1   **FRONT MATTER**

2

3   **Title**

4

5   ***The SARS-CoV-2 Spike Variant D614G Favors an Open Conformational***

6   ***State***

7

8   **Authors**

9

10   Rachael A. Mansbach,[1]†‡ Srirupa Chakraborty,[1,2]† Kien Nguyen,[1]†, David C. Montefiori,[3] Bette

11   Korber,[1] S. Gnanakaran[1]*

12

13   **Affiliations**

14   [1] Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM

15   87545

16   [2] Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545

17   [3] Duke Human Vaccine Institute & Department of Surgery, Durham, NC 27710

18

19   * Correspondence and requests for materials should be addressed to S. Gnanakaran

20   gnana@lanl.gov

21   † These authors contributed equally to this work

‡ Current affiliation: Physics Department, Concordia University, Montreal, Quebec, Canada, H4B IR6

Abstract

The COVID-19 pandemic underwent a rapid transition with the emergence of a dominant viral variant (from the "D-form" to the "G-form") that carried an amino acid substitution D614G in its "Spike" protein. The G-form is more infectious in vitro and associated with increased viral loads. To gain insight into the molecular-level underpinnings of these characteristics, we employed microsecond all-atom simulations. Here we show that changes in the protein energetics favor a higher population of infection-capable (open) states through release of hydrogen bonds of an asymmetry present in the D-form but not the G-form. Thus, the increased infectivity of the G-form is likely due to a higher rate of profitable binding encounters with the host receptor. It is also predicted to be more neutralization sensitive due to enhanced exposure of the receptor binding domain, a key target region for neutralizing antibodies. These results are significant for vaccine design.

MAIN TEXT

Introduction

The coronavirus SARS-CoV-2 is responsible for the COVID-19 pandemic, a global emergency. The virus infects human cells through a process initiated by the binding of the Spike protein on the viral surface to its receptor on human cells, the angiotensin-converting 2 (ACE2) protein. The Spike

47 protein comprises three identical protomers, non-covalently bonded protein subunits, which can

48 adopt different conformations (**Fig. 1**) that can enable ACE2 interactions through the exposure of

49 the Receptor Binding Domain (or RBD). Once bound, the S1 subunit (**Fig. 1A, red**) of the protein

50 detaches ("sheds"), and the S2 subunit (**Fig. 1A, blue**) triggers membrane fusion and mediates viral

51 entry. In the "all-down" Spike conformation, all three Receptor Binding Domains of the protomers

52 are packed together on the top, each one being in what is referred to as a "closed" conformation

53 (**Fig.1A**). Binding of the Receptor Binding Domain to ACE2 is thought to require the transition of

54 one of the protomers from a closed to a more accessible open conformation [1, 2] (**Fig. 1B**). This

55 infection-capable configuration is referred to as the "one-up" Spike conformation. The "two up"

56 and "three up" states, in which more than one protomer is open, are less stable and can be captured

57 in the related SARS-CoV1 Spike only by employing stabilizing mutations in SARS-CoV1 [3] or in

58 the SARS-CoV2 Spike by introducing non-native disulfide bonds [4]. A viral variant has recently

59 emerged, carrying a single amino acid substitution in Spike at residue 614 from an aspartic acid (D)

60 to a glycine (G) (D614G). This G-form (also referred to as G614) is now the globally prevalent

61 form and is potentially more transmissible than the originally observed D-form (also referred to as

62 D614) [5]. It is associated with increased viral nucleic acid in the upper respiratory tract, which

63 potentially indicates more viruses attaching to ACE2 receptors, [5] and has higher infectivity in

64 pseudotype virus assays in multiple cell types [5, 6]. Even though there have been several studies

65 on the infectivity and transmissibility of D614G substitution in laboratory settings, the real effect

66 of this substitution at the population level is still being investigated [7]. To identify the molecular

67 factors that determine these experimentally observed differences, we performed 20 microsecond-

68 length all-atom molecular dynamics simulations of the full Spike protein in explicit solvent.

69   This extensive atomistic study of the differences between the globally most dominant G-

70 form of the COVID-19 pandemic and the original D-form is of immediate relevance for ongoing

71 vaccine and antibody-based therapeutic studies. We generated four separate sets of simulations,

72 each with five replicas, to study both the D614 form and G614 forms, in both the all-down and one-

73 up states. Using these simulations, we compared interactions formed between twelve defined

74 regions of the Spike protein (**Fig. 1C**). To identify the factors that lead to distinct dynamics for each

75 of the four models, we compared residue-residue interactions and correlations, hydrogen bonding,

76 and exposure of ACE-2 binding and Receptor Binding Domain antibody epitope sites. We found

77 that the D614G substitution was directly associated with far-reaching alterations in interactions

78 between different protomers and indirectly associated with changes in binding site exposure.

79 Specifically, in the G614 form there is a rearrangement in residue-residue contacts and hydrogen

80 bonding in comparison to the D614 form that can lead to an increase in the population of the

81 infection-capable one-up state. A careful analysis of Receptor Binding Domain exposure showed

82 that both the ACE2 binding site and critical sites for the binding of neutralizing antibodies

83 ("epitopes") in the Receptor Binding Domain are more accessible in the one up state [1, 8, 9].

Results

Individual protomers of the Spike undergo distinct fluctuations (**Fig. S1**). To distinguish between protomers when discussing the results, we refer to the protomer that transitions between the up and down orientations as the "Up" protomer. The other two protomers that are in the down configurations are referred to as, "L-down" and "R-down", depending on whether the protomer is on the left or right of the Up protomer, respectively (**Fig. 1B**).

Protein contacts are more symmetric between protomers in the G614 one-up state

To elucidate how specific interactions are altered by the D614G substitution, we calculated residue-residue contacts [10] that are formed between subunits of all three protomers. Specifically, these contacts include (a) S1-S2 interactions within a protomer (i.e. intra-protomer contacts), and (b) S1-S1, S2-S2, and S1-S2 interactions where the subunits are from different protomers (i.e. inter-protomer contacts). Two residues were considered to be in contact if their minimum distance was within a cutoff of 6Å. Contacts that occurred with probabilities of at least 50% were defined as "persistent" contacts.

The average number of persistent intra-protomer contacts between S1 and S2 was not affected by the D614G substitution in either the open or closed Spike conformation (**Fig. 2A-B**, first three data points). Similarly, regardless of the conformational state, the inter-protomer interactions between S1-S1 and between S2-S2 subunits were largely unaffected by the substitution (**Fig. 3A-B**). In contrast, the inter-protomer S1-S2 interactions are significantly impacted by the D614G substitution in the open conformation (**Fig. 2A-B**, last three datapoints). Specifically, in the all-down case, both the G- and D-forms lead to persistent numbers of contacts that are similar (within error bars) across all three protomer interfaces. This inter-protomer contact signature of the

all-down Spike protein can be described as "symmetric." In contrast, the D614 form loses the symmetrical contact signature when the Spike protein adopts the one-up conformation (**Fig. 2B, Table 1**), while the one-up G-form complex maintains the symmetry in the number of persistent contacts between the three protomers. The interprotomer S1-S2 interaction energy follows a similar pattern, in which the interaction energies as approximated by Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) calculation (see Methods) are asymmetric in the one-up D-form, but symmetric in the one-up G-form (**Fig. S2**).

G-form Receptor Binding Domain dynamics influenced by symmetric inter-protomer S1-S2 interactions

To further explore the factors that lead to the observed contact asymmetry, we calculated the cross-correlation matrix C of the Cα atoms for the entire complex (**Figs. 4, S3-S4**). For an equilibrium system, the cross-correlation matrix is a measure of the extent to which different parts of the protein move in a coordinated fashion in the same direction (red) or in a coordinated fashion in opposing directions (blue). This interpretation holds best for small fluctuations, such as our equilibrium simulations, which obey the quasi-harmonic approximation. The greatest differences between the all-down state and the one-up state lie in the RBD-RBD and N terminal domain (NTD)-RBD correlations (**Fig. S3M-P**). Consistent with the contact analysis above, we observe greater symmetry in the G614 form one-up state, relative to the D614 form one-up state. The magnitude of the correlations between and within all three Receptor Binding Domains are highly similar for the G614 form (**Fig. 4A**), but the intra-domain correlations and the U-RBD/R-RBD (anti)-correlations are stronger than the other inter-domain correlations in the D614 form (**Fig. 4B,** four squares in the lower right). There is also a slight asymmetry in the inter-domain correlations of the G-form all-down system (**Fig. 5A** in the off-diagonal elements shows slight segregation of blue and red compared with **Fig. 5B** in the off-diagonal elements). In other words, there are slight anti-

correlations of the U protomer motions with respect to the other two in the G-form, compared to slight correlations between the R and L in the G-form, whereas there is almost no such correlation in the D-form. This observation is consistent with a subtle deviation from symmetry observed for the all-down G-form (**Fig. 2A, Table 1**). Overall, atomistic correlations indicate that in the one-up state the motions of the RBDs in the G614 form are more synchronized than the D614 form. Such significant positional correlations distal to the 614 substitution site are a signature of allostery [11, 12].

Contacts in C Terminal Domains 1 and 2 (528-685) and the Fusion Peptide and Fusion Peptide Region (816-911) Domains are major contributors to the symmetrization in the G614 form

To capture the specific regions where the inter-protomer S1-S2 contacts are most affected by the D614G substitution, we carried out a global differential contact analysis, in which we identified persistent contacts that existed in one form but not the other in the all-down or one-up states. Although the overall number of contacts does not change for most of the interfaces, specific contacts do, with different contacts being gained at different interfaces (**Fig 6, purple stars, S4**). Furthermore, the equalization of inter-protomer S1-S2 contacts in the one-up G614 form arises largely from (i) the overall loss of contacts between the C Terminal 1 (CT1) domain of the L-down protomer and the Fusion Peptide (FP) domain of the Up protomer and (ii) the overall gain of contacts between the C Terminal 1 and the C Terminal 2 (CT2) domains of the Up protomer and the Fusion Peptide domain of the R-down protomer (**Fig. S6**). Interestingly, the mutation of residues in the C Terminal 1, Fusion Peptide, and Fusion Peptide Region domains to more hydrophobic residues was experimentally linked to an increase of protomers adopting the up conformation. This provides experimental support for the importance of these regions in determining the conformational state of the Spike [4].

We next study the hydrogen bonding interactions of the speculated critical residues D/G 614 with T859 [5]. The occupancy of the D614 – T859 hydrogen bond increases for the one-up versus all-down state in the D614 form (**Fig. 4C**, right two panels), but because the hydrogen bond no longer exists in the G614 form (**Fig. 4C**, left two panels), this bias likewise does not exist. This is another type of "symmetrization". In addition, there is a concomitant increase in the flexibility of residues 610 to 650 in the G614 form (**Fig. S5G-J**). In general (**Figs S5-S6**), we do not observe compensatory hydrogen bond formation with THR 859. Thus, it is conceivable that this hydrogen bond reduction leads to a conformational relaxation.

Symmetrization of inter-protomer interactions in G614 form can lead to higher population of infection-capable (one-up) conformation

The distinct contact signatures associated with the D614G substitution may be associated with alterations in the relative population of the all-down and one-up ensembles of the Spike protein (**Fig. 2**). For the G-form, the symmetry in the number of contacts among the inter-protomer interfaces suggests that the energetics between the protomers are similar [13, 14]. The connection between contacts and energetics is further supported by our MMBPSA calculations, which demonstrate a similar pattern (**Fig. S2**). Since the G-form preserves these interactions in both the all-down and one-up conformations, the interaction energy between up- and down-protomers would be similar to the energetics between two down-protomers. Inspired by this rationale, we use an Ising model, an established and straightforward method [15] that has been applied to many different applications in physics, to predict how the D614G substitution can affect the relative stability of the open and closed Spike conformations. In this model, each protomer can adopt an "up" or "down" state, and each protomer interacts with its two neighbors with an energy that depends only on the states of both protomers. Defining the Spike as a periodic three-spin system, the Ising model suggests that symmetrization of the energetics of the G-form will lead to an ensemble ratio of 75:25

182   of the one-up state to the all-down state, as compared to the 50:50 population ratio for the original

183   D-form (see SI for details). This demonstrates how a single residue substitution can facilitate the

184   adoption of the open Spike configuration, thereby enhancing viral infectivity.

185   In summary, we find three contributions to symmetrizations associated with the D614G

186   substitution: (1) symmetrization in the inter-protomer contacts, (2) symmetrization in the

187   correlations between the RBDs, and (3) symmetrization of a specific inter-protomer hydrogen bond.

188   We hypothesize that these three symmetrizations lead to a decrease in the energy differences

189   between the all-down and one-up states, leading to an increase in infectivity through an increase in

190   the one-up population of the G-form.

191

192   **One-up state enhances exposure of RBD to ACE2 binding and epitope sites**

193   The Spike trimer is a glycoprotein, meaning it is covered in short flexible sugars called

194   glycans. We modeled the impact of those highly-flexible glycan ensembles on the exposure of the

195   protein surface, in particular focusing on the glycan shielding of Receptor Binding Domain epitope

196   regions. To do this we use a quantitative measure called the glycan encounter factor [16, 17]. Given

197   that glycans are highly dynamic, they transition between covering and revealing proximal protein

198   surfaces. A low glycan encounter factor value indicates a high exposure of the protein surface in a

199   given region. We perform a glycan encounter factor analysis here to assess changes in exposure

200   between all-down and one-up states and between the D- and G-forms. Independent of D- or G-

201   forms, the ACE2 binding site and epitope regions are significantly more exposed in the one-up state

202   than in the all-down state (**Fig. 7A-B)**. (For further details on local changes see **Fig. 8** and

203   **Supplementary Material**). Specifically, the exposure increases by ~40% at the Receptor Binding

204   Motif (residues 438-506). This is accompanied by larger exposure of both the ACE2 binding site

205   and Receptor Binding Domain targeting antibody epitopes. For example, there is a ~64% increase

206   in exposure of the C105 epitope region [8] when the Receptor Binding Domain adopts the open

state, as opposed to the closed orientation where it is buried by surrounding protein and glycans (**Fig. 7C-D**). Thus, in the one-up state, which is favored by the G614 form, key epitopes of the Receptor Binding Domain become more accessible compared to the D614 form.

In addition to affecting the binding site exposure, allostery-induced changes in rotations or tilting of Receptor Binding Domains between the D- and G-forms could lead to differential neutralization. Monoclonal antibody S309, isolated from a recovered SARS-CoV-infected subject, cross-neutralizes both SARS-CoV and SARS-CoV-2 [18]. Notably, S309 displays a striking increase in sensitivity for the G-form of SARS-CoV2 spike (IC50=0.073 µg/ml) as compared to the D-form (IC50=11.83 µg/ml) [19]. Structural and binding studies showed that its epitope lies outside the receptor binding motif of the Receptor Binding Domain, and that S309 binds to both the one-open and all-down conformations. Biolayer interferometry confirmed the absence of competition between S309 and ACE2 binding [18], suggesting that the molecular mechanism of action was not direct interference with ACE2 binding. Our glycan encounter factor analysis shows that there is no difference in epitope exposure between the D- and G-forms. However, there is a difference in epitope exposure between the one-up and all-down states: we identified two glycans, N343 and N331, that distinguish between S309 epitope exposure (**Fig. S7 A**).

We further compared the orthogonal rotational space sampled by the Receptor Binding Domain between the D- and G-forms (**Fig. S7 B-F**). To describe Receptor Binding Domain motions, we defined two axes of rotation that are roughly orthogonal to each other, "roll" and "pitch" (see Methods and **Fig. S7B**). Even though this space remains similar in the all-down state of the D- and G-forms (**Fig. S7 C**), it exhibited differences in the one-up state between the forms (**Fig. S7 D**). Using the known structure of S309 bound Receptor Binding Domain as a reference, mapping various Receptor Binding Domain configurations, we find that when the Receptor Binding Domain has a negative rotation (clockwise) more than 10 degrees towards the N Terminal Domain of the neighboring protomer on the right (-10 degrees and beyond in **Fig. 7 E-F and Fig. S7 C-D**),

232 it could hinder the binding of S309 due to steric clashes from the neighboring N Terminal Domain.

233 Considering that the MD simulation ensemble sampled higher populations in this sterically

234 hindered region for the D-form (~38%) compared to G-form (15%) (**Fig. S7 E-F**), it suggests that

235 S309 experiences less hinderance when binding to G-form compared to D-form in the one-up

236 conformation.

237   Finally, the paratope of S309 is aligned almost parallel to the Spike longitudinal axis when

238 the Receptor Binding Domain of one protomer is in the up conformation, making it more feasible

239 for the two arms of the antibody to cross-link with two different Spikes on the same viral surface

240 in tandem (**Supp. info, Fig. S8**). Thus, the increase in the population of one-up states that we

241 propose in the G-form would additionally favor mechanisms of neutralization such as cross-linking.

242

243 **Discussion**

244 The D614G substitution occurs on the inter-protomer interface of the Spike protein mediating

245 critical contacts. We have performed extensive all-atom MD simulations of the trimeric Spike

246 protein and assessed the effects of this substitution on both all-down and one-up conformations.

247 The fluctuations about the native contacts studied here provide a crucial foundation for how

248 molecular-level interactions could facilitate the adoption of one state versus another [20]. Although

249 the longer timescales of the up and down conformational transitions do not allow us to directly

250 probe the all-down to one-up dynamics, our findings are nonetheless of great significance.

251   Our simulations provide mechanistic details that indicate and can explain an increased

252 occupancy of the one-up state in the G614 form, which in turn accounts for both its enhanced

253 infectivity [5, 6] and neutralization sensitivity [19] and provides predictions for experiment, some

254 of which have already been verified [21, 22]. We found that the different interactions between

255 protomers in the D614 form are different in the one-up and all-down states. By contrast, in the G614

256 form there is a symmetrization in the number of inter-protomer contacts between S1 and S2

subunits, which is associated with allosteric fluctuations in RBD, distal to the 614 site. We have identified the regions of greatest difference between the two forms and the regions of high-magnitude correlations as areas of focus for experimental efforts.

Thermodynamically, the effect of symmetrization can be captured using a periodic three-spin Ising model. Using the observed 50:50 ratio of all-down to one-up conformations in the D614 form [4], we predict that symmetrization of the G-form energetics in comparison to the D-form leads to a 75:25 ratio of one-up to all-down. We do note a small tendency towards a concomitant deviation from symmetry in the all-down state of the G614 form, which would further increase this ratio. In parallel, Weissman and colleagues obtained structural evidence using negative stain electron microscopy that showed that the G614 form Spike populates the one-up state 82% of the time (versus 18% of the time all down), while the D614 form adopts this open conformation with only 42% frequency (versus 52% of the time all down), largely confirming our predictions [19]. Furthermore, they found that the G614 form is more sensitive to neutralization by a range of neutralizing agents: (i) sera raised from D614-form Spike and Spike Receptor Binding Domain vaccinations in mice, primates and humans, (ii) convalescent COVID-19 sera, and (iii) a subset of Receptor Binding Domain monoclonal antibodies.

A recent cryo-electron microscopy study points to a "looser" Spike protein with a higher probability of taking on an open structure with a more flexible Receptor Binding Domain [21]. According to that study, the G-form demonstrates increased infectivity in pseudovirus assays and across multiple species ACE2 orthologs, does not have any alterations in the synthesis or processing of the Spike trimer, and, despite the increased infectivity, does not have stronger binding to the ACE2 receptor. Therefore, this paper supports that the dominant factors in the increased infectivity are likely to be, as we argued, conformational changes and/or an increase in the population of open states in the G-form. They observe conformational changes in the S1 domain; however, they are not able to resolve the RBD. We also observe conformational changes in the S1 domain between D

282   and G forms, which we are able to localize to a difference in the rotational states occupied by the

283   RBD. As we show such a rotational effect of RBD in G-form can explain the striking increase in

284   sensitivity of antibody S309 for the G-form of SARS-CoV2 Spike as compared to the D-form.

285   The increased occupancy of the one-up state in the G614 form predicted by our simulations

286   is further supported by a recent study from Ralph Baric's group [22]. That study demonstrates

287   that D614G substitution exhibits significantly faster transmission phenotypes in vivo and shows

288   increased replication in the in the nasal epithelium and large airway epithelium. They conclude

289   that the infectivity, replication fitness and early transmission are enhanced in the G-form and their

290   data are consistent with a mechanistic model that accounts for both its enhanced infectivity and

291   neutralization sensitivity as proposed here; the increased occupancy of the one-up state in the

292   G614 form. The shift towards one-up Spike trimers would increase the likelihood of binding

293   events of RBD and ACE2, thus explaining the experimentally observed increase in infectivity (5)

294   and sensitivity to neutralization activity in vaccine sera [19, 22].

295   Our dynamic model also indicates that there is a transition in the glycan shield when going

296   from all-down to one-up; the glycan coverage disappears at the apex of the trimer when in the one-

297   up conformation. Thus, regardless of the D- or G-form, the one-up conformation exposes the RBD

298   for binding to the ACE2 receptor while simultaneously exposing more of the RBD protein surface

299   for antibody binding to RBD epitopes. There are no significant global changes coming from the

300   D614G substitution on the ACE2 binding site exposure in RBD once the Spike trimer is in the one-

301   up conformation. Thus, given the natural preference for a more open Spike conformation in the G-

302   form, it is possible that this form may have advantages as a vaccine antigen.

303   In summary, our findings suggest that the overall protein exposure remains globally similar

304   between D- and G-forms, but predicts a dramatic increase in the one-up state with the G-form,

305   meaning that both ACE2 binding and RBD-targeting antibody binding are likely to increase in the

306   one-up state. Therefore, a change towards a higher one-up state population is likely the dominant

307 effect of the D614G substitution. The mechanistic studies presented here, the structural data [19,

308 21], the experimentally determined increase in infectivity [5, 22], and the enhanced neutralizing

309 antibody sensitivity [19, 22] all come together to form a consistent story .

310

311 **Materials and Methods**

312 *Structure Preparation.*

313 Atomic structures derived from cryo-EM (PDB IDs: 6VXX and 6VYB; Walls et al. [9])

314 were used to prepare the all-down and one-up configurations for simulation. As described in Ref

315 [9], residues 986—987 were kept as Proline-Proline (PP), and Fusion Peptide residues 682—685

316 as SGAG instead of the viral form residues KV and RRAR, to maintain the stable soluble form of

317 the spike extracellular domain protein. In the original 6VXX and 6VYB models, several flexible

318 regions were unresolved, which vary in lengths from 2-27 residues. We used a data-driven

319 structure-based modeling approach to build a complete model that accurately captures secondary

320 structures of these missing regions, without introducing artificial 'knotting' of loops. For this, we

321 employed homology modeling using numerous SARS-COV1 structures as templates (see

322 **Supplementary Information**). Missing residues in the RBD were built using an ACE2-bound

323 SARS-CoV2 structure (PDB: 6M0J) [23]. Structure-based sequence alignment was performed

324 using the 3D-Coffee program [24] and homology modeling was performed using the MODELLER

325 9.20 suite [25]. We generated 10 models for each configuration and selected the top model based

326 on DOPE and Procheck scores [26, 27]. After testing that CryoEM fitting did not significantly

327 improve the model employing Chimera 1.13.0 [28] rigid fitting, and the MDFF [29] flexible fitting,

328 we arbitrarily chose as starting structures the all down model after 2 ns in vacuum with gscale = 0.3

329 and a further 2 ns in vacuum with gscale = 0.2 and the one up model after 2 ns in vacuum with

330 gscale = 0.3. To avoid artificial charges at the protein ends, we introduced N-terminal acetylated

331 and C-terminal N-methylamide capping groups. All histidines were modeled as the neutral tautomer

332   where the epsilon nitrogen is protonated. 15 disulfide bonds found in 6VXX and 6VYB were

333   modeled in the forcefield for simulation.

MD Simulation Details

335   All-atom explicit-solvent simulations were performed with the Gromacs v5.1.2 and 2018.3 software

336   packages [30, 31], using the CHARMM36m [32] forcefield and TIP3P water model [33]. Each

337   configuration was solvated and centered in a cubic box. The side length of the box was defined

338   such that there is at least 15Å padding around the molecule. Each system was neutralized with an

339   excess of 150mM KCL. Energy minimization was conducted using the steepest descent algorithm.

340   Equilibration simulations were performed under constant number-volume-temperature (NVT, 2ns)

341   and constant number-pressure-temperature (NPT, 10ns) ensembles. During the equilibration stages,

342   harmonic position restraints were imposed on all heavy atoms of the molecule. Temperature

343   coupling was achieved using Langevin dynamics at 310K with a relaxation time of 1ps [34]. The

344   Berendsen barostat with isotropic coupling was employed to maintain a constant pressure of 1 bar,

345   with a relaxation time of 4ps and compressibility of 4.5x10E-5 / bar [35]. Covalent bonds were

346   constrained by implementing the LINCS algorithm [36]. Van der Waals interactions were evaluated

347   using a cutoff where forces smoothly switch to zero between 1.0 and 1.2nm. Coulomb interactions

348   were calculated using the particle mesh Ewald (PME) method, with a cutoff of 1.2nm, a Fourier

349   spacing of 0.12nm, an interpolation order of 4, and a tolerance of 1x10E-5 [37]. Unrestrained

350   production simulations were performed in the NPT ensemble, with an integration time step of 2fs.

351   For each configuration, we simulated five replicas. Each replica was run for 1.1 µs. The first 100

352   ns of each run was considered as further equilibration and was not included for analysis. In total,

353   we generated 20 1-µs production simulations.

Contact analysis

355   We used the g_contacts plugin for Gromacs [10] to identify contact frequencies of all

356   residue-residue contacts formed, both intra- and inter-protomer. A contact was defined as any heavy

357 atom of one residue coming within 6Å of the heavy atom of another residue. For the contact

358 analysis, we defined *persistent* contacts as any of the identified contacts that appeared with a

359 frequency of greater than or equal to 0.5 – that is, that appeared in at least 50% of analyzed

360 simulation frames. Average contacts with region S2 excluded the CD1 domain because of

361 observations that it "frayed" in simulation and was extremely flexible, which we presume to be an

362 artefact due to the lack of embedding of the Spike in the viral membrane.

363 *Inter-protomer S1-S2 interaction energy*

364 The inter-protomer S1-S2 interaction energy was approximated using the Molecular

365 Mechanics – Poisson Boltzmann Surface Area (MM-PBSA) method [38-40]. This end-point

366 approach estimates binding energy from the combination of molecular mechanical electrostatic and

367 van der Waals energies, and the polar and non-polar solvation energy terms. While this method has

368 been known to over-estimate energies, it performs reasonably well for relative comparison of

369 binding strength [41, 42]. It must also be noted that conformational entropy was not included in the

370 energy calculation. MM-PBSA calculation was implemented through the 'g_mmpbsa' plugin [43]

371 for GROMACS. Because the PB calculations are computationally very costly, a selection of 400

372 snapshots equally distributed over the complete equilibrated ensemble was used for these

373 calculations in order to obtain robust sampling and standard errors.

374 *Correlation analysis*

375 The cross-correlation matrix is defined as

376
$$C_{ij} = \frac{<\Delta \boldsymbol{r}_i(t) \cdot \Delta \boldsymbol{r}_j(t)>}{\sqrt{<\Delta \boldsymbol{r}_i(t) \cdot \Delta \boldsymbol{r}_i(t)><\Delta \boldsymbol{r}_j(t) \cdot \Delta \boldsymbol{r}_j(t)>}} \, , \qquad (1)$$

377

378 which is the normalized covariance where $\Delta \boldsymbol{r}_i(t)$ is the fluctuations of atom i with respect to its

379 average coordinates. We used the -cov, -norm, and -dot flags of the program carma to perform the

380 analysis [44].

Hydrogen bond analysis

We performed the hydrogen bond analysis with VMD 1.9.3 Hbonds plugin [45], where the definition of a hydrogen bond was the requirement of donor-acceptor atoms being polar, distance cutoff of 4Å and angle cutoff of 30°. This is a significantly more stringent criterion than that employed for the contact identification. A cutoff of 10% occupancy was used to identify extant hydrogen bonds.

Glycan modeling and exposure calculation

There are 22 potential N-glycosylation sites (PNGS) in the Spike, each of which has a unique glycoform distribution. We chose glycan types with the highest occurrence at each site, as determined by mass spectrometry studies [46] for 19 of these glycans present within our model range (see **Fig. 8B-C** and **Table S2**). There are two putative O-glycan sites on the Spike; however, they have been reported to have less than 2% occupancy [47] and were therefore not included in our models.

We performed a Jarvis Patrick based cluster analysis where a structure is added to a cluster if it has at least 6 neighbors in common with another neighbor. Clustering was performed on all five replicas, for each of the four different systems, using the "gmx cluster" command with a cutoff of 0.2 nm to identify a subset of distinct conformations based on Cα RMSDs. We identified 63 clusters for D614 all-down, 78 clusters for D614 one-up, 54 clusters for D614G all-down, and 103 clusters for D614G one-up. We took the closest to the mean structure from each of the top twenty largest clusters of each system, and used these snapshots as the basis for building glycan models. Glycan ensemble modeling was performed using a previously established pipeline [16] with the ALLOSMOD [48] suite of MODELLER. An initial glycan conformation was added at each PNGS with randomized atomistic deviations from CHARMM36 ideal carbohydrate geometries, and random orientations. This was followed by simulated annealing relaxation of the entire glycoprotein. 50 glycosylated conformations were built on each of the 20 selected models, resulting

406 in an ensemble of 1000 distinct glycoprotein conformations for each of the four systems. Glycoform

407 for each PNGS was chosen as the one with the highest probability of occurrence from site-specific

408 mass spectrometry results [46].

409 The Glycan Encounter Factor (GEF) was calculated for each residue exposed on the surface

410 of the protein as established in our previous study [16]. It is defined as the number of glycan heavy

411 atoms encountered by an approaching probe of 6Å diameter mimicking a typical hairpin loop of

412 antibodies interacting with epitopes. Geometric mean of this value measured in the three cardinal

413 directions (perpendicular to the surface, and along the plane) was taken to cover different

414 orientations.

415 *Reorientational dynamics of Receptor Binding Domain*

416 Treating the RBD domain as a rigid ellipsoid, rotations along the Euler angles 'roll' and 'pitch'

417 were measured through the equilibrated simulation trajectories. The rotation axes are indicated in

418 **Fig. S7B**. 'Roll' was defined as rotation around the axis that is spanned by residues 392 and 401.

419 'Pitch' was defined as rotation around the axis that is spanned by residues 377 and 357 of the initial

420 model conformations. The 'roll' – 'pitch' potential of mean force, *PMF(X)*, was calculated using:

421 *PMF(X) = –RT*log[p(X)]*, where *X* denotes values along the roll and pitch directions, and *p(X)* is

422 the probability. Rotation of RBD-down ensemble included the L-RBD and R-RBD (down)

423 protomers.

424 S309 Fab binding to up-RBD was modeled by rigid-body alignment to closed-RBD and Fab

425 interactions from PDB structure 6WPT [18], using the backbone of residues 331 to 527 for least

426 squares fitting.

427 *Molecular modeling of antibody mediated cross-linking of Spikes*

428 We modeled an intact IgG structure based on PDB entry 1IGT [49], in combination with the S309

429 Fab domain to visualize the relative orientations between the Spike ectodomain and IgG (**Fig. S8**).

430 The S309 Fab and IgG Fab domains were aligned to model the intact IgG. These two had an RMSD

431 of ~1 angstrom for the aligned and matched sequence residues. The Fab domain residues were

432 replaced by S309 to obtain an accurate binding interface. In the RBD-down conformation, the

433 orientation of the two IgG arms is such that Spike trimers from the same virion can only bind if

434 there is significant reorientation and structural reorganization of the antibody arms (**Fig. S8 B**).

435 Since the inter-domain linkers in antibodies do impart flexible reorganizations [50], this pathway

436 cannot be ruled out. However, we have tested all three available intact IgG structures in PDB (1IGT,

437 1IGY, 1HZH), and they all indicate that significant reorientation would be necessary for inter-Spike

438 binding in the closed-RBD conformation.

439 *Limitations of the study*

440 Although the results from these extensive MD simulations represent an extensive investment of

441 computational resources, large-scale conformational shifts in the dynamics, such as the actual

442 transition between the all-down and one-up states, are beyond the microsecond timescales

443 considered here. Specifically, by using an all-atom explicit solvent model (as described in the

444 manuscript) it is currently not possible to simulate large-scale up-down transitions due to the

445 massive computational cost. According to statistical mechanics, the fluctuations about equilibrium

446 studied here provide the motions of the system under linear response, and it has been well

447 established that protein transitions between folded and unfolded states can be probed purely based

448 on native contacts [20].

449   The work performed here lays the necessary groundwork for future simulations, which of

450 computational necessity must be coarse-grained, to study these transitions directly as well as

451 indirectly as done here. To study large movements, one could employ more simplified energetic

452 descriptions (e.g. Go-models) that would effectively elucidate the role of structural features during

453 this process. However, these are additional questions that may be explored in future efforts and that

454 cannot be performed without a detailed understanding of the system modeled accurately at an all-

455 atom scale.

456   Also, the current structure does not include the heptad repeat 2, trans-membrane or

457 cytoplasmic regions that could differentially alter the presentation of Spike on the membrane in the

458 G-form. We cannot comment directly on the S1/S2 cleavage site conformation or the effect of two

459 prolines used for stabilization, as we used the sequence from Walls et. al [9] that was mutated to

460 stabilize a soluble protein. In addition, we cannot comment on effects such as hyperfusogenicity,

461 which may impact the virus's ability to fuse with the host cell. Such large-scale effects would

462 require a coarse-grained or steered MD approach to be able to probe the interactions of the full

463 trimer with the host cell and its receptor.

464

465 **H2: Supplementary Materials**

466   Materials and Methods

467   Analysis

468   Table S1. Template selection for SARS-COV2 Spike structural modeling

469   Table S2. SARS-CoV2 S glycoprotein N-glycan types.

470   Fig. S1. Root mean square deviations (RMSDs) and Root mean square fluctuations

471   (RMSFs) of the RBD.

472   Fig. S2. Global and local correlation matrices, differences, and fractional errors.

473   Fig. S3. Global correlation matrices for G-form.

474   Fig. S4. Details of changing inter-protomer contacts.

475   Fig. S5. Increased flexibility with partial loss of intra-protomer hydrogen bond.

476   Fig. S6. Inter-protomer contacts of interest

477   Figure S7. Structural analysis of RBD suggests S309 binding can be sterically hindered in

478   D-form up-RBD.

479   Figure S8. Complete IgG model suggests possibility of inter-Spike cross-linking for RBD-

480   up conformation.

481

References

483
484 1.   Wrapp, D., et al., *Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion
485   Conformation.* Science, 2020. **367**(6483): p. 1260-1263.
486 2.   Song, W., et al., *Cryo-EM structure of the SARS coronavirus spike glycoprotein in
487   complex with its host cell receptor ACE2.* PLoS Pathogens, 2018. **14**(8).
488 3.   Kirchdoerfer, R.N., et al., *Stabilized coronavirus spikes are resistant to conformational
489   changes induced by receptor recognition or proteolysis.* Scientific Reports, 2018. **8**(1): p.
490   1-11.
491 4.   Henderson, R., et al., *Controlling the SARS-CoV-2 Spike Glycoprotein Conformation.*
492   bioRxiv, 2020: p. 2020.05.18.102087-2020.05.18.102087.
493 5.   Korber, B., et al., *Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases
494   infectivity of the COVID-19 virus.* Cell, 2020.
495 6.   Zhang, L., et al., *The D614G mutation in the SARS-CoV-2 spike protein reduces S1
496   shedding and increases infectivity.* bioRxiv, 2020: p. 2020.06.12.148726-
497   2020.06.12.148726.
498 7.   Volz, E., et al., *Evaluating the effects of SARS-CoV-2 Spike mutation D614G on
499   transmissibility and pathogenicity.* Cell, 2020.
500 8.   Barnes, C.O., et al., *Structures of human antibodies bound to SARS-CoV-2 spike reveal
501   common epitopes and recurrent features of antibodies.* Cell, 2020.
502 9.   Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike
503   Glycoprotein.* Cell, 2020. **181**(2): p. 281--292.e6.
504 10. Blau, C. and H. Grubmuller, *g _ contacts : Fast contact search in bio-molecular ensemble
505   data ✩.* Computer Physics Communications, 2013. **184**(12): p. 2856-2859.
506 11. Vanwart, A.T., et al., *Exploring residue component contributions to dynamical network
507   models of allostery.* Journal of Chemical Theory and Computation, 2012. **8**(8): p. 2949-
508   2961.
509 12. Tsai, C.J. and R. Nussinov, *A Unified View of "How Allostery Works".* PLoS
510   Computational Biology, 2014. **10**(2): p. e1003394-e1003394.
511 13. Best, R.B., G. Hummer, and W.A. Eaton, *Native contacts determine protein folding
512   mechanisms in atomistic simulations.* Proceedings of the National Academy of Sciences,
513   2013. **110**(44): p. 17874-17879.
514 14. Hills, R.D., *Balancing Bond, Nonbond, and Gō-Like Terms in Coarse Grain Simulations
515   of Conformational Dynamics*, in *Protein Dynamics*. 2014, Springer. p. 123-140.
516 15. Goldenfeld, N., *Lectures on Phase Transitions and the Renormalization Group.* 1992,
517   Lexington, KY: Westview Press.
518 16. Chakraborty, S., et al., *Quantification of the Resilience and Vulnerability of HIV-1 Native
519   Glycan Shield at Atomistic Detail.* bioRxiv preprint bioRxiv:846071, 2020: p. 1-40.
520 17. Berndsen, Z.T., et al., *Visualization of the HIV-1 Env Glycan Shield Across Scales.*
521   Proceedings of the National Academy of Sciences (accepted), 2020.
522 18. Pinto, D., et al., *Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV
523   antibody.* Nature, 2020: p. 1-6 %@ 1476-4687.

524 19. Weissman, D., et al., *D614G Spike Mutation Increases SARS CoV-2 Susceptibility to*
525 *Neutralization.* Cell Host & Microbe, In press, 2020.
526 20. Best, R.B., G. Hummer, and W.A. Eaton, *Native contacts determine protein folding*
527 *mechanisms in atomistic simulations.* Proceedings of the National Academy of Sciences,
528 2013. **110**(44): p. 17874-17879 %@ 0027-8424.
529 21. Yurkovetskiy, L., et al., *Structural and Functional Analysis of the D614G SARS-CoV-2*
530 *Spike Protein Variant.* Cell, 2020.
531 22. Hou, Y.J., et al., *SARS-CoV-2 D614G Variant Exhibits Enhanced Replication ex vivo and*
532 *Earlier Transmission in vivo.* bioRxiv, 2020.
533 23. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the*
534 *ACE2 receptor.* Nature, 2020. **581**: p. 215-220.
535 24. Poirot, O., et al., *Expresso: automatic incorporation of structural information in multiple*
536 *sequence alignments using 3D-Coffee.* Nucleic Acids Research, 2006. **34**: p. 604-608.
537 25. Eswar, N., et al., *Comparative Protein Structure Modeling Using Modeller.* Current
538 Protocols in Bioinformatics, 2006. **15**(1): p. 1-5.
539 26. Shen, M.-Y. and A. Sali, *Statistical potential for assessment and prediction of protein*
540 *structures.* Protein science : a publication of the Protein Society, 2006. **15**(11): p. 2507-
541 2524.
542 27. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of*
543 *protein structures.* Journal of Applied Crystallography, 1993. **26**(2): p. 283-291.
544 28. Petterson, E.F., et al., *UCSF Chimera--a Visualization System for Exploratory Research*
545 *and Analysis.* J. Comput. Chem., 2004. **25**(13): p. 1605-12.
546 29. Trabuco, L.G., et al., *Molecular dynamics flexible fitting: A practical guide to combine*
547 *cryo-electron microscopy and X-ray crystallography.* Methods, 2009. **49**(2): p. 174-180.
548 30. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational
549 chemistry, 2005. **26**(16): p. 1701-18.
550 31. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through*
551 *multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
552 32. Huang, J., et al., *CHARMM36m: an improved force field for folded and intrinsically*
553 *disordered proteins.* Nature Methods, 2017. **14**(1): p. 71-73.
554 33. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid*
555 *water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
556 34. Goga, N., et al., *Efficient algorithms for langevin and DPD dynamics.* Journal of Chemical
557 Theory and Computation, 2012. **8**(10): p. 3637-3649.
558 35. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath.* The
559 Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
560 36. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations.* Journal of
561 Computational Chemistry, 1997. **18**(12): p. 1463-1472.
562 37. Essmann, U., et al., *A smooth particle mesh Ewald method.* The Journal of Chemical
563 Physics, 1995. **103**(19): p. 8577-8577.
564 38. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules:*
565 *combining molecular mechanics and continuum models.* Accounts of chemical research,
566 2000. **33**(12): p. 889-897.
567 39. Fogolari, F., A. Brigo, and H. Molinari, *Protocol for MM/PBSA molecular dynamics*
568 *simulations of proteins.* Biophysical journal, 2003. **85**(1): p. 159-166.
569 40. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the*
570 *ribosome.* Proceedings of the National Academy of Sciences, 2001. **98**(18): p. 10037-
571 10041.

572 41. Siebenmorgen, T. and M. Zacharias, *Computational prediction of protein–protein binding*
573 *affinities.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2020.
574 **10**(3): p. e1448.
575 42. Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1.*
576 *The accuracy of binding free energy calculations based on molecular dynamics*
577 *simulations.* Journal of chemical information and modeling, 2011. **51**(1): p. 69-82.
578 43. Kumari, R., et al., *g_mmpbsa□ A GROMACS tool for high-throughput MM-PBSA*
579 *calculations.* Journal of chemical information and modeling, 2014. **54**(7): p. 1951-1962.
580 44. Koukos, P.I. and N.M. Glykos, *Grcarma: A Fully Automated Task-Oriented Interface for*
581 *the Analysis of Molecular Dynamics Trajectories.* J. Comput. Chem., 2013. **34**(25): p.
582 2223-2312.
583 45. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of
584 Molecular Graphics, 1996. **14**(1): p. 33-38.
585 46. Watanabe, Y., et al., *Site-specific glycan analysis of the SARS-CoV-2 spike.* Science, 2020:
586 p. eabb9983-eabb9983.
587 47. Shajahan, A., et al., *Deducing the N- and O-glycosylation profile of the spike protein of*
588 *novel coronavirus SARS-CoV-2 [published online ahead of print, 2020 May 4].*
589 Glycobiology, 2020: p. 2020.04.01.020966-2020.04.01.020966.
590 48. Guttman, M., et al., *All-Atom Ensemble Modeling to Analyze Small-Angle X-Ray*
591 *Scattering of Glycosylated Proteins.* Structure/Folding and Design, 2013. **21**(3): p. 321-
592 331.
593 49. Harris, L.J., et al., *Refined structure of an intact IgG2a monoclonal antibody.*
594 Biochemistry, 1997. **36**(7): p. 1581-1597 %@ 0006-2960.
595 50. Stingaciu, L.R., et al., *Fast antibody fragment motion: flexible linkers act as entropic*
596 *spring.* Scientific reports, 2016. **6**: p. 22148 %@ 2045-2322.
597 51. Schrödinger, L.L.C., *The PyMOL Molecular Graphics System, Version 1.8.* 2015.
598 52. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic
599 acids research, 2015. **43**(W1): p. W389-94.
600 53. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: A unified platform for automated*
601 *protein structure and function prediction.* Nature Protocols, 2010. **5**(4): p. 725-738.
602 54. Neuman, B.W., et al., *A structural analysis of M protein in coronavirus assembly and*
603 *morphology.* Journal of structural biology, 2011. **174**(1): p. 11-22 %@ 1047-8477.
604 55. Hunter, J.D., *Matplotlib: A 2D Graphics Environment.* Computing in Science &
605 Engineering, 2007. **9**(3): p. 90-95.
606

607

608

613 EM data on G614. We are grateful to Paul Weber for securing extensive computational resources

614 that made this study possible.

615

625

626 **Author contributions:** R.A.M., S.C., K.N., and S.G. designed the study. S.C. and K.N. prepared

627 the structures. R.A.M. ran the simulations. S.C. calculated the GEF and MMPBSA energies. S.C.,

628 K.N., R.A.M, B.K., D.C.M. and S.G. performed data analysis and interpretation. R.A.M

629 and S.C. prepared the figures.

630 R.A.M. wrote the initial manuscript draft. R.A.M., S.C., K.N., B.K., D.C. M., and S.G. rewrote

631 and edited the manuscript. B.K. and S.G. secured the funding.

632 **Competing interests:** The authors declare no competing interests.

633 **Data and materials availability:** Molecular simulation data and raw materials are available from

634 the corresponding author upon reasonable request. All data required to evaluate the

635 conclusions in the paper are present in the paper and/or the Supplementary Materials.

636

637 **Figures and Tables**

A) S1/S2 in all down

S1

S2

B) Protomers in 1up

L down

Up

R down

C) Domain Definitions

NTD

RBD

CT0

CT1

CT2

S2S2'

CD

CD1

CH

FP

FPR

HR1

Domain		Residue Range
	NTD	14-305
	CT0	306-330
	RBD	331-527
	CT1	528-590
	CT2	591-685
	S2S2'	686-815
	FP	816-855
	FPR	856-911
	HR1	912-984
	CH	985-1034
	CD	1035-1080
	CD1	1081-1147

638
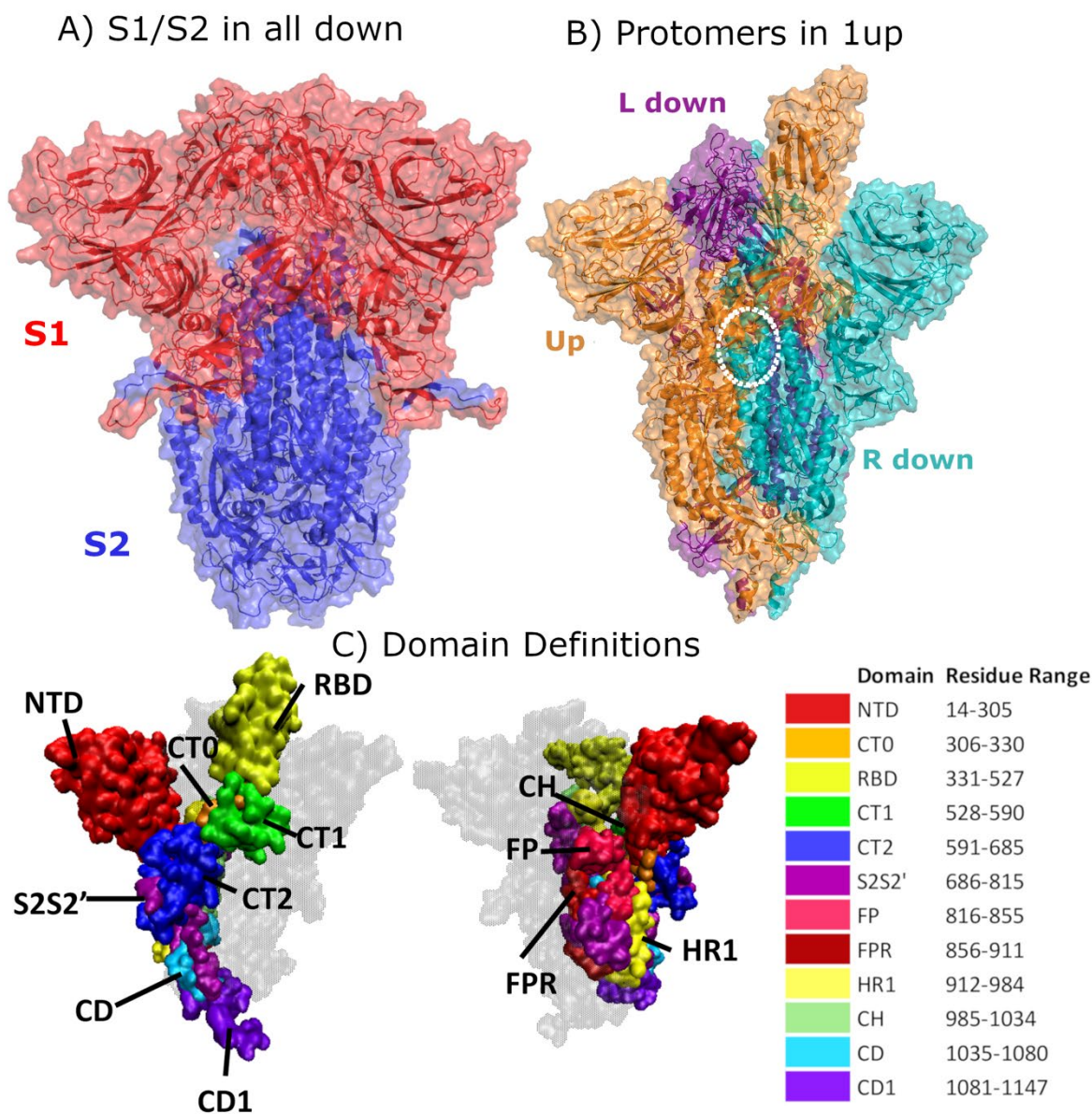
639   **Fig. 1. Structural Representation of the Spike protein. (A)** The Spike complex is shown in the "all-down" conformation. Its S1

640   and S2 subunits are depicted in red and blue. **(B)** The Spike complex is shown in the "one-up" conformation. It is composed of

641   three protomers, with the L-down protomer depicted in purple, the Up protomer in orange, and the R-down protomer in cyan. The

642   dashed white circle indicates a D614G site. Here, only one of the three D614G sites are shown. **(C)** Definition of twelve domains

643   and their residues used for analysis of simulations. It was necessary to define domains for every region for analysis, so some

644   regions may have arbitrarily assigned names or the defined regions may not follow canonical sequence ranges. We display the

645   domains highlighted in the Spike structure, shown from two different perspectives. Definitions of domain abbreviations: NTD, N-

646   Terminal Domain; CT0, C-Terminal domain 0; RBD, Receptor Binding Domain; CT1, C-Terminal domain 1; CT2, C-Terminal

647   domain 2; S2S2', S1/S2 Cleavage to S2'; FP, Fusion Peptide; FPR, Fusion Peptide Region; HR1, Heptad Repeat 1; CH, Center

648   Helix; CD, Connector Domain; and CD1, Connector Domain 1. Images in **(A)-(B)** were created with PyMol [51]; images in **(C)**

649   were prepared using VMD **[45].**

650

651
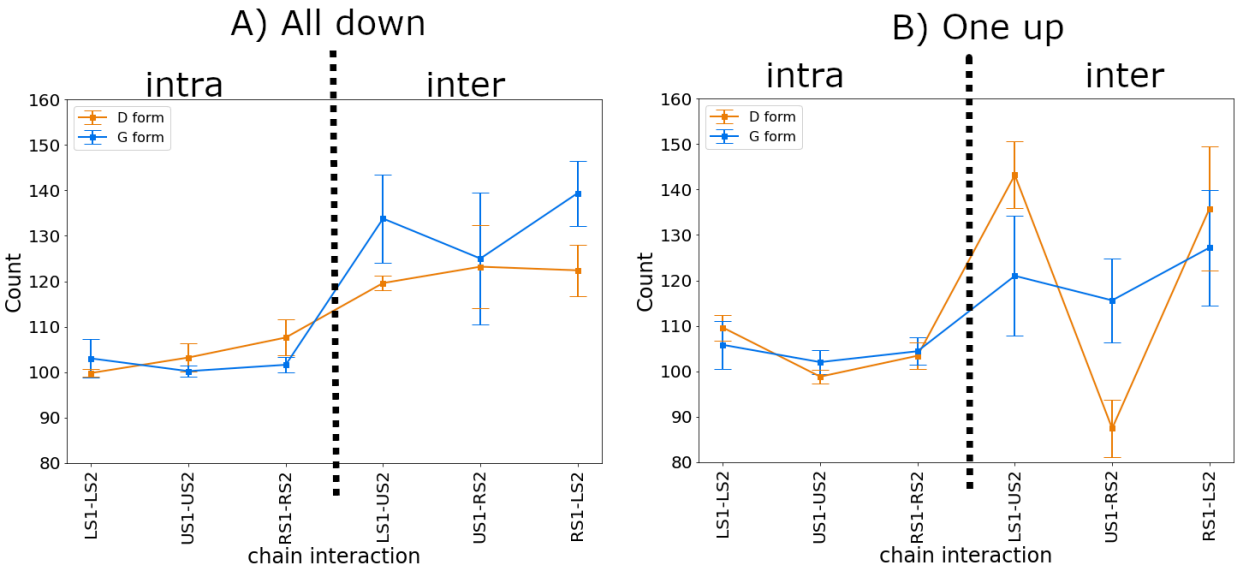
Fig. 2 D614G substitution alters inter-protomer interactions between S1 and S2 subunits. (A-B). Total number of contacts formed at S1-S2 interfaces **(A)** in the all-down system and **(B)** in the one-up system. Appropriate labelling and dashed lines indicate intra-protomer and inter-protomer S1-S2 interactions. For each set of simulations, error bars were calculated as standard error across five replicas. In the x-axis of all panels, 'L' denotes the L-down protomer, 'U' the Up protomer, and 'R' the R down protomer. For instance, LS1-U2 represents the interactions between the S1 region of the L-down protomer and the S2 region of the Up protomer.

658
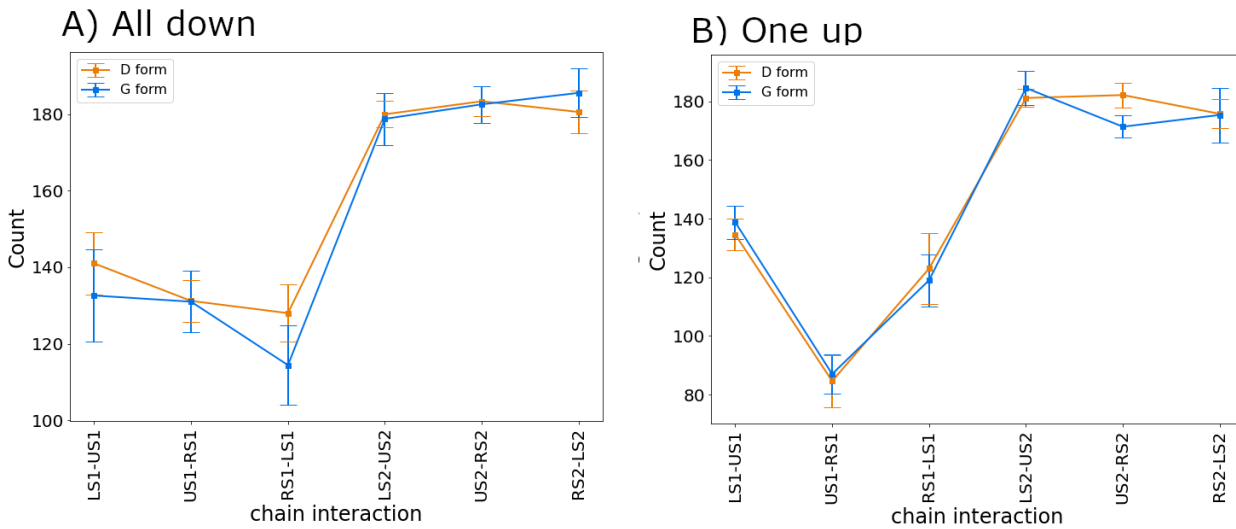
Fig. 3. S1-S1 and S2-S2 contacts. Average total number of contacts at the S1-S1 and S2-S2 interfaces in **(A)** the all-down system and **(B)** the one-up system. For each set of simulations, error bars were calculated as standard error across five replicas. In the x-axis of all panels, 'L' denotes the L-down protomer, 'U' the Up protomer, and 'R' the R-down protomer. For instance, LS1-US1 represents the interactions between the S1 region of the L-down protomer and the S1 region of the Up protomer.

663

G form all down	G form one up	D form all down	D form one up
LU: 130 ± 10	120 ± 10	119 ± 2	143 ± 7
UR: 130 ± 10	116 ± 9	123 ± 9	87 ± 6
RL: 139 ± 7	127 ± 10	139 ± 7	127 ± 10

Table 1. Average number of inter-protomer contacts between S1 and S2 regions.
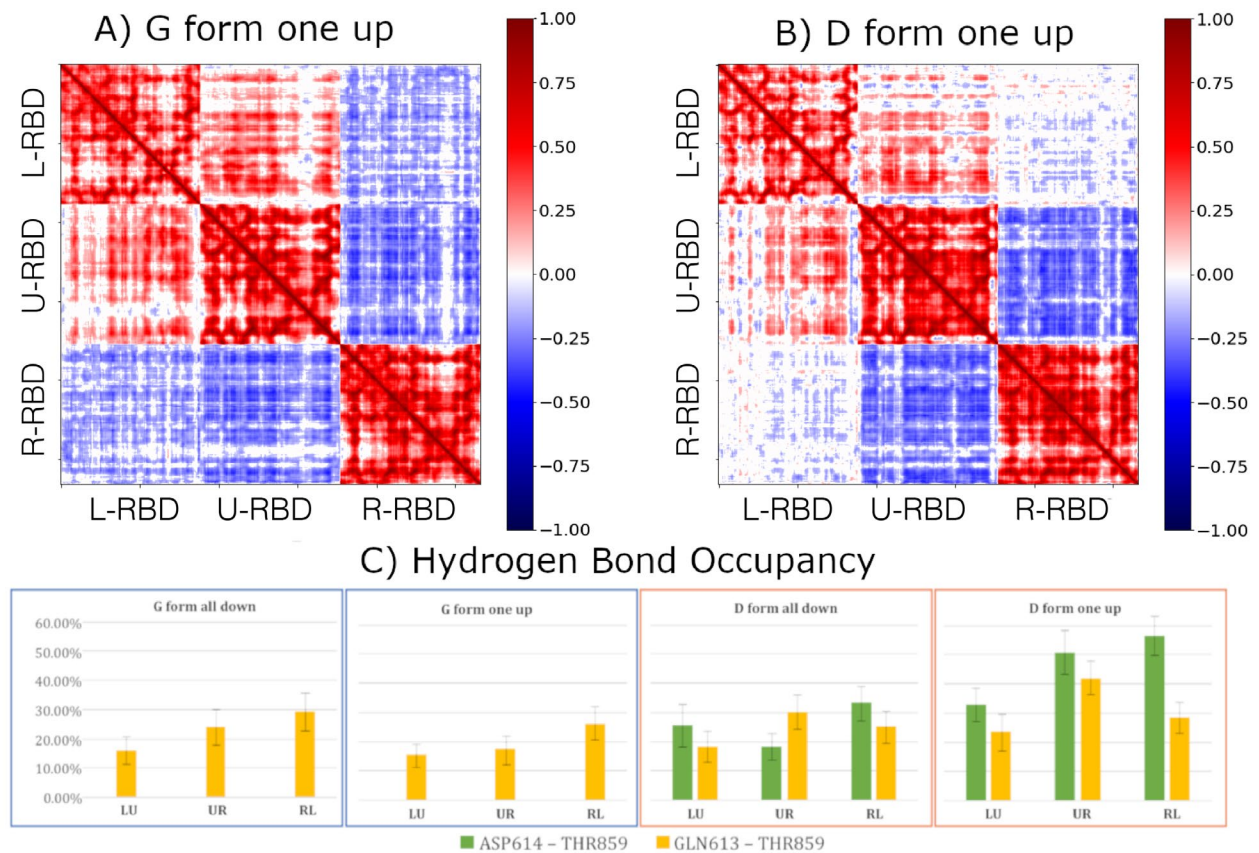
Fig. 4. Residue-residue cross correlation matrices and hydrogen bonding. (A) G614 form one-up **(B)** D614 form one-up, In the x and y axes, L-RBD, U-RBD, and R-RBD denote RBD regions of the L-down, Up, and R-down protomers. **(C)** Hydrogen bond occupancy for critical residue pairs located between protomers. Occupancies were calculated for the D/G 614-THR859 pair and the GLN613-THR859 pair. For each system, error bars were calculated as standard error over five replicas. In all subpanels, 'L' represents the L-down protomer, 'U' the Up protomer, and 'R' the R-down protomer. For example, "LU" denotes the bonding between L and U protomers.
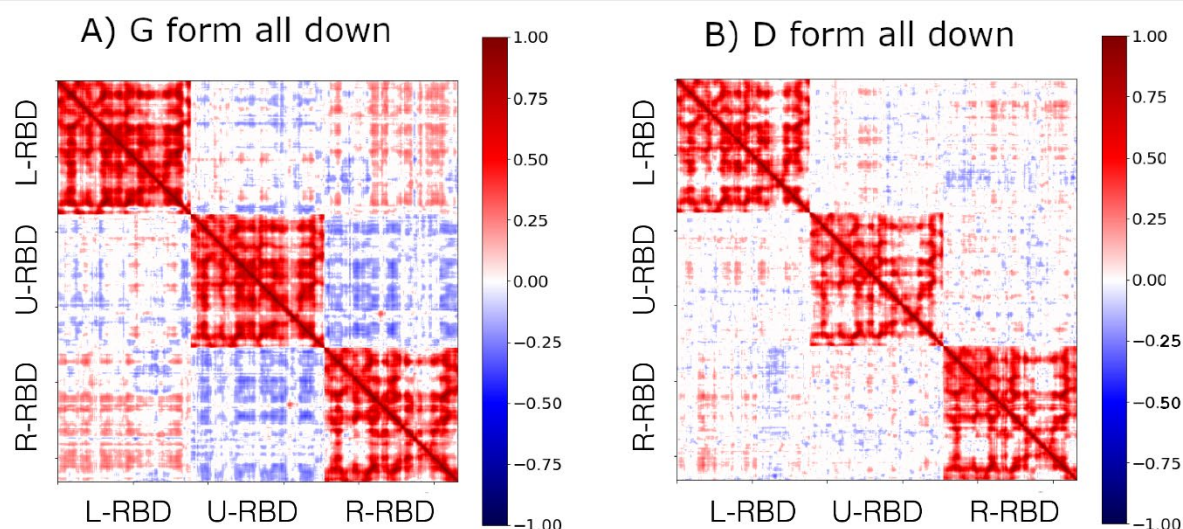
Fig. 5. Residue-residue cross correlation matrices. (A) G-form all-down **(B)** D-form all-down. In the x and y axes, L-RBD, U-RBD, and R-RBD denote RBD regions of the L-down, Up, and R-down protomers, respectively. Red denotes positive and blue denotes negative cross-correlations. Higher the intensity of the pixels, stronger the magnitude of correlations. The diagonal blocks (top-left to bottom-right) denote high intra-domain correlation, and therefore have intense red pixelation. Inter-domain off-diagonal blocks have relatively stronger (though slightly asymmetric) correlations in G-form down (left) as compared to D-form down (right).
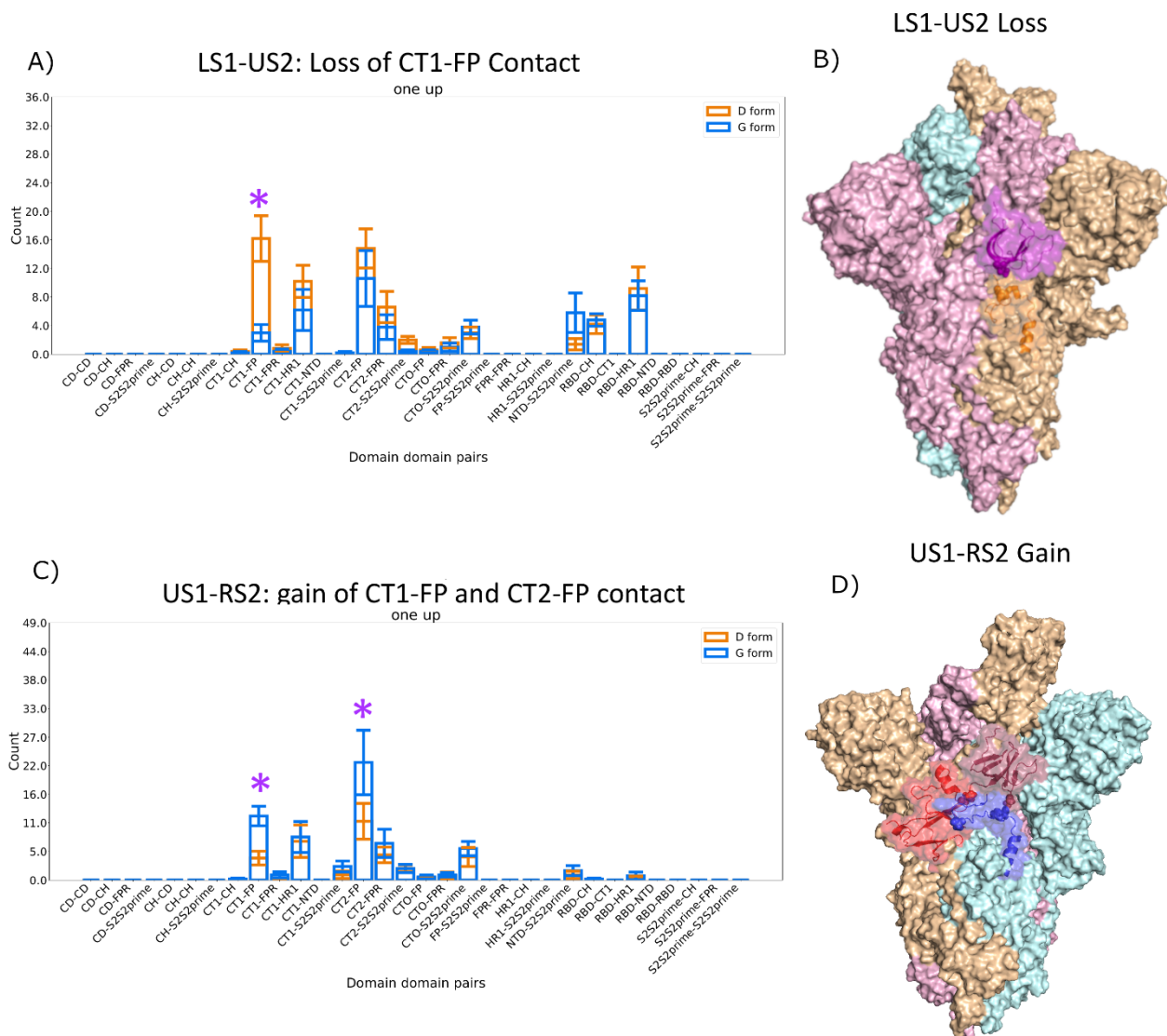
Fig. 6. Domain-domain contact gain and loss for LS1-US2 and US1-RS2 interfaces. **(A)/(C)** Average number of contacts gained between forms by each system in terms of domain-domain interactions. Each bar represents the average number of contacts gained by the system in question between the two domains listed on the *x* axis. To calculate this, we sum up the number of new contacts in each form in comparison to the other. Purple stars represent the three interfaces with the greatest differences. **(B)/(D)** Image of the two interfaces in question, with all three protomers shown as different colored surfaces, the domains involved in gain or loss shown as transparent surfaces with cartoon representations beneath, and the specific residues involved in the highest-frequency changing contact shown as spheres. Images rendered with PyMol [51].
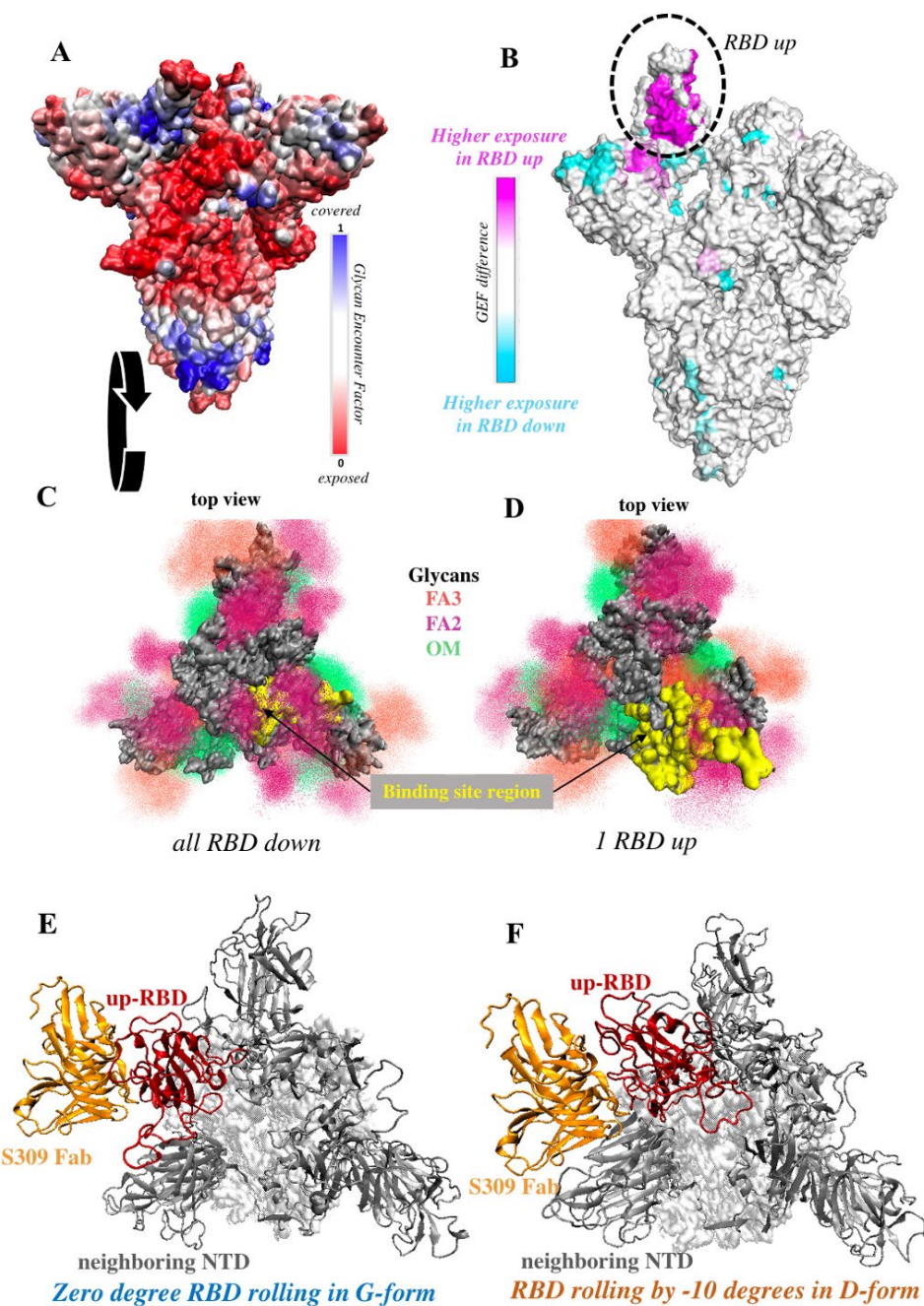
Fig. 7. RBD is significantly more exposed in the one-up conformation. (A) Surface representation of the all-down Spike complex. Colors ranging between red, white and blue represent values for the glycan encounter factor (GEF). GEF describes relative exposure and coverage of protein surfaces**. (B)** Surface representation of the one-up Spike. Colors ranging between cyan, white, and magenta indicate GEF differences between the closed and open structures. The RBD in the up orientation is indicated by the dotted circle, where the originally buried (magenta) region around 458-478 is now exposed. **(C-D)** Top view of the Spike protein apex in **(C)** all-down and **(D)** one-up conformations. Protein surface is shown in grey. Glycan ensembles covering the protein surface are represented by point densities. Fucosylated 2 and 3 antennae complex (FA2 and FA3) glycans and oligomannose (OM) glycans are depicted in dark red, orange, and green. Binding site region, which include receptor binding motif (residues 438-506) and C105 antibody binding residues (i.e. 403, 405, 406, 408, 409, 415-417, 420, 421, 449, 453, 455-460, 473-477, 486, 487, 489, 493-496, 498, 500-505), is

699   colored yellow. These binding site residues are significantly more exposed in the one-up conformation, compared to the all-down

700   case. **(E)** Representative configuration describing S309 Fab binding to an up-RBD in G-form that shows no rotation towards NTD

701   of the neighboring protomer on the right (~0 degree rolling angle) (top view of the spike is shown). This structure shows that the

702   Fab domain does not sterically clash with the Spike protein. **(F)** Representative configuration showing S309 Fab binding to an up-

703   RBD in D-form where the RBD is rotated towards the NTD of the neighboring protomer on the right (rolling angle is < -10 degrees;

704   same view as in E is shown). The Fab domain starts sterically clashing with the Spike protein NTD of the neighboring protomer.

705   See Supplemental Information, Methods, and **Fig. S7** for details on RBD rotations.
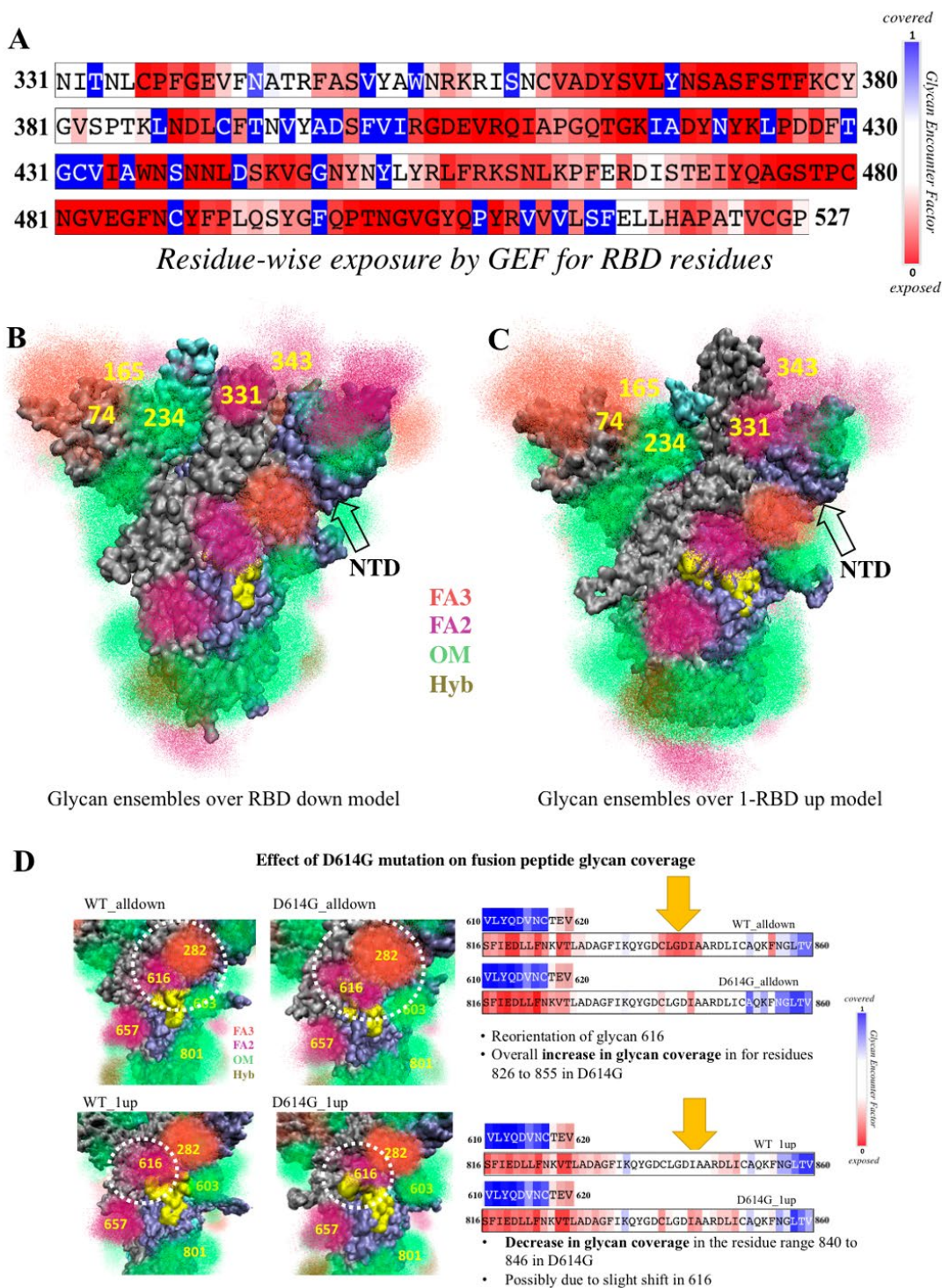
Fig. 8. Potential immunological effects of SARS-CoV2 Spike protein glycosylation. (**A**) Residue-wise GEF of the RBD

domain in D-form, down conformation. Buried residues are colored blue. Exposure increases as the color changes toward white

and red. (**B-C**) Glycan ensembles over Spike ectodomain, D-form. Spike protein surface given in grey. The surface glycans are

highly dynamic. Glycans overlaid from 250 different conformations are shown as points; this gives a "cloud" like rendition of the

area most heavily occupied where the points are dense, and space that glycans sample less frequency. Fucosylated 2 and 3

antennae complex (FA2 and FA3) glycans, oligomannose (OM) glycans and hybrid (Hyb) glycans are colored according to key.

(**B**) all-down and (**C**) one-up conformations shown. Glycans directly affecting RBD and NTD coverage change due to RBD

opening are marked. Removal of glycan 165 has experimentally been shown to increase the population of protomers in the single-protomer up state. Removal of glycan 234 has experimentally been shown to decrease the population of protomers in the single-protomer up state [4]. FP residue range 816 to 855 shown in yellow. **(D)** Effect of D614G substitution on FP coverage. N-glycans surrounding FP are numbered. Glycans at 282, 603 and 801 from the same protomer and glycan 616 from the neighboring protomer contribute to partial shielding of FP. On the left is the point cloud of glycans from 250 different conformations for the four different systems. White dashed circle shows the areas of change between D-form and G-form. Numbers label all glycans that touch the FP domain. On the right is the change in GEF superimposed on the relevant sequence.

References

1. Wrapp, D., et al., *Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation.* Science, 2020. **367**(6483): p. 1260-1263.
2. Song, W., et al., *Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2.* PLoS Pathogens, 2018. **14**(8).
3. Kirchdoerfer, R.N., et al., *Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis.* Scientific Reports, 2018. **8**(1): p. 1-11.
4. Henderson, R., et al., *Controlling the SARS-CoV-2 Spike Glycoprotein Conformation.* bioRxiv, 2020: p. 2020.05.18.102087-2020.05.18.102087.
5. Korber, B., et al., *Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus.* Cell, 2020.
6. Zhang, L., et al., *The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity.* bioRxiv, 2020: p. 2020.06.12.148726-2020.06.12.148726.
7. Volz, E., et al., *Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity.* Cell, 2020.
8. Barnes, C.O., et al., *Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies.* Cell, 2020.
9. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281--292.e6.
10. Blau, C. and H. Grubmuller, *g _ contacts : Fast contact search in bio-molecular ensemble data ✰.* Computer Physics Communications, 2013. **184**(12): p. 2856-2859.
11. Vanwart, A.T., et al., *Exploring residue component contributions to dynamical network models of allostery.* Journal of Chemical Theory and Computation, 2012. **8**(8): p. 2949-2961.
12. Tsai, C.J. and R. Nussinov, *A Unified View of "How Allostery Works".* PLoS Computational Biology, 2014. **10**(2): p. e1003394-e1003394.
13. Best, R.B., G. Hummer, and W.A. Eaton, *Native contacts determine protein folding mechanisms in atomistic simulations.* Proceedings of the National Academy of Sciences, 2013. **110**(44): p. 17874-17879.

757 14. Hills, R.D., *Balancing Bond, Nonbond, and Gō-Like Terms in Coarse Grain Simulations*
758 *of Conformational Dynamics*, in *Protein Dynamics*. 2014, Springer. p. 123-140.
759 15. Goldenfeld, N., *Lectures on Phase Transitions and the Renormalization Group*. 1992,
760 Lexington, KY: Westview Press.
761 16. Chakraborty, S., et al., *Quantification of the Resilience and Vulnerability of HIV-1 Native*
762 *Glycan Shield at Atomistic Detail.* bioRxiv preprint bioRxiv:846071, 2020: p. 1-40.
763 17. Berndsen, Z.T., et al., *Visualization of the HIV-1 Env Glycan Shield Across Scales.*
764 Proceedings of the National Academy of Sciences (accepted), 2020.
765 18. Pinto, D., et al., *Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV*
766 *antibody.* Nature, 2020: p. 1-6 %@ 1476-4687.
767 19. Weissman, D., et al., *D614G Spike Mutation Increases SARS CoV-2 Susceptibility to*
768 *Neutralization.* Submitted, 2020: p. 1-16.
769 20. Best, R.B., G. Hummer, and W.A. Eaton, *Native contacts determine protein folding*
770 *mechanisms in atomistic simulations.* Proceedings of the National Academy of Sciences,
771 2013. **110**(44): p. 17874-17879 %@ 0027-8424.
772 21. Yurkovetskiy, L., et al., *Structural and Functional Analysis of the D614G SARS-CoV-2*
773 *Spike Protein Variant.* Cell, 2020.
774 22. Hou, Y.J., et al., *SARS-CoV-2 D614G Variant Exhibits Enhanced Replication ex vivo and*
775 *Earlier Transmission in vivo.* bioRxiv, 2020.
776 23. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the*
777 *ACE2 receptor.* Nature, 2020. **581**: p. 215-220.
778 24. Poirot, O., et al., *Expresso: automatic incorporation of structural information in multiple*
779 *sequence alignments using 3D-Coffee.* Nucleic Acids Research, 2006. **34**: p. 604-608.
780 25. Eswar, N., et al., *Comparative Protein Structure Modeling Using Modeller.* Current
781 Protocols in Bioinformatics, 2006. **15**(1): p. 1-5.
782 26. Shen, M.-Y. and A. Sali, *Statistical potential for assessment and prediction of protein*
783 *structures.* Protein science : a publication of the Protein Society, 2006. **15**(11): p. 2507-
784 2524.
785 27. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of*
786 *protein structures.* Journal of Applied Crystallography, 1993. **26**(2): p. 283-291.
787 28. Petterson, E.F., et al., *UCSF Chimera--a Visualization System for Exploratory Research*
788 *and Analysis.* J. Comput. Chem., 2004. **25**(13): p. 1605-12.
789 29. Trabuco, L.G., et al., *Molecular dynamics flexible fitting: A practical guide to combine*
790 *cryo-electron microscopy and X-ray crystallography.* Methods, 2009. **49**(2): p. 174-180.
791 30. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational
792 chemistry, 2005. **26**(16): p. 1701-18.
793 31. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through*
794 *multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
795 32. Huang, J., et al., *CHARMM36m: an improved force field for folded and intrinsically*
796 *disordered proteins.* Nature Methods, 2017. **14**(1): p. 71-73.
797 33. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid*
798 *water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
799 34. Goga, N., et al., *Efficient algorithms for langevin and DPD dynamics.* Journal of Chemical
800 Theory and Computation, 2012. **8**(10): p. 3637-3649.
801 35. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath.* The
802 Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
803 36. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations.* Journal of
804 Computational Chemistry, 1997. **18**(12): p. 1463-1472.
805 37. Essmann, U., et al., *A smooth particle mesh Ewald method.* The Journal of Chemical
806 Physics, 1995. **103**(19): p. 8577-8577.

---

807 38. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.* Accounts of chemical research, 2000. **33**(12): p. 889-897.

810 39. Fogolari, F., A. Brigo, and H. Molinari, *Protocol for MM/PBSA molecular dynamics simulations of proteins.* Biophysical journal, 2003. **85**(1): p. 159-166.

812 40. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proceedings of the National Academy of Sciences, 2001. **98**(18): p. 10037-10041.

815 41. Siebenmorgen, T. and M. Zacharias, *Computational prediction of protein–protein binding affinities.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2020. **10**(3): p. e1448.

818 42. Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.* Journal of chemical information and modeling, 2011. **51**(1): p. 69-82.

821 43. Kumari, R., et al., *g_mmpbsa□ A GROMACS tool for high-throughput MM-PBSA calculations.* Journal of chemical information and modeling, 2014. **54**(7): p. 1951-1962.

823 44. Koukos, P.I. and N.M. Glykos, *Grcarma: A Fully Automated Task-Oriented Interface for the Analysis of Molecular Dynamics Trajectories.* J. Comput. Chem., 2013. **34**(25): p. 2223-2312.

826 45. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.

828 46. Watanabe, Y., et al., *Site-specific glycan analysis of the SARS-CoV-2 spike.* Science, 2020: p. eabb9983-eabb9983.

830 47. Shajahan, A., et al., *Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2 [published online ahead of print, 2020 May 4].* Glycobiology, 2020: p. 2020.04.01.020966-2020.04.01.020966.

833 48. Guttman, M., et al., *All-Atom Ensemble Modeling to Analyze Small-Angle X-Ray Scattering of Glycosylated Proteins.* Structure/Folding and Design, 2013. **21**(3): p. 321-331.

836 49. Harris, L.J., et al., *Refined structure of an intact IgG2a monoclonal antibody.* Biochemistry, 1997. **36**(7): p. 1581-1597 %@ 0006-2960.

838 50. Stingaciu, L.R., et al., *Fast antibody fragment motion: flexible linkers act as entropic spring.* Scientific reports, 2016. **6**: p. 22148 %@ 2045-2322.

840 51. Schrödinger, L.L.C., *The PyMOL Molecular Graphics System, Version 1.8.* 2015.

841 52. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic acids research, 2015. **43**(W1): p. W389-94.

843 53. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: A unified platform for automated protein structure and function prediction.* Nature Protocols, 2010. **5**(4): p. 725-738.

845 54. Neuman, B.W., et al., *A structural analysis of M protein in coronavirus assembly and morphology.* Journal of structural biology, 2011. **174**(1): p. 11-22 %@ 1047-8477.

847 55. Hunter, J.D., *Matplotlib: A 2D Graphics Environment.* Computing in Science & Engineering, 2007. **9**(3): p. 90-95.

849