 2898-2909 (2021).

60. Woo H*, et al.* Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J Phys Chem B* **124**, 7128-7137 (2020).

61. Lee J*, et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12**, 405-413 (2016).

62. Yang LW, Eyal E, Bahar I, Kitao A. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics* **25**, 606-614 (2009).

63. Morris GM*, et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785-2791 (2009).

64. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-461 (2010).

Acknowledgements (optional)

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: hp200153, hp200028). The research was supported in part by MEXT as "FLAGSHIP 2020 project", "Priority Issue on Post-K computer" (Building Innovative Drug Discovery Infrastructure Through Functional Control of Biomolecular Systems), "Program for Promoting Researches on the Supercomputer Fugaku" (Biomolecular dynamics in a living cell/MD-driven Precision Medicine), and MEXT/KAKENHI (Grant Numbers 19H05645 (to YS), 20K15737 (to H.M.D.), 19K12229 (to SR), and 19K06532 (to TM)) and by RIKEN Pioneering Research Projects (Dynamic Structural Biology/Glycolipidologue Initiative/Biology of Intracellular Environments) (to YS).

Ethics declarations

Competing interests

The authors declare no competing interests.

Supplementary Information

Supporting information is available for this paper.

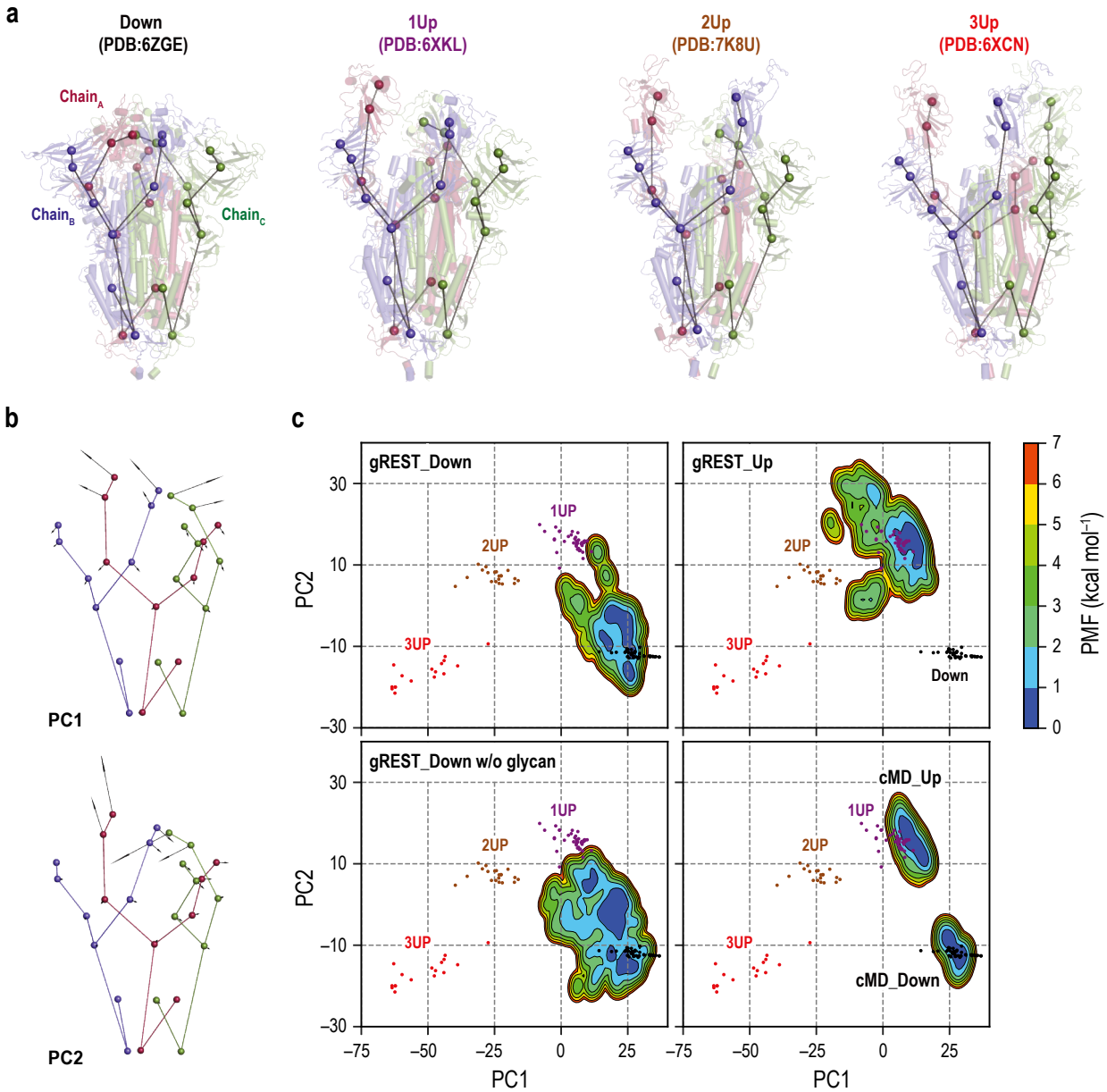

Figure 1. Comparisons of S-protein structures in MD simulations with Cryo-EM structures. a) 27 coarse-grained beads representations of four representative Cryo-EM structures: Down (PDB ID: 6ZGE), 1Up (6XKL), 2Up (7K8U), and 3Up (6XCN). Chain A, B, and C in S-protein are shown in red, blue, and green, respectively. **b)** The lowest two modes (PC1 and PC2) in principal component analysis of the 27 beads model of 126 Cryo-EM structures. PC1 and PC2 respectively represent a symmetric and an antisymmetric Down to Up motion of the receptor binding domains. 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 and 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.

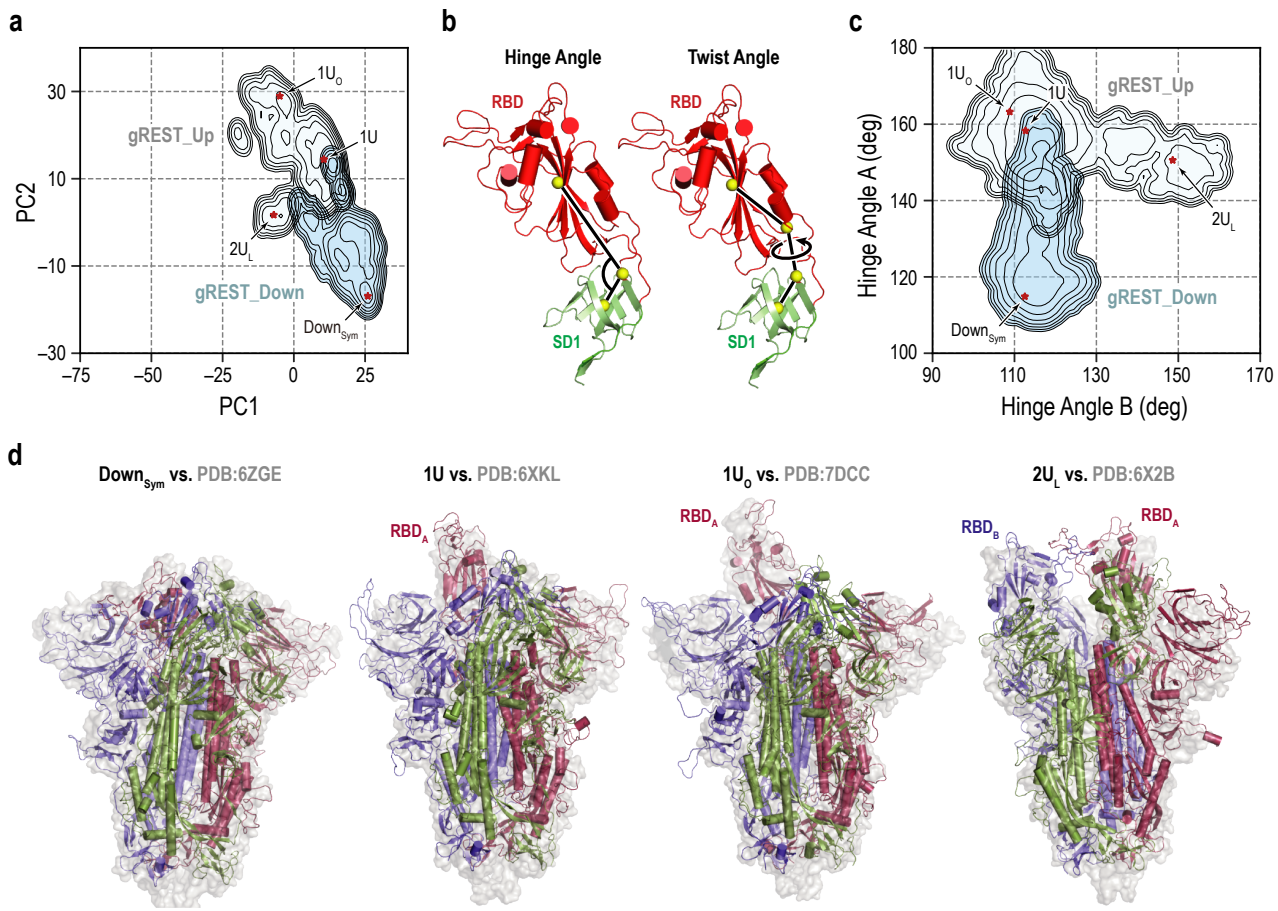

Figure 2. Representative RBD conformations from MD simulations. a) An overlay of the two free energy landscapes at 310 K along the PC1 and PC2 obtained from gREST_Down (light blue) and gREST_Up (light cyan) simulations. The red dots represent the positions of four representative RBD conformations: symmetric Down ($Down_{Sym}$), 1RBD Up (1U and $1U_O$) and 2RBDs Up like ($2U_L$) conformations. **b)** Definitions of the hinge and twist angles representing the RBD conformations. The hinge angle is determined by three center of masses (COMs, yellow spheres) of the core and top residues of SD1 (green, Cα atoms only) and the core residues of RBD (red, Cα 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_A and RBD_B 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 grey surface). Chain A, B, and C in S-protein are shown in red, blue, and green, respectively.

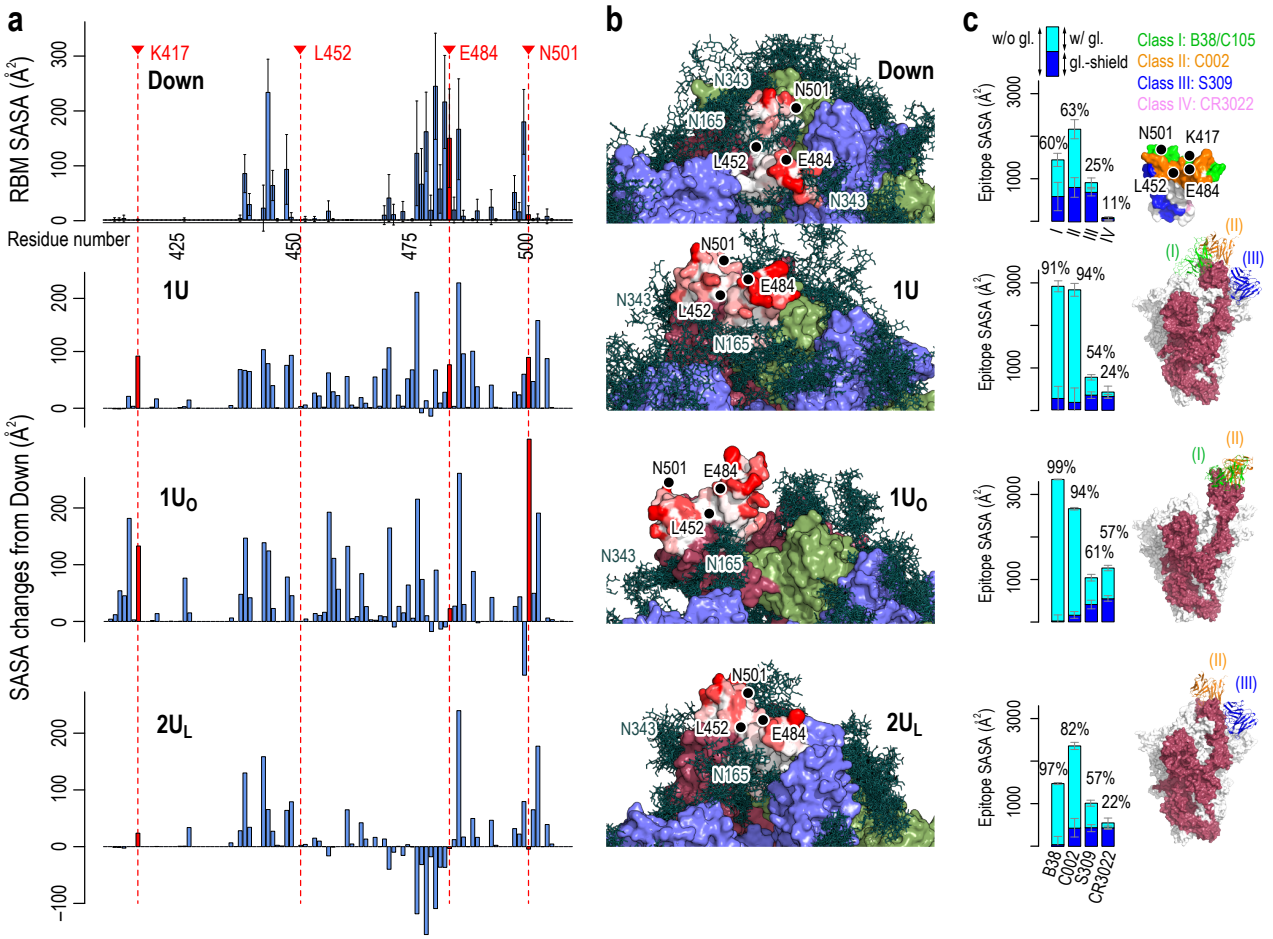

Figure 3. Accessibility of receptor binding motif (RBM). a) Per-residue solvent accessible surface area (SASA) values of the receptor binding motif (RBM, residues 410 to 510) in Down conformation (top) and their changes in Up conformations (bottom three). SASA values were calculated using the probe radius 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 to 260 Å^2, the maximum value in Down, the values higher than this are truncated for consistent color scheme). Chain 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 show 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).

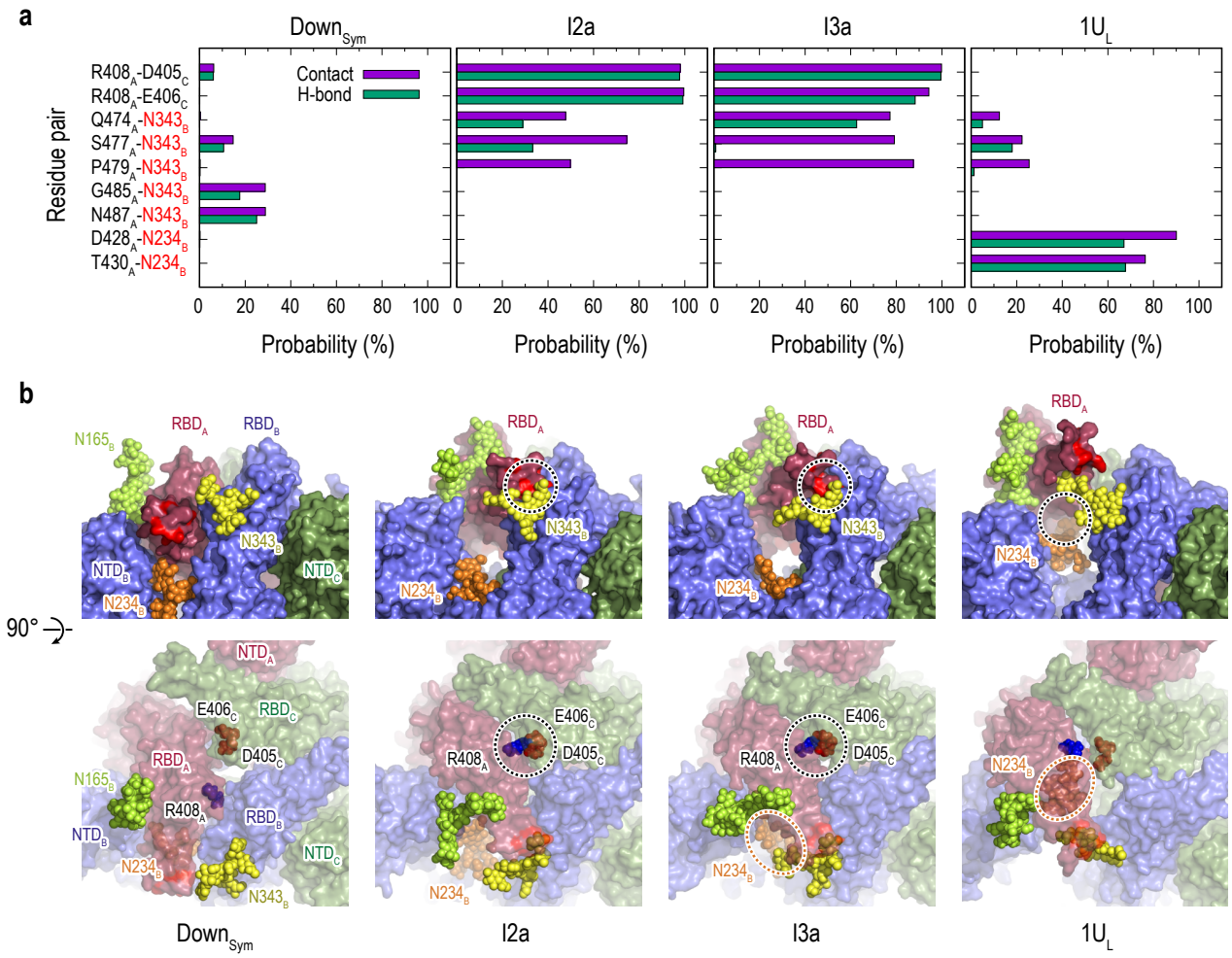

Figure 4. Protein-protein and protein-glycan interactions critical for Down-to-Up transition. a) Probability of finding the hydrogen bond (green) and contact (purple) pairs between protein residues or protein-glycans that markedly change along the transition pathway (D_{Sym}, I2a, I3a and $1U_L$). All hydrogen bond (probability of finding of >50%) and contact pairs (probability of finding of >70%) are shown in figure S19 and S20, respectively. **b)** Typical snapshots of the protein-protein and protein-glycan interactions along the transition pathway. Chain 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 $N343_B$–RBD_A contact are highlighted in red surface (top). The salt-bridges formed by $R408_A$ (blue), $E406_C$ (brown) and $D405_C$ (red) are also highlighted with black dashed circles (top and bottom), while the location of N234 glycan is highlighted with orange dashed circles.

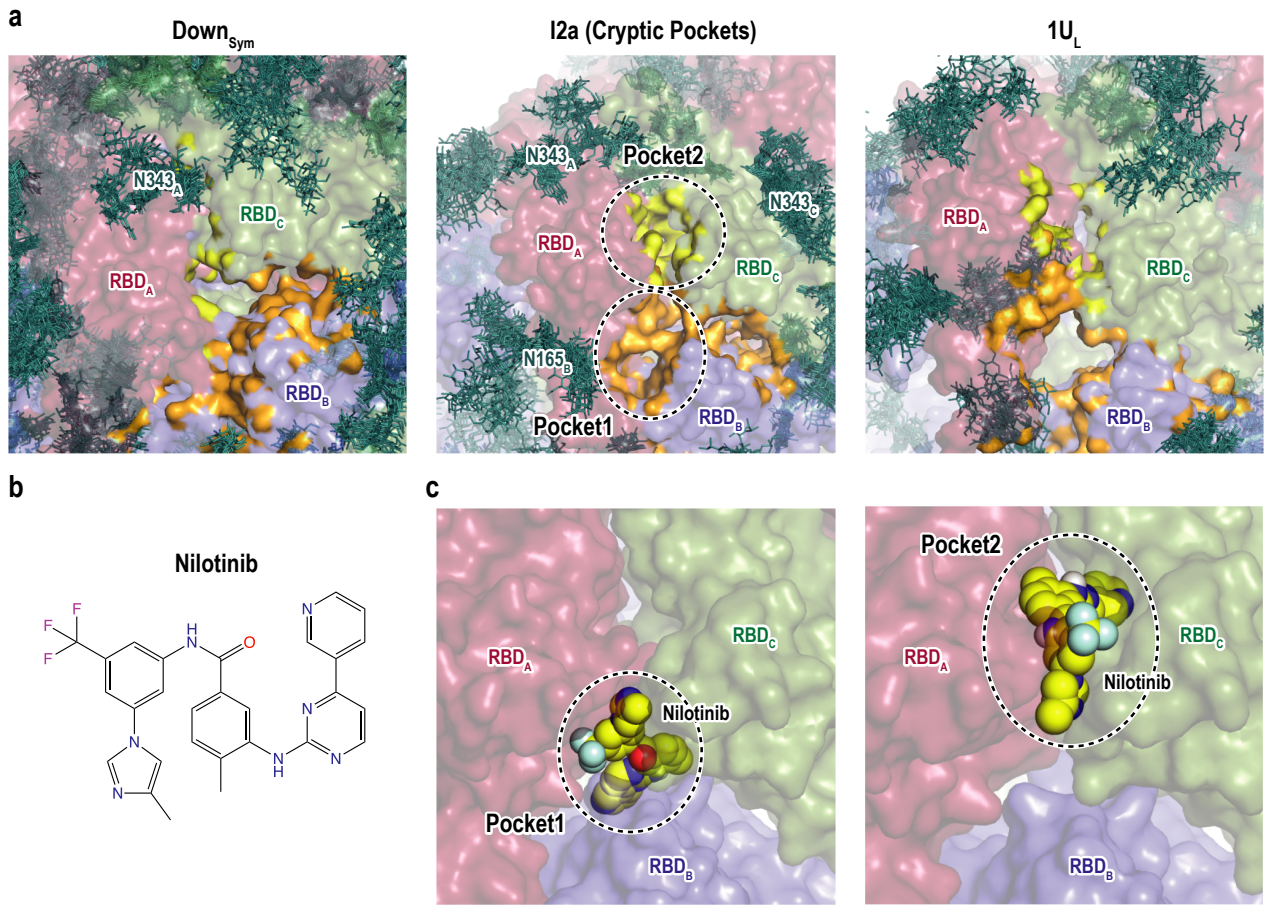

Figure 5. Druggable cryptic pockets. a) Snapshots of RBD interface in Down symmetric ($Down_{Sym}$), Intermediate 2a (I2a) and 1Up like ($1U_L$) conformations. Chain A, B, and C in the protein are shown in red, blue, and green surfaces, respectively, while a collection of glycans from 10 snapshots are 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)** Chemical structure of Nilotinib and **c)** the 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.