Article

Exploring dynamics and network analysis of spike glycoprotein of SARS-COV-2

Mahdi Ghorbani,[1,3] Bernard R. Brooks,[3] and Jeffery B. Klauda[1,2,*]

[1]Department of Chemical and Biomolecular Engineering and [2]Biophysics Graduate Program, University of Maryland, College Park, Maryland; and [3]Laboratory of Computational Biology, National, Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland

ABSTRACT The ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 continues to rage with devastating consequences on human health and global economy. The spike glycoprotein on the surface of coronavirus mediates its entry into host cells and is the target of all current antibody design efforts to neutralize the virus. The glycan shield of the spike helps the virus to evade the human immune response by providing a thick sugar-coated barrier against any antibody. To study the dynamic motion of glycans in the spike protein, we performed microsecond-long molecular dynamics simulation in two different states that correspond to the receptor binding domain in open or closed conformations. Analysis of this microsecond-long simulation revealed a scissoring motion on the N-terminal domain of neighboring monomers in the spike trimer. The roles of multiple glycans in shielding of spike protein in different regions were uncovered by a network analysis, in which the high betweenness centrality of glycans at the apex revealed their importance and function in the glycan shield. Microdomains of glycans were identified featuring a high degree of intracommunication in these microdomains. An antibody overlap analysis revealed the glycan microdomains as well as individual glycans that inhibit access to the antibody epitopes on the spike protein. Overall, the results of this study provide detailed understanding of the spike glycan shield, which may be utilized for therapeutic efforts against this crisis.

SIGNIFICANCE Severe acute respiratory syndrome coronavirus 2 spike protein, the target of most neutralizing antibodies and vaccine design efforts, utilizes a glycan shield that covers most of the spike surface and helps the virus to avoid the human immune response. A detailed understanding of dynamics of glycans in providing an effective shield against antibodies is necessary to improve design of potent antibodies and vaccines against coronavirus. This study uses microsecond-long molecular dynamics simulation and network analysis in graph theory to unravel the detailed structural and dynamic features of individual glycans as well as microdomains of glycans in the spike protein.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) has rapidly spread worldwide since early 2020 and has been considered one of the most challenging global health crises within the century. SARS-COV-2 has caused more than 100 million cases and more than 2 million deaths worldwide as of February 2021 (1–3). Drug and vaccine development are underway; multiple vaccines have entered clinical trials, and some of them are in the last stages of development (4–6).

SARS-COV-2 is a lipid-enveloped single-stranded RNA virus belonging to the betacoronavirus family, which also includes Middle East respiratory syndrome (MERS), SARS, and bat-related coronaviruses (7–10). A major characteristic of all coronaviruses is the spike protein (S), which protrudes outward from the viral membrane and plays a key role in the entry of the pathogen into the host cells by binding to human angiotensin converting enzyme-2 (h-ACE2) (7,11–13). The structure of each monomer of the trimeric spike protein in SARS-COV-2 (Fig. 1) can be divided into two subunits (S1 and S2), which can be cleaved at residues 685–686 (furin cleavage site) by TMPRSS protease after binding to host cell receptor h-ACE2 (13). The S1 subunit in the S trimer includes an N-terminal domain (NTD) and a receptor binding domain (RBD) that is responsible for binding to h-ACE2 (14). The S2 subunit contains fusion peptides, heptad repeats (HR1 and HR2), and a transmembrane (TM) and a
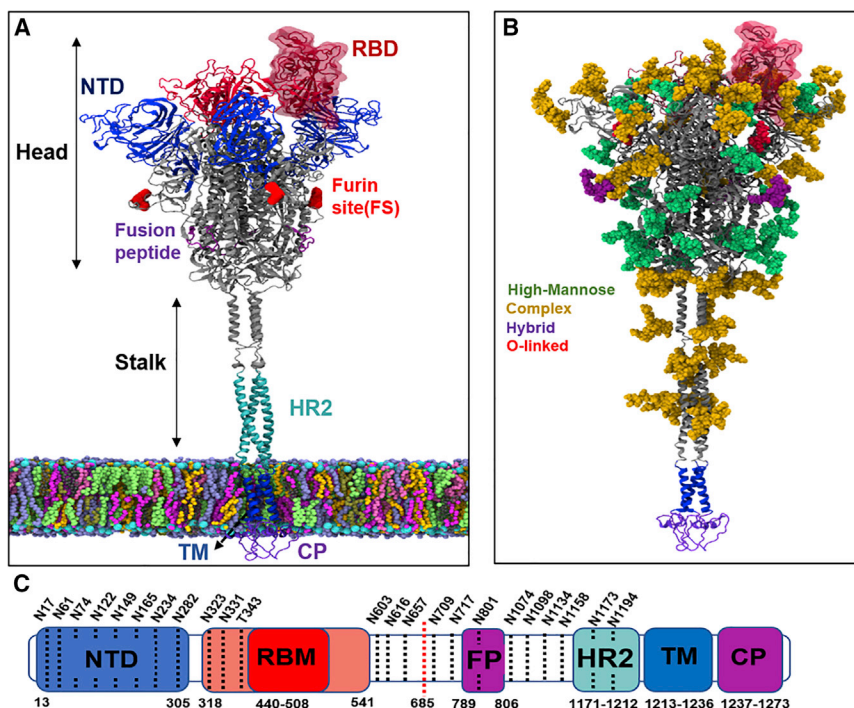
FIGURE 1 Structure of spike protein and its glycosylation pattern. (*A*) Different regions of spike protein including the N-terminal domain (NTD), receptor binding domain (RBD), furin cleavage site for cleaving between S1 and S2 subdomains (FS), fusion peptides (FPs), heptad repeat (HR2), and transmembrane (TM) and CP regions. The spike protein is divided into a head and a stalk region. (*B*) Glycans on the spike protein color coded based on their types. (*C*) Sequence of full-length spike protein with domain assignments. To see this figure in color, go online.

cytoplasmic (CP) domain. The homotrimeric spike protein is highly glycosylated, with 22 predicted N-linked glycosylation and four O-glycosylation sites per monomer, most of which are confirmed by cryo-electron microscopy studies (15–18). Glycosylation of proteins plays a crucial role in numerous biological process such as protein folding and evasion of immune response (19).

The RBD of SARS-COV-2 binds to its receptor h-ACE2 with a significantly higher affinity than the earlier SARS coronavirus, which is suggested to be the reason for its higher infection rate than SARS-COV in 2002–2003 (20). A receptor binding motif (RBM) in RBD makes all the contacts with the receptor by binding its concave surface to the convex exposed surface of h-ACE2 (21). The details of interaction between the RBD of SARS-COV-2 and h-ACE2 are studied by us and by other groups, which showed that RBD of SARS-COV-2 binds to h-ACE2 with ∼30 kcal/mol higher affinity than the SARS-COV RBD. Mutations at critical residues from the SARS to the SARS-COV-2 RBM are suggested to be the reason for the higher affinity (20,22). Natural mutations at the RBM of SARS-COV-2 have been found in multiple countries, which are not known to be linked with the severity of coronavirus (23). In an earlier study, we have shown that most of these mutations in the RBM do not affect the binding affinity of SARS-COV-2 RBD to h-ACE2, demonstrating the high tolerance of the interface for mutations. Our analysis showed the high contribution of critical residues K417, Q493, Q498, and G502 of RBM in SARS-COV-2 for its high binding affinity. In addition, our alanine scanning of interface residues in

nCOV-2019 RBD showed that alanine substitution at some residues such as G502, K417, and L455 can significantly decrease the binding affinity of the complex (20).

The spike protein in SARS-COV-2 is highly glycosylated, with 22 N-linked glycosylation sites and four O-linked glycosylation sites (15–17). Glycosylation of the spike proteins plays a major role in evading the humoral immune system by cloaking the S protein with a thick shield of glycans (24,25). For example, the HIV-1 envelope glycoprotein (Env) features ∼93 N-linked glycosylation sites with mostly high-mannose glycans (Man5–9) and covers most of the surface of the spike protein in HIV-1, comprising over half its mass (3,26,27). N-linked glycosylation starts with synthesis of precursor oligosaccharides, which are modified to high-mannose forms by glucosidases and then trimmed to complex forms in the Golgi by glucosyltransferases for signaling and other glycobiological functions (28). A higher degree of processing is usually indicative of exposure or accessibility of glycans to enzymes (3). Dense crowding glycan regions limit the activity of processing enzymes at these locations.

In this study, we report the microsecond-long molecular dynamics (MD) simulation of all-atom solvated fully glycosylated spike protein embedded in a viral membrane model in both RBD-up or open state (Protein Data Bank, PDB: 6VSB) (11) and RBD-down (PDB: 6VXX) (14) or closed state. The structure of the two states for glycosylated S protein in viral membrane were taken from the CHARMM-GUI website for this study (16). Details of modeling different regions were presented in detail by Im and co-workers (16).

Structural changes in the spike protein happen on an order of microseconds timescale, in which a scissoring motion is observed between the NTDs in the RBD-up conformation. We have used network analysis and centrality measures in graph theory to pinpoint the structural features of the glycan shield in the context of glycan-glycan interactions as well as binding of antibodies to the spike protein. A modularity algorithm helped us to find glycan microdomains featuring high glycan-glycan interactions with breaches between microdomains for antibodies to bind and neutralize the virus.

METHODS

MD simulation

Structures for glycosylated spike protein in the viral membrane with RBD-up and RBD-down states were taken from the CHARMM-GUI website (http://www.charmm-gui.org/docs/archive/covid-19). Im et al. (16) provided eight different structures of spike protein in open and closed conformations, in which two models were built for HR linker (HR2), two models for the HR2-TM domain, and two models for the CP region. We used model 1_2_1, in which the missing loops for RBD were built by template-based modeling and the N-terminal loop was constructed based on electron density map. Ab initio monomer structure prediction and ab initio trimer docking were used for the HR2 linker domain using PDB: 5SZS (29) from SARS-COV-1, for which two models were built (model 1 used here). Two models were constructed for the HR2-TM junction using PDB: 5JYN (30) as a template: a model using a template structure (model 2 used here) and another model with more structural differences. Finally, the Cys-rich CP domain was constructed using PDB: 5L5K (31) as a template. Moreover, the palmitoylation sites for this model are at residues C1236 and C1241. The list of disulfide bonds in both open and closed structures are shown in Table S1. The viral membrane composition used in this study is shown in Table S2. The CHARMM36 force field (32–34) was used for protein, lipids, and carbohydrates in this study. The glycan composition for each site in the spike protein represents the most abundant based on experimental mass spectroscopy data. The selected glycan sequences include 22 N-linked and 1 O-linked for each monomer of the trimeric spike, and the composition at each site is shown in Table S2 (15,17). GROMACS (35) software was used for MD simulation. Energy minimization was performed in 5000 steps using the steepest descent algorithm. A LINCS algorithm in all steps constrained the bonds containing hydrogen atoms. Equilibration was performed with a standard six-step equilibration script from CHARMM-GUI with restraints on protein, lipid, and glycan atoms (36). For the first three steps of equilibration, each step included 125 ps using the Berendsen thermostat at temperature 310.15 K and a coupling constant of 1 ps (37). In the last four equilibration steps, a Berendsen barostat (38) was used to maintain the pressure at 1 bar. For the production step, all restraints were removed, and the system was simulated under an NPT ensemble using the Parrinello-Rahman barostat (39) with a compressibility of 4.5×10^{-5} bar^{-1} and a coupling constant of 5 ps. The temperature was maintained at 310 K using a Nosé-Hoover thermostat with a temperature coupling constant of 1 ps (40). The production run lasted 1 μs for each system (RBD up and RBD down), with a 2 fs time step and the particle-mesh Ewald (41) for long-range electrostatic interactions using GROMACS 2018.3 package (35).

Solvent-accessible surface area

Solvent-accessible surface area (SASA) was calculated using VMD (42) with a probe radius of 7.2 Å, which represents the hypervariable region of antibodies (43). Multiple regions were chosen for SASA calculation:

RBM (residues 440–508), RBD (residues of RBD away from RBM 330–440 and 509–520), and NTD (residues 13–310).

Network analysis

In the glycan network, each glycan is represented with a node (69 nodes for 69 glycans). To assign edges between nodes in the graph, we first calculated the distance between heavy atoms of different glycans in the starting structure, and if the distance between two glycan heavy atoms is less than 50 Å, an edge is assigned between the two nodes. Next, to incorporate simulation data and dynamics of glycans into the network, we calculated the absolute value of average nonbonded interaction energy between every two glycans, normalized the values to be between 0 and 1, and assigned them as the edge weights. To simplify the network, we removed the edges with weights less than 0.05. This further reduced the number of edges from the starting graph. These normalized interaction energies represent the adjacency matrix for the graph from which different centrality measurements can be made. Two different centrality measurements were used to analyze the network of glycans. "Betweenness centrality" quantifies the number of times a node acts as a bridge along the shortest path connecting every two nodes in the graph and is an important indicator of the influence of the node within the network. "Eigenvector centrality" measures the node's importance in the network by considering the importance of the neighbors of that node. If the node is connected to many other nodes that are themselves well connected, that node is assigned a high eigenvector centrality score. For a graph with adjacency matrix A, the relative centrality score of the node i is defined as

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j$$

The sum is over all j such that nodes i and j are connected. To write this in a matrix form, let x be the vector containing the centrality scores and A be the adjacency matrix. Then, we can write

$$Ax = \lambda x$$

With a constraint that all the components of the eigenvector x be positive, there is only one eigenvalue that satisfies the equation, and therefore, a unique centrality score is assigned to each node in the graph. A modularity algorithm in Gephi was used to find glycan microdomains that have high number of edges. All network analysis was performed in NetworkX, and graph visualizations were done using Gephi (44–46).

Glycan-antibody overlap analysis

Antibodies that bind and neutralize spike protein are divided into three categories (47): antibodies that bind to the exposed part of the RBD (RBM-binder), antibodies that bind to epitopes away from the RBM in the RBD or RBD-binder, and antibodies that bind to epitopes on the NTD (NTD-binder). Three different antibodies were used in this study: B38 for the RBM-binder (PDB: 7BZ5) (48), S309 for the RBD-binder (PDB: 6WPT for open and PDB: 6WPS for closed state) (49), and 4A8 antibody as the NTD-binder (PDB: 7C2L) (50).

Data availability

The simulation trajectories, pdb, and psf of open and closed spike conformations embedded in viral membrane are uploaded to the MolSSI database and can be downloaded from https://molssi-bioexcel-covid-19-structure-therapeutics-hub.s3.amazonaws.com/KlaudaGroup/spike_6VXX_1_2_1_nowater.gz and https://molssi-bioexcel-covid-19-structure-therapeuticshub.s3.amazonaws.com/KlaudaGroup/spike_6VSB_1_2_1_nowater.gz.

RESULTS

Dynamical motions of the spike protein

The root mean-square deviation (RMSD) for each region of the spike protein in RBD-up and RBD-down states is represented in Fig. S1. The stalk region in both RBD-up and RBD-down states show a fluctuating RMSD, which is due to bending motion in this region of spike protein. A tilting in the head of spike is also observed in high-resolution cryo-electron tomography (cryo-ET) images as well as other recent MD studies of glycosylated spike in viral membrane (51,52) The bending dynamics in the stalk region is suggested to assist the virus in scanning the cell surface for receptor proteins more efficiently (52). The angle distribution for tilting of the head and stalk domains were calculated from total 2 μs simulation. The head region had an angle distribution of 18 ± 10°, and the stalk region angle distribution was 34.5 ± 6°. The head of the spike protein was shown to bend up to 90° toward the membrane. However, observing larger tilting angles requires more sampling of spike protein in viral membrane (53). A snapshot of the open system at 1 μs is represented in Fig. 2 A, which shows the incline between the head and the stalk domains of spike. The three chains in the spike protein demonstrate various types of motions characterized by different RMSD values (Fig. S2). Glycan root mean-square fluctuations (RMSFs) were calculated for heavy atoms in each glycan, and then we computed the mean and standard deviation for the glycan (Fig. 2 B for open and Fig. S3 for closed state).

Consistently, in all chains, few of the NTD glycans such as N74 show high RMSF, which is due to the high solvent exposure of this glycan in NTD compared to other regions. The glycans near the RBD in chain A (RBD up), such as N234 and T323, showed less fluctuation than other chains. N234 and T323 are sandwiched between the RBD and NTD of neighboring monomers in the trimeric spike protein. Glycans in the stalk region showed high RMSF values, demonstrating the effective shielding of the spike in this region for both RBD-up and RBD-down states.

A principal component analysis (PCA) was performed on the head region (residues 1–1140) of both RBD-up and RBD-down states to extract the fundamental motions of the trimeric protein. The first two PCs are visualized in a two-dimensional plot (Fig. 3, A and B). Both RBD-down and up states feature similar behaviors in their PCA plot. In RBD up, PC-1 captures 56% of conformational motion, whereas in RBD down, PC-1 captures only 44% of all conformational motion. This shows the higher conformational change in the RBD-up state compared with the RBD-down state. The first eigenvector was used to construct the porcupine plots to visualize the most dominant motions in RBD-up (Fig. S4 A) and RBD-down (Fig. S4 B) states. In the RBD-up state, a scissoring motion is observed between the NTD of chain A and NTD of chain B. Based on the PCA for the RBD-up state, the total simulation time was separated into three clusters: cluster 1, 0–200 ns; cluster 2, 200–600 ns; and cluster 3, 600–1000 ns. The distribution of distance between center of mass of NTD of different
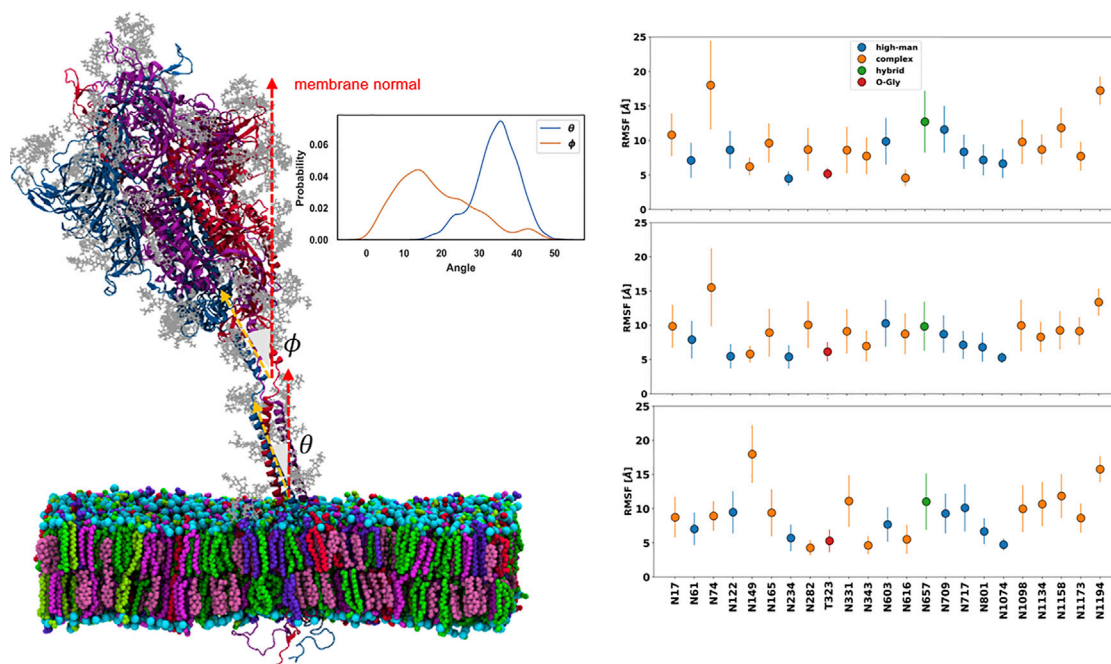

FIGURE 2 (A) Snapshot of the spike in open state after 1000 ns. Different monomers of the spike trimers are color coded with monomer A (up) in red, B in blue, and C in purple. (B) RMSF of glycans in the open state of spike for different chains A to C from top to bottom. Error bars show the standard deviation for heavy atoms within each glycan. To see this figure in color, go online.
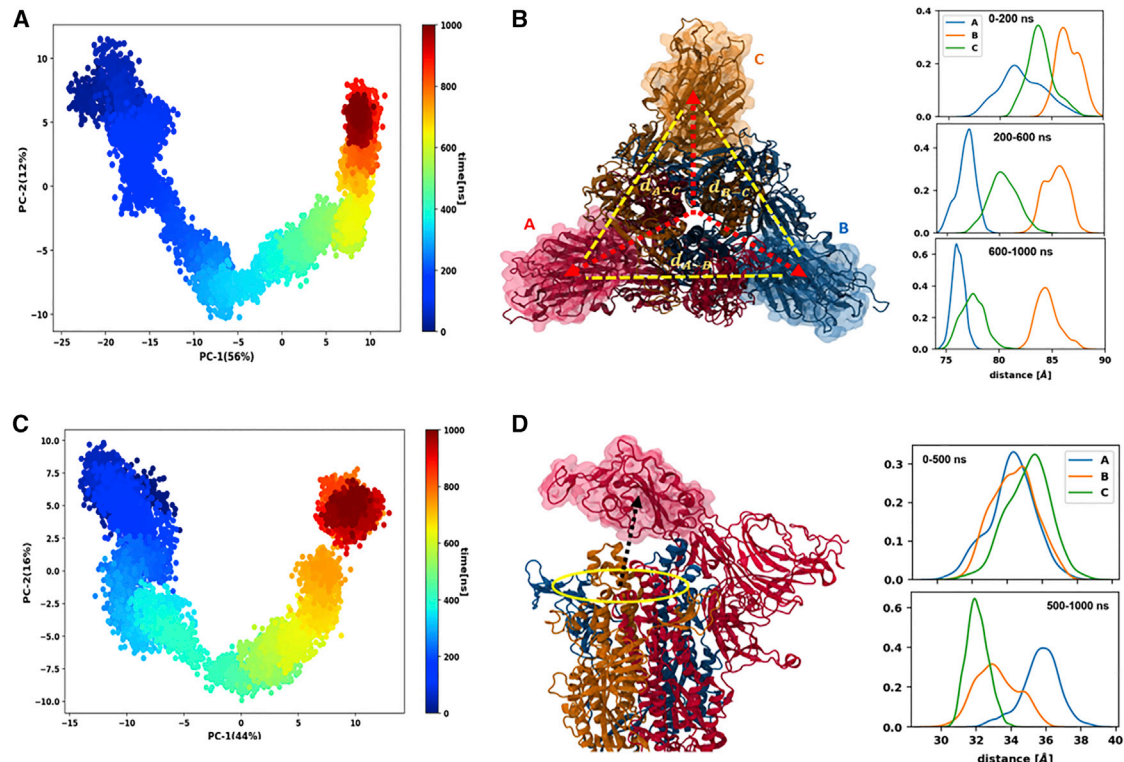
FIGURE 3 (*A*) Two-dimensional PCA for the open state of spike head. (*B*) Distribution of distance between the centers of mass of NTDs on different monomers in the spike trimer for clusters of simulation data 0–200 ns, 200–600 ns, and 600–1000 ns. (*C*) Two-dimensional PCA for the closed spike head. (*D*) Distribution of distance between the center of mass of RBDs of different monomers and the center of apex for 0–500 ns and 500–1000 ns of simulation. To see this figure in color, go online.

chains are calculated for the three different clusters and shown in Fig. 3 *B*. In the open state, the NTD of chain B goes toward the center of apex. The distribution of distance between the center of apex and each NTD is shown in Fig. S4 *C*. In the RBD-down state, PC-1 shows the motion in the RBD of chain A toward the open conformation (Fig. S4 *B*). The simulation for RBD-down states was separated into two clusters, and distribution of the distance between center of mass of RBD and the apex center was calculated for all the chains in both clusters. A clear separation is observed in which the RBD of chain A is more distant from the apex center in the second cluster (Fig. 3 *D*).

Occupancy of spike protein by glycans

Despite the highly dense glycan shield in the spike protein, there are breaches within the shield that antibodies can bind to neutralize the virus (43). A volume map of glycans in the spike protein in both up and down conformations are shown in Fig. 4, in which isosurfaces were visualized for glycans from the 1 μs MD simulation trajectory. The stalk region in the spike protein is highly shielded by the glycans, and it is unlikely for antibodies to bind to this region. The spike head, on the other hand, shows glycan holes, providing opportunities for antibodies to bind. Importantly, the NTDs of

all chains show epitopes free from glycans. The RBM of chain A in the up conformation is completely free from glycans and presents the least shielded domain of the spike. Furthermore, there are regions on the RBD away from the RBM that show epitopes for antibodies and are not shielded by glycans. We have computed the SASA of different regions of the spike for antibodies with a probe radius of 7.2 Å that represents the hypervariable domains of antibodies (43). The RBM of chain A in the up conformation shows the highest SASA in all chains, followed by the RBM of chains B and C. On the other hand, the RBD of chain A shows the lowest SASA among the three monomers. The NTDs of all chains show high SASA in all chains. In the closed state, the RBM and NTD of chain A show higher SASA than other chains, which is due to their conformational changes toward the open state. In summary, epitopes on the RBM, RBD, and NTD of spike protein show high SASA for antibodies to bind and neutralize the spike. These findings are consistent with observations from Amaro et al. (43). In their study, the RBM of the open conformation showed higher SASA than that of the closed conformation using different probe sizes. Moreover, using a probe radius of 7.2 Å, they showed that the RBD has a lower SASA than the RBM and NTD, which is consistent with the findings in our study.
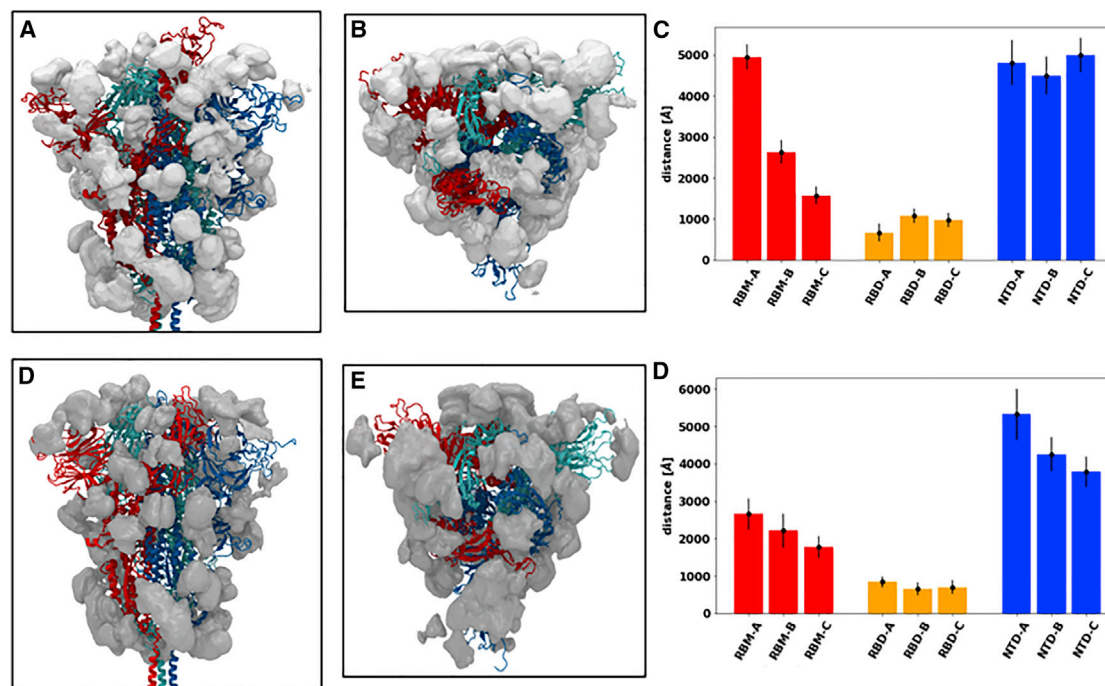
FIGURE 4 (*A*) Glycan occupancy (*gray surface*) of different regions in the spike head for RBD-up state. Monomer A (up) is shown as red, monomer B as blue, and C as cyan. (*B*) Top view of spike head in open state. (*C*) SASA calculated with probe radius of 7.2 Å for different regions of spike head. (*D*) Glycan occupancy at closed state. (*E*) Top view of closed state. (*F*) SASA of closed state. To see this figure in color, go online.

Glycan-glycan interaction and network analysis of glycans

A network analysis was carried out on the glycan shield of the spike protein in both RBD-up and RBD-down states to find the glycans that are most important for an effective shield. This approach has recently been applied to the glycan shield of HIV-1 spike protein (54–56). In the glycan network, each glycan represents a node in the graph, and two nodes are connected by an edge if the glycans have a distance less than 50 Å in the starting structure. Edges are weighted by the normalized absolute value of nonbonded interaction energy (vdw + electrostatic) between each pair of glycans. A network was built for the whole simulation time in both RBD-up and down states. All network analysis was performed in NetworkX, and graph visualizations were done using Gephi (45,46). The adjacency matrix for these two networks is calculated (Fig. S5). Two centrality measurements in graph theory were utilized to analyze the network of glycans. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path connecting every two nodes in the graph and is an important indicator of the influence of the node within the network. Eigenvector centrality measures the node's importance in the network by considering the importance of the neighbors of that node. If the node is connected to many other nodes that are themselves well connected, that node is assigned a high eigenvector centrality score (Fig. 5).

Glycans in the stalk region show high eigenvector centrality in both RBD-up and down states. This means that glycan-glycan interactions in the stalk region are strong and glycans in this region are well connected, which results in effective shielding of the stalk against antibodies. In contrast, connections in the spike head, and especially at the apex, are sparser because the eigenvector centralities are small for this region. Two apex glycans (N234A and N165B) near the RBD of chain A (up) in the open state show a high betweenness centrality (BC). These two glycans show a low BC in the closed state. Glycan N616B in RBD up and N603B in RBD down show the highest BC in their corresponding network, which shows the great impact of this glycan in the proper shielding of the spike protein. Glycans N603 and N616 connect lower head with upper head glycans and are highly central in the network.

Glycans in the head region are well separated from glycans that shield the stalk of the spike protein. Consequently, we performed network analysis of glycans in the head region of the spike protein for the total simulation time. Centrality measurements for open and closed conformations of the spike head are shown in Fig. S6, *A* and *B*. A modularity maximization algorithm (44) is used, which resulted in identifying five different glycan microdomains for the RBD-up state (Fig. 6 *A*) and four glycan microdomains in the RBD-down state (Fig. 6 *B*). Microdomains feature a high glycan-glycan interaction among them and a lower number
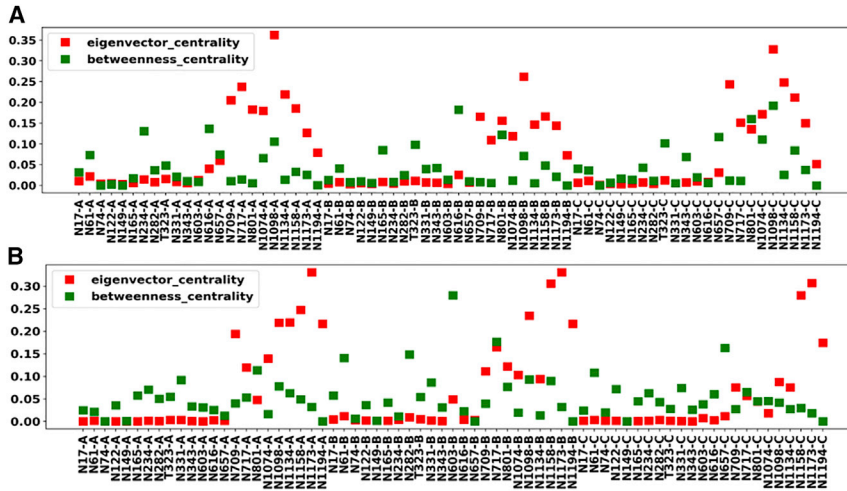
FIGURE 5 Centrality measurements (eigenvector and betweenness) for (*A*) RBD-up and (*B*) RBD-down states. To see this figure in color, go online.

of edges between different microdomains (54). The higher number of microdomains in the RBD-up state shows that the spike protein is more vulnerable when the RBD is in the up conformation. Glycans in the lower head all belong to the same microdomain (Cyan-I) in both the RBD-up

and RBD-down states. This demonstrates the effective shielding of the lower head by glycans regardless of the RBD conformation, as all these glycans belong to the same microdomain. Glycan microdomains in RBD-up and down conformations were mapped onto the spike head to
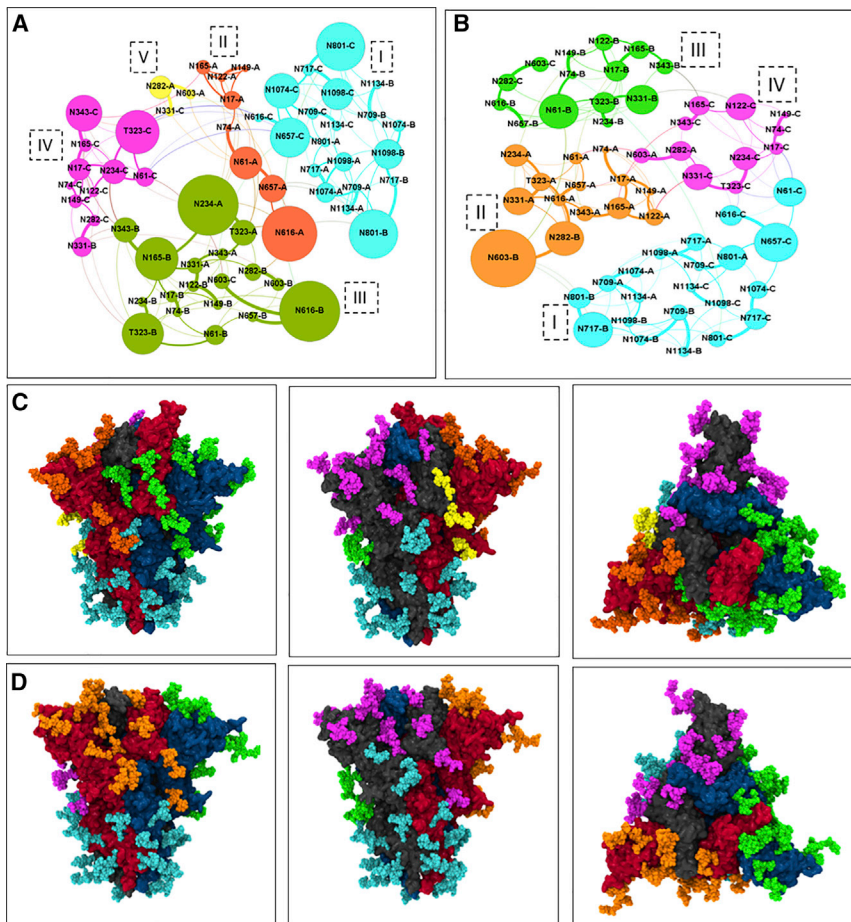

FIGURE 6 (*A*) Microdomains in the open state of spike head with each microdomain color coded. Glycans are connected, and the thickness of the edge shows the edge weight. (*B*) Microdomains in the closed state of spike head. (*C*) Microdomains color coded on the spike protein in open state. (*D*) Microdomains in the closed state of spike head. To see this figure in color, go online.

visualize these clusters on the protein (Fig. 6). Overall connectivity is lower near the RBD because this region is divided into three different microdomains in the RBD-down state and four microdomains in the RBD-up state. In the RBD-down state, most glycans have similar BC, and only one glycan, N603B, shows a relatively higher BC in the network. This is a high-mannose glycan, which connects the upper head with the lower head region in the glycan network and is crucial for effective shielding. In both the RBD-up and down states, glycans at the lower head also showed high eigenvector centrality, indicating the effective shield of this region. When the RBD is in up conformation, glycans near the RBD of chain A (up) can interact with glycans from the NTD of chain B. As a result, when RBD is open, glycans N234A, T323A, N331A, and N343A all belong to the microdomain that comprises glycans of chain B (shown as *green*). This leads to encompassing the RBD of chain (A) in the RBD-up state by the same microdomain (Green-III), which enhances the shielding of the RBD region away from the RBM. This is also demonstrated by the lower SASA of the region of the RBD of chain A away from the binding interface with ACE2 (Fig. 4 *B*). However, in the RBD-down state, the mentioned glycans of chain A are distant from glycans of chain B, and therefore, they belong to the microdomain that includes other glycans from chain A (shown as Orange-II). Furthermore, glycans N234A and N165B show high BC in the RBD-up state, which is due to their interaction with other glycans from the space left open from the RBD of chain A. Glycan N616, which is a fucosylated complex glycan, also shows a high BC in the RBD-up state. Interestingly, most of the glycans in the lower head region are oligomannose, whereas the glycans at the upper head region (apex) are mostly complex glycans. Complex glycans have a higher degree of processing by glycan-processing enzymes. This correlates with the higher number of microdomains in the upper head region, where connections between glycans are sparser than the lower head region. Glycan sparsity was shown before to correlate with the degree of processing.

Dynamic motions of the spike protein can affect the patterns of glycan-glycan interactions by bringing glycans of different domains in closer proximity. To study whether the conformational motions identified by PCA are coupled with any changes in centrality of glycans, we performed network analysis on the clusters of simulation found from PCA (three clusters for open and two for closed conformation). BCs were calculated for glycans in these clusters, and changes in betweenness centrality ΔBC were measured between these networks of glycans built on different clusters (Fig. S6, *C* and *D*). For the open state simulation, the scissoring motion in NTD is coupled with increasing the BC of N234A, N603A, and N165B. As the NTD of chains A and B come closer together, glycans in these two chains make stronger connections, especially near the RBD of chain A, where because of the open state, glycan N234A,

which is inserted in the space left open by the RBD in the open conformation, can freely interact with glycans of chain B. N603 is a high-mannose glycan in the middle regions of the spike, and the NTD scissoring motion grants a higher BC for this glycan in the RBD-up state. In the RBD-down state, the motion in the RBD of chain A to the up conformation increases the BC of glycans such as T323A and N657A and does not affect the BC of most other glycans. Glycan T323A is located at the tail of the RBD, and N657A is in close proximity of the RBD of chain A in the middle head region. The conformational change in RBD brings these glycans closer, resulting in a more compact network of glycans in the middle head region and higher BC of these glycans and other neighboring glycans in chain A (Fig. S6 *D*).

Antibody overlap analysis

Neutralizing antibodies look for breaches in the glycan shield, where the glycan densities are sparse (3,54). Within a microdomain, glycans are highly connected by a high number of edges in the network, and most antibodies bind the regions between these microdomains as comparatively sparse edges connect different microdomains (3). Therefore, these microdomains help identify susceptible regions of spike protein for immunological studies. Antibodies for spike protein were divided into three categories: RBM-binder, RBD-binder (region of RBD away from RBM), and NTD-binder. The antibodies chosen for each category are presented in the Methods section. To investigate the relation between binding of antibodies to known epitopes of the spike protein and the identified glycan microdomains, we utilized an antibody overlap analysis (3). Antibodies were first overlaid with the spike protein by fitting to their corresponding region in the spike protein. Next, we calculated the average number of clashes between each overlaid antibody and the glycan heavy atoms in each microdomain with a cutoff distance of 5 Å during 1 μs simulation. Results of this analysis are shown in Fig. 7 *A* for the RBD-up state and Fig. 7 *B* for the RBD-down state.

The RBM-binder antibody had the lowest number of clashes among all antibodies with microdomains. RBD-binder antibody of chain A had the highest number of clashes among all chains with glycans in microdomain III(G). RBD-binder antibodies of chains B and C have lower number of clashes with glycan microdomains than chain A, where the RBD is in the up state. Antibody binding to the NTD of chain C (NTD-C) in open state shows a high number of clashes with microdomain IV(P). Similarly, in the closed state, NTD-binder antibody in chain C (NTD-C) also shows a high number of clashes with cluster IV(P). The NTD of chain C also showed a lower SASA than the NTD of other chains. In the open state, microdomain III(G) comprises the high-BC glycans N234A, N165B, and T323B. The high BC of glycans in this microdomain correlates with its high number of clashes with antibodies
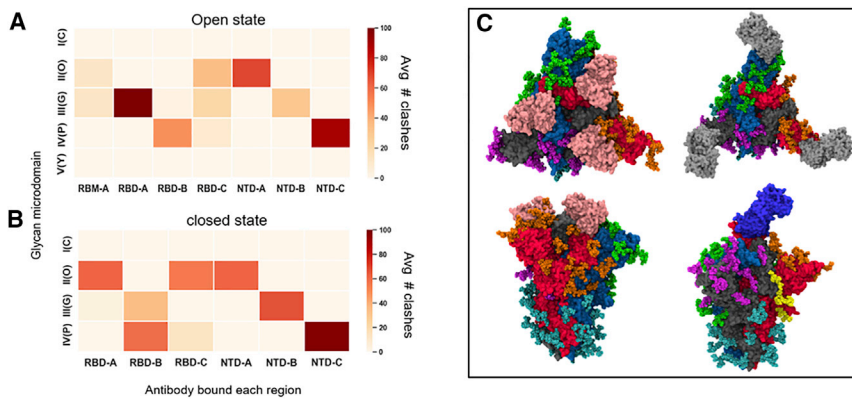
FIGURE 7 (*A*) Antibody overlap analysis for the open state. Microdomains are represented by their group and their color in Fig. 6 with Yellow-V (Y), Purple-IV (P), Green-III (G), Orange-II (O), and Cyan-I (C). The character after each region in the *x* axis specifies the chain on which the antibody was overlaid, and the number of clashes was calculated. (*B*) Antibody overlap analysis for closed state. Because all the RBDs are in closed conformation, we did not calculate clashes for RBM-A. (*C*) Overlaid antibodies with spike protein. Top left shows RBD-binder antibody (*pink*) with closed state spike. Top right shows NTD-binder antibody (*silver*) with closed spike. Lower left shows RBD-binder antibody (*pink*) with open state of spike, and lower right shows RBM-binder antibody (*blue*) with spike open. To see this figure in color, go online.

of RBD-A and NTD-A. In the RBD-down state, RBD-binder antibodies seem to have a similar number of clashes with different glycan microdomains. To identify glycans that have the most effect on antibody binding, we also quantified the number of clashes of antibodies with each glycan in different chains and averaged over the simulation time of 1 μs (Fig. S7). The RBM antibody has only a low number of clashes with the N165A glycan. In the open state, the RBD-A antibody has a high number of clashes with multiple glycans (N122B, N149B, N331A, and N343A). RBD antibodies bound to the other chains (RBD-B and RBD-C) show a lower number of clashes with glycans N122 and N165. The NTD antibody of chain C (NTD-C) in both the open and closed states shows a high number of clashes with glycans N74C and N149C.

DISCUSSION

Understanding the structure and dynamics of glycan shield in the spike protein of SARS-COV-2 is an indispensable requirement for any antibody and vaccine design endeavors (24,57,58). To this end, we have performed MD simulations of fully glycosylated spike protein of SARS-COV-2 in both the open and closed states. Analysis of dynamics for 2 μs trajectories showed a tilting motion in the stalk region, which was also demonstrated with experimental cryo-ET images (51–53) and suggested to aid the virus with screening the host cells for receptor proteins (Fig. 2 *A*). Glycan motions were characterized by RMSF (Fig. 2 *B*; Fig. S3), which showed higher values for stalk glycans demonstrating the high shielding potential of this normally solvent-exposed region. PCA of the head region of the open state of the spike demonstrated a scissoring motion between NTDs of neighboring monomers A (up) and B. This scissoring motion resulted in trimer asymmetry with the NTD of monomer B advancing toward the center of the apex region and the NTD of monomer A showing an angular motion centering on the apex center and toward the NTD of

monomer B. Based on the PCA, the simulation was divided into three different clusters with distribution of distances between the centers of mass of NTDs of different monomers showing different asymmetrical trimers in each cluster (Fig. 3 *B*; Fig. S4, *A* and *C*). A scissoring motion in the trimeric spike of HIV-1 on the submicrosecond timescale was observed by Lemmin et al. (54), which was suggested to be essential for receptor binding. The NTD scissoring motion is only observed in the open state, and this could suggest that the scissoring motion is a means employed by the virus to camouflage parts of the spike protein (such as regions of the RBD excluding the RBM) when the RBD is in the up conformation. The first PC for the RBD-down state was visualized (Fig. S4 *C*) and shows the conformational change of the RBD in chain A from the down toward the up conformation. The distribution of distance between the center of mass of the RBD in each monomer from the center of apex region (Fig. 3 *D*) exhibited this motion for two clusters of data, 0–500 and 500–1000 ns, separated based on the PCA of the open-state simulation.

The abundance of information in MD simulations of glycosylated spike protein may hinder identifying important biological features of the glycan shield. Therefore, a network analysis approach is used to identify collective behavior of glycans. The most central region of the spike, based on eigenvector centrality of the network, is shown to be the stalk domain, and the lower head region of the spike is where a dense array of glycans gives rise to resilience to enzymatic actions. Most of the glycans at the lower head region and upper stalk domain are high mannose, with the lower degree of processing that correlates with their high centralities in the graph. The high eigenvector centrality of the lower head and the stalk glycans also makes it hard for neutralizing enzymes to target, and to date, no epitopes for antibodies have been found that target this region (18,24,47).

Glycans on the head region demonstrated different behaviors depending on the RBD state (up or down conformation). Interestingly, two glycans at the open state, N234A and

N165B, show a high BC. N234A occupies the volume left open by the RBD in the open state, whereas in the closed state, it is directed toward the solvent. The highly conserved glycan N165B is inserted between the RBD of chain A (up) and the NTD of chain B, and in the open state, it occupies the volume left vacant by the RBD (A) in the up conformation. Amaro and co-workers (43) studied the fully glycosylated spike protein of SARS-COV-2 computationally and showed that N234A and N165B are crucial for stabilizing the RBD in the up conformation in the open state of the spike. Their simulation showed that mutating N234 and N165 to Ala destabilized the RBD in the open state. Furthermore, experimental negative stain electron microscopy and single-particle cryo-electron microscopy showed that the equilibrium population ratio between the open and closed states is 1:1. Deletion of N234 glycan shifts this ratio to 1:4 favoring the closed state, and deletion of glycan at N165 increased the population of the open state with a ratio of 2:1. It was shown that N234 glycan stabilizes the open state of the RBD and inhibits the up-to-down conformational change, and N165 glycan sterically inhibits the down-to-up conformational change of the RBD. Here, we have shown that these two glycans in the open state exhibit a high BC, which is due to their interaction with each other as well as other neighboring glycans at the RBD of chain A as well as the NTD of chain B. We have further demonstrated that the BC of glycans in the head region is coupled with the scissoring motion between the NTDs of monomers A and B. In addition, the scissoring motion gives rise to high BC for glycans in the middle region of the spike for glycans N616A and N616B. This is caused by the tighter packing of glycans in the asymmetric trimer. In the closed state, the BCs of most glycans do not change, which correlates with lower fluctuation of the RBD-down simulation. It was demonstrated for the spike protein of HIV virus that glycans with high BC display a high degree of interaction with other neighboring glycans and are less accessible to glycan-processing enzymes. These highly central glycans such as N603 and N234 are essential for maintaining the mannose character of the glycan shield. Regions with dense crowding glycans have steric constraints on glycans that limit their processing by carbohydrate processing enzymes. Experimental studies using mass spectroscopy along with site-directed glycan removal are needed to understand how deletion of these glycans such as N603, N616, and N234 can affect the glycan processing of the neighboring glycans on the spike protein.

Modularity maximization in network analysis allowed us to find five microdomains of glycans in the RBD-up state and four microdomains in the RBD-down state. The higher number of microdomains at the apex of the RBD-up state shows the greater vulnerability of the spike protein to antibodies in the open state. Glycans at the lower head in both open and closed states belong to the same microdomain (Cyan-I), which shows large number of edges between glycans in this region and thereby effective shielding. Apex glycans are divided into three microdomains in the closed state and four microdomains in the open state. Glycan microdomains were shown in the HIV-1 Env glycan shield to have a broad implication for anticipating immune escape (59). The antibody overlap analysis showed that RBM-binder antibody in open state (up) shows the lowest number of clashes with glycan microdomains (Fig. 7 A). The RBM of chain A also showed the highest SASA among other epitopes on spike protein, which shows its great potential for antibody design strategies. NTD-binder antibodies can also bind to epitopes on the NTD of spike protein. These antibodies showed high number of clashes with glycans N74 and N149 of the respective monomer that they bind. Experimental studies are needed to explore the effect of deletion of these glycans on the neutralization effect by antibodies. The glycans on the surface of spike protein exert a collective behavior, which is an important property that needs to be considered in the context of vaccine and antibody design.

CONCLUSION

In this work, we have studied the microsecond time dynamics and network analysis of the glycosylated spike protein in SARS-COV-2. To answer the need for quantification of the glycan shield of the spike protein of SARS-COV-2, we have utilized MD simulation and network analysis to aid in understanding the collective behavior of glycans. The role of glycans N234A and N165B as the central glycans in the network of glycans in the RBD-up system is discussed. Glycan microdomains are identified featuring high interaction inside them and lower interaction of glycans between different microdomains, which indicates most neutralizing antibodies would bind to regions between these microdomains. The higher number of microdomains in the open state suggests the higher vulnerability of the spike protein in the open state. An antibody overlap analysis identified the microdomains of glycans with higher number of clashes with antibodies. Collectively, this work presents insights to design antibodies and vaccines against coronavirus.

AUTHOR CONTRIBUTIONS

M.G., B.R.B., and J.B.K. designed research. M.G. performed research and analysis. M.G. and J.B.K. wrote the manuscript.

Ghorbani et al.

REFERENCES

1. Zhou, P., X.-L. Yang, …, C.-L. Huang. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 579:270–273.

2. Zhu, N., D. Zhang, …, R. Lu. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382:727–733.

3. Stewart-Jones, G. B., C. Soto, …, P. D. Kwong. 2016. Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. *Cell.* 165:813–826.

4. Dhama, K., K. Sharun, …, W. Chaicumpa. 2020. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Hum. Vaccin. Immunother.* 16:1232–1238.

5. Liu, C., Q. Zhou, …, D. Albaiu. 2020. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Cent. Sci.* 6:315–331.

6. Pang, J., M. X. Wang, …, L. Y. Hsu. 2020. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. *J. Clin. Med.* 9:623.

7. Wang, Q., Y. Zhang, …, J. Qi. 2020. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell.* 181:894–904.e9.

8. Wan, Y., J. Shang, …, F. Li. 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 94:e00127-20.

9. Chan, J. F.-W., K.-H. Kok, …, K.-Y. Yuen. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9:221–236.

10. Cui, J., F. Li, and Z.-L. Shi. 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17:181–192.

11. Wrapp, D., N. Wang, …, J. S. McLellan. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 367:1260–1263.

12. Yan, R., Y. Zhang, …, Q. Zhou. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science.* 367:1444–1448.

13. Davidson, A. D., M. K. Williamson, …, D. A. Matthews. 2020. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med.* 12:68.

14. Walls, A. C., Y.-J. Park, …, D. Veesler. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* 181:281–292.e6.

15. Watanabe, Y., J. D. Allen, …, M. Crispin. 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science.* 369:330–333.

16. Woo, H., S.-J. Park, …, W. Im. 2020. Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *J. Phys. Chem. B.* 124:7128–7137.

17. Shajahan, A., N. T. Supekar, …, P. Azadi. 2020. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology.* 30:981–988.

18. Zhao, P., J. L. Praissman, …, L. Wells. 2020. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microbe.* 28:586–601.e6.

19. Reily, C., T. J. Stewart, …, J. Novak. 2019. Glycosylation in health and disease. *Nat. Rev. Nephrol.* 15:346–366.

20. Ghorbani, M., B. R. Brooks, and J. B. Klauda. 2020. Critical sequence hotspots for binding of novel coronavirus to angiotensin converter enzyme as evaluated by molecular simulations. *J. Phys. Chem. B.* 124:10034–10047.

21. Lan, J., J. Ge, …, X. Wang. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 581:215–220.

22. Spinello, A., A. Saltalamacchia, and A. Magistrato. 2020. Is the rigidity of SARS-CoV-2 spike receptor-binding motif the hallmark for its enhanced infectivity? Insights from all-atom simulations. *J. Phys. Chem. Lett.* 11:4785–4790.

23. Elbe, S., and G. Buckland-Merrett. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* 1:33–46.

24. Watanabe, Y., Z. T. Berndsen, …, M. Crispin. 2020. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* 11:2688.

25. Watanabe, Y., T. A. Bowden, …, M. Crispin. 2019. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta Gen. Subj.* 1863:1480–1497.

26. Zhou, T., N. A. Doria-Rose, …, P. D. Kwong. 2017. Quantification of the impact of the HIV-1-glycan shield on antibody elicitation. *Cell Rep.* 19:719–732.

27. Crispin, M., A. B. Ward, and I. A. Wilson. 2018. Structure and immune recognition of the HIV glycan shield. *Annu. Rev. Biophys.* 47:499–523.

28. Lee, H. S., Y. Qi, and W. Im. 2015. Effects of N-glycosylation on protein conformation and dynamics: protein data bank analysis and molecular dynamics simulation study. *Sci. Rep.* 5:8926.

29. Walls, A. C., M. A. Tortorici, …, D. Veesler. 2016. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 23:899–905.

30. Dev, J., D. Park, …, J. J. Chou. 2016. Structural basis for membrane anchoring of HIV-1 envelope spike. *Science.* 353:172–175.

31. Kong, Y., B. J. Janssen, …, E. Y. Jones. 2016. Structural basis for plexin activation and regulation. *Neuron.* 91:548–560.

32. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

33. Klauda, J. B., R. M. Venable, …, R. W. Pastor. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B.* 114:7830–7843.

34. Guvench, O., S. S. Mallajosyula, …, A. D. Mackerell, Jr. 2011. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* 7:3162–3180.

35. Abraham, M. J., T. Murtola, …, E. Lindahl. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 1–2:19–25.

36. Jo, S., T. Kim, …, W. Im. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29:1859–1865.

37. Lemak, A. S., and N. K. Balabaev. 1994. On the Berendsen thermostat. *Mol. Simul.* 13:177–187.

38. Rühle, V. 2008. Pressure coupling/barostats https://www2.mpip-mainz.mpg.de/~andrienk/journal_club/barostats.pdf.

39. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.

40. Evans, D. J., and B. L. Holian. 1985. The nose–hoover thermostat. *J. Chem. Phys.* 83:4069–4074.

41. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

42. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

43. Casalino, L., Z. Gaieb, …, R. E. Amaro. 2020. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent. Sci.* 6:1722–1734.

44. Blondel, V. D., J.-L. Guillaume, …, E. Lefebvre. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008:P10008.

45. Hagberg, A., P. Swart, and D. S. Chult. 2008. Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Lab. (LANL), Los Alamos, New Mexico.

46. Bastian, M., S. Heymann, and M. Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. *Icwsm.* 8:361–362.

47. Liu, L., P. Wang, …, D. D. Ho. 2020. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature.* 584:450–456.

48. Wu, Y., F. Wang, …, L. Liu. 2020. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science.* 368:1274–1278.

49. Pinto, D., Y.-J. Park, …, A. De Marco. 2020. Structural and functional analysis of a potent sarbecovirus neutralizing antibody. *BioRxiv* https://doi.org/10.1101/2020.04.07.023903.

50. Chi, X., R. Yan, …, W. Chen. 2020. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science.* 369:650–655.

51. Sikora, M., S. von Bülow, …, G. Hummer. 2020. Map of SARS-CoV-2 spike epitopes not shielded by glycans. *bioRxiv* https://doi.org/10.1101/2020.07.03.186825.

52. Turoňová, B., M. Sikora, …, M. Beck. 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science.* 370:203–208.

53. Yao, H., Y. Song, …, S. Li. 2020. Molecular architecture of the SARS-CoV-2 virus. *Cell.* 183:730–738.e13.

54. Lemmin, T., C. Soto, …, P. D. Kwong. 2017. Microsecond dynamics and network analysis of the HIV-1 SOSIP Env trimer reveal collective behavior and conserved microdomains of the glycan shield. *Structure.* 25:1631–1639.e2.

55. Chakraborty, S., Z. T. Berndsen, …, S. Gnanakaran. 2020. Quantification of the resilience and vulnerability of HIV-1 native glycan shield at atomistic detail. *iScience.* 23:101836.

56. Berndsen, Z. T., S. Chakraborty, …, A. B. Ward. 2020. Visualization of the HIV-1 Env glycan shield across scales. *Proc. Natl. Acad. Sci. USA.* 117:28014–28025.

57. Yuan, M., N. C. Wu, …, I. A. Wilson. 2020. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science.* 368:630–633.

58. Chen, X., R. Li, …, L. Ye. 2020. Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. *Cell. Mol. Immunol.* 17:647–649.

59. Hargett, A. A., Q. Wei, …, M. B. Renfrow. 2018. Defining HIV-1 envelope N-glycan microdomains through site-specific heterogeneity profiles. *J. Virol.* 93:e01177-18.