Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor

Peng Zhao[1,#], Jeremy L. Praissman[1,#], Oliver C. Grant[1,#], Yongfei Cai[2], Tianshu Xiao[2], Katelyn E. Rosenbalm[1], Kazuhiro Aoki[1], Benjamin P. Kellman[3], Robert Bridger[1], Dan H. Barouch[4], Melinda A. Brindley[5], Nathan E. Lewis[3,6], Michael Tiemeyer[1], Bing Chen[2], Robert J. Woods[1], and Lance Wells[1]*

[1] Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, and Department of Chemistry, University of Georgia, Athens, Georgia, 30602, USA

[2] Division of Molecular Medicine, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, 02115, USA

[3] Departments of Pediatrics and Bioengineering, University of California, San Diego, La Jolla, California, 92093, USA

[4] Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, 02215, USA

[5] Department of Infectious Diseases, Department of Population Health, Center for Vaccines and Immunology, College of Veterinary Medicine, University of Georgia, Athens, GA 30602, USA

[6] Novo Nordisk Foundation Center for Biosustainability at UC San Diego, La Jolla, California, 92093, USA

[#] Authors contributed equally

* To whom correspondence should be addressed: Lance Wells, lwells@ccrc.uga.edu

SUMMARY

The current COVID-19 pandemic is caused by the SARS-CoV-2 betacoronavirus, which utilizes its highly glycosylated trimeric Spike protein to bind to the cell surface receptor ACE2 glycoprotein and facilitate host cell entry. We utilized glycomics-informed glycoproteomics to characterize site-specific microheterogeneity of glycosylation for a recombinant trimer spike mimetic immunogen and for a soluble version of human ACE2. We combined this information with bioinformatic analyses of natural variants and with existing 3D-structures of both glycoproteins to generate molecular dynamic simulations of each glycoprotein alone and interacting with one another. Our results highlight roles for glycans in sterically masking polypeptide epitopes and directly modulating Spike-ACE2 interactions. Furthermore, our results illustrate the impact of viral evolution and divergence on Spike glycosylation, as well as the influence of natural variants on ACE2 receptor glycosylation that, taken together, can facilitate immunogen design to achieve antibody neutralization and inform therapeutic strategies to inhibit viral infection.

Keywords: SARS-CoV-2, COVID-19, spike protein, coronavirus, ACE2, glycoprotein, glycosylation, mass spectrometry, molecular dynamics, 3D-modeling

INTRODUCTION

The SARS-CoV-2 coronavirus, a positive-sense single-stranded RNA virus, is responsible for the severe acute respiratory syndrome referred to as COVID-19 that was first reported in China in December of 2019 (1). In approximately six months, this betacoronavirus has spread globally with more than 8 million people testing positive worldwide resulting in greater than 470,000 deaths as of June 22nd, 2020 (https://coronavirus.jhu.edu/map.html). The SARS-CoV-2 coronavirus is highly similar (nearly 80% identical at the genomic level) to SARS-CoV-1, which was responsible for the severe acute respiratory syndrome outbreak that began in 2002 (2,3). Furthermore, human SARS-CoV-2 at the whole genome level is >95% identical to a bat coronavirus (RaTG13), the natural reservoir host for multiple coronaviruses (1,4,5). Given the rapid appearance and spread of this virus, there is no current validated vaccine or SARS-CoV-2-specific targeting therapy clinically approved although statins, heparin, and steroids look promising for lowering fatality rates and antivirals likely reduce the duration of symptomatic disease presentation (6-12).

SARS-CoV-2, like SARS-CoV-1, utilizes the host angiotensin converting enzyme II (ACE2) for binding and entry into host cells (13,14). Like many viruses, SARS-CoV-2 utilizes a spike glycoprotein trimer for recognition and binding to the host cell entry receptor and for membrane fusion (15). Given the importance of viral spike proteins for targeting and entry into host cells along with their location on the viral surface, spike proteins are often used as immunogens for vaccines to generate neutralizing antibodies and frequently targeted for inhibition by small molecules that might block host receptor binding and/or membrane fusion (15,16). In similar fashion, wildtype or catalytically-impaired ACE2 has also been investigated as a potential therapeutic biologic that might interfere with the infection cycle of coronaviruses (17,18). Thus, a detailed understanding of SARS-CoV-2 Spike binding to ACE2 is critical for elucidating mechanisms of viral binding and entry, as well as for undertaking the rational design of effective therapeutics.

The SARS-CoV-2 Spike glycoprotein consists of two subunits, a receptor binding subunit (S1) and a membrane fusion subunit (S2) (1,2). The Spike glycoprotein assembles into stable

homotrimers that together possess 66 canonical sequons for N-linked glycosylation (N-X-S/T, where X is any amino acid except P) as well as a number of potential O-linked glycosylation sites (19,20). Interestingly, coronaviruses virions bud into the lumen of the endoplasmic reticulum-Golgi intermediate compartment, ERGIC, raising unanswered questions regarding the precise mechanisms by which viral surface glycoproteins are processed as they traverse the secretory pathway (21,22). Nonetheless, it has been proposed that the virus acquires a glycan coat sufficient and similar enough to endogenous host protein glycosylation that it serves as a glycan shield, facilitating immune evasion by masking non-self viral peptides with self-glycans (15,20-22). In parallel to their potential masking functions, glycan-dependent epitopes can elicit specific, even neutralizing, antibody responses, as has been described for HIV-1 (15,23-26). Thus, understanding the glycosylation of the viral spike trimer is fundamental for the development of efficacious vaccines, neutralizing antibodies, and therapeutic inhibitors of infection.

ACE2 is an integral membrane metalloproteinase that regulates the renin-angiotensin system (27). Both SARS-CoV-1 and SARS-CoV-2 have co-opted ACE2 to function as the receptor by which these viruses attach and fuse with host cells (13,14). ACE2 is cleavable by ADAM proteases at the cell surface (28), resulting in the shedding of a soluble ectodomain which can be detected in apical secretions of various epithelial layers (gastric, airway, etc.) and in serum (29). The N-terminal extracellular domain of ACE2 contains 6 canonical sequons for N-linked glycosylation and several potential O-linked sites. Several nonsynonymous single-nucleotide polymorphisms (SNPs) in the ACE2 gene have been identified in the human population and could potentially alter ACE2 glycosylation and/or affinity of the receptor for the viral spike protein (30). Given that glycosylation can affect the half-life of circulating glycoproteins in addition to modulating the affinity of their interactions with receptors and immune/inflammatory signaling pathways (31,32), understanding the impact of glycosylation of ACE2 with respect to its binding of SARS-CoV-2 spike glycoprotein is of high importance. The proposed use of soluble extracellular domains of ACE2 as decoy, competitive inhibitors for SARS-CoV-2 infection

emphasizes the critical need for understanding the glycosylation profile of ACE2 so that optimally active biologics can be produced (17,18).

To accomplish the task of characterizing site-specific glycosylation of the trimer Spike of SARS-CoV-2 and the host receptor ACE2, we began by expressing and purifying a stabilized, soluble trimer Spike glycoprotein mimetic immunogen (S, (33)) and a soluble version of the ACE2 glycoprotein from a human cell line. We utilized multiple mass spectrometry-based approaches, including glycomic and glycoproteomic approaches, to determine occupancy and site-specific heterogeneity of N-linked glycans. We also identified sites of O-linked glycosylation and the heterogeneity of the O-linked glycans on S and ACE2. We leveraged this rich dataset, along with existing 3D-structures of both glycoproteins, to generate static and molecular dynamic models of S alone, and in complex with the glycosylated, soluble ACE 2 receptor. By combining bioinformatic characterization of viral evolution and variants of the Spike and ACE2 with molecular dynamic simulations of the glycosylated S-ACE2 interaction, we identified important roles for glycans in multiple processes, including receptor-viral binding and glycan-shielding of the Spike. Our rich characterization of the recombinant, glycosylated spike trimer mimetic immunogen of SARS-CoV-2 in complex with the soluble human ACE2 receptor provides a detailed platform for guiding rational vaccine, antibody, and inhibitor design.

RESULTS

Expression, Purification, and Characterization of SARS-CoV-2 Spike Glycoprotein Trimer and Soluble Human ACE2

A trimer-stabilized, soluble variant of the SARS-CoV-2 S protein (S) that contains 22 canonical N-linked glycosylation sequons and a soluble version of human ACE2 that contains 6 canonical N-linked glycosylation sequons (**Fig. 1A**) were purified from the media of transfected HEK293 cells and the quaternary structure confirmed by negative EM staining for the S trimer (**Fig. 1B**) and purity examined by SDS-PAGE Coomassie G-250 stained gels for both (**Fig. 1C**). In addition, proteolytic digestions followed by proteomic analyses confirmed that the proteins were highly

purified (data not shown). Finally, the N-terminus of both the mature S and the soluble mature ACE2 were empirically determined via proteolytic digestions and LC-MS/MS analyses. These results confirmed that both the secreted, mature forms of S protein and ACE2 begin with an N-terminal glutamine that has undergone condensation to form pyroglutamine at residue 14 and 18, respectively (**Figs. 1D and S1**). The N-terminal peptide observed for S also contains a glycan at Asn-0017 (**Fig. 1D**) and mass spectrometry analysis of non-reducing proteolytic digestions confirmed that Cys-0015 of S is in a disulfide linkage with Cys-0136 (**Fig. S2, Supplemental Table, Tab 2**). Given that SignalP (34) predicts signal sequence cleavage between Cys-0015 and Val-0016 but we observed cleavage between Ser-0013 and Gln-0014, we examined the possibility that an in-frame upstream Methionine to the proposed start Methionine might be used to initiate translation (**Fig. S3**). If one examines the predicted signal sequence cleavage using the in-frame Met that is encoded 9 amino acids upstream, SignalP now predicts cleavage between the Ser and Gln that we observed in our studies (**Fig. S3**). To examine whether this impacted S expression, we expressed constructs that contained or did not contain the upstream 27 nucleotides in a pseudovirus (VSV) system expressing SARS-CoV2 S (**Fig. S4**) and in our HEK293 system (data not shown). Both expression systems produced a similar amount of S regardless of which expression construct was utilized (**Fig. S4**). Thus, while the translation initiation start site has still not been fully defined, allowing for earlier translation in expression construct design did not have a significant impact on the generation of S.

Glycomics Informed Glycoproteomics Reveals Site-Specific Microheterogeneity of SARS-CoV-2 S Glycosylation

We utilized multiple approaches to examine glycosylation of the SARS-CoV-2 S trimer. First, the portfolio of glycans linked to SARS-CoV-2 S trimer immunogen was analyzed following their release from the polypeptide backbone. N-glycans were released from protein by treatment with PNGase F and O-glycans were subsequently released by beta-elimination. Following permethylation to enhance detection sensitivity and structural characterization, released glycans were analyzed by multi-stage mass spectrometry (MS^n) (35,36). Mass spectra were processed

by GRITS Toolbox and the resulting annotations were validated manually (37). Glycan assignments were grouped by type and by additional structural features for relative quantification of profile characteristics (**Fig. 2A**, **Supplemental Table, Tab 3**). This analysis quantified 49 N-glycans and revealed that 55% were of the complex type, 17% were of the hybrid type, and 28% were high-mannose. Among the complex and hybrid N-glycans, we observed a high degree of core fucosylation and significant abundance of bisected and LacDiNAc structures. We also detected 15 O-glycans released from the S trimer (**Fig. S5, Supplemental Table, Tab 4**).

To determine occupancy of N-linked glycans at each site, we employed a sequential deglycoslyation approach using Endoglycosidase H and PNGase F in the presence of ^{18}O-H_2O following tryptic digestion of S (26,38). Following LC-MS/MS analyses, the resulting data confirmed that 19 of the canonical sequons had occupancies greater than 95% (**Supplemental Table, Tab 5**). One canonical sequence, N0149, had insufficient spectral counts for quantification by this method but subsequent analyses described below suggested high occupancy. The 2 most C-terminal N-linked sites, N1173 and N1194, had reduced occupancy, 52% and 82% respectively. Reduced occupancy at these sites may reflect hindered en-bloc transfer by the oligosaccharyltransferase (OST) due to primary amino acid sequences at or near the N-linked sequon. Alternatively, this may reflect these two sites being post-translationally modified after release of the protein by the ribosome by a less efficient STT3B-containing OST, either due to activity or initial folding of the polypeptide, as opposed to co-translationally modified by the STT3A-containing OST (39). None of the non-canoncial sequons (3 N-X-C sites and 4 N-G-L/I/V sites, (40)) showed significant occupancy (>5%) except for N0501 that showed moderate (19%) conversion to ^{18}O-Asp that could be due to deamidation that is facilitated by glycine at the +1 position (**Supplemental Table, Tab 5,** (41)). Further analysis of this site (see below) by direct glycopeptide analyses allowed us to determine that N0501 undergoes deamidation but is not glycosylated. Thus, all, and only the, 22 canonical sequences for N-linked glycosylation (N-X-S/T) are utilized with only N1173 and N1194 demonstrating occupancies below 95%.

Next, we applied 3 different proteolytic digestion strategies to the SARS-CoV-2 S immunogen to maximize glycopeptide coverage by subsequent LC-MS/MS analyses. Extended gradient nanoflow reverse-phase LC-MS/MS was carried out on a ThermoFisher Lumos™ Tribrid™ instrument using Step-HCD fragmentation on each of the samples (see STAR methods for details, (23,24,26,38,42)). Following data analyses using pGlyco 2.2.2 (43), Byonic (44), and manual validation of glycan compositions against our released glycomics findings (**Fig. 2A**, **Supplemental Table, Tab 3**), we were able to determine the microheterogeneity at each of the 22 canonical sites (**Fig. 2B-2E, S6, Supplemental Table, Tab 6**). Notably, none of the non-canonical consensus sequences, including N0501, displayed any quantifiable glycans. The N-glycosites N0074 (**Fig. 2B**) and N0149 (**Fig. 2C**) are highly processed and display a typical mammalian N-glycan profile. N0149 is, however, modified with several hybrid N-glycan structures while N0074 is not. N0234 (**Fig. 2D**) and N0801 (**Fig. 2E**) have N-glycan profiles more similar to those found on other viruses such as HIV (15) that are dominated by high-mannose structures. N0234 (**Fig. 2D**) displays an abundance of Man7 - Man9 high-mannose structures suggesting stalled processing by early acting ER and cis-Golgi mannosidases. In contrast, N0801 (**Fig. 2E**) is processed more efficiently to Man5 high-mannose and hybrid structures suggesting that access to the glycan at this site by MGAT1 and α-Mannosidase II is hindered. In general, for all 22 sites (**Fig. 2B-2E, S6, Supplemental Table, Tab 6**), we observed under processing of complex glycan antennae (i.e. under-galactosylation and under-sialyation) and a high degree of core fucosylation in agreement with released glycan analyses (**Fig. 2A**, **Supplemental Table, Tab 3**). Based on the assignments and the spectral counts for each topology, we were able to determine the percent of total N-linked glycan types (high-mannose, hybrid, or complex) present at each site (**Figure 3, Supplemental Table, Tab 7**). Notably, 3 of the sites (N0234, N0709, and N0717) displayed more than 50% high-mannose glycans while 11 other sites (N0017, N0074, N0149, N0165, N0282, N0331, N0657, N1134, N1158, N1173, and N1194) were more than 90% complex when occupied. The other 8 sites were distributed between these 2 extremes. Notably, only 1 site (N0717 at 45%), which also had greater than 50% high-mannose (55%), had greater than 33%

hybrid structures. To further evaluate the heterogeneity, we grouped all the topologies into the 20 classes recently described by the Crispin laboratory and referred to here as the Oxford classification (**Fig. S7, Supplemental Table, Tab 8,** (19)). Among other features observed, this classification allowed us to observe that while most sites with high mannose structures were dominated by the Man5GlcNAc2 structure, N0234 and N0717 were dominated by the higher Man structures of Man8GlcNAc2 and Man7GlcNAc2, respectively (**Fig. S7, Supplemental Table, Tab 8**). Limited processing at N0234 is in agreement with a recent report suggesting that high mannose structures at this site help to stabilize the receptor-binding domain of S (www.biorxiv.org/content/10.1101/2020.06.11.146522v1). Furthermore, applying the Oxford classifications to our dataset clearly demonstrates that the 3 most C-terminal sites (N1158, N1173, and N1194), dominated by complex type glycans, were more often further processed (i.e. multiple antennae) and elaborated (i.e. galactosylation and sialyation) than other sites (**Fig. S7, Supplemental Table, Tab 8**).

We also analyzed our generated mass spectrometry data for the presence of O-linked glycans based on our glycomic findings (**Fig. S5, Supplemental Table, Tab 4**) and a recent manuscript suggesting significant levels of O-glycosylation of S1 and S2 when expressed independently (45). We were able to confirm sites of O-glycan modification with microheterogeneity observed for the vast majority of these sites (**Supplemental Table, Tab 9**). However, occupancy at each site, determined by spectral counts, was observed to be very low (below 4%) except for Thr0323 that had a modest 11% occupancy (**Fig. S8, Supplemental Table, Tab 10**).

3D Structural Modeling of Glycosylated SARS-CoV-2 Trimer Immunogen Enables Predictions of Epitope Accessibility and Other Key Features

A 3D structure of the S trimer was generated using a homology model of the S trimer described previously (based on PDB code 6VSB, (46). Onto this 3D structure, we installed specific glycans at each glycosylated sequon based on three separate sets of criteria, thereby generating three different glycoform models for comparison (**Supplemental Table, Tab 1**). We denote the first of these criteria sets as “Abundance,” under which the most abundant glycan structure detected by

glycomics, that matched the most abundant glycan composition detected by glycoproteomics at each individual site, was selected for modeling. The second set of criteria, which we refer to as "Oxford Class" (**Fig. S7, Supplemental Table, Tabs 1,8),** assigned to individual sites the most abundant glycan structure detected by glycomics, that was consistent with the most highly represented Oxford classification group detected at each site by glycoproteomics. The third set of criteria was designated as "Processed" and assigned the most highly trimmed, elaborated, or terminally decorated structure detected by glycomics, that was present at $\geq 1/3^{rd}$ of the abundance of the most highly abundant composition detected by glycoproteomics at each site (**Supplemental Table, Tab 1**). The glycoforms assembled by these three criteria were then subjected to multiple all atom MD simulations with explicit water. Information from analyses of these structures is presented in **Figure 4A** along with the sequence of the SARS-CoV-2 S protomer. We also determined variants in S that are emerging in the virus that have been sequenced to date (**Supplemental Table, Tab 11**). The inter-residue distances were measured between the most α-carbon-distal atoms of the N-glycan sites and Spike glycoprotein population variant sites in 3D space (**Figure 4B**). Notable from this analysis, there are several variants that don't ablate the N-linked sequon, but that are sufficiently close in 3-dimensional space to N-glycosites, such as D138H, H655Y, S939F, and L1203F, to warrant further investigation.

The percentage of simulation time that each S protein residue is accessible to a probe that approximates the size of an antibody variable domain was calculated for a model of the S trimer using the Abundance glycoforms (**Supplemental Table, Tab 1,** (47)). The predicted antibody accessibility is visualized across the sequence, as well as mapped onto the 3D surface, via color shading **(Figure 4A, 4C, Supplemental Movie A)**. Additionally, the Oxford class glycoforms (**Supplemental Table, Tab 1)**, which is arguably the most representative means for showing glycan microheterogeneity using a single glycan structure (**Supplemental Table, Tab 8**), model is shown with the sequence variant information (Figure 4D, **Supplemental Table, Tab 11**). A substantial number of these variants occur (directly by comparison to **Figure 4A** or visually by comparison to **Figure 4C**) in regions of high calculated epitope accessibility (e.g. N74K, T76I,

R78M, D138H, H146Y, S151I, D253G, V483A, etc.) suggesting potential selective pressure to avoid host immune response. Also, it is interesting to note that 3 of the emerging variants would eliminate N-linked sequons in S; N74K and T76I would eliminate N-glycosylation of N74 (found in the insert variable region 1 of CoV-2 S compared to CoV-1 S), and S151I eliminates N-glycosylation of N149 (found in the insert variable region 2) (**Fig. 4A, S9, Supplemental Table, Tab 11)**. Lastly, the SARS-CoV-2 S Processed glycoform model is shown (**Supplemental Table, Tab 1**), along with marking amino acid T0323 that has a modest (11% occupancy, **Fig. S8, Supplemental Table, Tab 10**) amount of O-glycosylation to represent the most heavily glycosylated form of S (**Figure 4E**).

Glycomics Informed Glycoproteomics Reveals Complex N-linked Glycosylation of ACE2

We also analyzed ACE2 glycosylation utilizing the same glycomic and glycoproteomic approaches described for S protein. Glycomic analyses of released N-linked glycans (**Fig. 5A, Supplemental Table, Tab 3**) revealed that the majority of glycans on ACE2 are complex with limited high-mannose and hybrid glycans. Glycoproteomic analyses revealed that occupancy was high (>75%) at all 6 sites and significant microheterogeneity dominated by complex N-glycans was observed for each site (**Fig. 5B-5G, S10, Supplemental Table, Tabs 5-8**). We also observed, consistent with the O-glycomics (**Fig. S5, Supplemental Table, Tab 4**), that Ser 155 and several S/T residues at the C-terminus of ACE2 outside of the peptidase domain were O-glycosylated but stoichiometry was extremely low (less than 2%, **Supplemental Table, Tab 9 and 10**).

3D Structural Modeling of Glycosylated, Soluble, ACE2 Highlighting Glycosylation and Variants.

We integrated our glycomics, glycoproteomics, and population variant analyses results with a 3D model of Ace 2 (based on PDB code 6M0J (48), see methods for details) to generate two versions of the soluble glycosylated ACE2 for visualization and molecular dynamics simulations. Information from these analyses is laid out first in **Figure 6A** along the primary structure (sequence) of the SARS-CoV-2 S protomer for reference. We visualized the ACE2 glycoprotein with the Abundance glycoform model simulated at each site as well as highlighting the naturally

occurring variants observed in the human population (**Fig. 6B, Supplemental Movie B, Supplemental Table, Tab 11**). Note, that the Abundance glycoform model and the Oxford class glycoform model for ACE2 are identical (**Supplemental Table, Tabs 1,8**). Notably, one site of N-linked glycosylation (N546) is predicted to not be present in 3 out of 10,000 humans based on naturally occurring variation in the human population (**Supplemental Table, Tab 11**). We also modeled ACE2 using the Processed glycoform model (**Fig. 6C**). In both models, the interaction domain with S is defined (**Fig. 6B-6C, Supplemental Movie B**) .

Molecular Dynamics Simulation of the Glycosylated Trimer Spike of SARS-CoV-2 in Complex with Glycosylated, Soluble, Human Ace 2 Reveals Protein and Glycan Interactions

Molecular dynamics simulations were performed to examine the co-complex (generated from a crystal structure of the ACE2-RBD co-complex, PDB code 6M0J, (48)) of glycosylated S with glycosylated ACE2 with the 3 different glycoforms models (Abundance, Oxford Class, and Processed, **Supplemental Table, Tab 1, Supplemental Simulations 1-3**). Information from these analyses is laid out along the primary structure (sequence) of the SARS-CoV-2 S protomer and ACE2 highlighting regions of glycan-protein interaction observed in the MD simulation (**Fig. 7A-B**, **Supplemental Simulations 1-3**). Interestingly, two glycans on Ace 2 (at N090 and N322), that are highlighted in **Figure 7B** and shown in a more close-up view in **Figure 7C**, are predicted to form interactions with the S protein. The N322 glycan interaction with the S trimer is outside of the receptor binding domain, and the interaction is observed across multiple simulations and throughout each simulation (**Fig. 7A**, **Supplemental Simulations 1-3**). The ACE2 glycan at N090 is close enough to the S trimer surface to repeatedly form interactions, however the glycan arms interact with multiple regions of the surface over the course of the simulations, reflecting the relatively high degree of glycan dynamics (**Fig. 7C-E, Supplemental Movie C**). Inter-molecule glycan-glycan interactions are also observed repeatedly between the glycan at N546 of ACE2 and those in the S protein at residues N0074 and N0165 (**Fig. 7D**). Finally, a full view of the ACE2-S complex with Oxford class glycoforms on both proteins illustrates the extensive glycosylation of the complex (**Fig. 7F, Supplemental Movie D**).

DISCUSSION

We have defined the glycomics-informed, site-specific microheterogeneity of 22 sites of N-linked glycosylation per monomer on a SARS-CoV-2 trimer and the 6 sites of N-linked glycosylation on a soluble version of its human ACE2 receptor using a combination of mass spectrometry approaches coupled with evolutionary and variant sequence analyses to provide a detailed understanding of the glycosylation states of these glycoproteins (**Figs. 1-6**). Our results suggest essential roles for glycosylation in mediating receptor binding, antigenic shielding, and potentially the evolution/divergence of these glycoproteins.

The highly glycosylated SARS-CoV-2 Spike protein, unlike several other viral proteins including HIV-1 (15) but in agreement with another recent report (19), presents significantly more processing of N-glycans towards complex glycosylation, suggesting that steric hinderance to processing enzymes is not a major factor at most sites (**Figs. 2-3**). However, the N-glycans still provide considerable shielding of the peptide backbone (**Fig. 4**). Although O-glycosylation has recently been reported for individually-expressed S1 and S2 domains of the Spike glycoprotein (45), in trimeric form the level of O-glycosylation is extremely low, with the highest level of occupancy being only 11% at T0323 (**Fig. 4E**). The soluble ACE2 protein examined here contains 6 sites of N-linked glycosylation dominated by complex type N-linked glycans (**Fig. 5**). O-glycans were also present on this glycoprotein but at very low levels of occupancy at all sites (<2%).

Our glycomics-informed glycoproteomics allowed us to assign defined sets of glycans to specific glycosylation sites on 3D-structures of S and ACE2 glycoproteins based on experimental evidence (**Figs. 4, 6**). Similar to most other glycoproteins, microheterogeneity is evident at most glycosylation sites of S and ACE2; each glycosylation site can be modified with one of several glycan structures, generating site-specific glycosylation portfolios. For modeling purposes, however, explicit structures must be placed at each glycosylation site. In order to capture the impact of microheterogenity on S and ACE2 molecular dynamics we chose to generate glycoforms for modeling that represented reasonable portfolios of glycan types. Using 3

glycoform models for S (Abundance, Oxford Class, and Processed) and 2 models for ACE2 (Abundance, which was equivalent to Oxford Class, and Processed), we generated 3 molecular dynamic simulations of the co-complexes of these 2 glycoproteins (**Fig. 7 and Supplemental Simulations 1-3**). The observed interactions over time allowed us to evaluate glycan-protein contacts between the 2 proteins as well as examine potential glycan-glycan interactions (**Fig. 7**). We observed glycan-mediated interactions between the S trimer and glycans at N090, N322 and N546 of ACE2. Thus, variations in glycan occupancy or processing at these sites, could alter the affinity of the SARS-CoV-2 – ACE2 interaction and modulate infectivity. It is well established that glycosylation states vary depending on tissue and cell type as well as in the case of humans, on age (49), underlying disease (50,51) and ethnicity (52). Thus glycosylation portfolios may in part be responsible for tissue tropism and individual susceptibility to infection. The importance of glycosylation for S binding to ACE2 is even more emphatically demonstrated by the direct glycan-glycan interactions observed (**Fig. 7D**) between S glycans (at N0074 and N0165) and an ACE2 receptor glycan (at N546), adding an additional layer of complexity for interpreting the impact of glycosylation on individual susceptibility.

Several emerging variants of the virus appear to be altering N-linked glycosylation occupancy by disrupting N-linked sequons. Interestingly, the 2 N-linked sequons in SARS-CoV-2 S directly impacted by variants, N0074 and N0149, are in divergent insert regions 1 and 2, respectively, of SARS-CoV-2 S compared to SARS-CoV-1 S (**Fig. 4A**). The N0074, in particular, is one of the S glycans that interact directly with ACE2 glycan (at N546), suggesting that glycan-glycan interactions may contribute to the unique infectivity differences between SARS-CoV-2 and SARS-CoV-1. These sequon variants will also be important to examine in terms of glycan shielding that could influence immunogenicity and efficacy of neutralizing antibodies, as well as interactions with the host cell receptor ACE2. Naturally-occurring amino acid-changing SNPs in the ACE2 gene generate a number of variants including 1 variant, with a frequency of 3 in 10,000 humans, that eliminates a site of N-linked glycosylation at N546 (**Fig. 6**). Understanding the impact of ACE2 variants on glycosylation and more importantly on S binding, especially for N546S which impacts

the glycan-glycan interaction between S and ACE2 , should be prioritized in light of efforts to develop ACE2 as a potential decoy therapeutic. Intelligent manipulation of ACE2 glycosylation may lead to more potent biologics capable of acting as better competitive inhibitors of S binding. The data presented here, and related similar recent findings (19), provide a framework to facilitate the production of immunogens, vaccines, antibodies, and inhibitors as well as providing additional information regarding mechanisms by which glycan microheterogeneity is achieved. However, considerable efforts still remain in order to fully understand the role of glycans in SARS-CoV-2 infection and pathogenicity. While HEK-expressed S and ACE2 provide a useful window for understanding human glycosylation of these proteins, glycoproteomic characterization following expression in cell lines of more direct relevance to disease and target tissue is sorely needed. While site occupancy will be unlikely to change in different cell lines, processing of N-linked glycans will almost certainly be altered in a cell-type dependent fashion. Thus, analyses of the Spike trimer extracted from pseudoviruses, virion-like particles, and ultimately from infectious SARS-CoV-2 virions harvested from airway cells or patients will provide the most accurate view of how trimer immunogens reflect the true glycosylation pattern of the virus. Detailed analyses of the impact of emerging variants in S and natural and designed-for-biologics variants of ACE2 on glycosylation and binding properties are important next steps for developing therapeutics. Finally, it will be important to monitor the slow evolution of the virus to determine if existing sites of glycosylation are lost or new sites emerge with selective pressure that might alter the efficacy of vaccines, neutralizing antibodies, and/or inhibitors.

ACKNOWLEDGEMENTS

The authors would like to thank ProteinMetrics for providing licenses for their software used here and the developers of pGlyco for productive discussions regarding their software. We would also like to thank Galit Alter of the Ragon Institute for facilitating this collaborative effort. This effort was facilitated by the ThermoFisher Scientific appointed Center of Excellence in Glycoproteomics at the Complex Carbohydrate Research Center at the University of Georgia (co-directed by MT

and LW). This research is supported in part by R35GM119850 (NEL), NNF10CC1016517 (NEL), R01AI139238 (MAB), R01AI147884-01A1S1 (BC), Massachusetts Consortium on Pathogen Readiness (BC), U01CA207824 (R.J.W), P41GM103390 (R.J.W.), P41GM103490 (MT and LW), U01GM125267 (MT), and R01GM130915 (LW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

Conceptualization: M.T., B.C., R.J.W. and L.W.; Methodology, Software, Validation, Formal Analysis, Investigation, Resources, and Data Curation: P.Z., J.L.P., O.C.G., Y.C., T.X., K.E.R., K.A., B.P.K., R.B., D.H.B., M.A.B., N.E.L., M.T., B.C., R.J.W., and L.W.; Writing-Original Draft: P.Z., J.L.P., and L.W.; Writing-Review & Editing: All authors; Visualization: P.Z., J.L.P., O.C.G, Y.C., M.T., B.C., R.J.W., and L.W.; Supervision, Project Administration, and Funding Acquisition: D.H.B., M.A.B., N.E.L., M.T., B.C., R.J.W., and L.W..

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Expression and Characterization of SARS-CoV-2 Spike Glycoprotein Trimer Immunogen and Soluble Human ACE2. A) Sequences of SARS-CoV-2 S immunogen and soluble human ACE2. The N-terminal pyroglutamines for both mature protein monomers are bolded, underlined, and shown in green. The canonical N-linked glycosylation sequons are bolded, underlined, and shown in red. Negative stain electron microscopy of the purified trimer (B) and Coomassie G-250 stained reducing SDS-PAGE gels (C) confirmed purity of the SARS-CoV-2 S protein trimer and of the soluble human ACE2. MWM = molecular weight markers. D) A representative Step-HCD fragmentation spectrum from mass spectrometry analysis of a tryptic

digest of S annotated manually based on search results from pGlyco 2.2. This spectrum defines the N-terminus of the mature protein monomer as (pyro-)glutamine 0014. A representative N-glycan consistent with this annotation and our glycomics data (Fig. 2) is overlaid using the Symbol Nomenclature For Glycans (SNFG) code. This complex glycan occurs at N0017. Note, that as expected, the cysteine is carbamidomethylated and the mass accuracy of the assigned peptide is 0.98 ppm. On the sequence of the N-terminal peptide and in the spectrum, the assigned b (blue) and y (red) ions are shown. In the spectrum, purple highlights glycan oxonium ions and green marks intact peptide fragment ions with various partial glycan sequences still attached. Note that the green labeled ions allow for limited topology to be extracted including defining that the fucose is on the core and not the antennae of the glycopeptide.

Figure 2. Glycomics Informed Glycoproteomics Reveals Substantial Site-Specific Microheterogeneity of N-linked Glycosylation on SARS-CoV-2 S. A) Glycans released from SARS-CoV-2 S protein trimer immunogen were permethylated and analyzed by MSn. Structures were assigned, grouped by type and structural features, and prevalence was determined based on ion current. The pie chart shows basic division by broad N-glycan type. The bar graph provides additional detail about the glycans detected. The most abundant structure with a unique categorization by glycomics for each N-glycan type in the pie chart, or above each feature category in the bar graph, is indicated. B – E) Glycopeptides were prepared from SARS-CoV-2 S protein trimer immunogen using multiple combinations of proteases, analyzed by LC-MSn, and the resulting data was searched using several different software packages. Four representative sites of N-linked glycosylation with specific features of interest were chosen and are presented here. N0074 (B) and N0149 (C) are shown that occur in variable insert regions of S compared to SARS-CoV and other related coronaviruses, and there are emerging variants of SARS-CoV-2 that disrupt these two sites of glycosylation in S. N0234 (D) contains the most high-mannose N-linked glycans. N0801 (D) is an example of glycosylation in the S2 region of the immunogen and displays a high degree of hybrid glycosylation compared to other sites. The abundance of each

458 composition is graphed in terms of assigned spectral counts. Representative glycans (as
459 determined by glycomics analysis) for several abundant compositions are shown in SNFG format.
460 The abbreviations used here and throughout the manuscript are N for HexNAc, H for Hexose, F
461 for Fucose, and A for Neu5Ac. Note that the graphs for the other 18 sites and other graphs
462 grouping the microheterogeneity observed by other properties are presented in Supplemental
463 Information.
464
465 **Figure 3. SARS-CoV-2 S Immunogen N-glycan Sites are Predominantly Modified by**
466 **Complex N-glycans.** N-glycan topologies were assigned to all 22 sites of the S protomer and
467 the spectral counts for each of the 3 types of N-glycans (high-mannose, hybrid, and complex) as
468 well as the unoccupied peptide spectral match counts at each site were summed and visualized
469 as pie charts. Note that only N1173 and N1194 show an appreciable amount of the unoccupied
470 amino acid.
471
472 **Figure 4. 3D Structural Modeling of Glycosylated SARS-CoV-2 Spike Trimer Immunogen**
473 **Reveals Predictions for Antigen Accessibility and Other Key Features.** Results from
474 glycomics and glycoproteomics experiments were combined with results from bioinformatics
475 analyses and used to model several versions of glycosylated SARS-CoV-2 S trimer immunogen.
476 A) Sequence of the SARS-CoV-2 S immunogen displaying computed antigen accessibility and
477 other information. Antigen accessibility is indicated by red shading across the amino acid
478 sequence. B) Emerging variants confirmed by independent sequencing experiments were
479 analyzed based on the 3D structure of SARS-CoV-2 S to generate a proximity chart to the
480 determined N-linked glycosylation sites. C) SARS-CoV-2 S trimer immunogen model from MD
481 simulation displaying abundance glycoforms and antigen accessibility shaded in red for most
482 accessible, white for partial, and black for inaccessible (see **supplemental movie A**). D) SARS-
483 CoV-2 S trimer immunogen model from MD simulation displaying oxford class glycoforms and
484 sequence variants. * indicates not visible while the box represents 3 amino acid variants that are

clustered together in 3D space. E) SARS-CoV-2 S trimer immunogen model from MD simulation displaying processed glycoforms plus shading of Thr-323 that has O-glycoslyation at low stoichiometry in yellow.

Figure 5: Glycomics Informed Glycoproteomics of Soluble Human ACE2 Reveals High Occupancy, Complex N-linked Glycosylation. A) Glycans released from soluble, purified ACE2 were permethylated and analyzed by MSn. Structures were assigned, grouped by type and structural features, and prevalence was determined based on ion current. The pie chart shows basic division by broad N-glycan type. The bar graph provides additional detail about the glycans detected. The most abundant structure with a unique categorization by glycomics for each N-glycan type in the pie chart, or above each feature category in the bar graph, is indicated. B – G) Glycopeptides were prepared from soluble human ACE2 using multiple combinations of proteases, analyzed by LC-MSn, and the resulting data was searched using several different software packages. All six sites of N-linked glycosylation are presented here. Displayed in the bar graphs are the individual compositions observed graphed in terms of assigned spectral counts. Representative glycans (as determined by glycomics analysis) for several abundant compositions are shown in SNFG format. The abbreviations used here and throughout the manuscript are N for HexNAc, H for Hexose, F for Fucose, and A for Neu5Ac. The pie chart (analogous to Figure 3 for SARS-CoV-2 S) for each site is displayed in the upper right corner of each panel. B) N053. C) N090. D) N103. E) N322. F) N432. G) N546, a site that does not exist in 3 in 10,000 people.

Figure 6: 3D Structural Modeling of Glycosylated Soluble Human ACE2. Results from glycomics and glycoproteomics experiments were combined with results from bioinformatics analyses and used to model several versions of glycosylated soluble human ACE2. A) Sequence of soluble human ACE2. The one variant with a reported MAF in NCBI dbSNP (MAF = 0.0003) that removes an N-glycan sequon (N546S) is highlighted in pink. B) Soluble human ACE2 model from MD simulations displaying abundance glycoforms, interaction surface with S, and sequence

variants. N546 variant is boxed that would remove N-linked glycosylation at that site (see **supplemental movie B**). C) Soluble human ACE2 model from MD simulations displaying processed glycoforms.

Figure 7: Interactions of Glycosylated Soluble Human ACE2 and Glycosylated SARS-CoV-2 S Trimer Immunogen Revealed By 3D-Structural Modeling and Molecular Dynamics Simulations. A) Primary Sequence of S and ACE2 shaded for Protein-Glycan interactions uncovered by molecular dynamic simulations. On S, red shading displays interactions with ACE2 N090 glycan, green displays interactions with ACE2 N322 glycan, and blue displays interaction with ACE2 N546 glycan. Combined colors on S represent interactions with multiple Ace glycans from the set. On ACE2, red shading displays interactions with S N0074 glycan, green shading displays interactions with S N0165 glycan. Yellow on ACE2 represents interactions with both glycans. B) Molecular dynamics simulation of glycosylated soluble human ACE2 and glycosylated SARS-CoV-2 S trimer immunogen interaction displayed with 180 degree rotation. ACE2 (top) is colored red with glycans in pink while S is colored white with glycans in dark grey. Highlighted are ACE2 glycans that interacts with S (see **supplemental simulations 1-3**). C) Zoom in images of ACE2-S interface highlighting Ace 2 glycan-S protein interactions. D) Zoom in of ACE2-S interface highlighting ACE2 and S glycan interactions using 3D-SNFG icons (53) with S protein (pink) as well as ACE2-S glycan-glycan interactions. E) Zoom in of dynamics trajectory of glycans at the interface of soluble human ACE2 and S (see **supplemental movie C**). F) Full view of molecular dynamic trajectory of oxford class glycans simulated on the ACE2-S complex (see **supplemental movie D**).

STAR METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
SARS-CoV-2 S protein	This Study	N/A

Human ACE2 protein	This Study	N/A
2x Laemmli sample buffer	Bio-Rad	Cat#161-0737
Invitrogen NuPAGE 4 to 12%, Bis-Tris, Mini Protein Gel	Thermo Fisher Scientific	Cat#NP0321PK2
Coomassie Brilliant Blue G-250 Dye	Thermo Fisher Scientific	Cat#20279
Dithiothreitol	Sigma Aldrich	Cat#43815
Iodoacetamide	Sigma Aldrich	Cat#I1149
Trypsin	Promega	Cat#V5111
Lys-C	Promega	Cat#V1671
Arg-C	Promega	Cat#V1881
Glu-C	Promega	Cat#V1651
Asp-N	Promega	Cat#VA1160
Endoglycosidase H	Promega	Cat#V4871
PNGaseF	Promega	Cat#V4831
Chymotrypsin	Athens Research and Technology	Cat#16-19-030820
Alpha lytic protease	New England BioLabs	Cat#P8113
18O water	Cambridge Isotope Laboratories	OLM-782-10-1
O-protease OpeRATOR	Genovis	Cat#G1-OP1-020
Deposited Data		
MS data for site-specific N-linked glycopeptides for SARS-Cov-2 S and human ACE2	This Study	PXD019937
MS data for site-specific O-linked glycopeptides for SARS-Cov-2 S and human ACE2	This Study	PXD019940
MS data for deglycosylated N-linked glycopeptides for SARS-Cov-2 S and human ACE2	This Study	PXD019938
MS data for disulfide bond analysis for SARS-Cov-2 S	This Study	PXD019939
Experimental Models: Cell Lines		
293-F Cells	Gibco	Cat#R79007
Experimental Models: Organisms/Strains		
VSV(G)-Pseudoviruses	This Study	
Software and Algorithms		
pGlyco	(43)	v2.2.2
Proteome Discoverer	Thermo Fisher Scientific	v1.4
Byonic	Protein Metrics Inc. (44)	v3.8.13
ProteoIQ	Premier Biosoft	v2.7
GRITS Toolbox	http://www.grits-toolbox.org (37)	v1.1
EMBOSS needle	(54)	v6.6.0
Biopython	https://biopython.org/	v1.76
Rpdb	https://rdrr.io/cran/Rpdb/	v2.3

SignalP	http://www.cbs.dtu.dk/services/SignalP/ (34)	v5.0
LibreOFFICE Writer	The Document Foundation	V6.4.4.2

LEAD CONTACT

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Peng Zhao (pengzhao@uga.edu) and/or Lance Wells (lwells@ccrc.uga.edu).

METHOD DETAILS

Expression, Purification, and Characterization of SARS-CoV-2 S and Human ACE2 Proteins

To express a stabilized ectodomain of Spike protein, a synthetic gene encoding residues 1−1208 of SARS-CoV-2 Spike with the furin cleavage site (residues 682–685) replaced by a "GGSG" sequence, proline substitutions at residues 986 and 987, and a foldon trimerization motif followed by a C-terminal 6xHisTag was created and cloned into the mammalian expression vector pCMV-IRES-puro (Codex BioSolutions, Inc, Gaithersburg, MD). The expression construct was transiently transfected in HEK 293F cells using polyethylenimine (Polysciences, Inc, Warrington, PA). Protein was purified from cell supernatants using Ni-NTA resin (Qiagen, Germany), the eluted fractions containing S protein were pooled, concentrated, and further purified by gel filtration chromatography on a Superose 6 column (GE Healthcare). Negative stain electron microscopy (EM) analysis was performed as described (55).

The DNA fragment encoding human ACE2 (1-615) with a 6xHis tag at C terminus was synthesized by Genscript and cloned to the vector pCMV-IRES-puro. The expression construct was transfected in HEK293F cells using polyethylenimine. The medium was discarded and replaced with FreeStyle 293 medium after 6-8 hours. After incubation in 37 °C with 5.5% CO2

for 5 days, the supernatant was collected and loaded to Ni-NTA resin for purification. The elution was concentrated and further purified by a Superdex 200 column.

In-Gel Analysis of SARS-CoV-2 S and Human ACE2 Proteins

A 3.5-μg aliquot of SARS-CoV-2 S protein as well as a 2-μg aliquot of human ACE2 were combined with Laemmli sample buffer, analyzed on a 4-12% Invitrogen NuPage Bis-Tris gel using the MES pH 6.5 running buffer, and stained with Coomassie Brilliant Blue G-250.

Analysis of N-linked and O-linked Glycans Released from SARS-Cov-2 S and Human ACE2 Proteins

Aliquots of approximately 25-50 μg of S or ACE2 protein were processed for glycan analysis as previously described (35,36). For N-linked glycan analysis, the proteins were reduced, alkylated, and digested with trypsin. Following trypsinization, glycopeptides were enriched by C18 Sep-Pak and subjected to PNGaseF digestion to release N-linked glycans. Following PNGaseF digestion, released glycans were separated from residual glycosylated peptides bearing O-linked glycans by C18 Sep-Pak. O-glycosylated peptides were eluted from the Sep-Pak and subjected to reductive β-elimination to release the O-glycans. Another 50 μg aliquot of each protein was denatured with SDS and digested with PNGaseF to remove N-linked glycans. The de-N-glycosylated, intact protein was then subjected to reductive β-elimination to release O-glycans. The profiles of O-glycans released from peptides or from intact protein were found to be comparable. N- and O-linked glycans released from glycoproteins were permethylated with methyliodide according to the method of Anumula and Taylor prior to MS analysis (56). Glycan structural analysis was performed using an LTQ-Orbitrap instrument (Orbitrap Discovery, ThermoFisher). Detection and relative quantification of the prevalence of individual glycans was accomplished using the total ion mapping

(TIM) and neutral loss scan (NL scan) functionality of the Xcalibur software package version 2.0 (Thermo Fisher Scientific) as previously described (35,36). Mass accuracy and detector response was tuned with a permethylated oligosaccharide standard in positive ion mode. For fragmentation by collision-induced dissociation (CID in MS2 and MSn), normalized collision energy of 45% was applied. Most permethylated glycans were identified as singly or doubly charged, sodiated species [M+Na] in positive mode. Peaks for all charge states were deconvoluted by the charge state and summed for quantification. All spectra were manually interpreted and annotated. The explicit identities of individual monosaccharide residues have been assigned based on known human biosynthetic pathways. Graphical representations of monosaccharide residues are consistent with the Symbol Nomenclature for Glycans (SNFG), which has been broadly adopted by the glycomics community (57). The MS-based glycomics data generated in these analyses and the associated annotations are presented in accordance with the MIRAGE standards and the Athens Guidelines (58).

Analysis of Disulfide Bonds for SARS-Cov-2 S Protein by LC-MS

Two 10-µg aliquots of SARS-CoV-2 S protein were reduced by incubating with 20% acetonitrile at room temperature and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The two aliquots of proteins were then digested respectively using alpha lytic protease, or a combination of trypsin, Lys-C and Glu-C. Following digestion, the proteins were deglycosylated by PNGaseF treatment. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the

608 temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z
609 200 to 2000 at 60k resolution, and MS/MS scans following electron transfer dissociation (ETD)
610 were collected in the orbitrap at 15k resolution. The raw spectra were analyzed by Byonic
611 (v3.8.13, Protein Metrics Inc.) with mass tolerance set as 20 ppm for both precursors and
612 fragments. The search output was filtered at 0.1% false discovery rate and 10 ppm mass error.
613 The spectra assigned as cross-linked peptides were manually evaluated for Cys0015.

614 **Analysis of Site-Specific N-linked Glycopeptides for SARS-Cov-2 S and Human ACE2**
615 **Proteins by LC-MS**

616 Four 3.5-µg aliquots of SARS-CoV-2 S protein were reduced by incubating with 10 mM of
617 dithiothreitol at 56 °C and alkylated by 27.5 mM of iodoacetamide at room temperature in dark.
618 The four aliquots of proteins were then digested respectively using alpha lytic protease,
619 chymotrypsin, a combination of trypsin and Glu-C, or a combination of Glu-C and AspN. Three
620 10-µg aliquots of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56 °C
621 and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The three aliquots of
622 proteins were then digested respectively using alpha lytic protease, chymotrypsin, or a
623 combination of trypsin and Lys-C. The resulting peptides were separated on an Acclaim
624 PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-electrospray ion source
625 of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200 nL/min. The
626 elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by
627 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the
628 temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z
629 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation
630 (HCD) with stepped collision energy (15%, 25%, 35%) were collected in the orbitrap at 15k
631 resolution. pGlyco v2.2.2 (43) was used for database searches with mass tolerance set as 20
632 ppm for both precursors and fragments. The database search output was filtered to reach a 1%
633 false discovery rate for glycans and 10% for peptides. Quantitation was performed by

calculating spectral counts for each glycan composition at each site. Any N-linked glycan compositions identified by only one spectra were removed from quantitation. N-linked glycan compositions were categorized into the 21 oxford classes: HexNAc(2)Hex(9~5)Fuc(0~1) was classified as M9 to M5 respectively; HexNAc(2)Hex(4~1)Fuc(0~1) was classified as M1-M4; HexNAc(3~6)Hex(5~9)Fuc(0)NeuAc(0~1) was classified as Hybrid with HexNAc(3~6)Hex(5~9)Fuc(1~2)NeuAc(0~1) classified as F-Hybrid; Complex-type glycans are classified based on the number of antenna and fucosylation:
HexNAc(3)Hex(3~4)Fuc(0)NeuAc(0~1) is assigned as A1 with
HexNAc(3)Hex(3~4)Fuc(1~2)NeuAc(0~1) assigned as F-A1;
HexNAc(4)Hex(3~5)Fuc(0)NeuAc(0~2) is assigned as A2/A1B with
HexNAc(4)Hex(3~5)Fuc(1~5)NeuAc(0~2) assigned as F-A2/A1B;
HexNAc(5)Hex(3~6)Fuc(0)NeuAc(0~3) is assigned as A3/A2B with
HexNAc(5)Hex(3~6)Fuc(1~3)NeuAc(0~3) assigned as F-A3/A2B;
HexNAc(6)Hex(3~7)Fuc(0)NeuAc(0~4) is assigned as A4/A3B with
HexNAc(6)Hex(3~7)Fuc(1~3)NeuAc(0~4) assigned as F-A4/A3B;
HexNAc(7)Hex(3~8)Fuc(0)NeuAc(0~1) is assigned as A5/A4B with
HexNAc(7)Hex(3~8)Fuc(1~3)NeuAc(0~1) as F-A5/A4B; HexNAc(8)Hex(3~9)Fuc(0) is assigned as A6/A5B with HexNAc(8)Hex(3~9)Fuc(1) assigned as F-A6/A5B.

Analysis of Deglycosylated SARS-Cov-2 S and Human ACE2 Proteins by LC-MS

Three 3.5-μg aliquots of SARS-CoV-2 S protein were reduced by incubating with 10 mM of dithiothreitol at 56 °C and alkylated by 27.5 mM of iodoacetamide at room temperature in dark. The three aliquots were then digested respectively using chymotrypsin, Asp-N, or a combination of trypsin and Glu-C. Two 10-μg aliquots of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56 °C and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The two aliquots were then digested respectively using chymotrypsin, or a combination of trypsin and Lys-C. Following digestion, the proteins were deglycosylated by Endoglycosidase

H followed by PNGaseF treatment in the presence of 18O water. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following collision-induced dissociation (CID) at 38% collision energy were collected in the ion trap. The spectra were analyzed using SEQUEST (Proteome Discoverer 1.4) with mass tolerance set as 20 ppm for precursors and 0.5 Da for fragments. The search output was filtered using ProteoIQ (v2.7) to reach a 1% false discovery rate at protein level and 10% at peptide level. Occupancy of each N-linked glycosylation site was calculated using spectral counts assigned to the 18O-Asp-containing (PNGaseF-cleaved) and/or HexNAc-modified (EndoH-cleaved) peptides and their unmodified counterparts.

Analysis of Site-Specific O-linked Glycopeptides for SARS-Cov-2 S and Human ACE2 Proteins by LC-MS

Three 10-µg aliquots of SARS-CoV-2 S protein and one 10-µg aliquot of ACE2 protein were reduced by incubating with 5 mM of dithiothreitol at 56 °C and alkylated by 13.75 mM of iodoacetamide at room temperature in dark. The four aliquots were then digested respectively using trypsin, Lys-C, Arg-C, or a combination of trypsin and Lys-C. Following digestion, the proteins were deglycosylated by PNGaseF treatment and then digested with O-protease OpeRATOR®. The resulting peptides were separated on an Acclaim PepMap RSLC C18 column (75 µm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer at a flow rate of 200 nL/min. The elution gradient consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the

temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation (HCD) with stepped collision energy (15%, 25%, 35%) or electron transfer dissociation (ETD) were collected in the orbitrap at 15k resolution. The raw spectra were analyzed by Byonic (v3.8.13) with mass tolerance set as 20 ppm for both precursors and fragments. MS/MS filtering was applied to only allow for spectra where the oxonium ions of HexNAc were observed. The search output was filtered at 0.1% false discovery rate and 10 ppm mass error. The spectra assigned as O-linked glycopeptides were manually evaluated. Quantitation was performed by calculating spectral counts for each glycan composition at each site. Any O-linked glycan compositions identified by only one spectra were removed from quantitation. Occupancy of each O-linked glycosylation site was calculated using spectral counts assigned to any glycosylated peptides and their unmodified counterparts from searches without MS/MS filtering.

Sequence Analysis of SARS-CoV-2 S and Human ACE2 Proteins

The genomes of SARS-CoV as well as bat and pangolin coronavirus sequences reported to be closely related to SARS-CoV-2 were downloaded from NCBI. The S protein sequences from all of those genomes were aligned using EMBOSS needle v6.6.0 (54) via the EMBL-EBI provided web service (59). Manual analysis was performed in the regions containing canonical N-glycosylation sequons (N-X-S/T). For further sequence analysis of SARS-CoV-2 S variants, the genomes of SARS-CoV-2 were downloaded from NCBI and GISAID and further processed using Biopython 1.76 to extract all sequences annotated as “surface glycoprotein” and to remove any incomplete sequence as well as any sequence containing unassigned amino acids. For sequence analysis of human ACE2 variants, the single nucleotide polymorphisms (SNPs) of ACE2 were extracted from the NCBI dbSNP database and filtered for missense mutation entries with a reported minor allele frequency. Manual analysis was performed on both SARS-CoV-2 S and human ACE2 variants to further examine the regions containing canonical N-glycosylation

sequons (N-X-S/T). LibreOffice Writer was used to shade regions on the linear sequence of S and ACE2.

3D Structural Modeling and Molecular Dynamic Simulation of Glycosylated SARS-CoV-2 S and Human ACE2 Proteins

SARS-CoV2 spike (S) protein structure and ACE2 co-complex – A 3D structure of the prefusion form of the S protein (RefSeq: YP_009724390.1, UniProt: P0DTC2 SPIKE_SARS2), based on a Cryo-EM structure (PDB code 6VSB) (46), was obtained from the SWISS-MODEL server (swissmodel.expasy.org). The model has 95% coverage (residues 27 to 1146) of the S protein. The receptor binding domain (RBD) in the "open" conformation was replaced with the RBD from an ACE2 co-complex (PDB code 6M0J) by grafting residues C336 to V524.

Glycoform generation – 3D structures of the three glycoforms (abundance, oxford class, processed) were generated for the SARS-CoV2 S protein alone, and in complex with the glycosylated ACE2 protein. The glycoprotein builder available at GLYCAM-Web (www.glycam.org) was employed together with an in-house program that adjusts the asparagine side chain torsion angles and glycosidic linkages within known low-energy ranges (60) to relieve any atomic overlaps with the core protein, as described previously (61,62).

Energy minimization and Molecular dynamics (MD) simulations **–** Each glycosylated structure was placed in a periodic box of TIP3P water molecules with a 10 Å buffer between the solute and the box edge. Energy minimization of all atoms was performed for 20,000 steps (10,000 steepest decent, followed by 10,000 conjugant gradient) under constant pressure (1 atm) and temperature (300 K) conditions. All MD simulations were performed under nPT conditions with the CUDA implementation of the PMEMD (63,64) simulation code, as present in the Amber14 software suite (University of California, San Diego). The GLYCAM06j force field (65) and Amber14SB force field (66) were employed for the carbohydrate and protein moieties, respectively. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, while a Langevin thermostat with a collision frequency of 2 ps^{-1} was employed for

737 temperature regulation. A nonbonded interaction cut-off of 8 Å was employed. Long-range
738 electrostatics were treated with the particle-mesh Ewald (PME) method (67). Covalent bonds
739 involving hydrogen were constrained with the SHAKE algorithm, allowing an integration time
740 step of 2 fs to be employed. The energy minimized coordinates were equilibrated at 300K over
741 400 ps with restraints on the solute heavy atoms. Each system was then equilibrated with
742 restraints on the Cα atoms of the protein for 1ns, prior to initiating 4 independent 250 ns
743 production MD simulations with random starting seeds for a total time of 1 μs per system, with
744 no restraints applied.
745 *Antigenic surface analysis.* A series of 3D structure snapshots of the simulation were taken at 1
746 ns intervals and analysed in terms of their ability to interact with a spherical probe based on the
747 average size of hypervariable loops present in an antibody complementarity determining region
748 (CDR), as described recently (https://www.biorxiv.org/content/10.1101/2020.04.07.030445v2).
749 The percentage of simulation time each residue was exposed to the AbASA probe was
750 calculated and plotted onto both the 3D structure and primary sequence.
751
752 **Analysis of SARS-CoV-2 Spike VSV pseudoparticles (ppVSV-SARS-2-S)**
753 293T cells were transfected with an expression plasmid encoding SARS-CoV-2 Spike
754 (pcDNAintron-SARS-2-SΔ19). To increase cell surface expression, the last 19 amino acids
755 containing the Golgi retention signal were removed. Two SΔ19 constructs were compared, one
756 started with Met1 and the other with Met2. Twenty-four hours following transfection, cells were
757 transduced with ppVSVΔG-VSV-G (particles that were pseudotyped with VSV-G in trans). One
758 hour following transduction cells were extensively washed and media was replaced.
759 Supernatant containing particles were collected 12-24 hour following transduction and cleared
760 through centrifugation. Cleared supernatant was frozen at -80°C for future use. Target cells
761 VeroE6 were seeded in 24-well plates ($5x10^5$ cells/mL) at a density of 80% coverage. The
762 following day, ppVSV-SARS-2-S/GFP particles were transduced into target cells for 60 minutes,
763 particles pseudotyped with VSV-G, Lassa virus GP, or no glycoprotein were included as

764 controls. 24 hours following transduction, transduced cells were released from the plate with
765 trypsin, fixed with 4% formaldehyde, and GFP-positive virus-transduced cells were quantified
766 using flow cytometry (Bectin Dickson BD-LSRII). To quantify the ability of various SARS-CoV-2
767 S mutants to mediate fusion, effector cells (HEK293T) were transiently transfected with the
768 indicated pcDNAintron-SARS-2-S expression vector or measles virus H and F (68). Effector
769 cells were infected with MVA-T7 four hours following transduction to produce the T7 polymerase
770 (69). Target cells naturally expressing the receptor ACE2 (Vero) or ACE2 negative cells
771 (HEK293T) were transfected with pTM1-luciferase, which encodes for firefly luciferase under the
772 control of a T7 promoter (70). 24 hours following transfection, the target cells were lifted and
773 added to the effector cells at a 1:1 ratio. 4 hours following co-cultivation, cells were washed,
774 lysed and luciferase levels were quantified using Promega's Steady-Glo substrate. To visualize
775 cell-to-cell fusion, Vero cells were co-transfected with pGFP and the pcDNAintron-SARS-2-S
776 constructs. 24 hours following transfection, syncytia was visualized by fluorescence microscopy.

777 **DATA AVAILABILITY**

778 The mass spectrometry proteomics data are available via ProteomeXchange with identifiers
779 listed in the KEY RESOURCES TABLE.

780 S**UPPLEMENTAL INFORMATION**

781 Tables (1, 11 tabs), Figures (10), Movies (4), and Simulations (3).
782 **SUPPLEMENTAL LEGEND:**

783 **Supplemental Table, Tab 1**. Glycans modeled as Abundance, Oxford Class, and Processed.

784 **Supplemental Table, Tab 2.** Cys0015-Cys0136 Disulfide Linked Peptide for SARS-CoV-2 S.

785 **Supplemental Table Tab 3**. Detection of N-linked glycans released from SARS-CoV-2 S and
786 human ACE2. Relative abundance (prevalence) of each species is calculated based on peak
787 intensity in full MS.

788 **Supplemental Table Tab 4**. Detection of O-linked glycans released from SARS-CoV-2 S and
789 human ACE2. Relative abundance (prevalence) of each species is calculated based on peak
790 intensity in full MS.
791 **Supplemental Table Tab 5**. N-linked glycan occupancy at each site of SARS-CoV-2 S and
792 human ACE2. Occupancy is calculated using spectral counts assigned to the 18O-Asp-
793 containing (PNGaseF-cleaved) and/or HexNAc-modified (EndoH-cleaved) peptides and their
794 unmodified counterparts. Sequon refers to the Asn-x-Ser/Thr/Cys, Asn-Gly-x sequences.
795 **Supplemental Table Tab 6**. N-linked glycan compositions identified at each site of SARS-CoV-
796 2 S and human ACE2. Asn(N)# indicates the numbers of asparagines in protein sequences. In
797 compositions: N=HexNAc, H=Hexose (Hex), F=Fucose (Fuc), and A=Neu5Ac. In fucosylation:
798 NoFuc=No Fuc identified; 1Core=One Fuc identified at core position; 1Term=One Fuc identified
799 at terminal position; 1Core and 1Term=One Fuc identified as a mixture of core and terminal
800 positions; 1Core1Term=Two Fuc identified and one is at core and the other is at terminal;
801 2Term=Two Fuc identified at terminal positions; 1Core1Term and 2Term=Two Fuc identified as
802 a mixture of core and terminal positions; 1Core2Term=Three Fuc identified and one is at core
803 and the others are at terminal; 3Term=Three Fuc identified at terminal positions; 1Core2Term
804 and 3Term=Three Fuc identified as a mixture of core and terminal positions; 1Core3Term=Four
805 Fuc identified and one is at core and the others are at terminal; 4Term=Four Fuc identified at
806 terminal positions; 1Core3Term and 4Term=Four Fuc identified as a mixture of core and
807 terminal positions; 1Core4Term=Five Fuc identified and one is at core and the others are at
808 terminal.
809 **Supplemental Table Tab 7**. N-linked glycan types identified at each site of SARS-CoV-2 S and
810 human ACE2. All N-linked glycans are categorized into 3 types: high-mannose, hybrid and
811 complex.
812 **Supplemental Table Tab 8**. N-linked glycan oxford classes identified at each site of SARS-
813 CoV-2 S and human ACE2. All N-linked glycan compositions are categorized into 21 classes:

M9 to M5 respectively is defined as HexNAc(2)Hex(9~5)Fuc(0~1); M1-M4 is defined as HexNAc(2)Hex(4~1)Fuc(0~1); Hybrid is defined as HexNAc(3~6)Hex(5~9)Fuc(0)NeuAc(0~1) and F-Hybrid is defined as HexNAc(3~6)Hex(5~9)Fuc(1~2)NeuAc(0~1). Complex-type glycans are classified based on the number of antenna and fucosylation:
HexNAc(3)Hex(3~4)Fuc(0)NeuAc(0~1) is assigned as A1 with
HexNAc(3)Hex(3~4)Fuc(1~2)NeuAc(0~1) assigned as F-A1;
HexNAc(4)Hex(3~5)Fuc(0)NeuAc(0~2) biantennary is assigned as A2/A1B with
HexNAc(4)Hex(3~5)Fuc(1~5)NeuAc(0~2) biantennary or hybrid with LacDiNAc assigned as F-A2/A1B; HexNAc(5)Hex(3~6)Fuc(0)NeuAc(0~3) bisected biantennary/triantennary is assigned as A3/A2B with HexNAc(5)Hex(3~6)Fuc(1~3)NeuAc(0~3) bisected biantennary/triantennary or biantennary with LacDiNAc assigned as F-A3/A2B; HexNAc(6)Hex(3~7)Fuc(0)NeuAc(0~4) tetraantennary or triantennary with LacDiNAc is assigned as A4/A3B with
HexNAc(6)Hex(3~7)Fuc(1~3)NeuAc(0~4) biantennary with LacDiNAc or bisected biantennary/triantennary with LacDiNAc or tetraantennary assigned as F-A4/A3B;
HexNAc(7)Hex(3~8)Fuc(0)NeuAc(0~1) triantennary with LacDiNAc is assigned as A5/A4B with
HexNAc(7)Hex(3~8)Fuc(1~3)NeuAc(0~1) triantennary with LacDiNAc assigned as F-A5/A4B;
HexNAc(8)Hex(3~9)Fuc(0) is assigned as A6/A5B with HexNAc(8)Hex(3~9)Fuc(1) assigned as F-A6/A5B.

Supplemental Table Tab 9. O-linked glycan compositions identified at each site of SARS-CoV-2 S and human ACE2. Ser/Thr# indicates the numbers of serines or threonines in protein sequences. In compositions: N=HexNAc, H=Hexose (Hex), F=Fucose (Fuc), and A=Neu5Ac.

Supplemental Table Tab 10. O-linked glycan occupancy at each site of SARS-CoV-2 S and human ACE2. Occupancy is calculated using spectral counts assigned to the glycosylated peptides and their unmodified counterparts.

Supplemental Table Tab 11. SARS-CoV-2 S and human ACE2 variants.

Supplemental Figure 1. Defining N-terminus of ACE2 as pyro-glutamine at site Q0018. Representative HCD MS2 spectrum shown.

Supplemental Figure 2. Disulfide bond formed between Cysteines 0015 and 0136 of SARS-CoV-2 S. Representative EThcD MS2 spectrum shown.

Supplemental Figure 3. Signal P Prediction of Two Different Start Methionines for S

Supplemental Figure 4. Functional characterization of various S constructs in Pseudovirus. A) Syncytia produced by SARS-CoV-2 S constructs in VeroE6 cells co-transfected with a GFP plasmid to visualize cell-to-cell fusion. Quantification of fusion using a luciferase complementation assay in 293T (B) or VeroE6 cells (C). D) Transduction efficiency in Vero E6 cells of ppVSV-GFP particles coated in the indicated glycoprotein. Results suggest that start methionine does not alter fusion or efficiency.

Supplemental Figure 5. Detection of O-linked glycans released from SARS-CoV-2 S and human ACE2. The detected O-glycans were categorized based on their structures and types. Relative abundance (prevalence) of each species is calculated based on peak intensity in full MS.

Supplemental Figure 6. Multi-panel. N-linked glycan compositions identified at each site of SARS-CoV-2 S not displayed in Figure 2. The y-axis represents spectral counts of each identified composition and indicated structures in red numbering are displayed.

Supplemental Figure 7. N-linked glycan oxford classes identified at each of 22 glycosylation sites of SARS-CoV-2 S. Percentages are calculated based on spectral counts. All N-linked glycan compositions are categorized into 21 classes: M9 to M5 respectively is defined as HexNAc(2)Hex(9~5)Fuc(0~1); M1-M4 is defined as HexNAc(2)Hex(4~1)Fuc(0~1); Hybrid is defined as HexNAc(3~6)Hex(5~9)Fuc(0)NeuAc(0~1) and F-Hybrid is defined as HexNAc(3~6)Hex(5~9)Fuc(1~2)NeuAc(0~1). Complex-type glycans are classified based on the number of antenna and fucosylation: HexNAc(3)Hex(3~4)Fuc(0)NeuAc(0~1) is assigned as A1 with HexNAc(3)Hex(3~4)Fuc(1~2)NeuAc(0~1) assigned as F-A1;

865 HexNAc(4)Hex(3~5)Fuc(0)NeuAc(0~2) biantennary is assigned as A2/A1B with
866 HexNAc(4)Hex(3~5)Fuc(1~5)NeuAc(0~2) biantennary or hybrid with LacDiNAc assigned as F-
867 A2/A1B; HexNAc(5)Hex(3~6)Fuc(0)NeuAc(0~3) bisected biantennary/triantennary is assigned
868 as A3/A2B with HexNAc(5)Hex(3~6)Fuc(1~3)NeuAc(0~3) bisected biantennary/triantennary or
869 biantennary with LacDiNAc assigned as F-A3/A2B; HexNAc(6)Hex(3~7)Fuc(0)NeuAc(0~4)
870 tetraantennary or triantennary with LacDiNAc is assigned as A4/A3B with
871 HexNAc(6)Hex(3~7)Fuc(1~3)NeuAc(0~4) biantennary with LacDiNAc or bisected
872 biantennary/triantennary with LacDiNAc or tetraantennary assigned as F-A4/A3B;
873 HexNAc(7)Hex(3~8)Fuc(0)NeuAc(0~1) triantennary with LacDiNAc is assigned as A5/A4B with
874 HexNAc(7)Hex(3~8)Fuc(1~3)NeuAc(0~1) triantennary with LacDiNAc assigned as F-A5/A4B;
875 HexNAc(8)Hex(3~9)Fuc(0) is assigned as A6/A5B with HexNAc(8)Hex(3~9)Fuc(1) assigned as
876 F-A6/A5B.
877 **Supplemental Figure 8**. Multipanel. Legend and 6 O-linked glycans detected at site T0323 of
878 SARS-CoV-2 S. Representative Step-HCD spectra shown for each.
879 **Supplemental Figure 9**. Two panel. Sequence alignments of SARS-CoV-1 and SARS-CoV-2 S
880 variants. Alignment of multiple S proteins from related coronaviruses.
881 **Supplemental Figure 10**. N-linked glycan oxford classes identified at each site of human
882 ACE2. Percentages are calculated based on spectral counts. All N-linked glycan compositions
883 are categorized into 21 classes: M9 to M5 respectively is defined as
884 HexNAc(2)Hex(9~5)Fuc(0~1); M1-M4 is defined as HexNAc(2)Hex(4~1)Fuc(0~1); Hybrid is
885 defined as HexNAc(3~6)Hex(5~9)Fuc(0)NeuAc(0~1) and F-Hybrid is defined as
886 HexNAc(3~6)Hex(5~9)Fuc(1~2)NeuAc(0~1). Complex-type glycans are classified based on the
887 number of antenna and fucosylation: HexNAc(3)Hex(3~4)Fuc(0)NeuAc(0~1) is assigned as A1
888 with HexNAc(3)Hex(3~4)Fuc(1~2)NeuAc(0~1) assigned as F-A1;
889 HexNAc(4)Hex(3~5)Fuc(0)NeuAc(0~2) biantennary is assigned as A2/A1B with
890 HexNAc(4)Hex(3~5)Fuc(1~5)NeuAc(0~2) biantennary or hybrid with LacDiNAc assigned as F-

A2/A1B; HexNAc(5)Hex(3~6)Fuc(0)NeuAc(0~3) bisected biantennary/triantennary is assigned as A3/A2B with HexNAc(5)Hex(3~6)Fuc(1~3)NeuAc(0~3) bisected biantennary/triantennary or biantennary with LacDiNAc assigned as F-A3/A2B; HexNAc(6)Hex(3~7)Fuc(0)NeuAc(0~4) tetraantennary or triantennary with LacDiNAc is assigned as A4/A3B with HexNAc(6)Hex(3~7)Fuc(1~3)NeuAc(0~4) biantennary with LacDiNAc or bisected biantennary/triantennary with LacDiNAc or tetraantennary assigned as F-A4/A3B; HexNAc(7)Hex(3~8)Fuc(0)NeuAc(0~1) triantennary with LacDiNAc is assigned as A5/A4B with HexNAc(7)Hex(3~8)Fuc(1~3)NeuAc(0~1) triantennary with LacDiNAc assigned as F-A5/A4B; HexNAc(8)Hex(3~9)Fuc(0) is assigned as A6/A5B with HexNAc(8)Hex(3~9)Fuc(1) assigned as F-A6/A5B.

Supplemental Movie A: Linked to Figure 4C, Glycosylated S antigen accessibility

Supplemental Movie B: Linked to Figure 6B, Glycosylated ACE2 with variants

Supplemental Movie C: Linked to Figure 7E, Interface of ACE2-S Complex

Supplemental Movie D: Linked to Figure 7F, the glycosylated ACE2-S Complex

Simulation 1: Linked to Figure 7B, Abundance glycoforms on ACE2-S Complex

Simulation 2: Linked to Figure 7B, Oxford class glycoforms on ACE2-S Complex

Simulation 3: Linked to Figure 7B, Processed glycoforms on ACE2-S Complex

REFERENCES

1. 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., and Shi, Z. L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273
2. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W. J., Wang, D., Xu, W., Holmes, E. C., Gao, G. F., Wu, G., Chen, W., Shi, W., and Tan, W. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565-574

3. Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, and Guan, Y. (2003) Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353-1358
4. Xia, X. (2020) Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. *Mol Biol Evol*
5. Zhang, T., Wu, Q., and Zhang, Z. (2020) Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol* **30**, 1346-1351 e1342
6. Alijotas-Reig, J., Esteve-Valverde, E., Belizna, C., Selva-O'Callaghan, A., Pardos-Gea, J., Quintana, A., Mekinian, A., Anunciacion-Llunell, A., and Miro-Mur, F. (2020) Immunomodulatory therapy for the management of severe COVID-19. Beyond the anti-viral therapy: A comprehensive review. *Autoimmun Rev* **19**, 102569
7. Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., Hohmann, E., Chu, H. Y., Luetkemeyer, A., Kline, S., Lopez de Castilla, D., Finberg, R. W., Dierberg, K., Tapson, V., Hsieh, L., Patterson, T. F., Paredes, R., Sweeney, D. A., Short, W. R., Touloumi, G., Lye, D. C., Ohmagari, N., Oh, M. D., Ruiz-Palacios, G. M., Benfield, T., Fatkenheuer, G., Kortepeter, M. G., Atmar, R. L., Creech, C. B., Lundgren, J., Babiker, A. G., Pett, S., Neaton, J. D., Burgess, T. H., Bonnett, T., Green, M., Makowski, M., Osinusi, A., Nayak, S., Lane, H. C., and Members, A.-S. G. (2020) Remdesivir for the Treatment of Covid-19 - Preliminary Report. *N Engl J Med*
8. Beun, R., Kusadasi, N., Sikma, M., Westerink, J., and Huisman, A. (2020) Thromboembolic events and apparent heparin resistance in patients infected with SARS-CoV-2. *Int J Lab Hematol* **42 Suppl 1**, 19-20
9. Dashti-Khavidaki, S., and Khalili, H. (2020) Considerations for Statin Therapy in Patients with COVID-19. *Pharmacotherapy* **40**, 484-486
10. Fedson, D. S., Opal, S. M., and Rordam, O. M. (2020) Hiding in Plain Sight: an Approach to Treating Patients with Severe COVID-19 Infection. *mBio* **11**
11. Shi, Q., Zhou, Q., Wang, X., Liao, J., Yu, Y., Wang, Z., Lu, S., Ma, Y., Xun, Y., Luo, X., Li, W., Fukuoka, T., Ahn, H. S., Lee, M. S., Luo, Z., Liu, E., Chen, Y., Li, Q., Yang, K., Guan, Q., Evidence, C.-., and Recommendations Working, G. (2020) Potential effectiveness and safety of antiviral agents in children with coronavirus disease 2019: a rapid review and meta-analysis. *Ann Transl Med* **8**, 624
12. Tang, N., Bai, H., Chen, X., Gong, J., Li, D., and Sun, Z. (2020) Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J Thromb Haemost* **18**, 1094-1099
13. 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., and Pohlmann, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278
14. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454
15. Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019) Exploitation of glycosylation in enveloped virus pathobiology. *Biochim Biophys Acta Gen Subj* **1863**, 1480-1497

16. Li, F. (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **3**, 237-261
17. Lei, C., Qian, K., Li, T., Zhang, S., Fu, W., Ding, M., and Hu, S. (2020) Neutralization of SARS-CoV-2 spike pseudotyped virus by recombinant ACE2-Ig. *Nat Commun* **11**, 2070
18. Monteil, V., Kwon, H., Prado, P., Hagelkruys, A., Wimmer, R. A., Stahl, M., Leopoldi, A., Garreta, E., Hurtado Del Pozo, C., Prosper, F., Romero, J. P., Wirnsberger, G., Zhang, H., Slutsky, A. S., Conder, R., Montserrat, N., Mirazimi, A., and Penninger, J. M. (2020) Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **181**, 905-913 e907
19. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*
20. Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., McLellan, J. S., Wilson, I. A., Bowden, T. A., Ward, A. B., and Crispin, M. (2020) Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat Commun* **11**, 2688
21. Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A., Weber, F., and Kochs, G. (2007) The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**, 304-315
22. Ujike, M., and Taguchi, F. (2015) Incorporation of spike and membrane glycoproteins into coronavirus virions. *Viruses* **7**, 1700-1725
23. Duan, H., Chen, X., Boyington, J. C., Cheng, C., Zhang, Y., Jafari, A. J., Stephens, T., Tsybovsky, Y., Kalyuzhniy, O., Zhao, P., Menis, S., Nason, M. C., Normandin, E., Mukhamedova, M., DeKosky, B. J., Wells, L., Schief, W. R., Tian, M., Alt, F. W., Kwong, P. D., and Mascola, J. R. (2018) Glycan Masking Focuses Immune Responses to the HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor Antibodies. *Immunity* **49**, 301-311 e305
24. Escolano, A., Gristick, H. B., Abernathy, M. E., Merkenschlager, J., Gautam, R., Oliveira, T. Y., Pai, J., West, A. P., Jr., Barnes, C. O., Cohen, A. A., Wang, H., Golijanin, J., Yost, D., Keeffe, J. R., Wang, Z., Zhao, P., Yao, K. H., Bauer, J., Nogueira, L., Gao, H., Voll, A. V., Montefiori, D. C., Seaman, M. S., Gazumyan, A., Silva, M., McGuire, A. T., Stamatatos, L., Irvine, D. J., Wells, L., Martin, M. A., Bjorkman, P. J., and Nussenzweig, M. C. (2019) Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. *Nature* **570**, 468-473
25. Seabright, G. E., Cottrell, C. A., van Gils, M. J., D'Addabbo, A., Harvey, D. J., Behrens, A. J., Allen, J. D., Watanabe, Y., Scaringi, N., Polveroni, T. M., Maker, A., Vasiljevic, S., de Val, N., Sanders, R. W., Ward, A. B., and Crispin, M. (2020) Networks of HIV-1 Envelope Glycans Maintain Antibody Epitopes in the Face of Glycan Additions and Deletions. *Structure*
26. Yu, W. H., Zhao, P., Draghi, M., Arevalo, C., Karsten, C. B., Suscovich, T. J., Gunn, B., Streeck, H., Brass, A. L., Tiemeyer, M., Seaman, M., Mascola, J. R., Wells, L., Lauffenburger, D. A., and Alter, G. (2018) Exploiting glycan topography for computational design of Env glycoprotein antigenicity. *PLoS Comput Biol* **14**, e1006093

27. Tikellis, C., Bernardi, S., and Burns, W. C. (2011) Angiotensin-converting enzyme 2 is a key modulator of the renin-angiotensin system in cardiovascular and renal disease. *Curr Opin Nephrol Hypertens* **20**, 62-68
28. Lambert, D. W., Yarski, M., Warner, F. J., Thornhill, P., Parkin, E. T., Smith, A. I., Hooper, N. M., and Turner, A. J. (2005) Tumor necrosis factor-alpha convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2). *J Biol Chem* **280**, 30113-30119
29. Epelman, S., Shrestha, K., Troughton, R. W., Francis, G. S., Sen, S., Klein, A. L., and Tang, W. H. (2009) Soluble angiotensin-converting enzyme 2 in human heart failure: relation with myocardial function and clinical outcomes. *J Card Fail* **15**, 565-571
30. Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S. K., Huang, I. C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., and Farzan, M. (2005) Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* **24**, 1634-1643
31. Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat Rev Immunol* **8**, 874-887
32. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3-49
33. Yu, J., Tostanoski, L. H., Peter, L., Mercado, N. B., McMahan, K., Mahrokhian, S. H., Nkolola, J. P., Liu, J., Li, Z., Chandrashekar, A., Martinez, D. R., Loos, C., Atyeo, C., Fischinger, S., Burke, J. S., Slein, M. D., Chen, Y., Zuiani, A., FJ, N. L., Travers, M., Habibi, S., Pessaint, L., Van Ry, A., Blade, K., Brown, R., Cook, A., Finneyfrock, B., Dodson, A., Teow, E., Velasco, J., Zahn, R., Wegmann, F., Bondzie, E. A., Dagotto, G., Gebre, M. S., He, X., Jacob-Dolan, C., Kirilova, M., Kordana, N., Lin, Z., Maxfield, L. F., Nampanya, F., Nityanandam, R., Ventura, J. D., Wan, H., Cai, Y., Chen, B., Schmidt, A. G., Wesemann, D. R., Baric, R. S., Alter, G., Andersen, H., Lewis, M. G., and Barouch, D. H. (2020) DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science*
34. Almagro Armenteros, J. J., Tsirigos, K. D., Sonderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423
35. Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007) Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* **282**, 9127-9142
36. Aoki, K., Porterfield, M., Lee, S. S., Dong, B., Nguyen, K., McGlamry, K. H., and Tiemeyer, M. (2008) The diversity of O-linked glycans expressed during Drosophila melanogaster development reflects stage- and tissue-specific requirements for cell signaling. *J Biol Chem* **283**, 30385-30400
37. Weatherly, D. B., Arpinar, F. S., Porterfield, M., Tiemeyer, M., York, W. S., and Ranzinger, R. (2019) GRITS Toolbox-a freely available software for processing, annotating and archiving glycomics mass spectrometry data. *Glycobiology* **29**, 452-460
38. Wang, S., Voronin, Y., Zhao, P., Ishihara, M., Mehta, N., Porterfield, M., Chen, Y., Bartley, C., Hu, G., Han, D., Wells, L., Tiemeyer, M., and Lu, S. (2020) Glycan Profiles of gp120 Protein Vaccines from Four Major HIV-1 Subtypes Produced from Different Host Cell Lines under Non-GMP or GMP Conditions. *J Virol* **94**

39. Ruiz-Canada, C., Kelleher, D. J., and Gilmore, R. (2009) Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* **136**, 272-283
40. Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907
41. Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L., and Larsen, M. R. (2012) Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res* **11**, 1949-1957
42. Zhou, T., Doria-Rose, N. A., Cheng, C., Stewart-Jones, G. B. E., Chuang, G. Y., Chambers, M., Druz, A., Geng, H., McKee, K., Kwon, Y. D., O'Dell, S., Sastry, M., Schmidt, S. D., Xu, K., Chen, L., Chen, R. E., Louder, M. K., Pancera, M., Wanninger, T. G., Zhang, B., Zheng, A., Farney, S. K., Foulds, K. E., Georgiev, I. S., Joyce, M. G., Lemmin, T., Narpala, S., Rawi, R., Soto, C., Todd, J. P., Shen, C. H., Tsybovsky, Y., Yang, Y., Zhao, P., Haynes, B. F., Stamatatos, L., Tiemeyer, M., Wells, L., Scorpio, D. G., Shapiro, L., McDermott, A. B., Mascola, J. R., and Kwong, P. D. (2017) Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Rep* **19**, 719-732
43. Liu, M. Q., Zeng, W. F., Fang, P., Cao, W. Q., Liu, C., Yan, G. Q., Zhang, Y., Peng, C., Wu, J. Q., Zhang, X. J., Tu, H. J., Chi, H., Sun, R. X., Cao, Y., Dong, M. Q., Jiang, B. Y., Huang, J. M., Shen, H. L., Wong, C. C. L., He, S. M., and Yang, P. Y. (2017) pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun* **8**, 438
44. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* **Chapter 13**, Unit13 20
45. Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020) Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*
46. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263
47. Ferreira, R. C., Grant, O. C., Moyo, T., Dorfman, J. R., Woods, R. J., Travers, S. A., and Wood, N. T. (2018) Structural Rearrangements Maintain the Glycan Shield of an HIV-1 Envelope Trimer After the Loss of a Glycan. *Sci Rep* **8**, 15031
48. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220
49. Kristic, J., Vuckovic, F., Menni, C., Klaric, L., Keser, T., Beceheli, I., Pucic-Bakovic, M., Novokmet, M., Mangino, M., Thaqi, K., Rudan, P., Novokmet, N., Sarac, J., Missoni, S., Kolcic, I., Polasek, O., Rudan, I., Campbell, H., Hayward, C., Aulchenko, Y., Valdes, A., Wilson, J. F., Gornik, O., Primorac, D., Zoldos, V., Spector, T., and Lauc, G. (2014) Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* **69**, 779-789
50. Pavic, T., Dilber, D., Kifer, D., Selak, N., Keser, T., Ljubicic, D., Vukic Dugac, A., Lauc, G., Rumora, L., and Gornik, O. (2018) N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. *J Transl Med* **16**, 323

51. Rudman, N., Gornik, O., and Lauc, G. (2019) Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS Lett* **593**, 1598-1615
52. Gebrehiwot, A. G., Melka, D. S., Kassaye, Y. M., Rehan, I. F., Rangappa, S., Hinou, H., Kamiyama, T., and Nishimura, S. I. (2018) Healthy human serum N-glycan profiling reveals the influence of ethnic variation on the identified cancer-relevant glycan biomarkers. *PLoS One* **13**, e0209515
53. Thieker, D. F., Hadden, J. A., Schulten, K., and Woods, R. J. (2016) 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology* **26**, 786-787
54. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277
55. Shaik, M. M., Peng, H., Lu, J., Rits-Volloch, S., Xu, C., Liao, M., and Chen, B. (2019) Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* **565**, 318-323
56. Anumula, K. R., and Taylor, P. B. (1992) A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates. *Anal Biochem* **203**, 101-108
57. 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., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324
58. Wells, L., Hart, G. W., and Athens Guidelines for the Publication of Glycomics, D. (2013) Glycomics: building upon proteomics to advance glycosciences. *Mol Cell Proteomics* **12**, 833-835
59. Madeira, F., Madhusoodanan, N., Lee, J., Tivey, A. R. N., and Lopez, R. (2019) Using EMBL-EBI Services via Web Interface and Programmatically via Web Services. *Curr Protoc Bioinformatics* **66**, e74
60. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., and Woods, R. J. (2014) Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J Comput Chem* **35**, 526-539
61. Grant, O. C., Xue, X., Ra, D., Khatamian, A., Foley, B. L., and Woods, R. J. (2016) Gly-Spec: a webtool for predicting glycan specificity by integrating glycan array screening data and 3D structure. *Glycobiology* **26**, 1027-1028
62. Peng, W., de Vries, R. P., Grant, O. C., Thompson, A. J., McBride, R., Tsogtbaatar, B., Lee, P. S., Razi, N., Wilson, I. A., Woods, R. J., and Paulson, J. C. (2017) Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. *Cell Host Microbe* **21**, 23-34
63. Gotz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* **8**, 1542-1555
64. Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **9**, 3878-3888

65. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622-655
66. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-3713
67. Darden, T. A., and Pedersen, L. G. (1993) Molecular modeling: an experimental tool. *Environ Health Perspect* **101**, 410-412
68. Brindley, M. A., Chaudhury, S., and Plemper, R. K. (2015) Measles virus glycoprotein complexes preassemble intracellularly and relax during transport to the cell surface in preparation for fusion. *J Virol* **89**, 1230-1241
69. Paal, M., Heel, T., Schneider, R., and Auer, B. (2009) A novel Ecotin-Ubiquitin-Tag (ECUT) for efficient, soluble peptide production in the periplasm of Escherichia coli. *Microb Cell Fact* **8**, 7
70. Brindley, M. A., and Plemper, R. K. (2010) Blue native PAGE and biomolecular complementation reveal a tetrameric or higher-order oligomer organization of the physiological measles virus attachment protein H. *J Virol* **84**, 12174-12184

SARS-CoV-2 S Trimer Immunogen

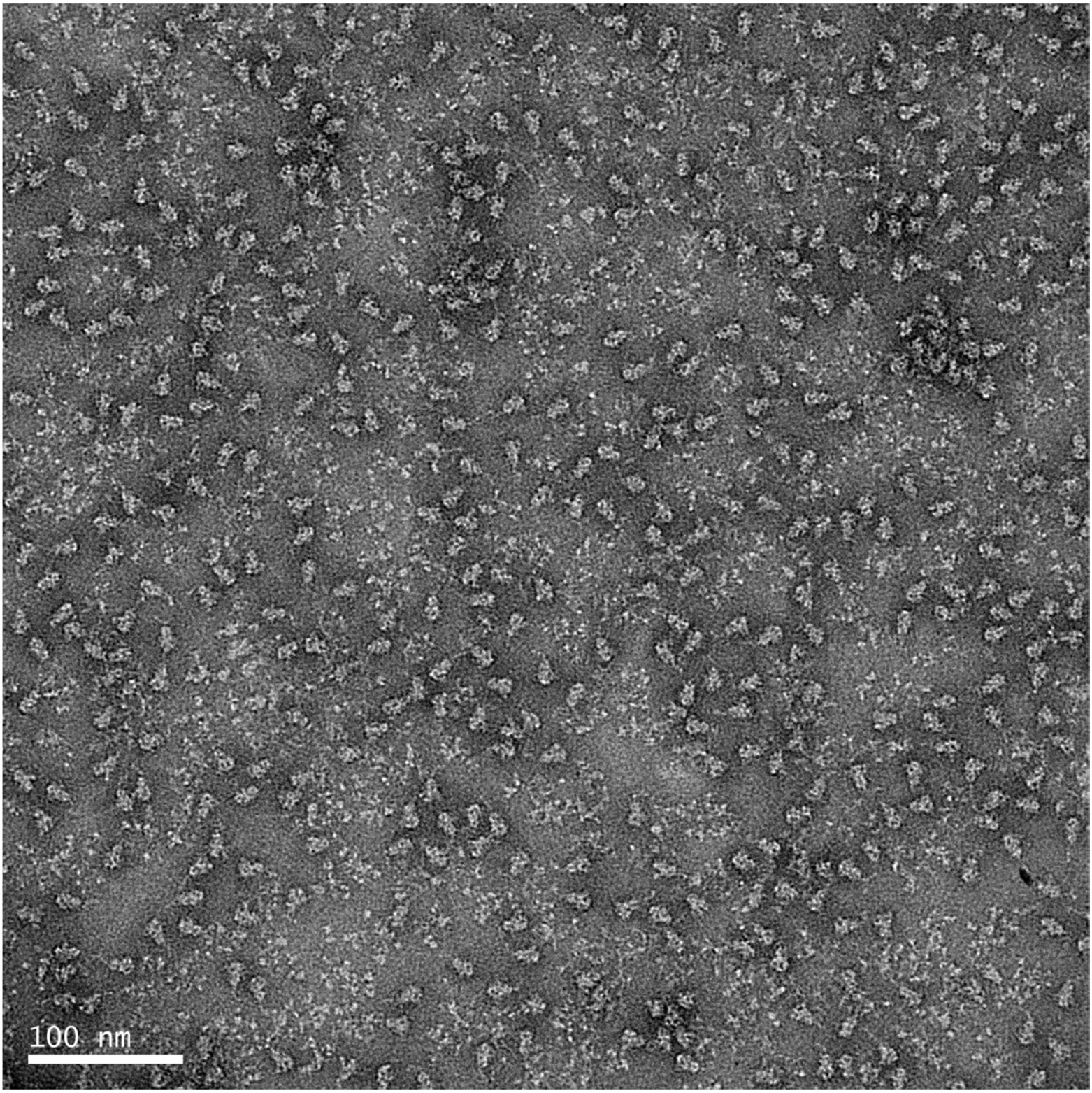
~~MFVFLVLLPL VSS~~QCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS 60
NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV 120
NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFLMDLE 180
GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT 240
LLALHRSYLT PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK 300
CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN 360
CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF VIRGDEVRQI APGQTGKIAD 420
YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST EIYQAGSTPC 480
NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN 540
FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP 600
GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY 660
ECDIPIGAGI CASYQTQTNS PGGSGSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI 720
SVTTEILPVS MTKTSVDCTM YICGDSTECS NLLLQYGSFC TQLNRALTGI AVEQDKNTQE 780
VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC 840
LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM 900
QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN 960
TLVKQLSSNF GAISSVLNDI LSRLDPPEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA 1020
SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA 1080
ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP 1140
LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL 1200
QELGKYEQ*GS GGYIPEAPRD GQAYVRKDGE WVLLSTFLGG SHHHHHH* 1247

Soluble Human ACE2

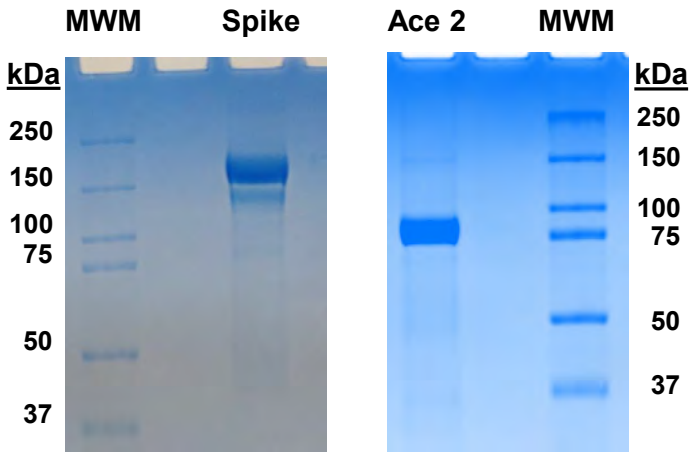
~~MSSSSWLLLS LVAVTAA~~QST IEEQAKTFLD KFNHEAEDLF YQSSLASWNY NTNITEENVQ 60
NMNNAGDKWS AFLKEQSTLA QMYPLQEIQN LTVKLQLQAL QQNGSSVLSE DKSKRLNTIL 120
NTMSTIYSTG KVCNPDNPQE CLLLEPGLNE IMANSLDYNE RLWAWESWRS EVGKQLRPLY 180
EEYVVLKNEM ARANHYEDYG DYWRGDYEVN GVDGYDYSRG QLIEDVEHTF EEIKPLYEHL 240
HAYVRAKLMN AYPSYISPIG CLPAHLLGDM WGRFWTNLYS LTVPFGQKPN IDVTDAMVDQ 300
AWDAQRIFKE AEKFFVSVGL PNMTQGFWEN SMLTDPGNVQ KAVCHPTAWD LGKGDFRILM 360
CTKVTMDDFL TAHHEMGHIQ YDMAYAAQPF LLRNGANEGF HEAVGEIMSL SAATPKHLKS 420
IGLLSPDFQE DNETEINFLL KQALTIVGTL PFTYMLEKWR WMVFKGEIPK DQWMKKWWEM 480
KREIVGVVEP VPHDETYCDP ASLFHVSNDY SFIRYYTRTL YQFQFQEALC QAAKHEGPLH 540
KCDISNSTEA GQKLFNMLRL GKSEPWTLAL ENVVGAKNMN VRPLLNYFEP LFTWLKDQNK 600
NSFVGWSTDW SPYAD*SGGSH HHHHH* 625

Q = Pyroglutamate, Mature N-terminus **N = Canonical N-Glycosylation Sequon Asparagine** ***Italics* = "Tag"**
~~Strikethrough~~ = Signal Peptide

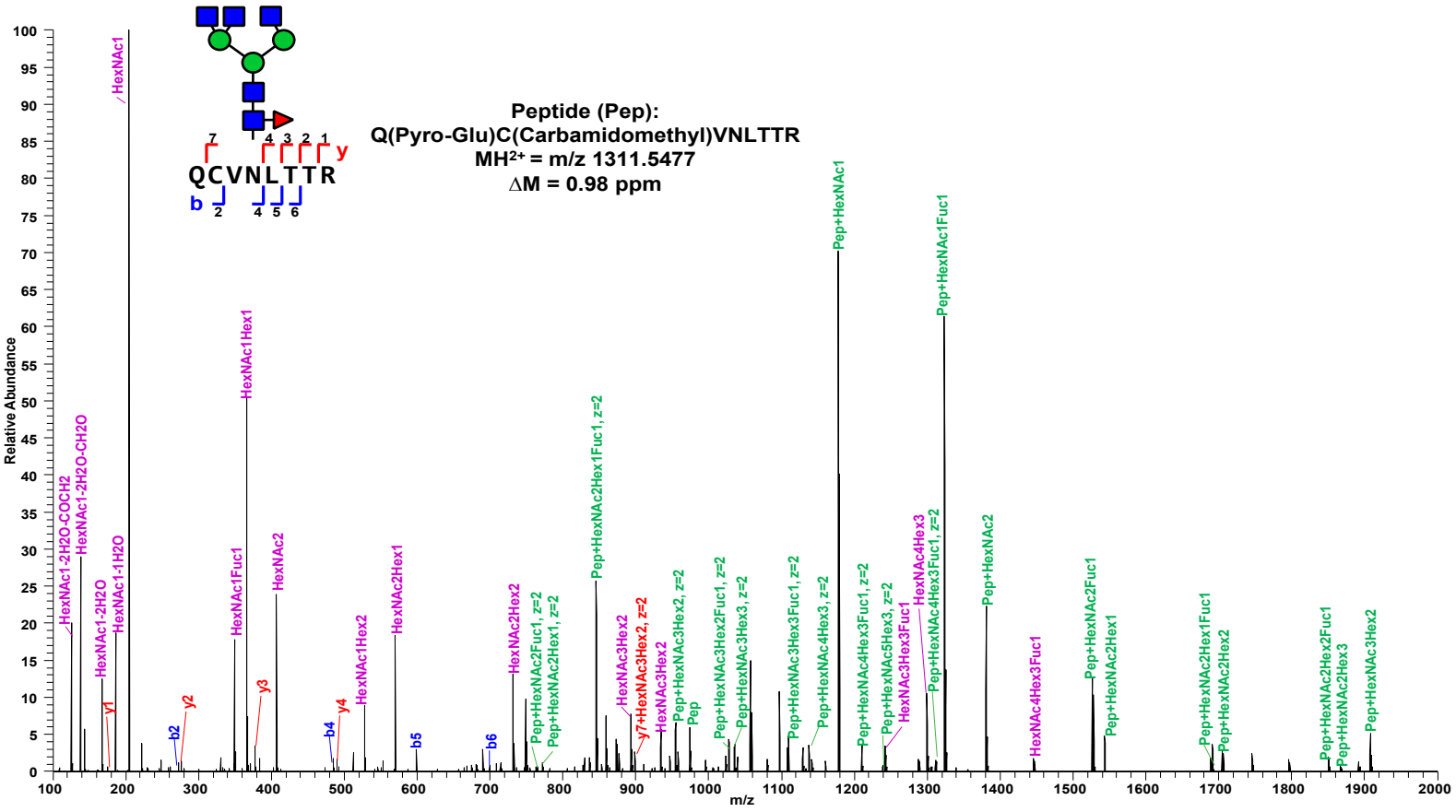
1A



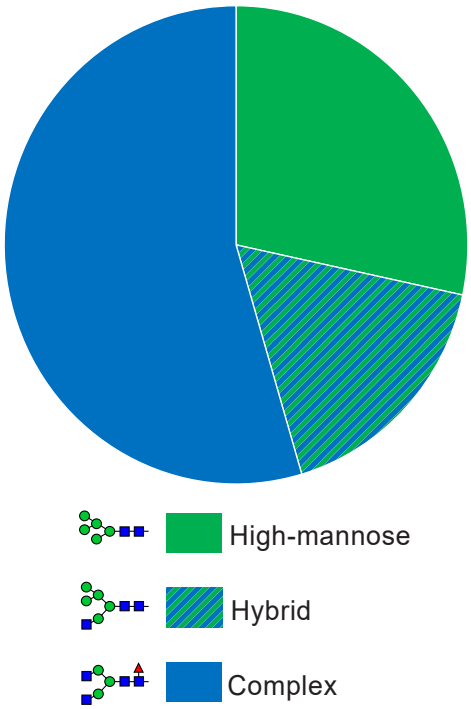
1B

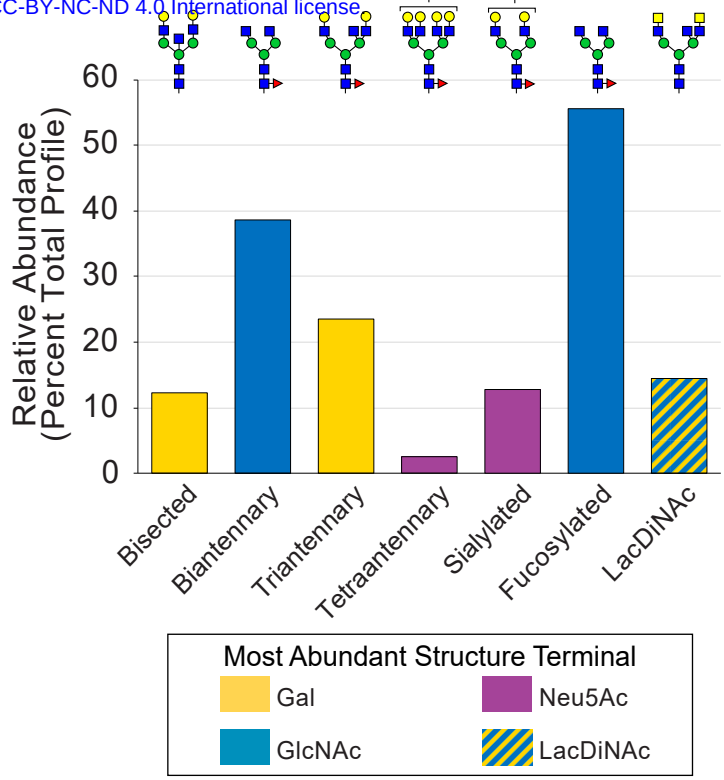


1C

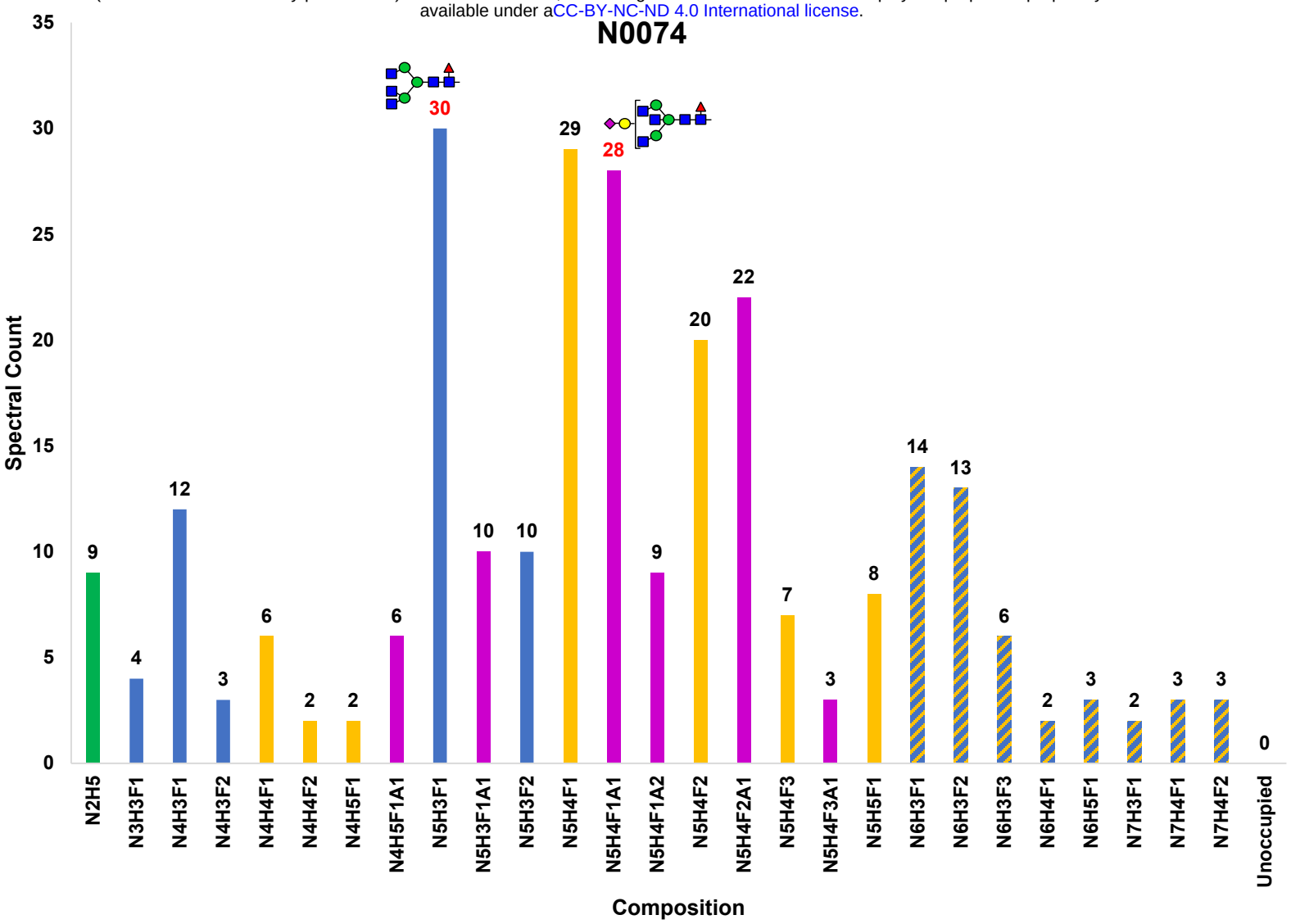


1D

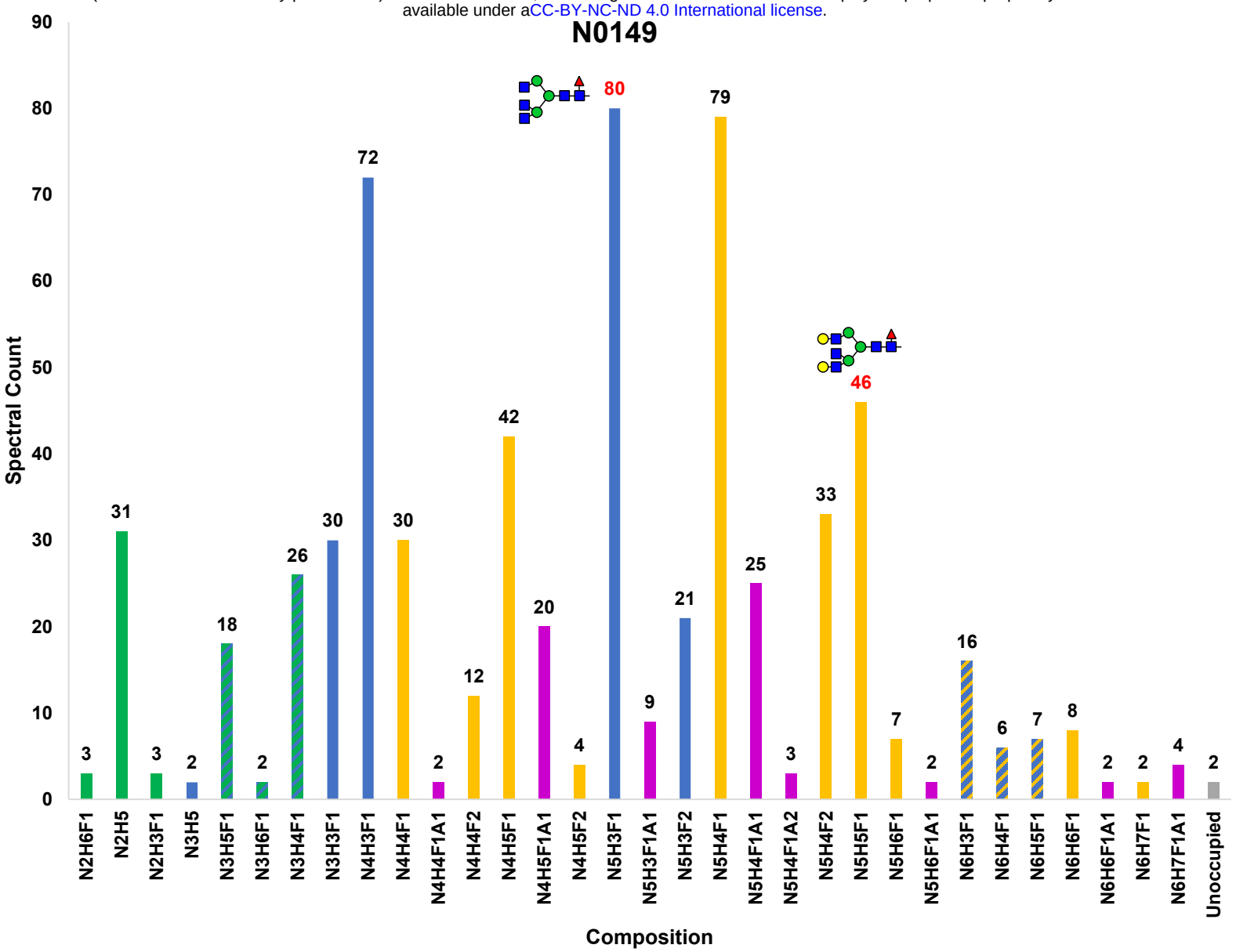




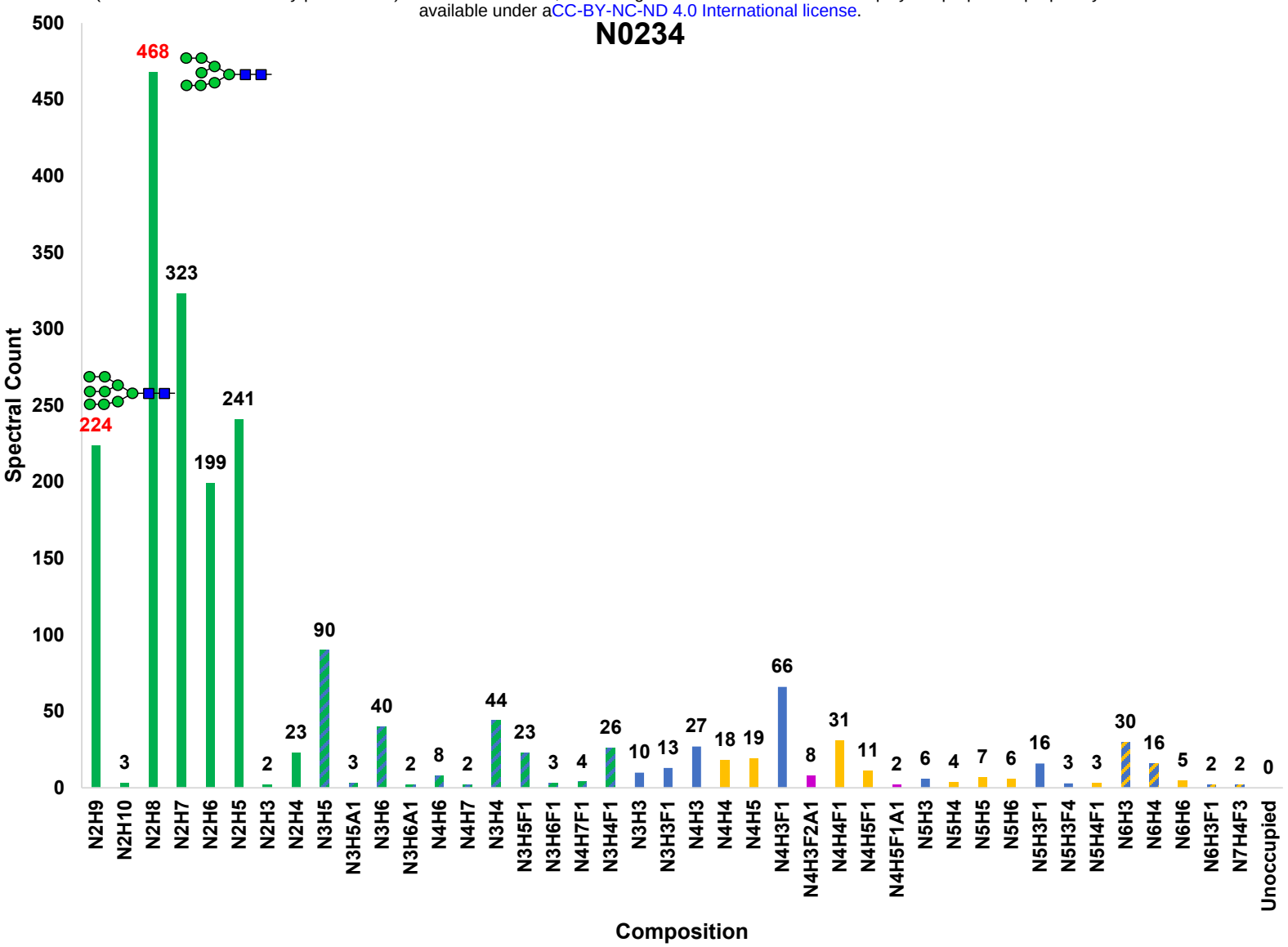
2A



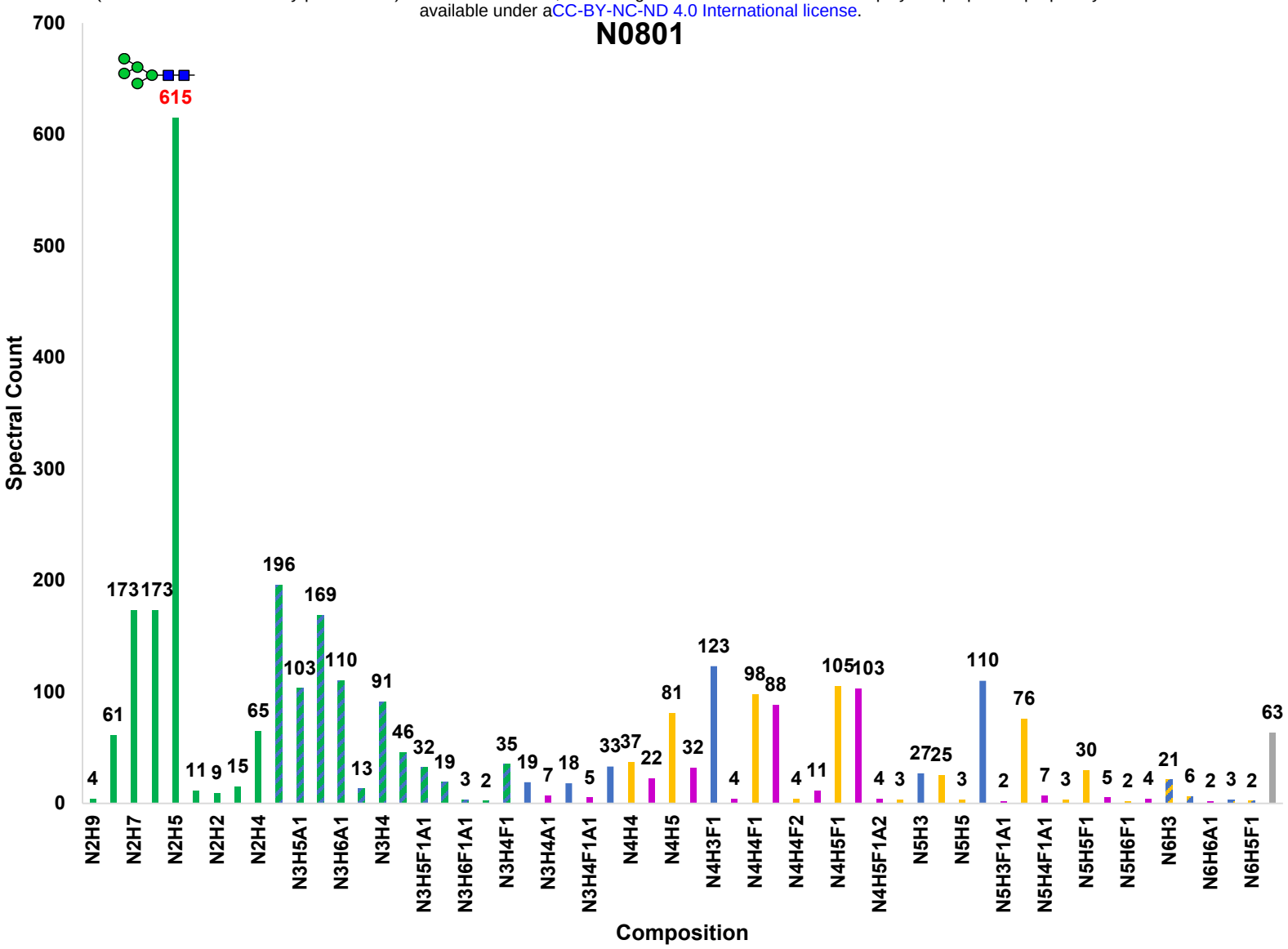
2B



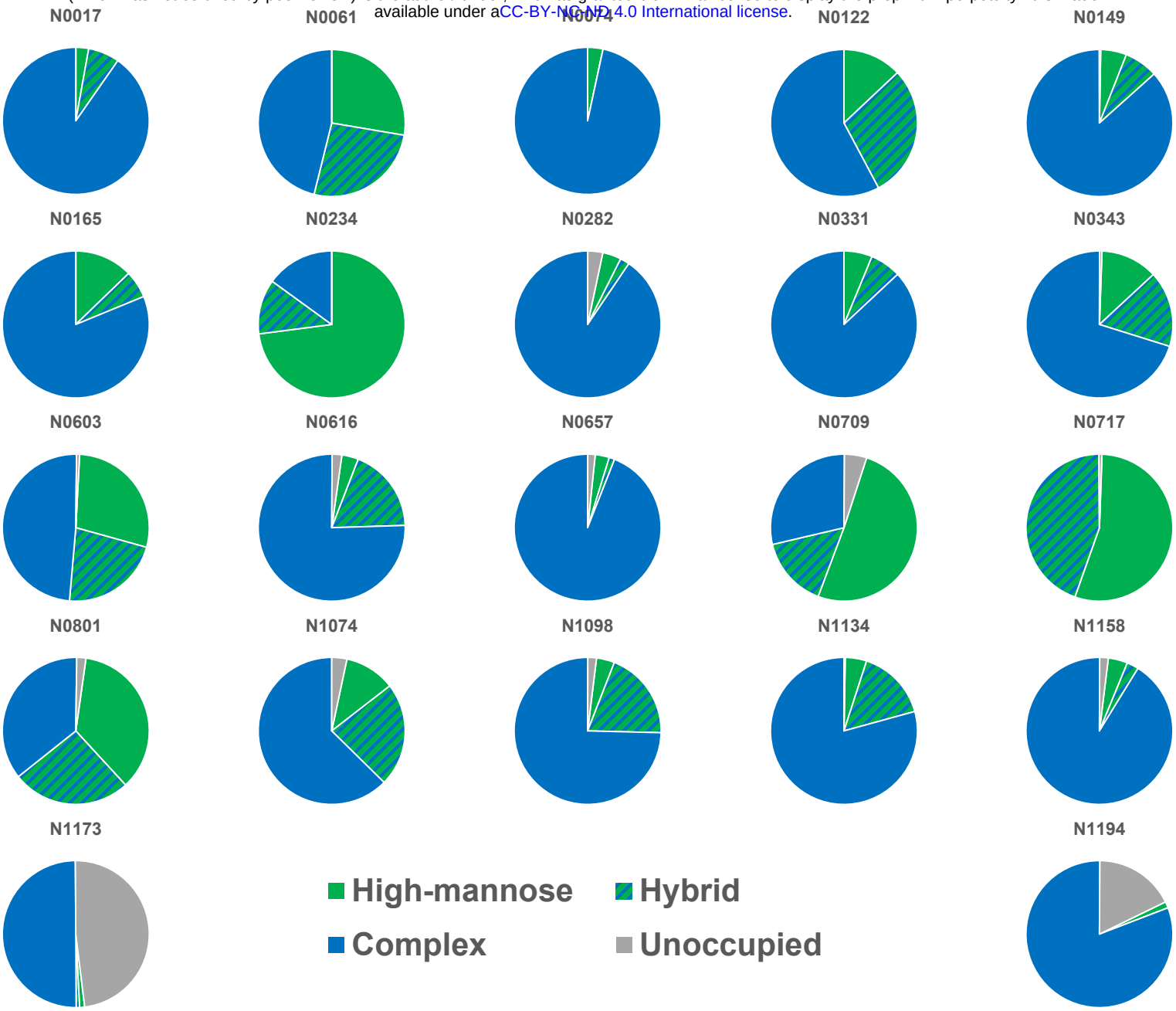
2C

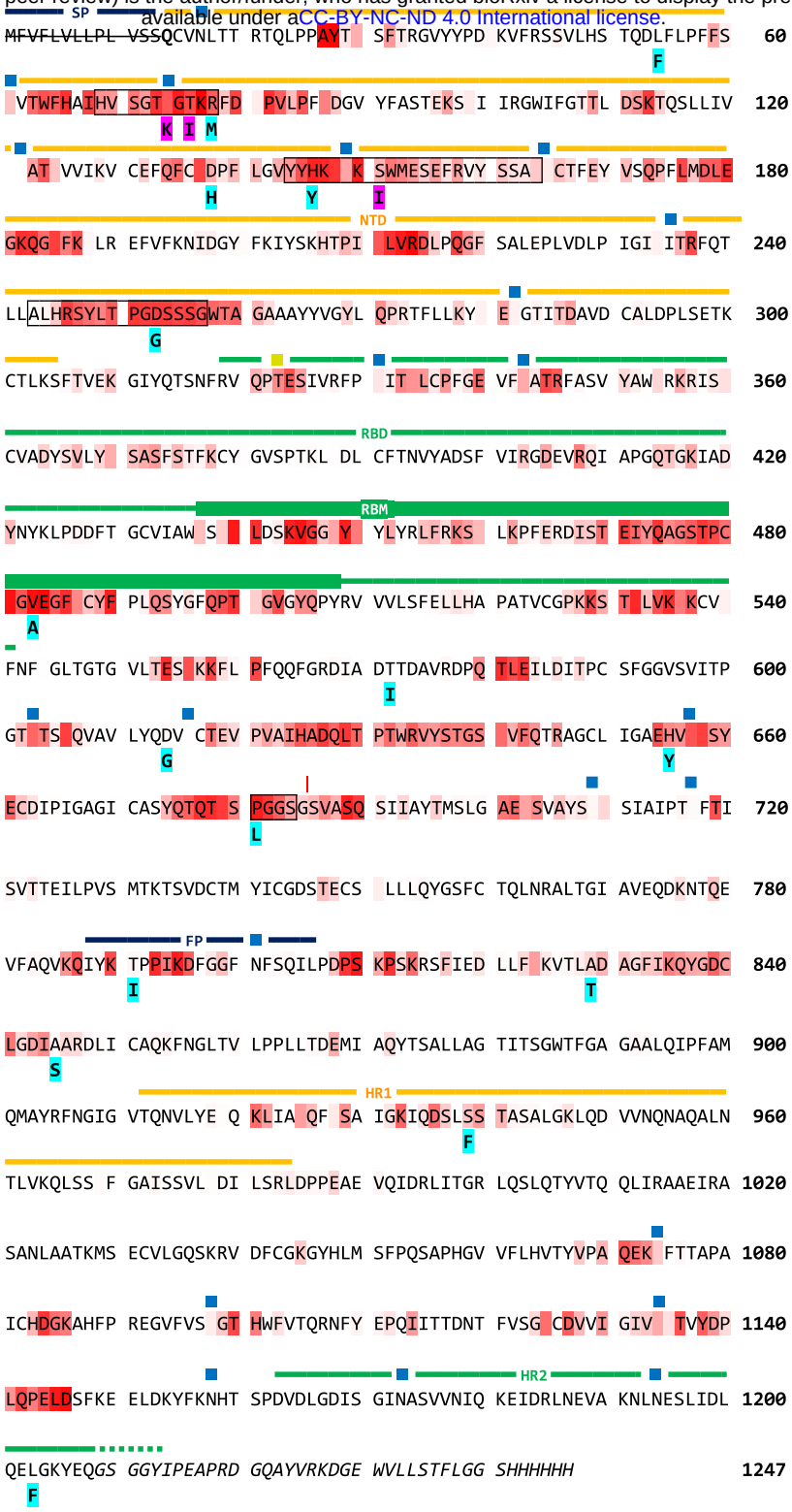


2D



2E

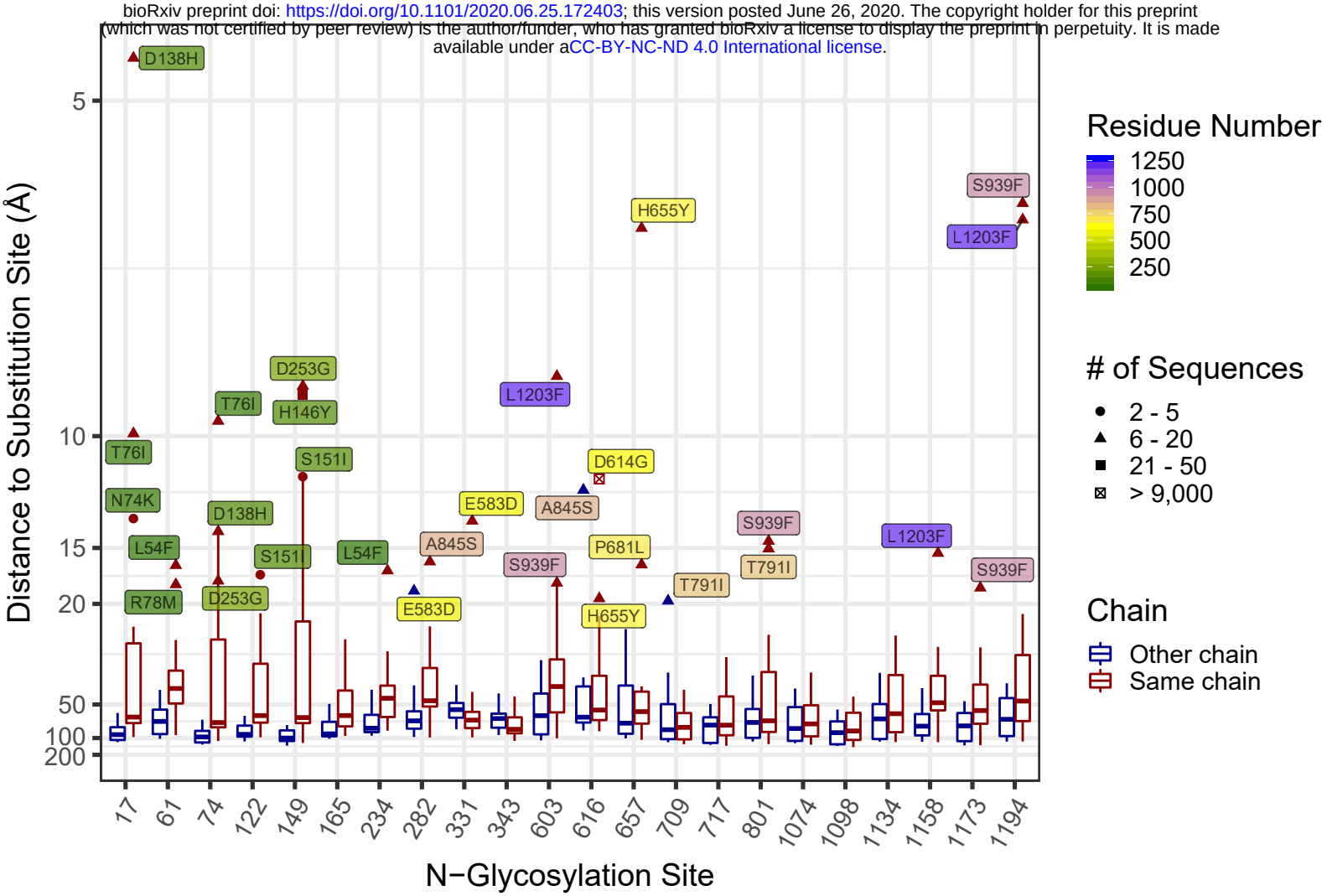




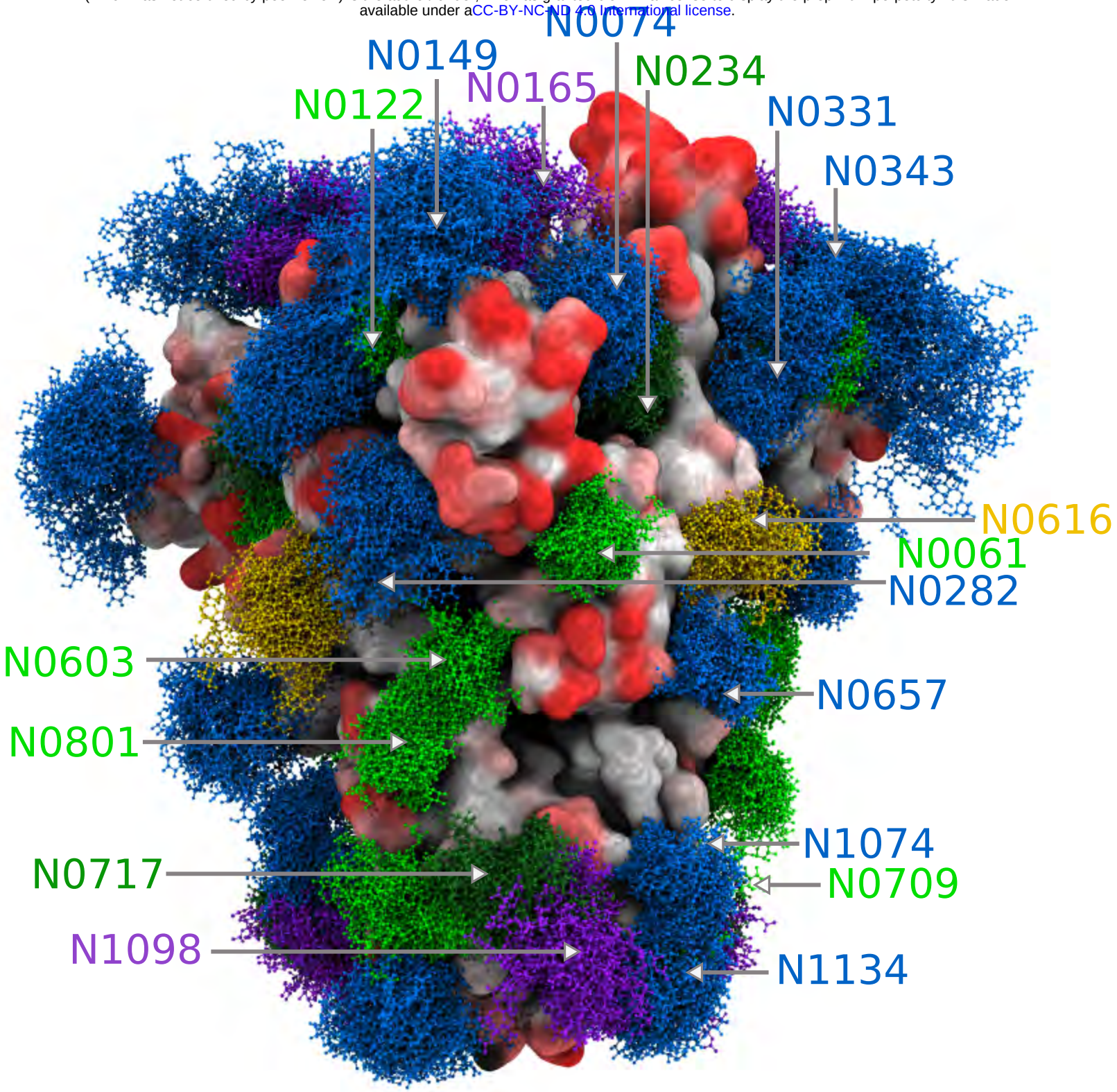
Key

SP	= Signal Peptide	NTD	= N-Terminal Domain
RBD	= Receptor Binding Domain	RBM	= Receptor Binding Motif
FP	= Fusion Peptide	HR1	= Heptad Repeat 1
HR2	= Heptad Repeat 2		= Heptad Repeat 2 WT extent
	= N-glycosite		= O-glycosite
X	= Variant, no sequon impact	X	= Variant, sequon ablated
|	= Furin cleavage site in WT	...XXX...	= Divergent region w/ insert
Italics	= "Tag"	~~Strikethrough~~	= Signal Peptide
...XXX...	= Antigen accessibility		

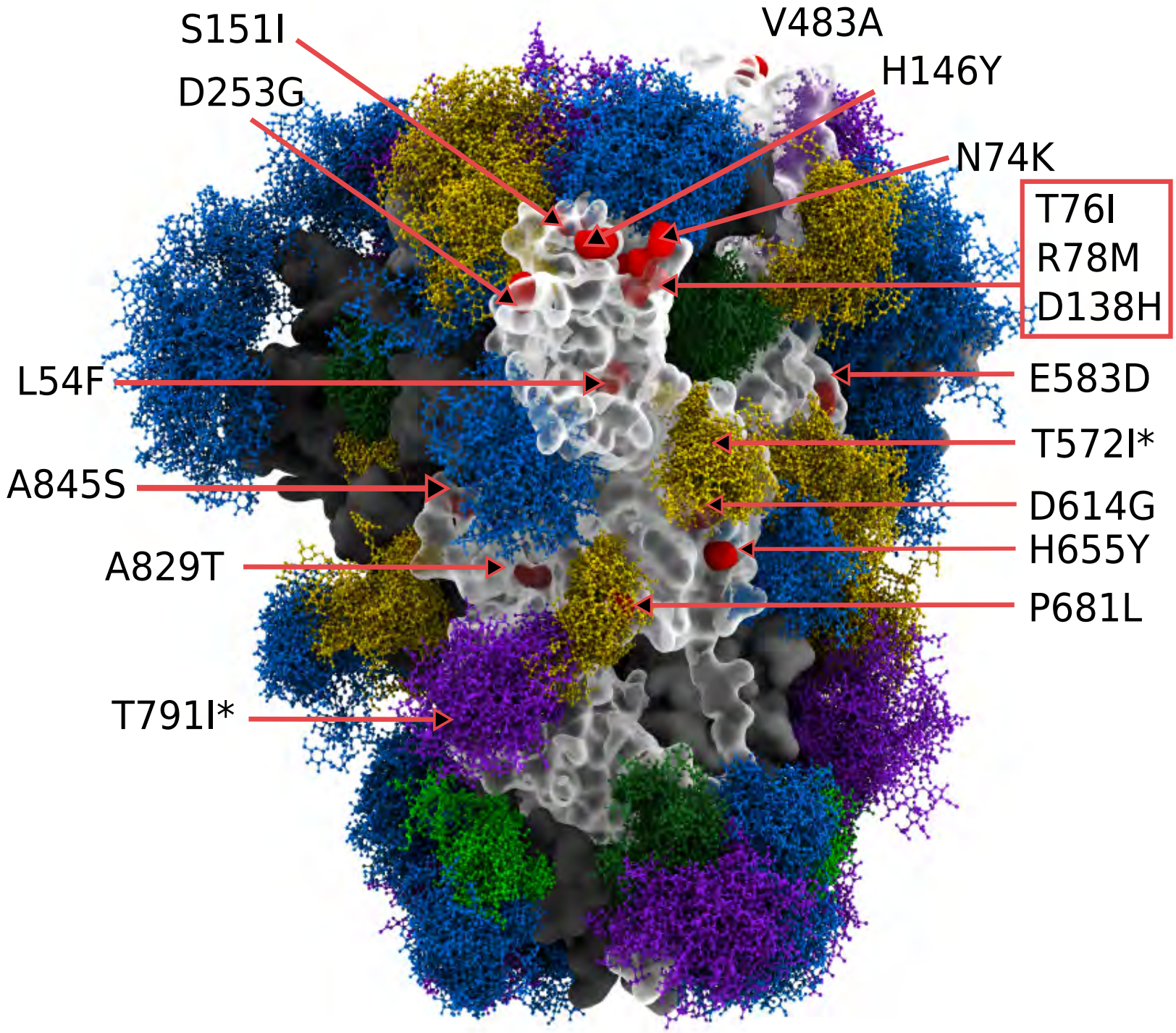
4A



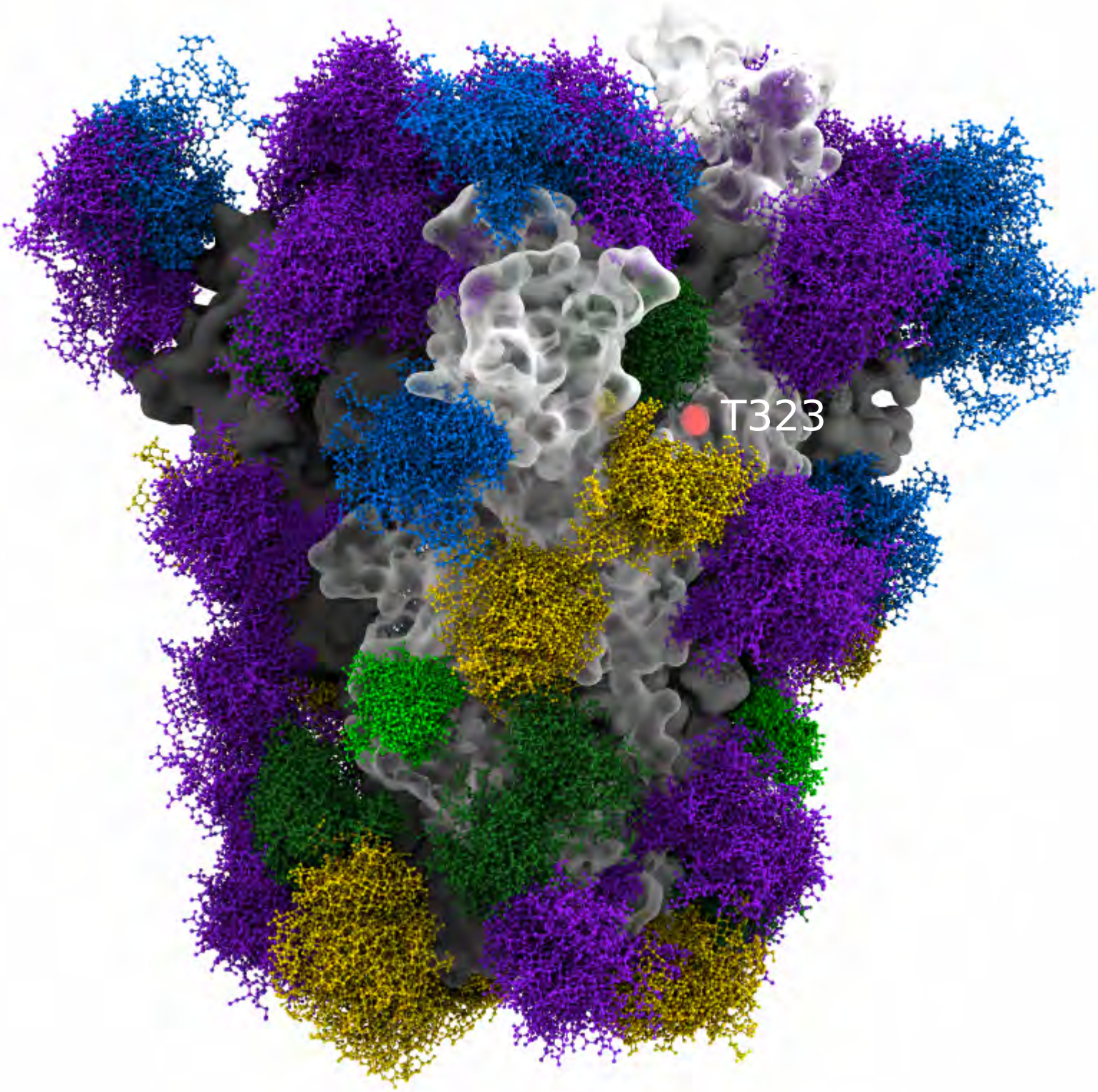
4B



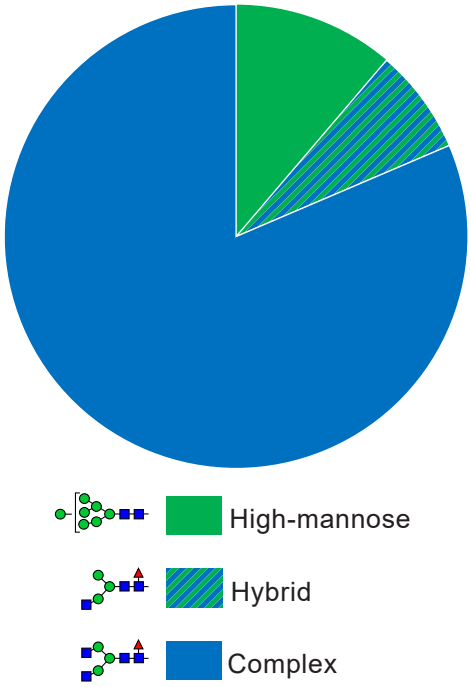
4C

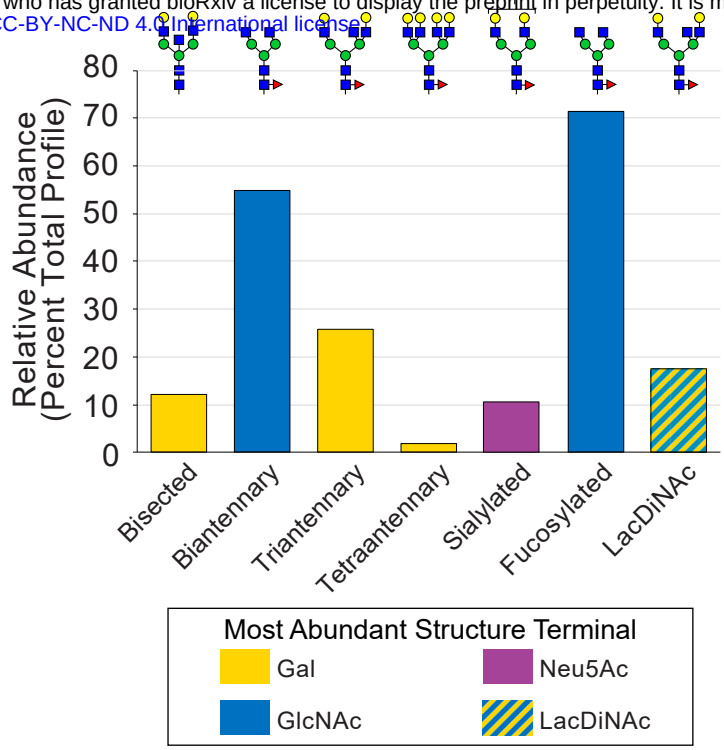


4D

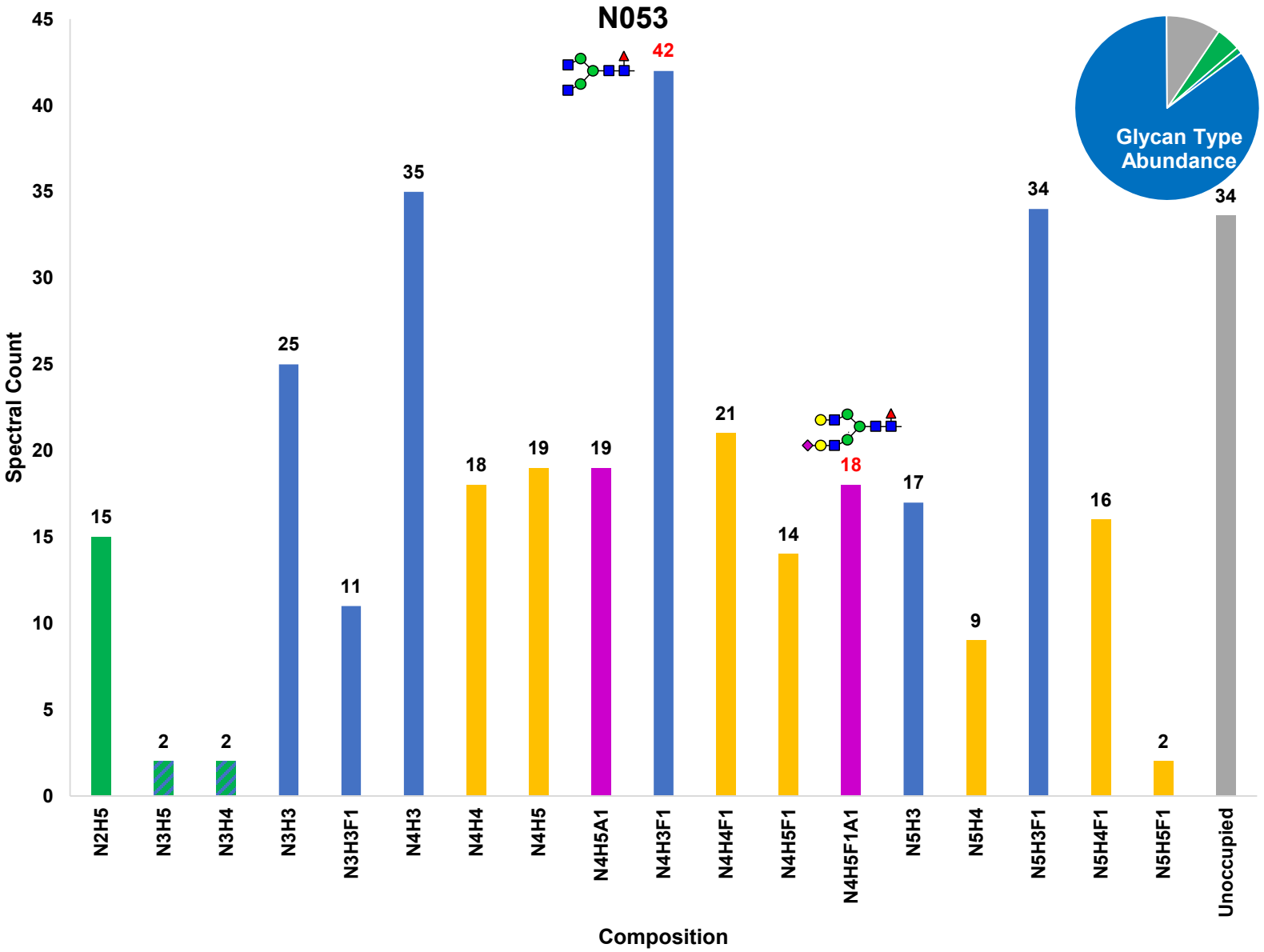


4E

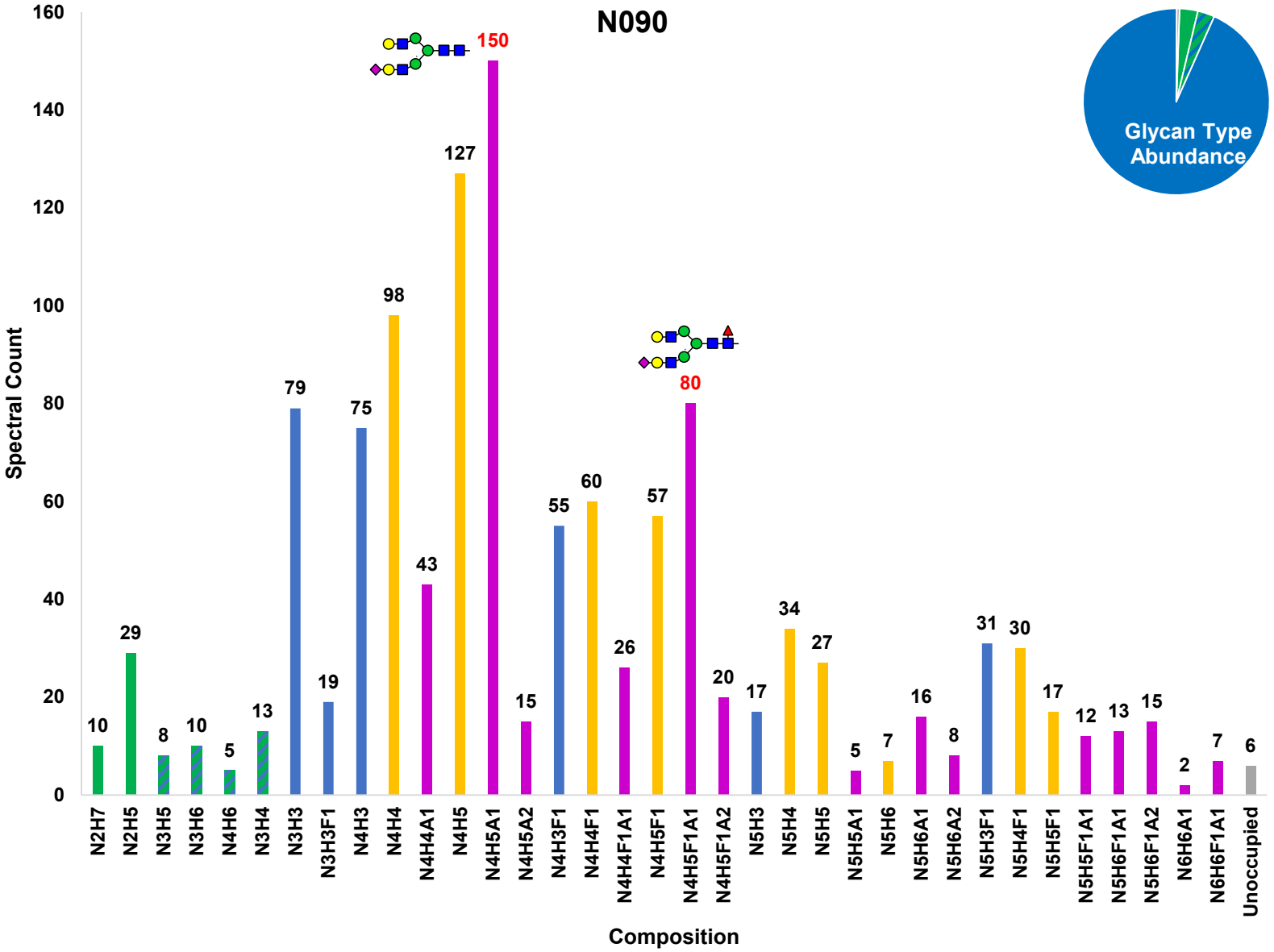




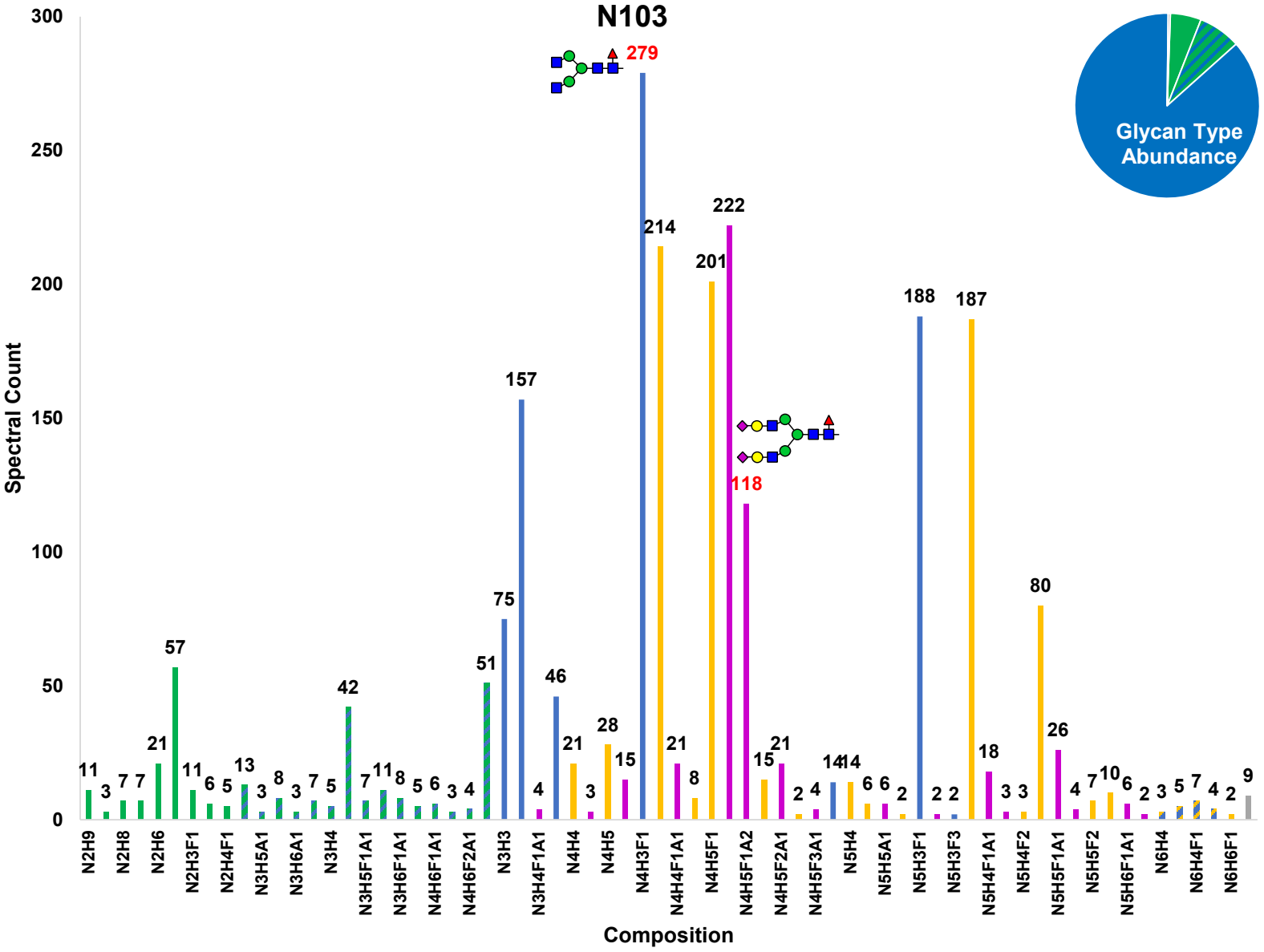
5A



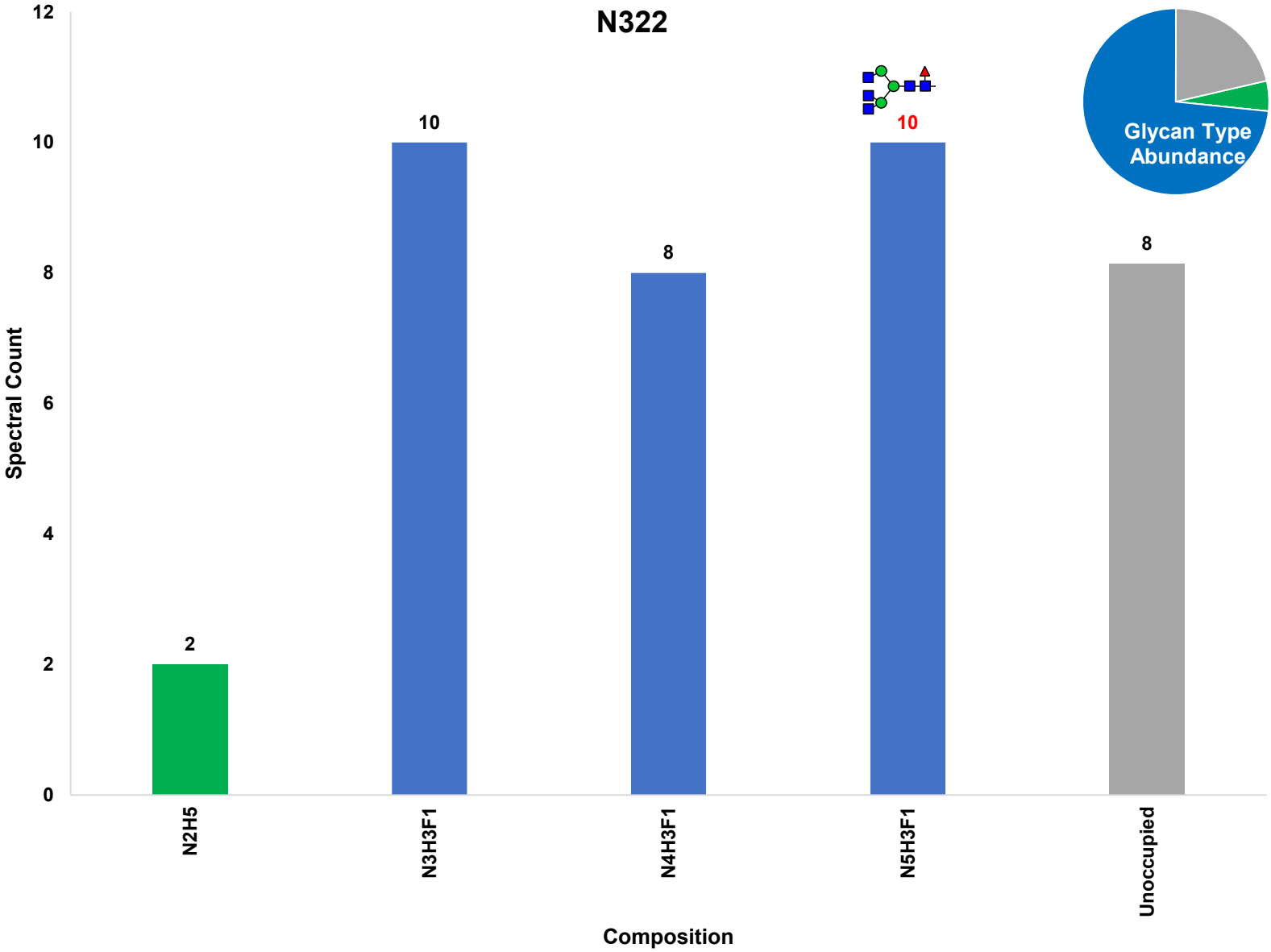
5B



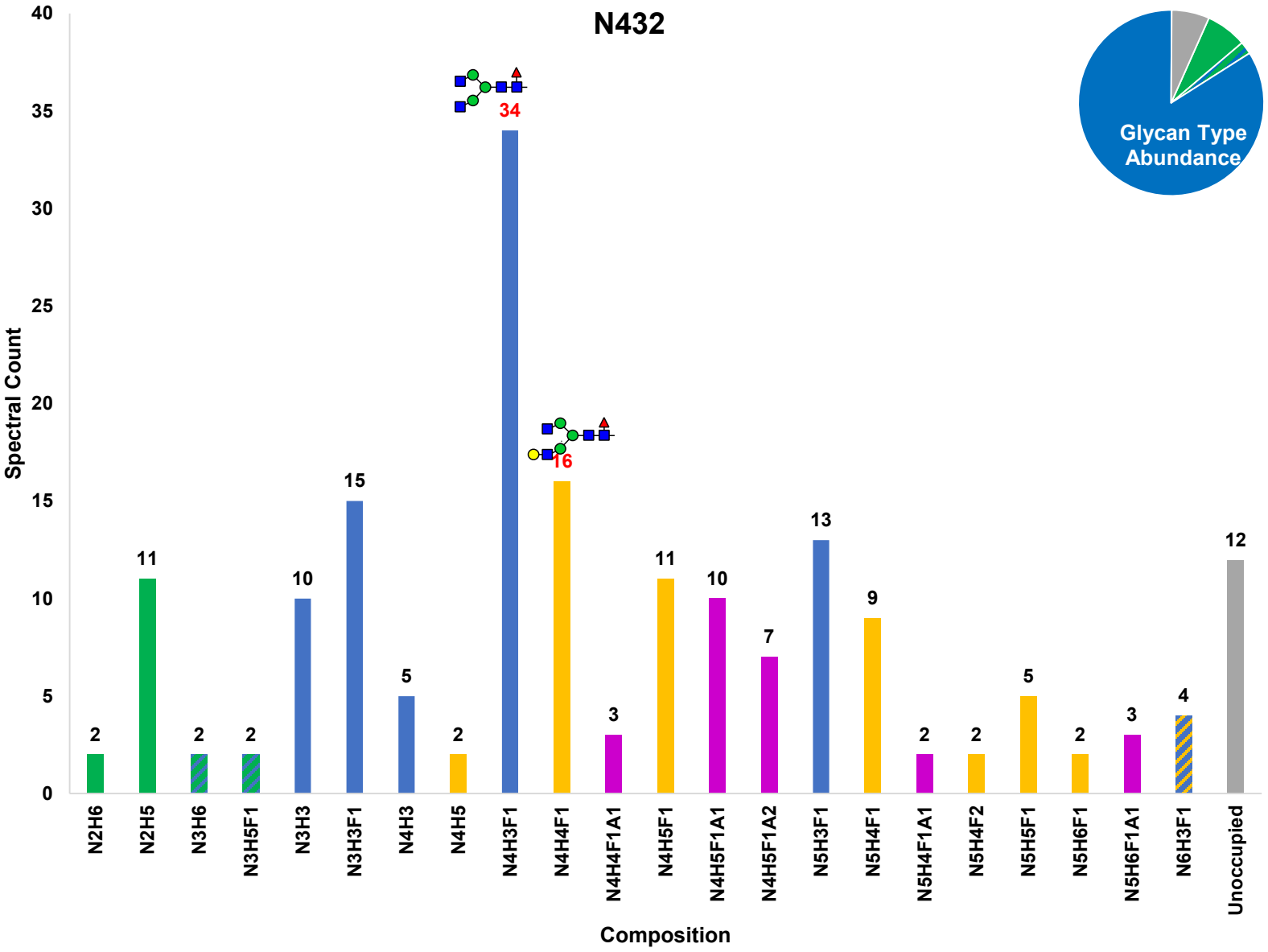
5C



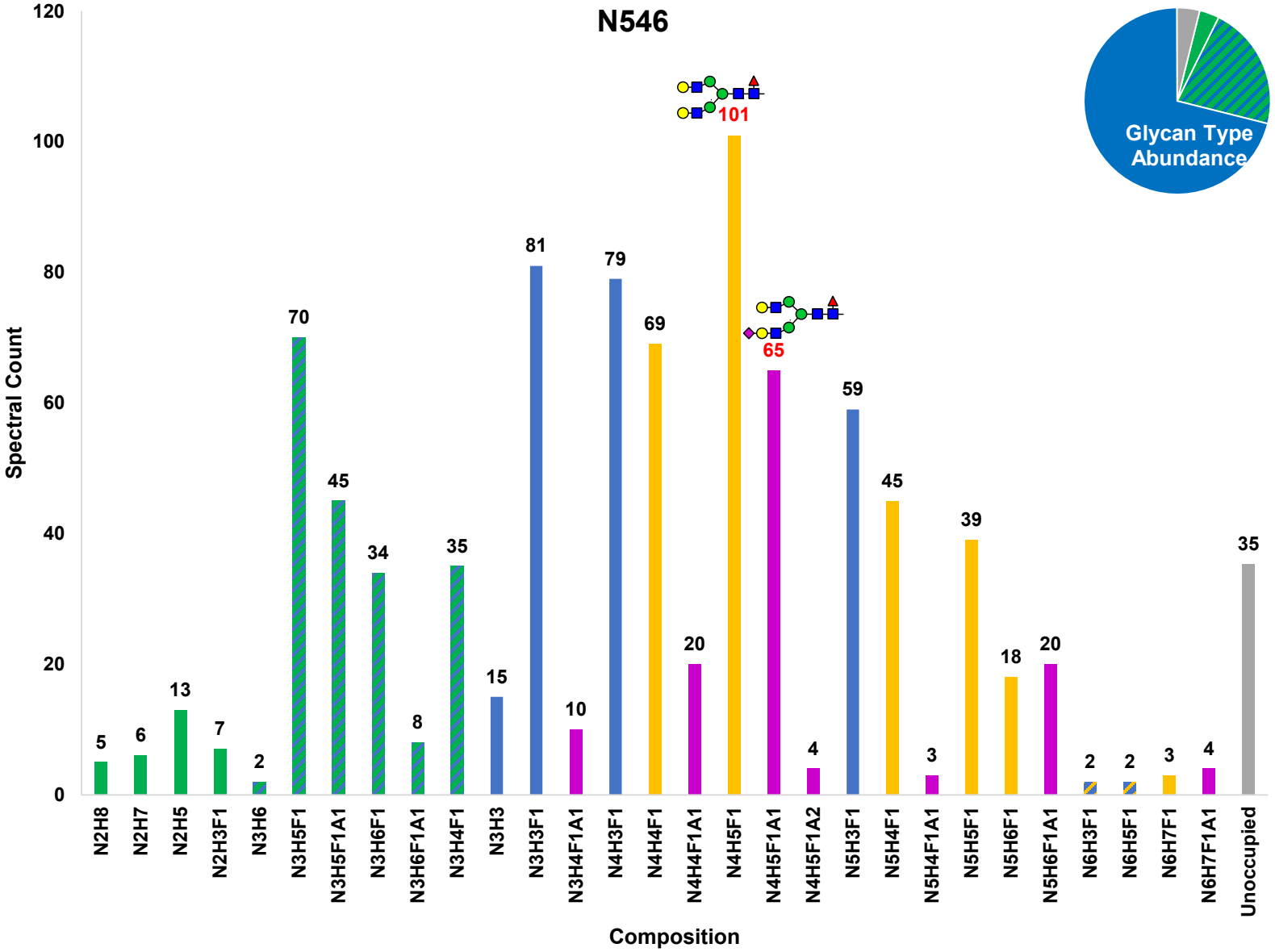
5D



5E



5F

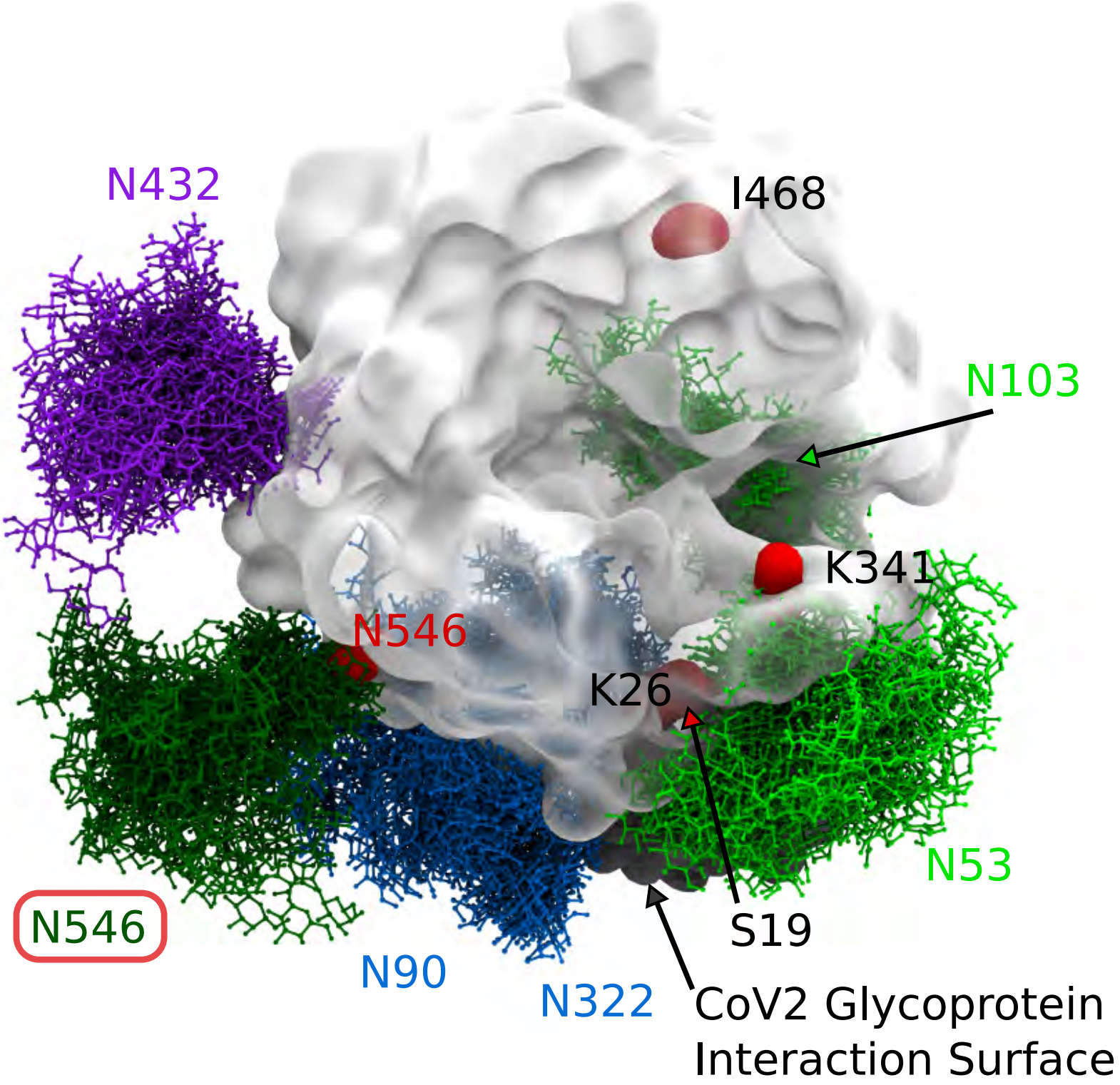


5G

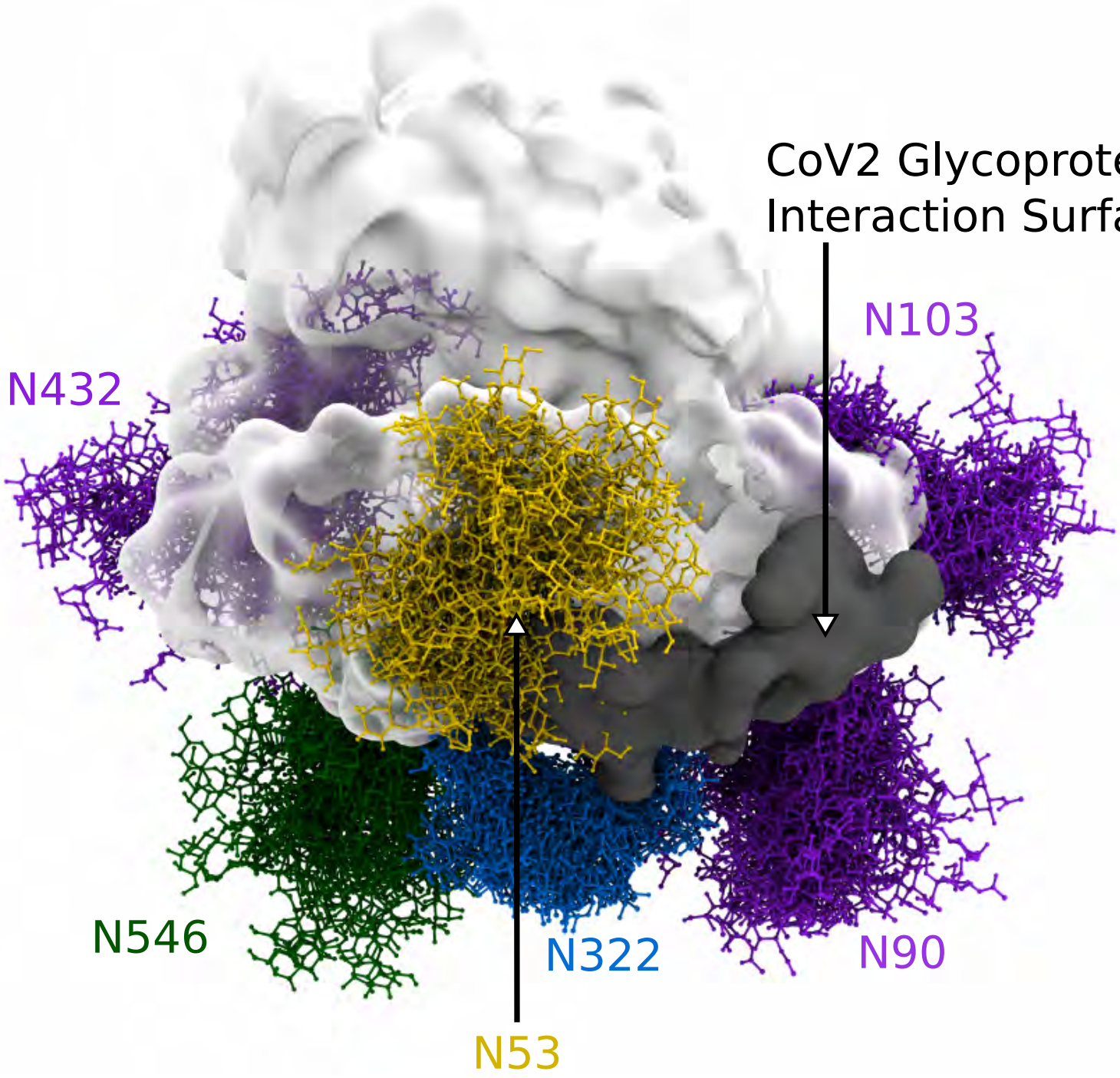
Key

- SP = Signal Peptide
- PD = Peptidase M2, ACE Domain
- = N-glycosite
- X or X = Variant, no sequon impact
- X = Variant, sequon ablated
- ~~Strikethrough~~ = Signal Peptide
- *Italics* = "Tag"

6A



6B



6C

SARS-CoV-2 S

```
SP
MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS TQDLFLPFFS   60
                                                              F
NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI IRGWIFGTTL DSKTQSLLIV  120
                K I M
NNATNVVIKV CEFQFCNDPF LGVYYHKNNK SWMESEFRVY SSANNCTFEY VSQPFLMDLE  180
                  H       Y     I
                                NTD
GKQGNFKNLR EFVFKNIDGY FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT  240

LLALHRSYLT PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK  300
            G
CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN  360
                                                RBD
CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF VIRGDEVRQI APGQTGKIAD  420
                                  RBM
YNYKLPDDFT GCVIAWNSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST EIYQAGSTPC  480

NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN  540
  A
FNFNGLTGTG VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP  600
                                  I
GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL IGAEHVNNSY  660
            G                                           Y
                                  |
ECDIPIGAGI CASYQTQTNS PGGSGSVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI  720
                      L

SVTTEILPVS MTKTSVDCTM YICGDSTECS NLLLQYGSFC TQLNRALTGI AVEQDKNTQE  780

          FP
VFAQVKQIYK TPPIKDFGGF NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC  840
          I                                           T

LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM  900
    S
                                HR1
QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN  960
                                  F
TLVKQLSSNF GAISSVLNDI LSRLDPPEAE VQIDRLITGR LQSLQTYVTQ QLIRAAEIRA 1020

SANLAATKMS ECVLGQSKRV DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA 1080

ICHDGKAHFP REGVFVSNGT HWFVTQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP 1140

                          HR2
LQPELDSFKE ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL 1200

QELGKYEQGS GGYIPEAPRD GQAYVRKDGE WVLLSTFLGG SHHHHHH              1247
  F
```

ACE2

```
SP
MSSSSWLLLS LVAVTAAQST IEEQAKTFLD KFNHEAEDLF YQSSLASWNY NTNITEENVQ   60
                  P       R
NMNNAGDKWS AFLKEQSTLA QMYPLQEIQN LTVKLQLQAL QQNGSSVLSE DKSKRLNTIL  120

NTMSTIYSTG KVCNPDNPQE CLLLEPGLNE IMANSLDYNE RLWAWESWRS EVGKQLRPLY  180

EEYVVLKNEM ARANHYEDYG DYWRGDYEVN GVDGYDYSRG QLIEDVEHTF EEIKPLYEHL  240

                                PD
HAYVRAKLMN AYPSYISPIG CLPAHLLGDM WGRFWTNLYS LTVPFGQKPN IDVTDAMVDQ  300

AWDAQRIFKE AEKFFVSVGL PNMTQGFWEN SMLTDPGNVQ KAVCHPTAWD LGKGDFRILM  360
                                              R
CTKVTMDDFL TAHHEMGHIQ YDMAYAAQPF LLRNGANEGF HEAVGEIMSL SAATPKHLKS  420

IGLLSPDFQE DNETEINFLL KQALTIVGTL PFTYMLEKWR WMVFKGEIPK DQWMKKWWEM  480
                                                V
KREIVGVVEP VPHDETYCDP ASLFHVSNDY SFIRYYTRTL YQFQFQEALC QAAKHEGPLH  540

KCDISNSTEA GQKLFNMLRL GKSEPWTLAL ENVVGAKNMN VRPLLNYFEP LFTWLKDQNK  600
      S
NSFVGWSTDW SPYADSGGSH HHHHH                                       625
```

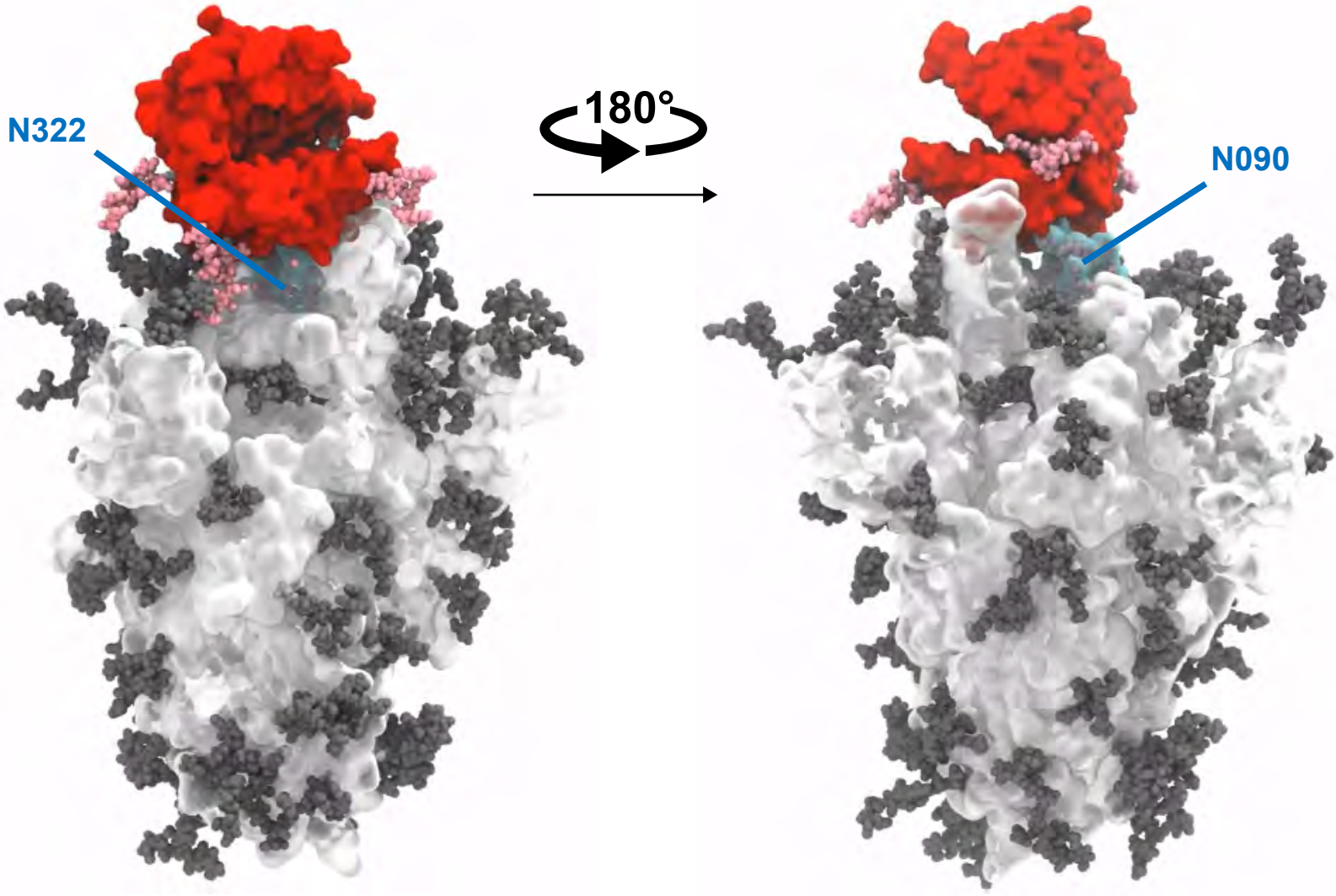
SARS-CoV-2 S Key

SP	= Signal Peptide	NTD	= N-Terminal Domain
RBD	= Receptor Binding Domain	RBM	= Receptor Binding Motif
FP	= Fusion Peptide	HR1	= Heptad Repeat 1
HR2	= Heptad Repeat 2		= Heptad Repeat 2 WT extent
	= N-glycosite		= O-glycosite
X	= Variant, no sequon impact	X	= Variant, sequon ablated
|	= Furin cleavage site in WT	...XXX...	= Divergent region w/ insert
Italics	= "Tag"	~~Strikethrough~~	= Signal Peptide
...XXX...	= ACE2 glycan interaction		

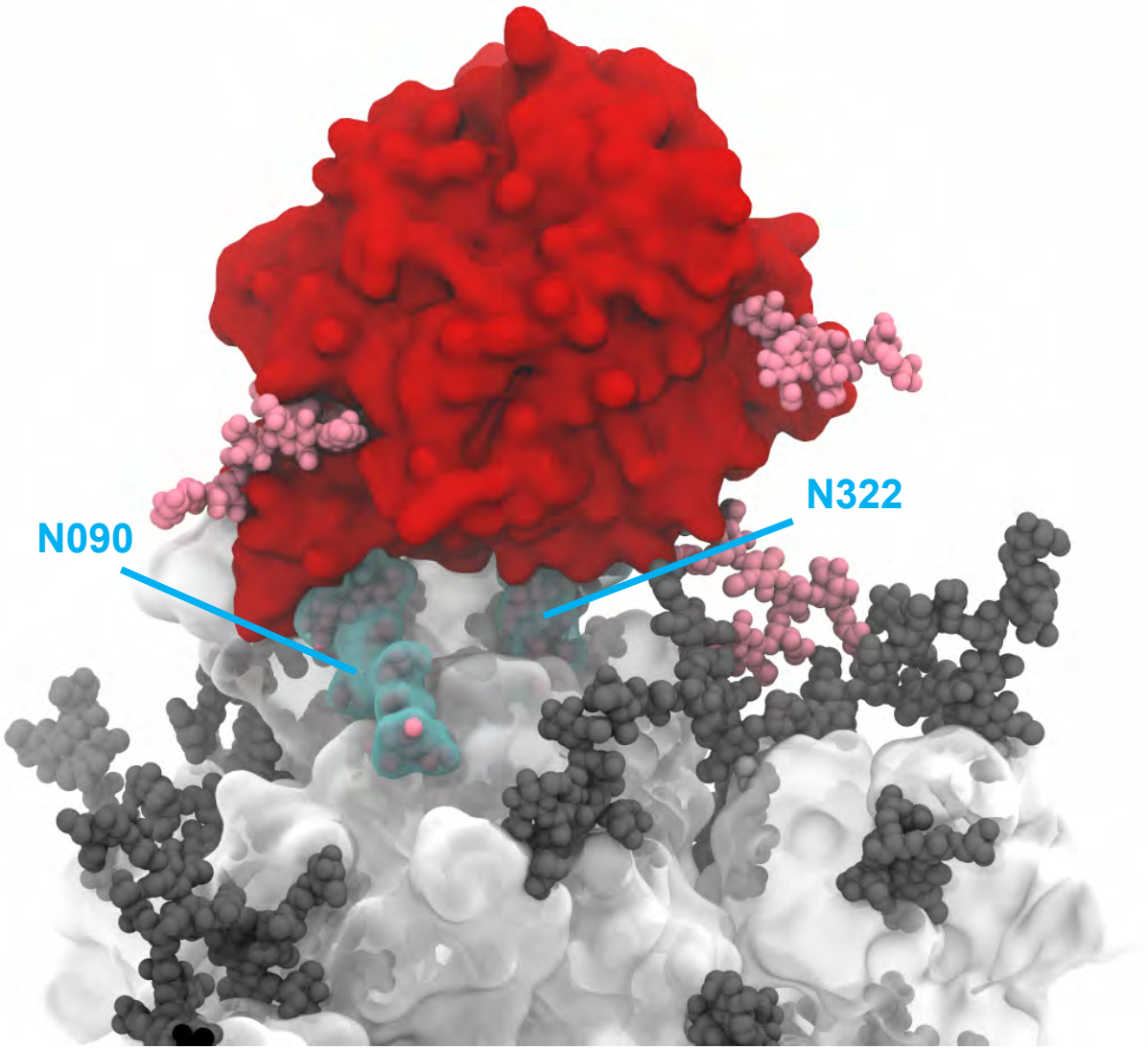
ACE2 Key

SP	= Signal Peptide	PD	= Peptidase M2, ACE Domain
	= N-glycosite	X or X	= Variant, no sequon impact
X	= Variant, sequon ablated	...XXX...	= CoV-2 S glycan interaction
~~Strikethrough~~	= Signal Peptide	*Italics*	= "Tag"

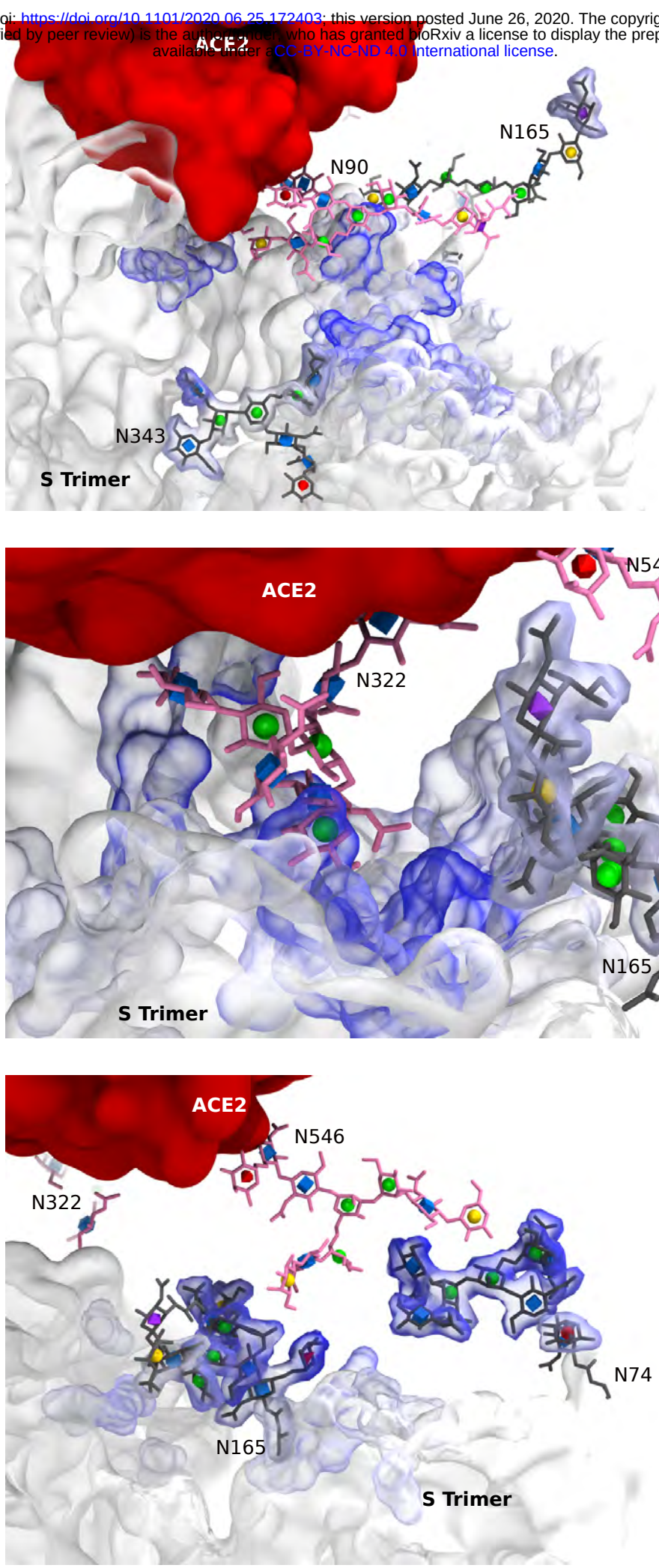
7A



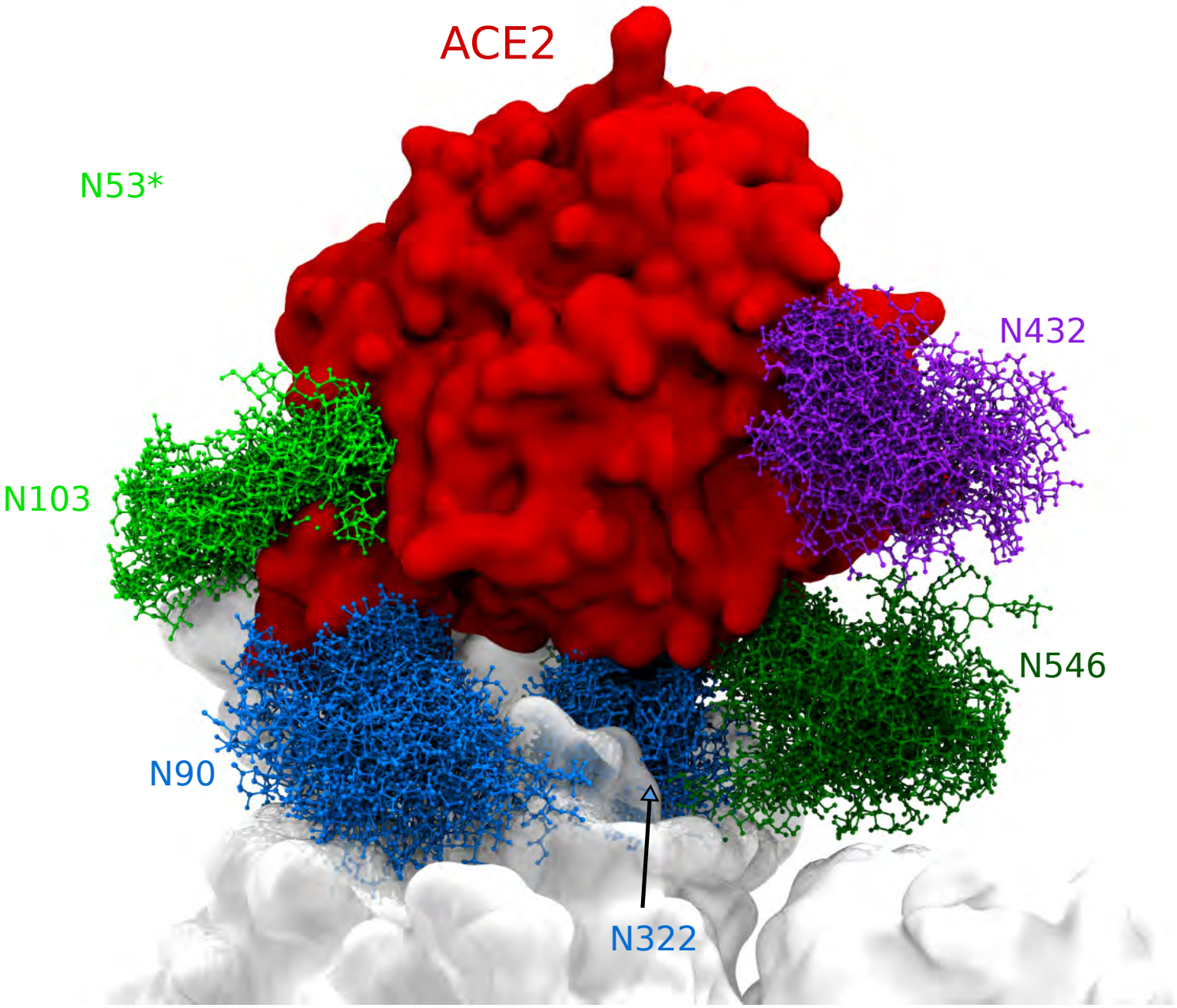
7B



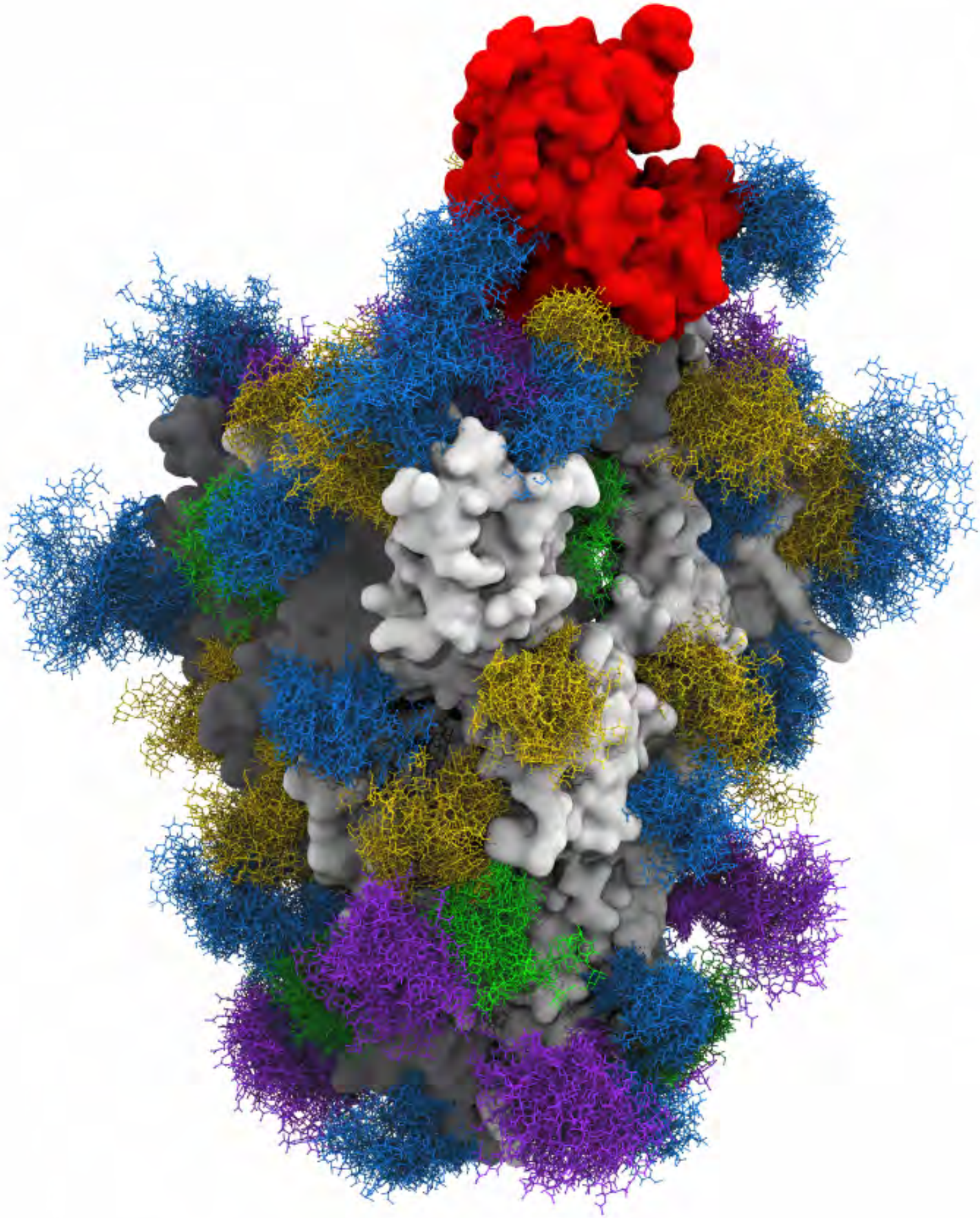
7C



7D



7E



7F