

# Enhanced Surface Accessibility of SARS-CoV-2 Omicron Spike Protein Due to an Altered Glycosylation Profile

Dongxia Wang,<sup>\*,†</sup> Zijian Zhang,<sup>†</sup> Jakub Baudys, Christopher Haynes, Sarah H. Osman, Bin Zhou, John R. Barr, and James C. Gumbart<sup>\*</sup>



Cite This: *ACS Infect. Dis.* 2024, 10, 2032–2046



Read Online

ACCESS |

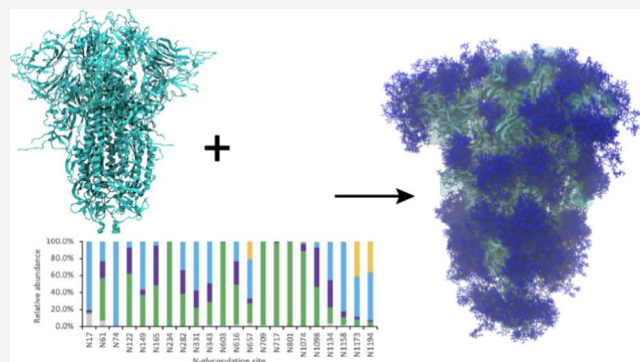
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** SARS-CoV-2 spike (S) proteins undergo extensive glycosylation, aiding in proper folding, enhancing stability, and evading host immune surveillance. In this study, we used mass spectrometric analysis to elucidate the N-glycosylation characteristics and disulfide bonding of recombinant spike proteins derived from the SARS-CoV-2 Omicron variant (B.1.1.529) in comparison with the D614G spike variant. Furthermore, we conducted microsecond-long molecular dynamics simulations on spike proteins to resolve how the different N-glycans impact spike conformational sampling in the two variants. Our findings reveal that the Omicron spike protein maintains an overall resemblance to the D614G spike variant in terms of site-specific glycan processing and disulfide bond formation. Nonetheless, alterations in glycans were observed at certain N-glycosylation sites. These changes, in synergy with mutations within the Omicron spike protein, result in increased surface accessibility of the macromolecule, including the ectodomain, receptor-binding domain, and N-terminal domain. Additionally, mutagenesis and pull-down assays reveal the role of glycosylation of a specific sequon (N149); furthermore, the correlation of MD simulation and HDX-MS identified several high-dynamic areas of the spike proteins. These insights contribute to our understanding of the interplay between structure and function, thereby advancing effective vaccination and therapeutic strategies.

**KEYWORDS:** SARS-CoV-2 Omicron spike, N-glycosylation, Disulfide bond, Molecular dynamics simulation



The COVID-19 pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), continues to pose a significant global health challenge.<sup>1–3</sup> Since its appearance, the virus has been evolving through multiple genetic changes, leading to the emergence of variants with diverse characteristics.<sup>4–6</sup> Omicron is the most recently circulating variant, and its subvariants are still the dominant strains spreading worldwide during 2023.<sup>7,8</sup> The Omicron lineages have attracted substantial attention due to their unusually high number of mutations (Figure 1) in the spike protein, which has resulted in increased transmissibility, milder symptoms, and low severity in infected patients.<sup>9,10</sup>

The spike protein (S) of SARS-CoV-2 virus plays a pivotal role in viral entry and infection.<sup>11,12</sup> It has been characterized that this membrane-anchored S protein forms a homotrimer structure, and each of the protomers contains a S1 and S2 subunits responsible for different functions.<sup>13,14</sup> The S1 subunit comprises the N-terminal domain (NTD) and receptor binding domain (RBD), a key component mediating the attachment of the virus to the host cell receptor angiotensin-converting enzyme 2 (ACE2), while the S2 subunit, containing the fusion peptide and other domains, is

responsible for viral membrane fusion.<sup>12,15,16</sup> The spike glycoprotein is the target for neutralizing antibodies and therefore is the primary target for vaccine development and therapeutic interventions.<sup>12,17</sup>

N-glycosylation, the attachment of carbohydrate moieties to asparagine residues, is a vital post-translational modification that influences protein folding, stability, and immune recognition.<sup>18</sup> The S protein of the SARS-CoV-2 virus is heavily glycosylated with 22 potential N-glycosylation sites (sequons, Figure 1) on each S protomer. These glycans can shield antigenic sites from immune surveillance, affecting viral neutralization and immune evasion strategies.<sup>19–22</sup> Previous studies have also demonstrated that the glycans play other roles beyond shielding the S protein from host immune recognition.<sup>23,24</sup> Since the outbreak of the COVID-19

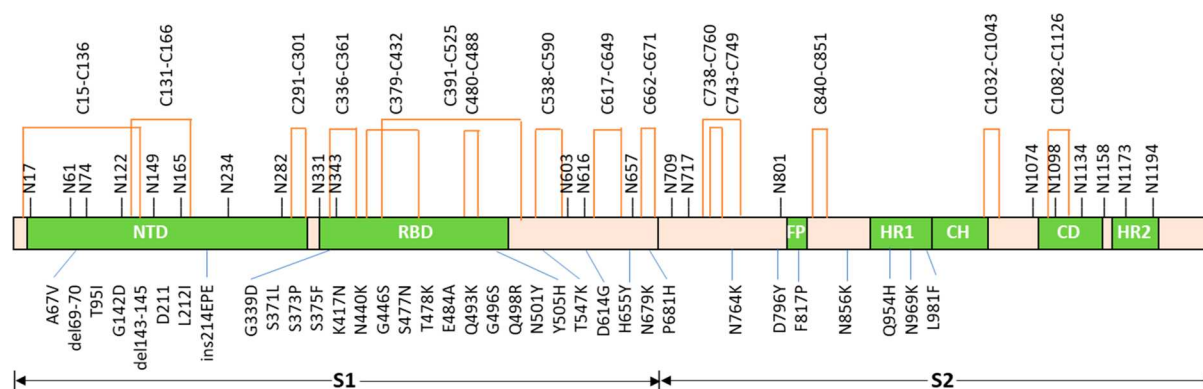
Received: January 8, 2024

Revised: April 27, 2024

Accepted: April 30, 2024

Published: May 10, 2024





**Figure 1.** Schematic structure of the SARS-CoV-2 Omicron (B.1.1.529) spike protein, including amino acid substitution (below), N-glycosylation sites (above), and disulfide bonds (top). Domains include: the N-terminal domain (NTD); receptor binding domain (RBD); fusion peptide (FP); heptad repeat1 (HR1); central helix (CH) region; connector domain (CD); and heptad repeat 2 (HR2).

pandemic, the glycosylation profile of the SARS-CoV-2 S glycoproteins on the initial strain and evolving variants has been intensively studied, particularly through the use of advanced liquid chromatography coupled to tandem mass spectrometry (LC-MS) techniques<sup>20,23,25–42</sup> with recombinantly expressed proteins of ectodomain or subunit constructs and even viral derived S protein.<sup>14</sup> The analysis of glycopeptides derived from S proteins allows the determination of N-glycan profiles for all 22 conserved sequons as well as two novel sequons in the Gamma spike protein.

Disulfide bonds are another important modification critical for maintaining the structural integrity of proteins. These covalent bonds between cysteine residues contribute to protein folding, stability, and overall conformation.<sup>43</sup> The primary sequence of the ectodomain of the SARS-CoV-2 S protein has 30 cysteine residues (Figure 1). Structural and mass spectrometry analysis have revealed that these cysteines form 15 disulfide bonds.<sup>12,13,40</sup> Disulfide bonds in S proteins play an important role in their structure and function. Studies have shown that thiol-based drugs can impair the binding of S protein to ACE2,<sup>44–46</sup> and engineered disulfide bonds can trap S protein in an RBD “down” conformation.<sup>47,48</sup> The disulfide bond between C840 and C851 in the fusion peptide of S protein also can facilitate the binding between this peptide and cell membrane.<sup>49</sup> A recent study has revealed that mutations in the RBD domain of the Omicron S protein affect the stability of two disulfide bonds, elevating the vulnerability of this S variant to reduction.<sup>50</sup>

Our aim in the present study was to better understand the structural characteristics and changes of the SARS-CoV-2 Omicron S protein (S-Omicron) relative to the S protein of the D614G variant (S-D614G). To do that, we used LC-MS to analyze the N-glycosylation profile and disulfide bonds for each, revealing distinct distributions of glycans at some sequons. Molecular dynamics simulations of models of S-Omicron and S-D614G trimers based on the LC-MS results further elucidated changes in protein conformational sampling as a result. In summary, we provide insights into the potential consequences of structural changes in variants on the viral structure, immune recognition, and therapeutic strategies.

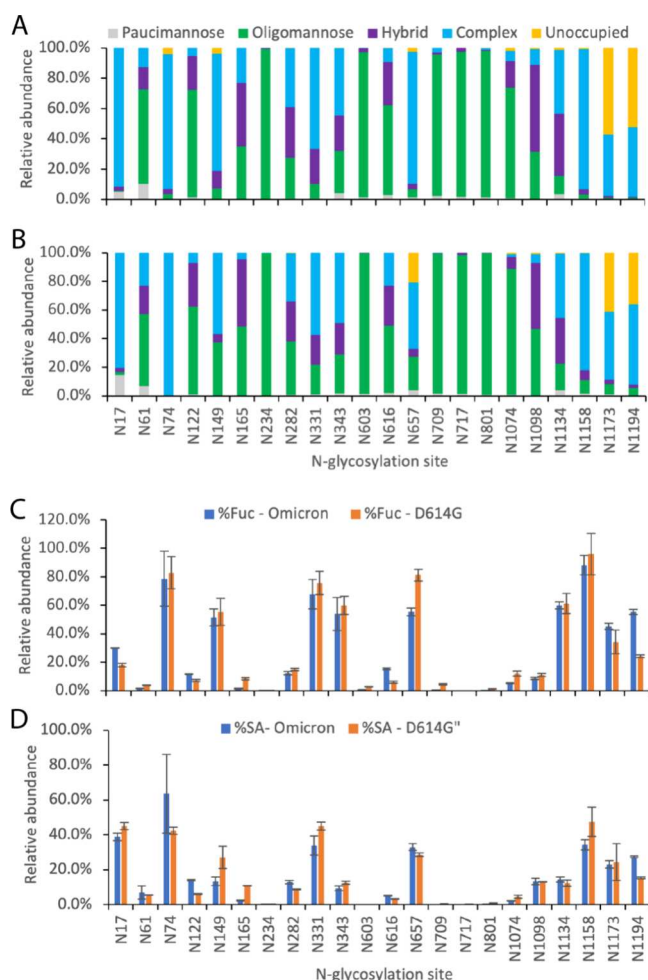
## RESULTS

### Characterization of N-Glycosylation on S Proteins.

Multiple enzyme digestion is a method commonly used to characterize post-translational modifications by the “bottom-

up” MS/proteomics approach. Because of nonspecific cleavage and production of undetected short or long peptides, sequence coverage of a protein characterized by this technique is usually limited by digestion with a single protease, such as trypsin, thus hindering identification and quantification of the peptides bearing target post-translational modifications. Microheterogeneity (multiple glycans on one site) of glycan distribution on glycoproteins adds to the difficulty of site-specific glycosylation analysis. Therefore, almost all researchers, including us, have used two or more proteases to digest proteins to generate peptides with a single glycosylation site and uniform sequences for characterizing glycosylation on SARS-CoV-2 spike proteins.<sup>27–29,31–33,35,37–41,51</sup> During the analysis of N-glycosylation of some SARS-CoV-2 spike variants of concern, we became aware that different enzyme sets could generate results with significant variation for a single sequon. We therefore developed a set of criteria to select the best data set for individual glycosylation sites.<sup>37</sup> Similar approaches were applied in this study. Five different enzyme combinations, including Lys-C/Lys-C, Lys-C/trypsin, Lys-C/chymotrypsin, Lys-C/ $\alpha$ -lytic protease, and Asp-N/Chymotrypsin, were able to provide optimal results of the N-glycosylation profiles at all 22 sites (Table S1; full workflow in Figure S1). As expected, consistent glycan distributions were obtained from the glycopeptides derived from different enzyme digests for many sequons. However, some digestion pairs generated inconsistent results for the same site. For instance, three digestions with Lys-C/Lys-C, Lys-C/trypsin, and Lys-C/chymotrypsin yielded similar results in terms of the number of N-glycans and the relative abundance of the glycans of different processed levels for N282, but the other set of data obtained from Lys-C/ $\alpha$ -lytic protease showed significant differences (Table S1). Our data also showed that the optimal protease combination for a specific sequon on S-Omicron was not necessarily the best one for other SARS-CoV-2 variants of concern spikes, presumably because the different amino acid substitutions among these variants alters the accessibility of some enzymatic cleavage sites altering the efficiency of the enzymes. These results underline the importance of applying multiple protease digests and obtaining consistent data from at least two enzyme combinations whenever possible for confirmation.

Figure 2 shows the distribution of different types of glycans and their contents of fucose and sialic acid groups on all individual sequons of the Omicron spike trimer and the control



**Figure 2.** N-glycosylation profiles of S-D614G (A) and S-Omicron (B), and the relative abundance of fucosylated (Fuc) (C) and sialylated (SA) glycans (D) in these two proteins.

sample, S-D614G. The majority of the 22 sequons in both S-Omicron and S-D614G were occupied predominantly by either underprocessed oligomannose (2 HexNAc and >4 Hex groups) or fully processed complex (more than 3 HexNAc and various Hex) glycans. Whereas complex glycans occupied approximately half or more of the sites N17, N74, N149, N331, N343, N657, N1158, N1173 (S-Omicron) and N1194, the oligomannose glycans dominated the sites of N61, N122, N165 (in S-Omicron), N234, N603, N616, N709, N717, N801, N1074, and N1098 (S-Omicron). A few sites, including N1098 and N165, were detected to have approximately half of hybrid (3 HexNAc) glycans. Approximately half of the N1173 and N1194 sites were not occupied, and a very low level of paucimannose (1–2 HexNAc and less than 4 Hex) was detected at some sequons.

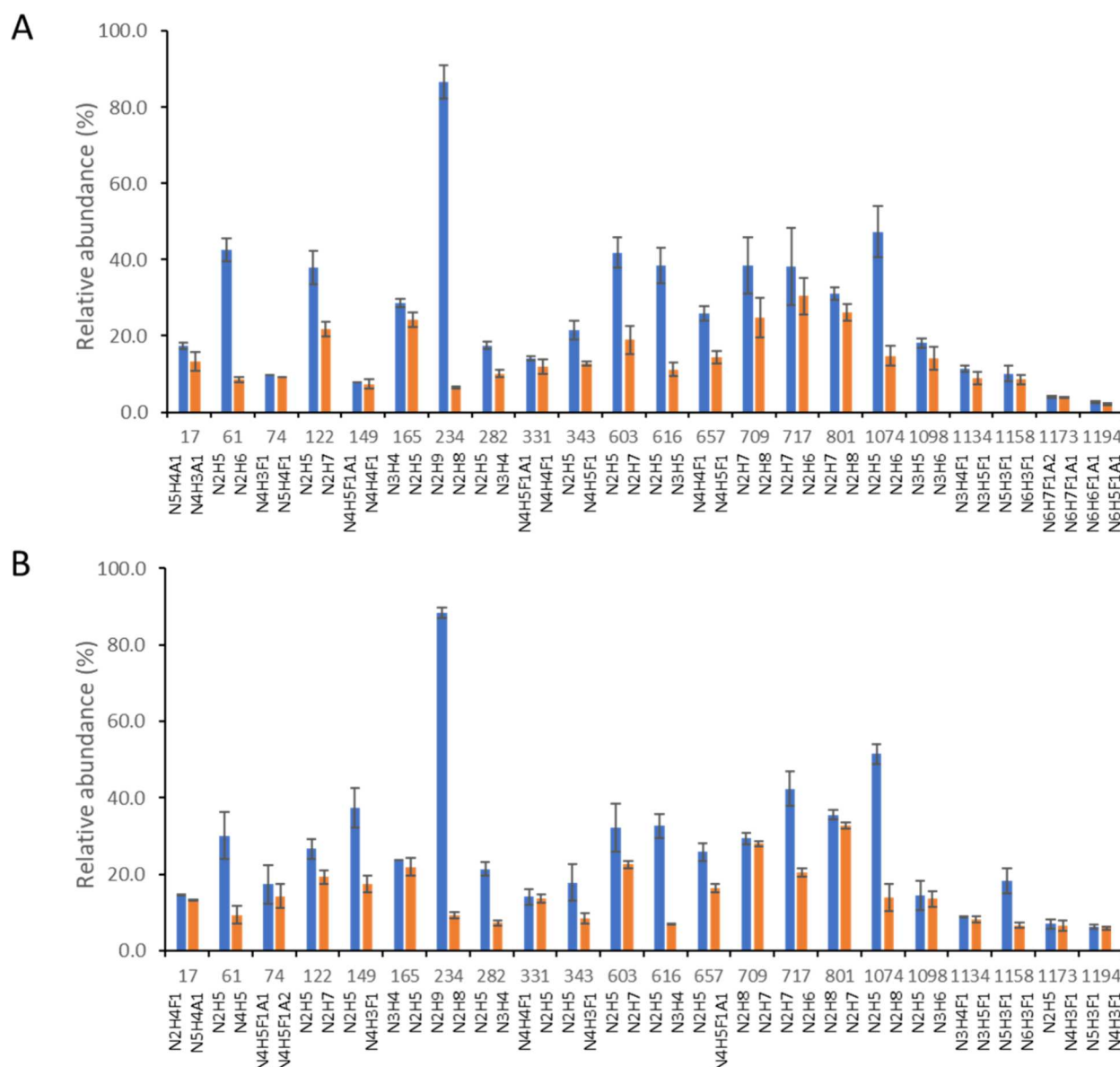
Inconsistent or conflicting results regarding the N-glycosylation profile of the ectodomain SARS-CoV-2 S protein have been reported from various studies.<sup>27</sup> This might be caused partially by the location of the sequons and subtle changes in the structure of the trimeric S protein as well as the conditions under which the recombinant proteins were produced. For example, in terms of the oligomannose content, this report and all previous studies have determined that N234 is occupied predominantly by this type of glycan. This is true no matter the differences in protein source, type of variants,

MS techniques, and data processes used in these studies, presumably due to the relatively buried location of this residue.<sup>13</sup> Using human-cell-expressed ectodomain S protein, many laboratories (including ours in this study) have determined that some other sequons, such as N61, N603, N709, N717, N801, and N1074, are heavily occupied by oligomannose-type glycans.<sup>14,25,26,30,31,39,41</sup> The fact that these residues are in the area between the head and the stalk of an S trimer structure might suggest that steric effects exist in these areas, preventing access of glycosylation enzymes. On the other hand, two recent reports have revealed that only N61 and N234 are almost fully modified by oligomannose glycans on the S variants investigated.<sup>33,37</sup> This might be because all of the proteins in these two studies were obtained from the same manufacturer. The constructs or protein preparation conditions of these samples might also be different from those of other sources, leading to subtle structure alterations in the middle area between the head and stalk of the S trimer.

Previous reports from various laboratories have consistently showed a high degree of similarity in the N-glycosylation profile of human embryonic kidney (HEK)-cell-expressed recombinant ectodomains of the spike proteins from different variants of concern, including Wuhan-hu-1, D614G, Alpha, Beta, Gamma, Delta, Omicron, and spike protein derived from the virus itself.<sup>14,26,31,33,37</sup> This similarity was also observed in our study when comparing the glycan patterns of S-Omicron and S-D614G (Figure 2A and 2B). Additionally, we found comparable relative abundance of fucosylated and sialylated glycans (Figure 2C and 2D), along with the similar distribution of the most abundant glycans (Figure 3), further indicating that the virus might have almost fully optimized the potential of N-glycosylation to evade immune responses.<sup>31</sup>

Despite the overall similarity of the N-glycosylation profiles between S-Omicron and S-D614G, significant differences were observed for some sequons. For example, the content of the complex glycans on N149 decreased from 77% on S-D614G to 57% in S-Omicron, whereas the oligomannose increased from about 7% to 37%, and the major oligomannose glycan of HexNAc2Hex5 displayed approximately 40% occupancy at this site of S-Omicron (Figure 3B). This might be attributed to the variant-specific mutations on S-Omicron near N149, including G142D and del143–145 (Figure 1), which create a steric microenvironment preventing the access of glycan processing enzymes to this residue. A similar trend was observed at the N657 site, where the complex glycans decreased from 87% in S-D614G to 46% in S-Omicron and the level of oligomannose increased by approximately 4-fold. In addition, the two most abundant glycan groups at N657 were different in these two proteins (Figure 3). A substitution in S-Omicron on a nearby residue, H655Y, might contribute to this shift. In contrast, other N-glycosylation sites near variant-specific mutations on S-Omicron did not cause significant change in the processing state of modified N-glycans. For instance, the profile of the glycan types on the sequons of N74, N331, N343, and N801 did not change between the two spike proteins, although some Omicron mutations, including del69–70, G339D, and D796Y, occurred only a few residues away upstream or downstream of these N-glycosylation sites. This suggests that the glycosylation processing on these sites was not altered by any structural changes caused by nearby mutations on the S-Omicron.

An unusual situation was observed for the glycosylation at the N149 site during the MS based analysis of glycopeptides derived from SARS-CoV-2 spike proteins. N149 is a surface



**Figure 3.** Relative abundance (%) of the most abundant (blue bar) and next most abundant (orange bar) glycans at individual N-glycosylation sites of the spike proteins of SARS-CoV-2 D614G (A) and Omicron (B) variants. N, H, F, and A represent N-acetyl hexosamine (HexNAc), hexose (Hex), fucose (Fuc), and N-acetylneuraminic acid (NeuAc) groups, respectively, in glycan compositions.

residue in the NTD of the spikes (Figure 1) and is within the area of antibody binding NTD supersite.<sup>52</sup> Multiple residues representing specific cleavage sites for certain common proteases such as trypsin (K/R), chymotrypsin (F/Y/W/L), and  $\alpha$ -lytic proteases (A/V/S/T) are in the sequences flanking N149 (Figure S2A). This implies that it should be easy to produce detectable N149-containing glycopeptides by various enzymes or protease combinations. However, no high quality data were produced for quantifying the glycan abundance at this site from the analysis of D614G, Alpha, Beta, Gamma and Delta spike variants in our previous study,<sup>37,38</sup> and this phenomenon also has been observed by other laboratories.<sup>30</sup> In this report the sequential digestion by Asp-N and chymotrypsin under selected conditions allowed quantitative characterization of N149 glycosylation on S-Omicron and S-D614G with the production of glycopeptides of DHKNN-(glycans)KSWMESEF and YHKNN(glycans)KSWMESEF, respectively (Figure S2A). On the other hand, from the

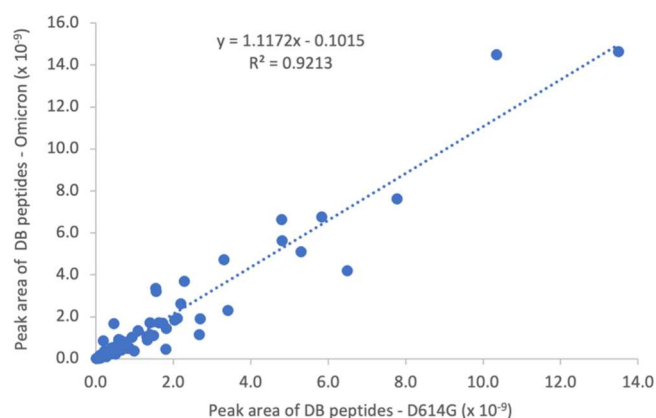
digestions by Lys-C/Lys-C, Lys-C/trypsin, and Lys-C/ $\alpha$ -lytic protease, a comparable number of the N-glycans at this site were determined for S-Omicron. In the Lys-C/Lys-C experiments, the additional fresh Lys-C was added for the second digestion to minimize the potential loss of enzyme activity during the first high-temperature digestion. However, only one glycan was detected from D614G (Table S1), suggesting that mutations of the G142D and del143–145 on S-Omicron (Figure 1) might lead to the formation of more detectable glycopeptides from this protein than from D614G and other variants of concern.

To further understand the role of the N-glycosylation at N149 on the interaction of the spike protein with receptor and antibody, we prepared two spike mutants, S-Omicron-N149Q and S-D614G-N149Q, and evaluated their binding capability to a human ACE2 and a monoclonal antibody (4A8) against the NTD of spike protein. 1D gel analysis of pull-down experiments showed that comparable amounts of four proteins,

either wild type S proteins or N149Q mutants, were able to bind to Fc tagged ACE2 proteins that were immobilized on protein G magnetic beads (Figure S2B, lanes 5, 8, 11, and 14), revealing that the carbohydrate groups at this site were not essential for the binding between S and the receptor proteins. We observed that 4A8-bound S-D614G-N149Q increased by approximately 23% in comparison with S-D614G (Figure S2C, lanes 1 and 2, and Figure S2D), suggesting the potential impairment by this glycan-modified residue on antibody recognition. Because the residues of H146 and Y145 of S protein have direct contact with mAb residues and glycosylated N149 is close to the complex interface revealed by the Cryo-EM structure of the S protein-4A8 complex,<sup>53</sup> increased binding of S-D614G-N149Q to 4A8 suggests that the N-glycans on N149 might participate in the interactions between two molecules. In contrast, no binding of wildtype and mutated S-Omicron proteins to the 4A8 mAb was observed, presumably due to the loss of contact residues because of variant-specific mutations in S-Omicron, including deletion of residues V143 to Y145 and/or substitution of G142D. Further investigation is needed to understand the role of N149 glycosylation in this Omicron variant.

**Analysis of Disulfide Bonds.** Fifteen disulfide bonds formed by 30 cysteine residues in the ectodomain of the S proteins have been visualized by three-dimensional structures and mass spectrometry.<sup>13,40</sup> Although these disulfide bonds are conserved in all variants of concerns, Yao et al. have revealed that some disulfides in the RBD of the Omicron S protein, such as C480–C488 and C379–C432, are susceptible to reduction that could affect binding capacity and stability of the protein.<sup>50</sup> To understand whether mutations of Omicron S affect the formation of disulfide bonds, we examined and compared the disulfide bonds of the S proteins of Omicron with the D614G variant using mass spectrometry analysis with four various enzyme digestion methods.

As depicted in Figure S3, the peptides containing disulfide bonds and free cysteines could be unambiguously detected through HCD-induced fragmentation of protein digests. This approach allowed the detection of all disulfide bonds, except C15–C136, C131–C166, and C617–649, with one or more interpeptide or intrapeptide disulfide-bond-containing peptides for each bond (Table S2). Figure 4 shows that the abundance of disulfide-bond related peptides on the Omicron and D614G S proteins were well correlated. It suggests that these spike proteins possess similar overall disulfide bond structures and that the large number of variant-specific mutations in Omicron S might not lead to a significant alteration to its disulfide bond linkages. Additionally, peptides with free cysteine modified by N-ethylmaleimide (NEM) were detected on most of the cysteine residues (data not shown), indicating that a fraction of each of these residues does not form proper disulfide bonds before sample preparation. This could be because of incomplete formation or partial reduction of these disulfide bonds during protein expression or storage. However, accurate quantification of individual disulfide bonds was limited by relatively high variations for some disulfide bonds represented by multiple peptides (Table S2). This presumably would be caused by more nonspecific cleavage for disulfide-bonded peptides than for linear peptides during enzyme digestion, because four cleavages for each interpeptide disulfide bonded molecule are required and the spatial structure of such dipeptides may also hinder the effective access of proteases. Additionally, disulfide-bond scrambling (the formation of non-



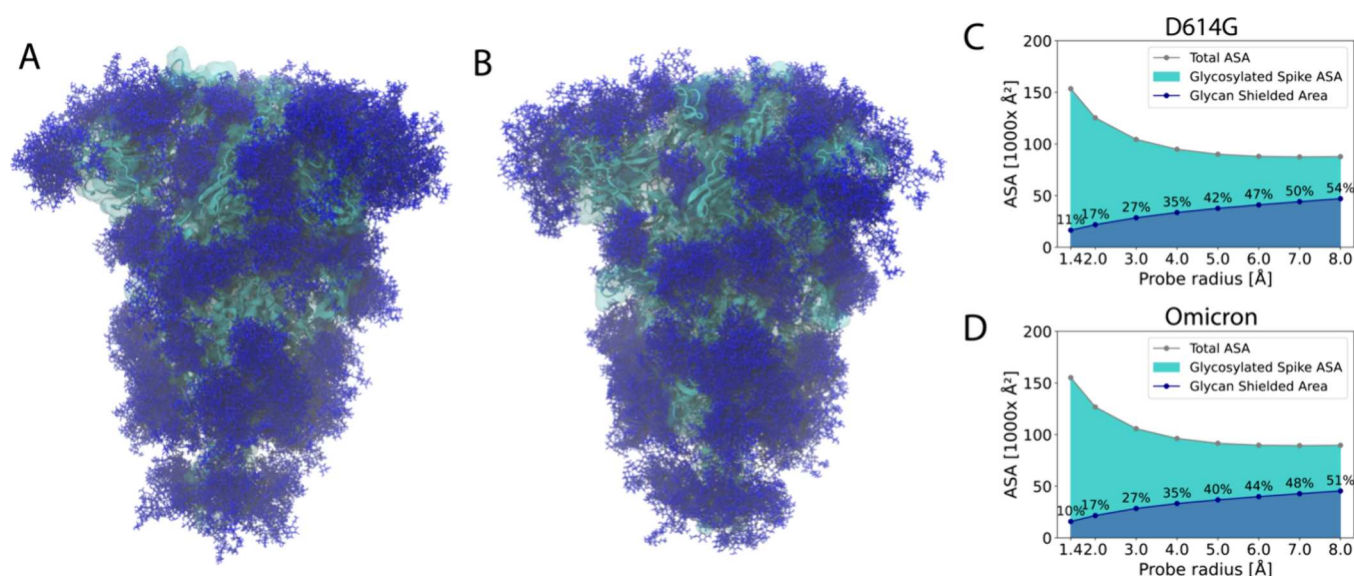
**Figure 4.** Correlation of the peak area of the disulfide-bond-containing peptides detected from SARS-CoV-2 spike proteins of the Omicron and D614G variants. DB represents disulfide bond. Peptides containing disulfide bonds were generated by digesting NEM-treated spike proteins with various enzymes and identified by mass spectrometry as described in the Materials and Methods section. The data was fitted with linear regression.

native disulfide linkages), a common phenomenon that was also observed in this study (data not shown), could obstruct the quantitative analysis of specific disulfide bonds. More optimization of sample preparation and the digestion procedure is needed to generate uniform disulfide-bonded peptides for more accurate disulfide bond quantification.

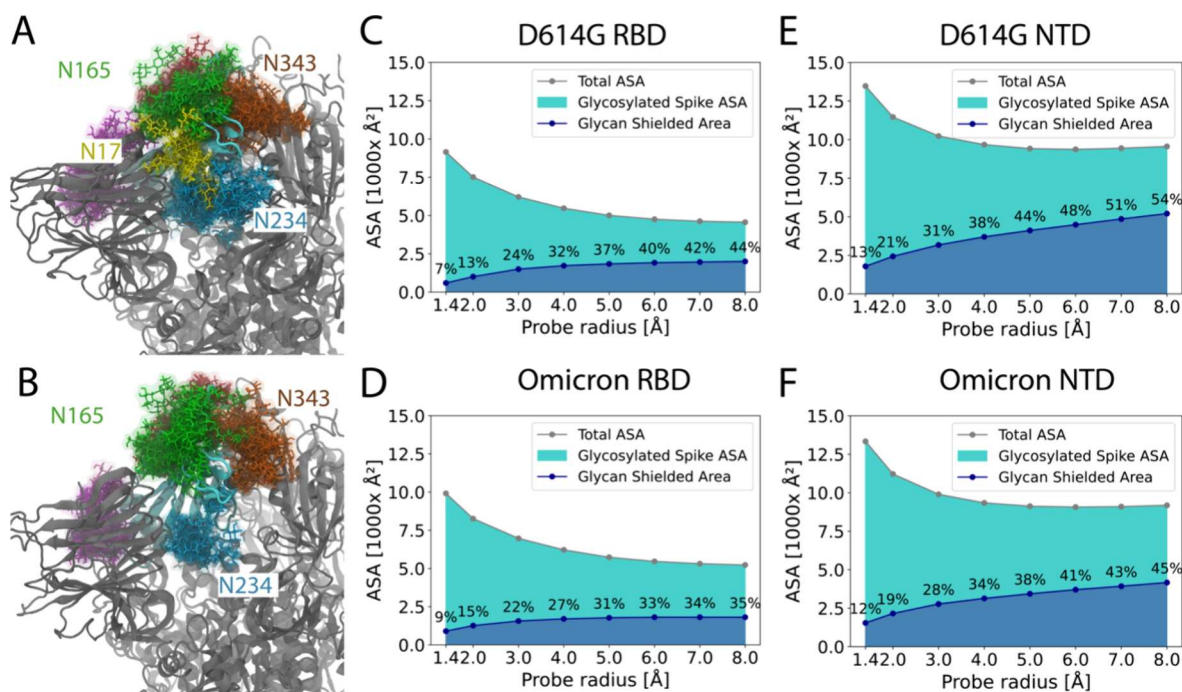
**Molecular Dynamics Simulations.** Glycans, vital components of the S protein, fulfill various key functions, including stabilizing the protein structure and promoting immune evasion.<sup>23,54–57</sup> According to the latest update (June 13, 2023) from the CoV-AbDab (Coronavirus Antibody Database),<sup>58</sup> of the total 12,536 antibody entries, nearly all (12,431) target the S protein, signifying its central role in antibody-based interventions. Glycosylation frequently aids the virus in circumventing the immune system by creating a shield over critical epitopes on the S protein, thereby reducing the antibodies' effectiveness.

To understand how the differences in the glycosylation profiles of the Omicron and D614G variants affect glycan shielding and other roles of glycans, we built all-atom models for both variants' S proteins, incorporating glycans (Figure S4). The selection and construction of specific glycans integrated into these models were guided by glycosylation profile data derived from the mass spectrometry analysis detailed above. Subsequently, we carried out three 1.4- $\mu$ s molecular dynamics simulations for each system; we also simulated the same protein models with no glycans present. This comprehensive approach allows us to resolve the role of glycosylation in viral immune evasion strategies and can help guide the development of more effective antibody-based interventions against different SARS-CoV-2 variants.

Inspired by the work of Casalino et al.,<sup>23</sup> we provide an overall view of the S protein's glycan shield (Figure 5). Figure 5A and 5B show the superimposition of glycans over the course of the simulations, demonstrating the range of potential conformations. Despite the use of smaller intervals in previous studies,<sup>23</sup> this representation still offers a realistic depiction of the glycan shielding. Given that the process of antibody binding takes place over microseconds, the chosen interval of 0.25  $\mu$ s strikes a suitable balance between computational feasibility and accuracy of the model.



**Figure 5.** Glycan shield of SARS-CoV-2 S protein. The D614G (A) and Omicron (B) S proteins are depicted in cyan using a cartoon representation overlaid with a transparent surface. The superimposed glycans are represented as dark blue sticks. These glycan conformations were captured at intervals of  $0.25 \mu\text{s}$  throughout the net  $4.2 \mu\text{s}$  of simulation trajectories ( $3 \times 1.4 \mu\text{s}$ ) for the D614G and Omicron S proteins. The accessible surface area (ASA) of D614G (C) and Omicron (D) S proteins were evaluated using a variety of probe sizes, spanning from 1.4 to 8 Å.



**Figure 6.** Glycan shielding of the receptor binding domain (RBD) and N-terminal domains (NTD) in the SARS-CoV-2 S protein. The RBDs of chain A in the D614G (A) and Omicron (B) S proteins are shown in cyan. Every glycan structure presented is from chain B. All glycan residues within 5 Å of the RBDs are depicted by differently colored stick representations. These glycan structures are superimposed at intervals of  $0.25 \mu\text{s}$  along the respective simulation trajectories. The ASA of RBDs in the D614G (C) and Omicron (D) S proteins was evaluated using a variety of probe sizes, spanning from 1.4 to 8 Å. The ASA of NTDs was also evaluated for both D614G (E) and Omicron (F) S proteins.

To comprehensively estimate the accessibility of S proteins under the glycan shield and identify the differences between the two variants, we performed accessible surface area (ASA) calculations using probe radii ranging from 1.4 to 8 Å. These different probe radii allow us to estimate accessibility for molecules of different dimensions—from smaller ones, such as water, to larger entities, such as antibodies. The 1.4-Å probe radius is typically used to mimic water's accessibility. In

assessing the accessible surface area for antibodies, we employed larger probe radii, reaching 8 Å. The probe radii are proven to be effective when compared with the 5 Å to 10 Å range used in previous research.<sup>20,57,59,60</sup> We observed that even without glycans, larger molecules have less accessibility to the S proteins than smaller ones (Figure 5C,D). Furthermore, the presence of glycans exacerbates this, making it increasingly difficult for larger molecules to access the S proteins. When

comparing the Omicron and D614G S proteins, the Omicron variant exhibits 3% lower shielded area for large (radius of 8 Å) molecules.

Among the various epitopes present on the S protein of SARS-CoV-2, the RBD is notably targeted by the largest number of monoclonal antibodies.<sup>61,62</sup> In fact, according to the CoV-AbDab, out of the 12,431 antibodies that target the S protein, 8,393 are specifically directed toward the RBD.<sup>58</sup> To gain further insight into the epitopes of the RBD, we specifically examined the glycan shield present on the RBD (Figure 6A–D). Intriguingly, our observations reveal that the RBD in chain A is enveloped by numerous glycans originating from chain B (Figure 6A,B). This pattern of glycan shielding is also evident when examining the RBDs in chains B and C, further illuminating the complex interplay of these structures in the virus's immune evasion strategies.

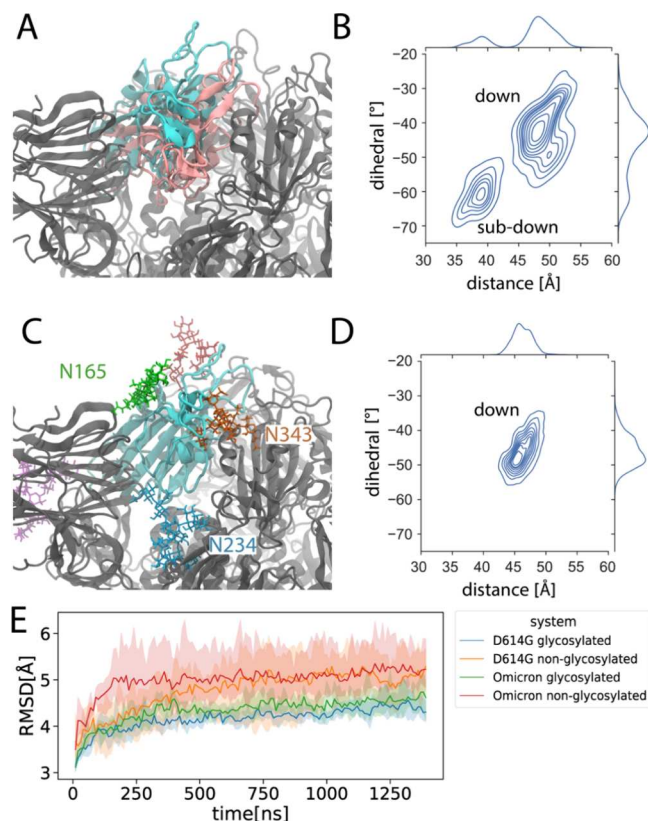
In the case of the D614G variant, the glycan at position N17 from chain B wraps around the RBD, which is not observed in the Omicron variant. The most abundant glycan at position N17 is H2H4F1 in the Omicron and N5H4A1 in D614G. This results in a shorter N17 glycan in the Omicron than in the D614G, which is likely why the N17 glycan wraps around the RBD in D614G but not in the Omicron. The absence of the N17 glycan around the RBD in Omicron results in residues from Ser469 to Val483 having a larger accessible area in this variant (Figures 6A,B, S6). It is noteworthy that the glycosylation site at N17 in the Delta variant is absent, which can be attributed to the T19R mutation. An absence of glycans at the N17 position in the Gamma variant's S protein also has been noted previously.<sup>31,33</sup>

To further identify potential epitopes within the RBD, we analyzed the ASA of each residue, taking into account the presence of glycans (Figure S7). In this analysis, residues ranging from Ser469 to Val483 in the Omicron variant again exhibited a larger ASA. Also noteworthy, the S477N and T478K mutations in the Omicron variant further amplify the ASA of the RBD in its S protein. The combination of a shorter or absent glycan at N17, along with protein mutations, makes this region more exposed to solvent. In a broader perspective, when comparing the RBDs in the Omicron and the Omicron variants, the glycan shield over the RBD is less extensive in the Omicron variant than in D614G, especially for larger molecules (Figure 6C,D). Specifically, for molecules with a radius of 8 Å, the Omicron variant RBD has 35% of the area shielded by glycans, whereas the D614G variant has 44% of its RBD area shielded.

Besides the RBD, another potential epitope is the N-terminal domain (NTD). The NTD in the Omicron variant is less shielded by glycans than is the NTD in the D614G variant (Figure 6). In particular, for molecules with a radius of 8 Å, the Omicron variant has 45% of the area covered by glycans, whereas the D614G variant has 54% of its area shielded by glycans. In the Omicron and D614G variants, the most frequently observed glycans at position N149 are N2H5 and N4H5F1A1, respectively, leading to a shorter N149 glycan in the Omicron and, along with alterations at N17, to a reduction in glycan shielding on its N-terminal domain (NTD). Figure S8 shows the extent of glycan shielding of the NTDs. Figure S9 shows the full-length glycans at positions N17 and N149. We observed that the secondary structure in the Omicron variant, specifically two  $\beta$ -strands and a loop between them spanning from residue 140 to 158, exhibited decreased stability. This

instability might be attributed to the deletion of residues V143, Y144 and Y145, as well as the G142D mutation in this region.

In addition to promoting immune evasion, glycans play a pivotal role in stabilizing the S-protein structure (Figure 7). In



**Figure 7.** Subdown RBD conformation in the Omicron S protein in the absence of glycans. (A) Initial RBD conformation (PDB: 7TNW<sup>63</sup>) and subdown RBD conformation are depicted in cyan and pink, respectively, in the simulations conducted without glycans. (B) A two-dimensional kernel density estimate (KDE) plot visualizes the spread of RBD conformations within the trajectories of the unglycosylated systems, with two collective variables used previously<sup>57</sup> (defined in Methods), a distance and a dihedral angle, characterizing the RBD conformations. (C) The RBD conformation in simulations that incorporate glycans. Glycans surrounding the RBD are depicted in differently colored stick representations. (D) The corresponding KDE plot illustrates the distribution of RBD conformations in the glycosylated simulations. (E) RMSD of D614G and Omicron spike proteins from MD simulations. The graph illustrates RMSD values for both the glycosylated and unglycosylated forms over time, with the shaded areas representing the 95% confidence intervals across three replicate simulations.

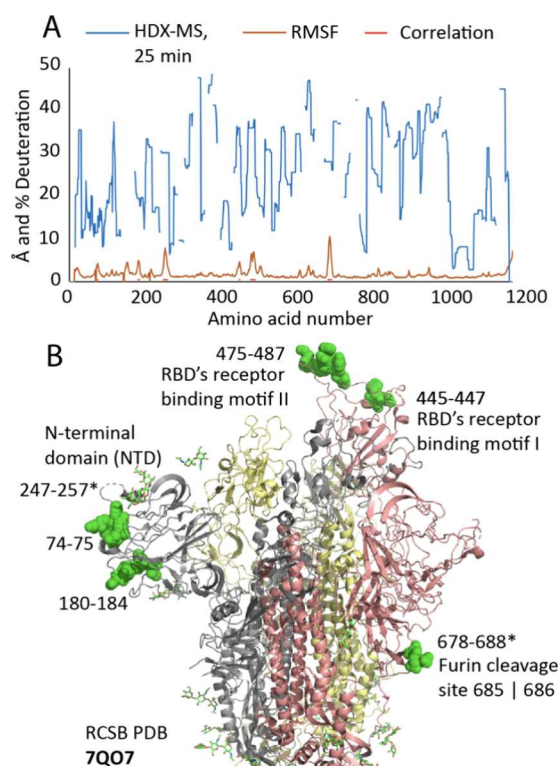
two out of the three 1.4- $\mu$ s simulations of the Omicron S protein without glycans, which started in a down (closed) conformation, we identified a “sub-down” RBD conformation (Figure 7A). This subdown RBD conformation does not manifest in any simulations conducted in the presence of glycans (Figure 7C). To quantify the RBD conformations, we introduced two collective variables, distance and dihedral angle, consistent with definitions used previously.<sup>57</sup> We used two-dimensional KDE plots to visually represent the distribution of these two variables concurrently (Figure 7B,D). To provide context, an “up” or open conformation of the RBD is defined by a distance of 70 Å and a dihedral angle of 0°. In the simulations devoid of glycans, we discerned two

clusters, one representing the “down” or closed RBD conformation and the other depicting the subdown conformation (Figure 7B). Both conformations are shown in Figure 7A. In contrast, the distribution of RBD conformations in the presence of glycans, as seen in all three replicas, is exclusively located within the closed-state cluster. Furthermore, it appears more intense and concentrated around the distribution center, indicating that the presence of glycans stabilizes the RBD conformation. Additionally, we observed glycans N165, N234, and N343 enveloping the RBD (Figure 7C), in agreement with previous observations.<sup>57</sup> This further supports the idea that the glycans surrounding the RBD stabilize it, preventing the emergence of a subdown conformation. However, the subdown RBD conformation is not observed in the unglycosylated D164G S protein (Figure S10). Analysis (Figure 7E) revealed a consistent trend: the glycosylated S proteins of both the D614G and Omicron variants exhibit reduced root mean square deviation (RMSD) values compared to their unglycosylated counterparts, indicating increased structural stability when glycans are present.

**Correlation of MD Simulations and Hydrogen/Deuterium Exchange Mass Spectrometry.** One quantity that can be extracted from MD simulations is the root mean squared fluctuation (RMSF) of each residue, which is an estimate of the residue's flexibility during the simulation. Hydrogen/deuterium exchange mass spectrometry (HDX-MS) measures changes in a protein's deuterated water uptake, which reflects changes in hydrogen bond stability. The implication of both techniques is that changes in the protein's conformational dynamics are being detected, e.g., high RMSF values may correlate with high percentage deuteriation. We tested this hypothesis by comparing RMSF and HDX-MS data per residue for the Omicron variant with an  $x$ - $y$  scatterplot (data not shown) and filtering for RMSF > 4 Å. We note that some regions of high deuteriation did not correspond to high RMSF, possibly due to the disparate time scales explored by the methods (microseconds for MD and minutes for HDX-MS). Nonetheless, six regions satisfied both criteria: residues 74–75, 180–184, 247–257, 445–447, 475–487, and 678–688 (Figure 8A). The first three regions are adjacent to each other in the NTD, the next two regions are the receptor binding motif (RBM) of the RBD, and the last region includes the furin cleavage site between residues 685 and 686. Visualizing these six regions on the Omicron S protein structure (7QO7, Figure 8B) indicates they are all on the surface, and the key roles of the RBMs and furin cleavage site in SARS-CoV-2 infection are well-documented.<sup>12,13</sup> Therefore, the correlation of RMSF in MD simulations and HDX-MS data indicates that the S protein's more dynamic regions also have important biological functions.

## DISCUSSION

In this comprehensive study, we used advanced LC-MS techniques to conduct an analysis of the N-glycosylation and disulfide bond profiles of the recombinant spike ectodomain protein derived from the Omicron variant while using the D614G S protein as a control. To ensure the accuracy and validity of our results, we used the proteins expressed and purified under identical experimental conditions and selected N-glycosylation data only when multiple data sets from distinct digestion experiments showed consistent outcomes. Through these strategies, we determined the distribution and relative



**Figure 8.** (A) Root mean square fluctuations (RMSF) and percentage of hydrogen/deuterium exchange mass spectrometry (HDX-MS) deuteriation percentages for residues in the Omicron S protein. The RMSF, obtained from MD simulations, shows the average level of fluctuations in the protein's structure across different chains and replicas. HDX-MS labeling indicates the rate at which N-amide hydrogen atoms in the protein exchange with deuterium, shown as heat map per-residue deuteriation (blue line, 1500 s, averages of triplicate measurements). Gaps in the HDX-MS sequence coverage are shown as breaks in both lines. Correlation of the two data sets used an  $x$ - $y$  scatterplot (RMSF vs % deuteriation, data not shown) and a cutoff for RMSF > 4 Å. This identified Omicron areas 74–75, 180–184, 247–257, 445–447, 475–487, and 678–688 shown as a red bar. (B) Model of Omicron S glycoprotein trimer (PDB 7QO7) localizing correlations between HDX-MS % deuteriation and MD RMSF. The S protein monomers are gray, pink and yellow, and with green glycans. Correlated areas are colored space-filling green. The gray monomer's NTD includes 74–75, 180–184 and 247–257. The pink monomer's RBD includes 445–447 and 475–487 and its S2 domain includes 678–688. Asterisks on areas 247–257 and 678–688 indicate that part of that sequence is missing from PDB 7QO7.

abundance across all 22 potential N-glycosylation sites of the S-Omicron and S-D614G proteins. We also detected disulfide bonds within the two proteins. Our data reveal that the Omicron S protein aligns closely with the glycan processing patterns of S-D614G at the majority of sequons. Meanwhile, significant differences in glycan types were observed at specific sites, namely, N61, N149, N657 and N1098. In terms of the predominant glycan abundances at various sequons, distinctions were evident at multiple sites, including N17, N149, N331, and N657. Our investigation also detected 12 out of 15 disulfide bonds with close similarity between two proteins by comparing the intensities of disulfide-formed and free sulfur-containing peptides. Moreover, we conducted MD simulations of both protein variants, with and without the inclusion of major glycans, shedding light on the effects of N-glycosylation on the protein structure. We combined advanced mass

spectrometric techniques with MD simulation models of D614G and Omicron S proteins in two ways. First, the results of the bottom-up glycoproteomics analysis of both S proteins were used to select specific glycans at each sequon for MD modeling to provide biologically accurate information. Second, after MD simulations and analysis, the computer-based results were compared to structural analyses from the laboratory, including HDX-MS analysis. Our simulations revealed that S-Omicron showed diminished glycan shielding in comparison to that of S-D614G. The difference was particularly pronounced within the RBD and NTD regions, which encompass potential epitope areas. The comparison between the results of MD simulations and HDX-MS analysis showed that areas identified by both methods as dynamic (in motion), such as the receptor binding domain and furin cleavage site, have well-documented roles in S-protein function. These findings enhance our understanding of the intricate interplay among glycosylation, protein structure, and immune evasion strategies. As the global pursuit of effective vaccination and therapeutic strategies continues, our research contributes vital insights into this ongoing endeavor.

## MATERIALS AND METHODS

**Materials.** All chemicals were obtained from Sigma-Aldrich (St. Louis, MO) except where otherwise indicated. Endoproteases including trypsin, chymotrypsin, Lys-C, and Asp-N were purchased from Promega (Madison, WI), and  $\alpha$ -lytic protease was obtained from New England BioLabs (Ipswich, MA). Phosphate-buffered saline (PBS) tablets, LC-MS grade water, 0.1% formic acid, methanol, and acetonitrile with 0.1% formic acid were from Fisher Scientific (PA). Deuterium oxide (99.9%) was obtained from Cambridge Isotope Laboratories (MA). Urea and tris(carboxyethyl)phosphine (TCEP) were purchased from Sigma-Aldrich (MO). Reagent and sample vials for HDX-MS were purchased from Trajan Scientific and Medical (NC) and Thermo Scientific (CA), respectively. The monoclonal antibody, 4A8 (catalog no. AB0247), was purchased from Absolute Antibody (Shirley, MA).

**Preparation of Recombinant Spike and ACE2 Proteins and Analysis Workflow.** The recombinant ectodomains of the spike proteins of SARS-CoV-2 Omicron and D614G variants as well as the recombinant Fc-tagged ACE2 were prepared in HEK cells under identical expression and purification conditions. In particular, the poly histidine-tagged and six-proline-stabilized ectodomains of the SARS-CoV-2 S proteins including S-Omicron, S-D614G, S-Omicron-N149Q, and S-D614G-N149Q were expressed in human Expi293F cells using the Expi293 Expression System (ThermoFisher Scientific) and purified using a HisTrap FF column (GE Life Sciences)<sup>64</sup> followed by gel filtration on a Superose 6 Increase 10/300 GL column (GE Life Sciences). The sequences of the proteins include the mutated furin cleavage site and six proline substitutions at F817P, A892P, A899P, A942P, K986P, and V987P to stabilize the prefusion conformation of the trimeric spike proteins. Alkylation, reduction, and digestion of the samples were conducted at pH 7.9 for the N-glycosylation analysis. Sample preparation for detection of disulfide bonds was performed at pH 6.5 to reduce the formation of scrambled disulfide bonds (Figure S1). The proteins also were treated with NEM to permanently block sulfhydryls of free cysteine residues to prevent them from oxidation and formation of artificial disulfide bonds. The most abundant N-glycan at each sequon of the two S proteins

obtained from the N-glycosylation analysis was incorporated into the protein models for MD simulations.

**Digestion of Spike Proteins.** The enzymatic digestion of recombinant proteins was conducted as previously reported.<sup>37</sup> In brief, aliquots of 1–2  $\mu$ g full-length ectodomain spike protein were denatured and reduced at 60 °C for 30 min in a solution containing 50 mM ammonium bicarbonate (pH 7.9), 0.05% RapiGest SF surfactant (Waters Corporation) and 5 mM DTT. Samples were alkylated using 15 mM iodoacetamide for 30 min in the dark at room temperature with gentle mixing. Digestions with a single enzyme were conducted at 37 °C overnight at an enzyme-to-protein ratio of 1:10 (w/w). The proteins were digested by two enzymes sequentially with a high temperature–short time condition followed by a regular temperature–overnight digestion to facilitate the detection of glycopeptides.<sup>36–38,65,66</sup> The first digestion was performed at 52 °C for 60 min with the first protease (1:3 w/w). The second digestion was conducted at 37 °C overnight at an enzyme-to-substrate ratio of 1:15 (w/w). The proteolytic reactions were quenched, and the RapiGest was precipitated by adding 5% trifluoroacetic acid to decrease the pH to below 3. The mixture was then incubated at 37 °C for 30 min. The solutions were centrifuged at 2000g for 10 min, and the supernatants (20  $\mu$ L) were transferred into new sample vials. Each digestion was performed in three replicates.

For the characterization of disulfide bonds, proteins were alkylated to block free sulfhydryls by 1 mM Pierce N-ethylmaleimide (ThermoFisher Scientific) in alkylation buffer containing 8 M urea and 100 mM Tris (pH 6.5) at 37 °C for 2 h. DTT reduction was not used in order to preserve the disulfide bonds in the proteins. The pretreated proteins were then precipitated by adding chilled acetone to the reaction solutions, and the pellets were resuspended with 100 mM ammonium citrate (pH 6.5), followed by sequential digestion.

**Mass Spectrometry Analysis.** LC-MS analysis of protein digests was performed using an Orbitrap Eclipse Tribrid mass spectrometer coupled to an UltiMate3000 RSLCnano chromatography system (Thermo Scientific) as described previously.<sup>37</sup> The peptides were injected into an integrated separation column/nanospray device (Thermo Scientific EASY-Spray PepMap RSLC C<sub>18</sub>, 75  $\mu$ m i.d.  $\times$  15 cm length, 3  $\mu$ m 100 Å particles), coupled to an EASY-Spray ion source. Mobile phase A (0.1% formic acid) and mobile phase B (0.1% formic acid in 80% acetonitrile) were mixed based on the following gradient with the flow rate of 300 nL/min: 4% B for 8 min; 4–10% B for 2 min; 10–35% B for 33 min; 35–60% B for 2 min; and 60–95% B in 1 min. The spray voltage was set to 1.8 kV, and the temperatures of the integrated column/nanospray device and the ion transfer tube were set at 55 and 275 °C, respectively.

MS data acquisition was accomplished in positive ion mode using a signature ion triggered electron-transfer/higher-energy collision dissociation (ET<sub>h</sub>CD) method. The full MS precursor scans were acquired by the Orbitrap at a resolution of 120,000 at  $m/z$  200, from  $m/z$  375–2000 with the automatic gain control (AGC) target setting as “standard” and the maximum injection time as “auto”. After the MS1 survey scan, a data-dependent MS2 scan was acquired over a 3-s cycle time using high-energy collision dissociation (HCD) at a resolution of 30,000, a mass range of  $m/z$  120–2000, and a normalized collision energy (NCE) of 28%. Signature ions representing glycan oxonium fragments were used to trigger electron-transfer dissociation (ETD) fragmentation. If one of three

common glycan signature ions— $m/z$  204.0867 (HexNAc), 138.0545 (HexNAc fragment), or 366.1396 (HexNAcHex)—was detected in the HCD spectrum within a 15 ppm mass accuracy, an additional precursor isolation and EThcD acquisition were performed. Settings for that included a resolution of 50,000 at  $m/z$  200, with the range of  $m/z$  150–2000, 500% normalized AGC target, 150 ms maximum injection time, and 35% supplemental activation NCE.

**Data Analysis.** MS/MS data for the analysis of glycopeptides were processed using the PMi-Byonic (version 3.7) node within Proteome Discover (ThermoFisher Scientific). Data were searched using the Protein Metrics 182 human N-glycan library (included in the Byonic program) for potential glycan modifications. The search parameters for enzyme digestion were set to semispecific, three allowed missed cleavage sites, and 6 and 20 ppm mass tolerance for precursors and fragment ions, respectively. Carbamidomethylation of cysteine was set as a fixed modification with variable modifications set to include deamidation at Asn and Gln and oxidation of Met. MS2 spectra of identified glycopeptides with Byonic scores higher than 150 were considered valid identifications. Identified glycopeptide and unoccupied peptide abundances were determined using precursor ion peak intensity with normalization on the total peptide amount per file. Relative abundance of each type of glycan at each site was calculated as the normalized peak intensity ratio of the peptides bearing a particular glycan type over the sum of total glycopeptide intensity. The glycan abundance was represented as the mean of three replicates along with the standard deviation of the mean. Data for disulfide bonded peptides were processed using PMi-Byos (Protein Metrics). The search parameters for enzyme digestion of disulfide bonded peptides in Byos were set to fully specific, four allowed missed cleavage sites, and 6 and 15 ppm mass tolerance for precursors and fragment ions, respectively. Oxidation of Met and Trp, deamidation of Asn, and thioether formation by derivatization with NEM on Cys were considered variable modifications. N-glycans were not included in the disulfide search due to the increased complexity of data generated. Peptide quantitation was performed using the Byologic module with fully specific digestion at specific sites corresponding to each digestion enzyme combination.

**Binding Assay and 1D Gel Electrophoresis.** The binding of S proteins to human ACE2 and the monoclonal antibody (mAb) was carried out on magnetic beads. The Fc-tagged ACE2 or mAb was first immobilized on Protein G Dynabeads (ThermoFisher Scientific). The S protein solutions were then incubated with an on-bead receptor or antibody at room temperature for 60 min. Bound S protein/ACE2 or S protein/mAb complexes were analyzed on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) on a NuPAGE Novex Bis-Tris gel following manufacturer's instructions (Invitrogen). Protein solutions were mixed with 4× sample buffer and deionized water (1:3 v/v) and heated at 80 °C for 10 min. The supernatants were loaded on a 4–12% gradient gel, and the gel was run in MOPS buffer at 200 V for 45 min. The gels were stained with the Invitrogen SYPRO Ruby protein gel stain (ThermoFisher Scientific).

**Hydrogen/Deuterium Exchange Mass Spectrometry.** A DHR-PAL system (Trajan) was used for sample preparation, with sample tray at 20 °C, quench tray at 4 °C, valve chamber and prechiller at 4 °C, and digestion chamber at 8 °C. Purified spike ectodomain (0.4  $\mu\text{g}/\mu\text{L}$ ) was mixed 1:5 (v/v) with PBS

prepared using water (equilibration buffer, measured pH 7.22) or deuterated water (labeling buffer, measured pH 7.56). Labeling times were 0, 60, 240, and 960 s with 3 technical replicates at each time. Samples were quenched with an equal volume of 2 M urea 0.5 M TCEP pH 2.5 and held at 4 °C for 2 min. An UltiMate3000 UPLC system (binary nano pump and loading pump, Thermo Scientific) was used for subsequent online sample handling. Automated valve switching passed the quenched sample over a  $2.1 \times 20$  mm Nepenthesin-2/Pepsin mixed digestion column (AffiPro, CZ) at 100  $\mu\text{L}/\text{min}$  water 0.1% formic acid for 2 min, trapping the resulting peptides on a  $2.1 \times 5$  mm Fully Porous C18 guard column (Phenomenex, CA), then desalted peptides at 300  $\mu\text{L}/\text{min}$  for 4 min. Peptides were eluted and resolved by a gradient from 13 to 65% mobile phase B (95:5:0.1 Acetonitrile/water/formic acid) over 23 min on a  $1 \times 100$  mm Luna Omega 1.6  $\mu\text{m}$  100 Å C18 column (Phenomenex, CA). A Tribrid Eclipse Orbitrap (OT) mass spectrometer (Thermo Scientific) with HESI-2 electrospray ion source and high-flow needle was operated in positive ion mode to detect peptides for 31 min. In all samples, precursor scans of resolution 120 K (at  $m/z$  200) in the range 375–2000  $m/z$  were acquired. For each MS1 scan, the top-10 abundant precursor features were selected for data-dependent MS2 scans, selecting for an intensity threshold of 30K counts, monoisotopic peptide precursors, charge states  $2^+$  to  $8^+$  with an isolation window of 1.2  $m/z$ , and not repeating precursor ions more than twice within 15 s. Precursors were fragmented with HCD at 28% normalized collision energy (NCE) and centroid scanned in the OT with standard automatic gain control (AGC) target, automatic injection time, scan range 120–2000  $m/z$ , and resolution 30K. If at least one of three selected oxonium ions was detected (HexNAc 204.0867, HexNAc fragment 138.0545, or HexNAcHex 366.1396) with 15 ppm mass tolerance, then EThcD OT-MS2 scans were acquired of that precursor ion. Supplemental HCD was at 20% NCE with profile scans from 150 to 2000  $m/z$  at a resolution of 50K using custom AGC target (500%) and fill time (90 ms). At the end of the analytical gradient, solvent transitioned to 90% mobile phase B for 6 min, and halfway through that time the analytical and trapping columns were put into back-flow washing mode by automated valve switching. After re-equilibration of the analytical column at 13% mobile phase B the injection cycle ended at 45 min. As recommended,<sup>67</sup> the entire batch of control and heat denatured samples (all time points and technical replicates) was randomized for HDX-MS acquisition to minimize batch effects on interpreted differences in protein state, labeling time, and replicates.

**HDX-MS Data Processing.** Protein Metrics Inc. (CA) Byos HDX 4.6–37 searched the 3 data files from equilibration buffer samples (0 s labeling time, 3 technical replicates) to identify (glyco)peptides using MS2 spectra. The search database included spike and both proteases. Both the HCD and EThcD tandem mass spectra contributed to peptide spectral matching. Putative (glyco)peptides were then searched in data files from all samples at the MS level (and appropriate retention times) to identify both unlabeled and deuterated peptides and visualize their isotopic envelopes. Initial spike results (1271 peptides) were narrowed by 1) default software filters (MS2 score >15, minimum alt\_rank\_score/primary\_rank\_score >0.99, maximum precursor  $m/z$  error  $\pm 40$  ppm, maximum retention time deviation  $\pm 5$  min) leaving 1056 peptides, and 2) removing peptides with MS2 score <150 leaving 106 of 205 glycopeptides, removing peptides with more

than  $\pm 10\%$  average, maximum, or minimum “deuteration” in 0 s samples, and removing peptides causing standard deviations  $>10\%$  at any labeled time-point, leaving 561 peptides. Additional manual curation involved adjustment of the extracted ion chromatogram (XIC) window used to integrate MS data and generate an isotopic envelope, optimizing the intensity and specificity of that envelope. Peptides with inadequate intensity XICs to estimate deuteration were discarded.

**Molecular Dynamics Simulations.** We selected S-protein structures for simulation by aiming for those at the highest resolution and with the fewest artificial mutations. The wild type (WT) Omicron S protein was initialized from the structure in the RCSB Protein Data Bank (PDB) ID: 7TNW.<sup>63</sup> Missing residues in 7TNW from 245 to 247 and 255 to 259 were copied from PDB ID: 7WK2.<sup>68</sup> Residue 141 was mutated to leucine according to the sequence of the Omicron S protein used in this study. On the basis of its local electrostatic environment, His625 was set as HSD (histidine delta; protonated on its  $\delta$  nitrogen); all other histidines were set as HSE (histidine epsilon; protonated on the  $\epsilon$  nitrogen). Using the WT S protein as a starting point, we also built a common laboratory version of the Omicron S protein. Here, residues 682–685 were mutated to the sequence GSAS, and residues 817, 892, 899, 942, 986, and 987 were mutated to proline. The lab version of D614G S protein was initialized from the PDB ID: 7KRQ.<sup>69</sup> Based on its local electrostatic environment, His49 was set as HSD and all other His were set as HSE. Residues 817, 892, 899, 942, 986, and 987 were mutated to proline.

Disulfide bonds between residues Cys15 and Cys136, Cys131 and Cys166, Cys291 and Cys301, Cys336 and Cys361, Cys379 and Cys432, Cys391 and Cys525, Cys480 and Cys488, Cys538 and Cys590, Cys617 and Cys649, Cys662 and Cys671, Cys738 and Cys760, Cys743 and Cys749, Cys840 and Cys851, Cys1032 and Cys1043, and Cys1082 and Cys1126 were added in all systems.

Glycosylation sites are located on N17, N61, N74, N122, N149, N165, N234, N282, N331, N343, N603, N616, N657, N709, N717, N801, N1074, N1098, N1134, and N1158 in all glycosylated systems. Overall, 20 N-linked glycans are present in each protomer, resulting in a total of 60 glycans for one S-protein trimer model. The glycan at each site with the highest population in the mass spectroscopy data (Figure 3) was added to the site. Among the possible glycan structures with the same constitution according to mass spectroscopy, we selected the most populous one with the highest “hit score” from the GlyGen Web site (<https://www.glygen.org/glycan-search>).<sup>70</sup> The glycan 3D structures are generated by the GLYCAM Web server developed by the Woods group (<http://glycam.org>).<sup>71,72</sup> Table S1 and Figure S11 show the glycans.

Missing hydrogen atoms were added to all systems, after which they were solvated in  $(220 \text{ \AA})^3$  and  $(215 \text{ \AA})^3$  water boxes, with and without glycans, respectively. We added sodium ( $\text{Na}^+$ ) and Chloride ( $\text{Cl}^-$ ) ions to achieve a salt concentration of 150 mM. We used Visual Molecular Dynamics (VMD)<sup>73</sup> to construct all protein systems. In total, we built six distinct systems: WT Omicron S protein with and without glycans, laboratory-engineered Omicron S protein with and without glycans, and laboratory-engineered D614G S protein with and without glycans. The final system sizes were  $\sim 1$  million atoms.

Simulations were conducted in NAMD<sup>74</sup> using the CHARMM36m protein force field,<sup>75</sup> CHARMM36 glycan force field,<sup>76</sup> and TIP3P water model.<sup>77</sup> We used hydrogen mass repartitioning (HMR) along with a uniform 4 fs time step.<sup>78,79</sup> The simulations were run at a constant temperature (310 K) and constant isotropic pressure (1 atm), maintained by a Langevin thermostat and piston,<sup>80</sup> respectively. Long-range electrostatics were calculated at every time step using the particle-mesh Ewald method.<sup>81</sup> We set a short-range cutoff for Lennard-Jones interactions at 12 Å with a switching function starting at 10 Å. We added extra bonds to hold the three  $\alpha$ -helix structures (residues 1141 to 1162) together, imitating how they are held together by the virus membrane envelope, which was not modeled here. We started with restraining protein backbones while equilibrating side chains and glycans for 1 ns. After the initial equilibration process, each of the six protein systems was simulated for 1.4  $\mu\text{s}$  with three replicates per system, resulting in an aggregated total of 25.2  $\mu\text{s}$  of simulation data.

The accessible surface area (ASA) was quantified using the “measure sasa” command in VMD.<sup>73</sup> In line with a previous study,<sup>23</sup> the ASA of the protein with and without glycans was measured separately. Subsequently, the glycan-shielded area was calculated by subtracting ASA with glycans from ASA without glycans. The ASA was evaluated every 30 ns throughout the simulation trajectories, and subsequently, the average values were computed.

Kernel density estimation (KDE) plots were constructed using the Seaborn library<sup>82</sup> in Python. For determining the position of the RBD with respect to the spike, two collective variables are defined as follows: a distance is measured between the centers of mass for RBD-A (336–518) and SD1-B (531–592), and a dihedral angle is measured using the centers of mass from the domains RBD-A (336–518), SD1-A (531–592), SD2-A (593–677), and NTD-A (27–307).

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The data sets generated and analyzed in the scope of this study are available from the corresponding authors upon request.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsinfecdis.4c00015>.

Additional tables of N-glycosylation and disulfide-bond analysis, figures of MS/MS spectra, glycans used in simulations, and MD simulation analysis, including glycan shielding of different areas, accessible surface area, and RBD conformations (PDF)

### Accession Codes

Raw LC-MS/MS data have been deposited on the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository server. The accession ID is MSV000094147.

## ■ AUTHOR INFORMATION

### Corresponding Authors

Dongxia Wang – National Center for Environmental Health, Division of Laboratory Sciences, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30322, United States; [orcid.org/0000-0002-0993-1018](https://orcid.org/0000-0002-0993-1018); Phone: 770-488-0446; Email: [dov2@cdc.gov](mailto:dov2@cdc.gov); Fax: 770-488-0509

James C. Gumbart – School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;

orcid.org/0000-0002-1510-7842; Phone: 404-385-0797;  
Email: gumbart@physics.gatech.edu

## Authors

**Zijian Zhang** – School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

**Jakub Baudys** – National Center for Environmental Health, Division of Laboratory Sciences, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30322, United States

**Christopher Haynes** – National Center for Environmental Health, Division of Laboratory Sciences, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30322, United States; orcid.org/0000-0002-5590-1784

**Sarah H. Osman** – National Center for Environmental Health, Division of Laboratory Sciences, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30322, United States

**Bin Zhou** – National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30322, United States

**John R. Barr** – National Center for Environmental Health, Division of Laboratory Sciences, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30322, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsinfectdis.4c00015>

## Author Contributions

<sup>1</sup>(D.W., Z.Z.) These authors made equal contributions to this work.

## Notes

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade names and commercial sources is for identification only and does not constitute endorsement by the U.S. Department of Health and Human Services, or the U.S. Centers for Disease Control and Prevention (CDC).

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

J.C.G. acknowledges support from the CDC under contract 75D30123P17466. Computational resources were provided through the Extreme Science and Engineering Discovery Environment (ACCESS; TG-MCB130173), which is supported by the National Science Foundation (NSF; #2138259, #2138286, #2138307, #2137603, and #2138296). This work also used the Hive cluster, which is supported by the NSF (1828187) and is managed by the Partnership for an Advanced Computing Environment (PACE) at GT.

## REFERENCES

- (1) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, 395 (10223), 497–506.
- (2) Zhou, P.; Yang, X. L.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.;

Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, 579 (7798), 270–273.

(3) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, 579 (7798), 265–269.

(4) Konings, F.; Perkins, M. D.; Kuhn, J. H.; Pallen, M. J.; Alm, E. J.; Archer, B. N.; Barakat, A.; Bedford, T.; Bhiman, J. N.; Caly, L.; Carter, L. L.; Cullinane, A.; de Oliveira, T.; Druce, J.; El Masry, I.; Evans, R.; Gao, G. F.; Gorbalenya, A. E.; Hamblion, E.; Herring, B. L.; Hodcroft, E.; Holmes, E. C.; Kakkar, M.; Khare, S.; Koopmans, M. P. G.; Korber, B.; Leite, J.; MacCannell, D.; Marklewitz, M.; Maurer-Stroh, S.; Rico, J. A. M.; Munster, V. J.; Neher, R.; Munnink, B. O.; Pavlin, B. I.; Peiris, M.; Poon, L.; Pybus, O.; Rambaut, A.; Resende, P.; Subissi, L.; Thiel, V.; Tong, S.; van der Werf, S.; von Gottberg, A.; Ziebuhr, J.; Van Kerkhove, M. D. SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. *Nat. Microbiol.* **2021**, 6 (7), 821–823.

(5) Harvey, W. T.; Carabelli, A. M.; Jackson, B.; Gupta, R. K.; Thomson, E. C.; Harrison, E. M.; Ludden, C.; Reeve, R.; Rambaut, A.; COVID-19 Genomics UK (COG-UK) Consortium; Peacock, S. J.; Robertson, D. L. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, 19 (7), 409–424.

(6) Tao, K.; Tzou, P. L.; Nouhin, J.; Gupta, R. K.; de Oliveira, T.; Kosakovsky Pond, S. L.; Fera, D.; Shafer, R. W. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* **2021**, 22 (12), 757–773.

(7) Viana, R.; Moyo, S.; Amoako, D. G.; Tegally, H.; Scheepers, C.; Althaus, C. L.; Anyaneji, U. J.; Bester, P. A.; Boni, M. F.; Chand, M.; Choga, W. T.; Colquhoun, R.; Davids, M.; Deforche, K.; Doolabh, D.; du Plessis, L.; Engelbrecht, S.; Everatt, J.; Giandhari, J.; Giovanetti, M.; Hardie, D.; Hill, V.; Hsiao, N. Y.; Iranzadeh, A.; Ismail, A.; Joseph, C.; Joseph, R.; Koopile, L.; Kosakovsky Pond, S. L.; Kraemer, M. U. G.; Kuate-Lere, L.; Laguda-Akingba, O.; Lesetedi-Mafoko, O.; Lessells, R. J.; Lockman, S.; Lucaci, A. G.; Maharaj, A.; Mahlangu, B.; Maponga, T.; Mahlakwane, K.; Makatini, Z.; Marais, G.; Maruapula, D.; Masupu, K.; Matshaba, M.; Mayaphi, S.; Mbhele, N.; Mbulawa, M. B.; Mendes, A.; Mlisana, K.; Mnguni, A.; Mohale, T.; Moir, M.; Moruisi, K.; Mosepele, M.; Motsatsi, G.; Motswaledi, M. S.; Mphoyakgosi, T.; Msomi, N.; Mwangi, P. N.; Naidoo, Y.; Ntuli, N.; Nyaga, M.; Olubayo, L.; Pillay, S.; Radibe, B.; Ramphal, Y.; Ramphal, U.; San, J. E.; Scott, L.; Shapiro, R.; Singh, L.; Smith-Lawrence, P.; Stevens, W.; Strydom, A.; Subramoney, K.; Tebeila, N.; Tshiabula, D.; Tsui, J.; van Wyk, S.; Weaver, S.; Wibmer, C. K.; Wilkinson, E.; Wolter, N.; Zarebski, A. E.; Zuze, B.; Goedhals, D.; Preiser, W.; Treurnicht, F.; Venter, M.; Williamson, C.; Pybus, O. G.; Bhiman, J.; Glass, A.; Martin, D. P.; Rambaut, A.; Gaseitsiwe, S.; von Gottberg, A.; de Oliveira, T. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **2022**, 603 (7902), 679–686.

(8) Tegally, H.; Moir, M.; Everatt, J.; Giovanetti, M.; Scheepers, C.; Wilkinson, E.; Subramoney, K.; Makatini, Z.; Moyo, S.; Amoako, D. G.; Baxter, C.; Althaus, C. L.; Anyaneji, U. J.; Kekana, D.; Viana, R.; Giandhari, J.; Lessells, R. J.; Maponga, T.; Maruapula, D.; Choga, W.; Matshaba, M.; Mbulawa, M. B.; et al. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat. Med.* **2022**, 28 (9), 1785–1790.

(9) Baker, J. M.; Nakayama, J. Y.; O'Hegarty, M.; McGowan, A.; Teran, R. A.; Bart, S. M.; Mosack, K.; Roberts, N.; Campos, B.; Paegle, A.; McGee, J.; Herrera, R.; English, K.; Barrios, C.; Davis, A.; Roloff, C.; Sosa, L. E.; Brockmeyer, J.; Page, L.; Bauer, A.; Weiner, J. J.; Khubbar, M.; Bhattacharyya, S.; Kirking, H. L.; Tate, J. E. SARS-CoV-2 B.1.1.529 (Omicron) Variant Transmission Within Households - Four U.S. Jurisdictions, November 2021-February 2022. *Morb. Mortal. Wkly. Rep.* **2022**, 71 (9), 341–346.

(10) Lewnard, J. A.; Hong, V. X.; Patel, M. M.; Kahn, R.; Lipsitch, M.; Tartof, S. Y. Clinical outcomes associated with SARS-CoV-2

Omicron (B.1.1.529) variant and BA.1/BA.1.1 or BA.2 subvariant infection in Southern California. *Nat. Med.* **2022**, *28* (9), 1933–1943.

(11) Letko, M.; Marzi, A.; Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **2020**, *5* (4), 562–569.

(12) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veelsler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *183* (6), 1735.

(13) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260–1263.

(14) Yao, H.; Song, Y.; Chen, Y.; Wu, N.; Xu, J.; Sun, C.; Zhang, J.; Weng, T.; Zhang, Z.; Wu, Z.; Cheng, L.; Shi, D.; Lu, X.; Lei, J.; Crispin, M.; Shi, Y.; Li, L.; Li, S. Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **2020**, *183* (3), 730–738.e13.

(15) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367* (6485), 1444–1448.

(16) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Kruger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N. H.; Nitsche, A.; Muller, M. A.; Drosten, C.; Pohlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271–280.e8.

(17) Piccoli, L.; Park, Y. J.; Tortorici, M. A.; Czudnochowski, N.; Walls, A. C.; Beltramello, M.; Silacci-Fregni, C.; Pinto, D.; Rosen, L. E.; Bowen, J. E.; Acton, O. J.; Jaconi, S.; Guarino, B.; Minola, A.; Zatta, F.; Sprugasci, N.; Bassi, J.; Peter, A.; De Marco, A.; Nix, J. C.; Mele, F.; Jovic, S.; Rodriguez, B. F.; Gupta, S. V.; Jin, F.; Piumatti, G.; Lo Presti, G.; Pellanda, A. F.; Biggiogero, M.; Tarkowski, M.; Pizzuto, M. S.; Camerini, E.; Havenar-Daughton, C.; Smithey, M.; Hong, D.; Lepori, V.; Albanese, E.; Ceschi, A.; Bernasconi, E.; Elzi, L.; Ferrari, P.; Garzoni, C.; Riva, A.; Snell, G.; Sallusto, F.; Fink, K.; Virgin, H. W.; Lanzavecchia, A.; Corti, D.; Veelsler, D. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **2020**, *183* (4), 1024–1042.e21.

(18) Varki, A. Biological roles of glycans. *Glycobiology* **2017**, *27* (1), 3–49.

(19) Huang, H. Y.; Liao, H. Y.; Chen, X.; Wang, S. W.; Cheng, C. W.; Shahed-Al-Mahmud, M.; Liu, Y. M.; Mohapatra, A.; Chen, T. H.; Lo, J. M.; Wu, Y. M.; Ma, H. H.; Chang, Y. H.; Tsai, H. Y.; Chou, Y. C.; Hsueh, Y. P.; Tsai, C. Y.; Huang, P. Y.; Chang, S. Y.; Chao, T. L.; Kao, H. C.; Tsai, Y. M.; Chen, Y. H.; Wu, C. Y.; Jan, J. T.; Cheng, T. R.; Lin, K. I.; Ma, C.; Wong, C. H. Vaccination with SARS-CoV-2 spike protein lacking glycan shields elicits enhanced protective responses in animal models. *Sci. Transl. Med.* **2022**, *14* (639), No. eabm0899.

(20) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **2020**, *10* (1), 14991.

(21) Zhao, X.; Chen, H.; Wang, H. Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Front. Mol. Biosci.* **2021**, *8*, No. 629873.

(22) Allen, J. D.; Ivory, D. P.; Song, S. G.; He, W. T.; Capozzola, T.; Yong, P.; Burton, D. R.; Andrabi, R.; Crispin, M. The diversity of the glycan shield of sarbecoviruses related to SARS-CoV-2. *Cell. Rep.* **2023**, *42* (4), No. 112307.

(23) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6* (10), 1722–1734.

(24) Maity, S.; Acharya, A. Many Roles of Carbohydrates: A Computational Spotlight on the Coronavirus S Protein Binding. *ACS Appl. Bio. Mater.* **2024**, *7*, 646–656.

(25) Allen, J. D.; Chawla, H.; Samsudin, F.; Zuzic, L.; Shivgan, A. T.; Watanabe, Y.; He, W. T.; Callaghan, S.; Song, G.; Yong, P.; Brouwer, P. J. M.; Song, Y.; Cai, Y.; Duyvesteyn, H. M. E.; Malinauskas, T.;

Kint, J.; Pino, P.; Wurm, M. J.; Frank, M.; Chen, B.; Stuart, D. I.; Sanders, R. W.; Andrabi, R.; Burton, D. R.; Li, S.; Bond, P. J.; Crispin, M. Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. *Biochemistry* **2021**, *60* (27), 2153–2169.

(26) Baboo, S.; Diedrich, J. K.; Torres, J. L.; Copps, J.; Singh, B.; Garrett, P. T.; Ward, A. B.; Paulson, J. C.; Yates, J. R., Evolving spike-protein N-glycosylation in SARS-CoV-2 variants. *bioRxiv*, May 9, 2023. DOI: 10.1101/2023.05.08.539897.

(27) Campos, D.; Girgis, M.; Sanda, M. Site-specific glycosylation of SARS-CoV-2: Big challenges in mass spectrometry analysis. *Proteomics* **2022**, *22* (15–16), No. e2100322.

(28) Chawla, H.; Fadda, E.; Crispin, M. Principles of SARS-CoV-2 glycosylation. *Curr. Opin. Struct. Biol.* **2022**, *75*, No. 102402.

(29) Cipollo, J. F.; Parsons, L. M. Glycomics and glycoproteomics of viruses: Mass spectrometry applications and insights toward structure-function relationships. *Mass Spectrom. Rev.* **2020**, *39* (4), 371–409.

(30) Kuo, C. W.; Yang, T. J.; Chien, Y. C.; Yu, P. Y.; Hsu, S. D.; Khoo, K. H. Distinct shifts in site-specific glycosylation pattern of SARS-CoV-2 spike proteins associated with arising mutations in the D614G and Alpha variants. *Glycobiology* **2022**, *32* (1), 60–72.

(31) Newby, M. L.; Fogarty, C. A.; Allen, J. D.; Butler, J.; Fadda, E.; Crispin, M. Variations within the Glycan Shield of SARS-CoV-2 Impact Viral Spike Dynamics. *J. Mol. Biol.* **2023**, *435* (4), No. 167928.

(32) Sanda, M.; Morrison, L.; Goldman, R. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal. Chem.* **2021**, *93* (4), 2003–2009.

(33) Shajahan, A.; Pepi, L. E.; Kumar, B.; Murray, N. B.; Azadi, P. Site specific N- and O-glycosylation mapping of the spike proteins of SARS-CoV-2 variants of concern. *Sci. Rep.* **2023**, *13* (1), 10053.

(34) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **2020**, *30* (12), 981–988.

(35) Tian, Y.; Parsons, L. M.; Jankowska, E.; Cipollo, J. F. Site-Specific Glycosylation Patterns of the SARS-CoV-2 Spike Protein Derived From Recombinant Protein and Viral WA1 and D614G Strains. *Front. Chem.* **2021**, *9*, No. 767448.

(36) Wang, D.; Baudys, J.; Bundy, J. L.; Solano, M.; Keppel, T.; Barr, J. R. Comprehensive Analysis of the Glycan Complement of SARS-CoV-2 Spike Proteins Using Signature Ions-Triggered Electron-Transfer/Higher-Energy Collisional Dissociation (ETHCD) Mass Spectrometry. *Anal. Chem.* **2020**, *92* (21), 14730–14739.

(37) Wang, D.; Baudys, J.; Osman, S. H.; Barr, J. R. Analysis of the N-glycosylation profiles of the spike proteins from the Alpha, Beta, Gamma, and Delta variants of SARS-CoV-2. *Anal. Bioanal. Chem.* **2023**, *415* (19), 4779–4793.

(38) Wang, D.; Zhou, B.; Keppel, T. R.; Solano, M.; Baudys, J.; Goldstein, J.; Finn, M. G.; Fan, X.; Chapman, A. P.; Bundy, J. L.; Woolfitt, A. R.; Osman, S. H.; Pirkle, J. L.; Wentworth, D. E.; Barr, J. R. N-glycosylation profiles of the SARS-CoV-2 spike D614G mutant and its ancestral protein characterized by advanced mass spectrometry. *Sci. Rep.* **2021**, *11* (1), 23561.

(39) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369* (6501), 330–333.

(40) Zhang, S.; Go, E. P.; Ding, H.; Anang, S.; Kappes, J. C.; Desaire, H.; Sodroski, J. G. Analysis of Glycosylation and Disulfide Bonding of Wild-Type SARS-CoV-2 Spike Glycoprotein. *J. Virol.* **2022**, *96* (3), No. e0162621.

(41) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L. Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **2020**, *28* (4), 586.

(42) Kuo, C.-W.; Chang, N.-E.; Yu, P.-Y.; Yang, T.-J.; Hsu, S.-T. D.; Khoo, K.-H. An N-glycopeptide MS/MS data analysis workflow leveraging two complementary glycoproteomic software tools for

more confident identification and assignments. *Proteomics* **2023**, 23 (20), No. 2300143.

(43) Sevier, C. S.; Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002**, 3 (11), 836–847.

(44) Shi, Y.; Zeida, A.; Edwards, C. E.; Mallory, M. L.; Sastre, S.; Machado, M. R.; Pickles, R. J.; Fu, L.; Liu, K.; Yang, J.; Baric, R. S.; Boucher, R. C.; Radi, R.; Carroll, K. S. Thiol-based chemical probes exhibit antiviral activity against SARS-CoV-2 via allosteric disulfide disruption in the spike glycoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, 119 (6), No. e2120419119.

(45) Khanna, K.; Raymond, W. W.; Jin, J.; Charbit, A. R.; Gitlin, I.; Tang, M.; Werts, A. D.; Barrett, E. G.; Cox, J. M.; Birch, S. M.; Martinelli, R.; Sperber, H. S.; Franz, S.; Duff, T.; Hoffmann, M.; Healy, A. M.; Oscarson, S.; Pohlmann, S.; Pillai, S. K.; Simmons, G.; Fahy, J. V. Exploring antiviral and anti-inflammatory effects of thiol drugs in COVID-19. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2022**, 323 (3), L372–L389.

(46) Grishin, A. M.; Dolgova, N. V.; Landreth, S.; Fiset, O.; Pickering, I. J.; George, G. N.; Falzarano, D.; Cygler, M. Disulfide Bonds Play a Critical Role in the Structure and Function of the Receptor-binding Domain of the SARS-CoV-2 Spike Antigen. *J. Mol. Biol.* **2022**, 434 (2), No. 167357.

(47) Zhang, X.; Li, Z.; Zhang, Y.; Liu, Y.; Wang, J.; Liu, B.; Chen, Q.; Wang, Q.; Fu, L.; Wang, P.; Zhong, X.; Jin, L.; Yan, Q.; Chen, L.; He, J.; Zhao, J.; Xiong, X. Disulfide stabilization reveals conserved dynamic features between SARS-CoV-1 and SARS-CoV-2 spikes. *Life Sci. Alliance* **2023**, 6 (9), No. e202201796.

(48) Qu, K.; Chen, Q.; Ciazynska, K. A.; Liu, B.; Zhang, X.; Wang, J.; He, Y.; Guan, J.; He, J.; Liu, T.; Zhang, X.; Carter, A. P.; Xiong, X.; Briggs, J. A. G. Engineered disulfide reveals structural dynamics of locked SARS-CoV-2 spike. *PLoS Pathog.* **2022**, 18 (7), No. e1010583.

(49) Shen, H.; Wu, Z. Effect of Disulfide Bridge on the Binding of SARS-CoV-2 Fusion Peptide to Cell Membrane: A Coarse-Grained Study. *ACS Omega* **2022**, 7 (41), 36762–36775.

(50) Yao, Z.; Geng, B.; Marcon, E.; Pu, S.; Tang, H.; Merluza, J.; Bello, A.; Snider, J.; Lu, P.; Wood, H.; Staglar, I. Omicron Spike Protein is Vulnerable to Reduction. *J. Mol. Biol.* **2023**, 435 (13), No. 168128.

(51) Miller, R. M.; Perkins, G. L.; Bush, D.; Tartiere, A.; DeGraan-Weber, N. Glycopeptide characterization of Sf9-derived SARS-CoV-2 spike protein recombinant vaccine candidates expedited by the use of glycopeptide libraries. *Rapid Commun. Mass Spectrom.* **2023**, 37 (5), No. e9452.

(52) Cerutti, G.; Guo, Y.; Zhou, T.; Gorman, J.; Lee, M.; Rapp, M.; Reddem, E. R.; Yu, J.; Bahna, F.; Bimela, J.; Huang, Y.; Katsamba, P. S.; Liu, L.; Nair, M. S.; Rawi, R.; Olia, A. S.; Wang, P.; Zhang, B.; Chuang, G. Y.; Ho, D. D.; Sheng, Z.; Kwong, P. D.; Shapiro, L. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **2021**, 29 (5), 819.

(53) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; Chen, Z.; Guo, Y.; Zhang, J.; Li, Y.; Song, X.; Chen, Y.; Xia, L.; Fu, L.; Hou, L.; Xu, J.; Yu, C.; Li, J.; Zhou, Q.; Chen, W. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **2020**, 369 (6504), 650–655.

(54) Lin, W.-S.; Chen, I. C.; Chen, H.-C.; Lee, Y.-C.; Wu, S.-C. Glycan Masking of Epitopes in the NTD and RBD of the Spike Protein Elicits Broadly Neutralizing Antibodies Against SARS-CoV-2 Variants. *Front. Immunol.* **2021**, 12, No. 795741.

(55) Tate, M.; Job, E.; Deng, Y.-M.; Gunalan, V.; Maurer-Stroh, S.; Reading, P. Playing Hide and Seek: How Glycosylation of the Influenza Virus Hemagglutinin Can Modulate the Immune Response to Infection. *Viruses* **2014**, 6 (3), 1294–1316.

(56) Li, Q.; Wu, J.; Nie, J.; Zhang, L.; Hao, H.; Liu, S.; Zhao, C.; Zhang, Q.; Liu, H.; Nie, L.; Qin, H.; Wang, M.; Lu, Q.; Li, X.; Sun, Q.; Liu, J.; Zhang, L.; Li, X.; Huang, W.; Wang, Y. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020**, 182 (5), 1284.

(57) Pang, Y. T.; Acharya, A.; Lynch, D. L.; Pavlova, A.; Gumbart, J. C. SARS-CoV-2 spike opening dynamics and energetics reveal the individual roles of glycans and their collective impact. *Commun. Biol.* **2022**, 5 (1), 1170.

(58) Raybould, M. I. J.; Kovaltsuk, A.; Marks, C.; Deane, C. M. CoV-AbDab: the coronavirus antibody database. *Bioinformatics* **2021**, 37 (5), 734–735.

(59) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; Hurley, M. F. D.; Harbison, A. M.; Fogarty, C. A.; Coffland, J. E.; Fadda, E.; Voelz, V. A.; Chodera, J. D.; Bowman, G. R. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* **2021**, 13 (7), 651–659.

(60) Urbanowicz, R. A.; Wang, R.; Schiel, J. E.; Keck, Z.-y.; Kerzic, M. C.; Lau, P.; Rangarajan, S.; Garagusi, K. J.; Tan, L.; Guest, J. D.; Ball, J. K.; Pierce, B. G.; Mariuzza, R. A.; Fong, S. K. H.; Fuerst, T. R. Antigenicity and Immunogenicity of Differentially Glycosylated Hepatitis C Virus E2 Envelope Proteins Expressed in Mammalian and Insect Cells. *J. Virol.* **2019**, 93 (7), e01403–18.

(61) Corti, D.; Purcell, L. A.; Snell, G.; Veessler, D. Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell* **2021**, 184 (12), 3086–3108.

(62) Valdes-Balbin, Y.; Santana-Mederos, D.; Paquet, F.; Fernandez, S.; Climent, Y.; Chiodo, F.; Rodríguez, L.; Sanchez Ramirez, B.; Leon, K.; Hernandez, T.; Castellanos-Serra, L.; Garrido, R.; Chen, G.-W.; Garcia-Rivera, D.; Rivera, D. G.; Verez-Bencomo, V. Molecular Aspects Concerning the Use of the SARS-CoV-2 Receptor Binding Domain as a Target for Preventive Vaccines. *ACS Cent. Sci.* **2021**, 7 (5), 757–767.

(63) Zhang, J.; Cai, Y.; Lavine, C. L.; Peng, H.; Zhu, H.; Anand, K.; Tong, P.; Gautam, A.; Mayer, M. L.; Rits-Volloch, S.; Wang, S.; Sliz, P.; Wesemann, D. R.; Yang, W.; Seaman, M. S.; Lu, J.; Xiao, T.; Chen, B. Structural and functional impact by SARS-CoV-2 Omicron spike mutations. *Cell Rep.* **2022**, 39 (4), No. 110729.

(64) Zhou, B.; Thi Nhu Thao, T.; Hoffmann, D.; Taddeo, A.; Ebert, N.; Labrousse, F.; Pohlmann, A.; King, J.; Steiner, S.; Kelly, J. N.; Portmann, J.; Halwe, N. J.; Ulrich, L.; Trueb, B. S.; Fan, X.; Hoffmann, B.; Wang, L.; Thomann, L.; Lin, X.; Stalder, H.; Pozzi, B.; de Brot, S.; Jiang, N.; Cui, D.; Hossain, J.; Wilson, M.; Keller, M.; Stark, T. J.; Barnes, J. R.; Dijkman, R.; Jores, J.; Benarafa, C.; Wentworth, D. E.; Thiel, V.; Beer, M. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* **2021**, 592, 122–127.

(65) Wang, D.; Baudys, J.; Rees, J.; Marshall, K. M.; Kalb, S. R.; Parks, B. A.; Nowaczyk, L.; Pirkle, J. L.; Barr, J. R. Subtyping Botulinum Neurotoxins by Sequential Multiple Endoproteases In-Gel Digestion Coupled with Mass Spectrometry. *Anal. Chem.* **2012**, 84 (11), 4652–4658.

(66) Capelo, J. L.; Carreira, R.; Diniz, M.; Fernandes, L.; Galesio, M.; Lodeiro, C.; Santos, H. M.; Vale, G. Overview on modern approaches to speed up protein identification workflows relying on enzymatic cleavage and mass spectrometry-based techniques. *Anal. Chim. Acta* **2009**, 650 (2), 151–159.

(67) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Iacob, R. E.; Jorgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrop, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **2019**, 16 (7), 595–602.

(68) Hong, Q.; Han, W.; Li, J.; Xu, S.; Wang, Y.; Xu, C.; Li, Z.; Wang, Y.; Zhang, C.; Huang, Z.; Cong, Y. Molecular basis of receptor binding and antibody neutralization of Omicron. *Nature* **2022**, 604 (7906), 546–552.

- (69) Zhang, J.; Cai, Y.; Xiao, T.; Lu, J.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rits-Volloch, S.; Zhu, H.; Woosley, A. N.; Yang, W.; Sliz, P.; Chen, B. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science*. **2021**, 372 (6541), 525–530.
- (70) York, W. S.; Mazumder, R.; Ranzinger, R.; Edwards, N.; Kahsay, R.; Aoki-Kinoshita, K. F.; Campbell, M. P.; Cummings, R. D.; Feizi, T.; Martin, M.; Natale, D. A.; Packer, N. H.; Woods, R. J.; Agarwal, G.; Arpinar, S.; Bhat, S.; Blake, J.; Castro, L. J. G.; Fochtman, B.; Gildersleeve, J.; et al. GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology*. **2020**, 30 (2), 72–73.
- (71) Thieker, D. F.; Hadden, J. A.; Schulten, K.; Woods, R. J. 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology*. **2016**, 26 (8), 786–787.
- (72) Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegthart, J. F. G.; Lütke, T.; Perez, S. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*. **2015**, 25 (12), 1323–1324.
- (73) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, 14 (1), 33–38.
- (74) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, 26 (16), 1781–1802.
- (75) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, 14 (1), 71–73.
- (76) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, 5 (9), 2353–2370.
- (77) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79 (2), 926–935.
- (78) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, 11 (4), 1864–1874.
- (79) Balusek, C.; Hwang, H.; Lau, C. H.; Lundquist, K.; Hazel, A.; Pavlova, A.; Lynch, D. L.; Reggio, P. H.; Wang, Y.; Gumbart, J. C. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2019**, 15 (8), 4673–4686.
- (80) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, 103 (11), 4613–4621.
- (81) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98 (12), 10089–10092.
- (82) Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **2021**, 6 (60), 3021.