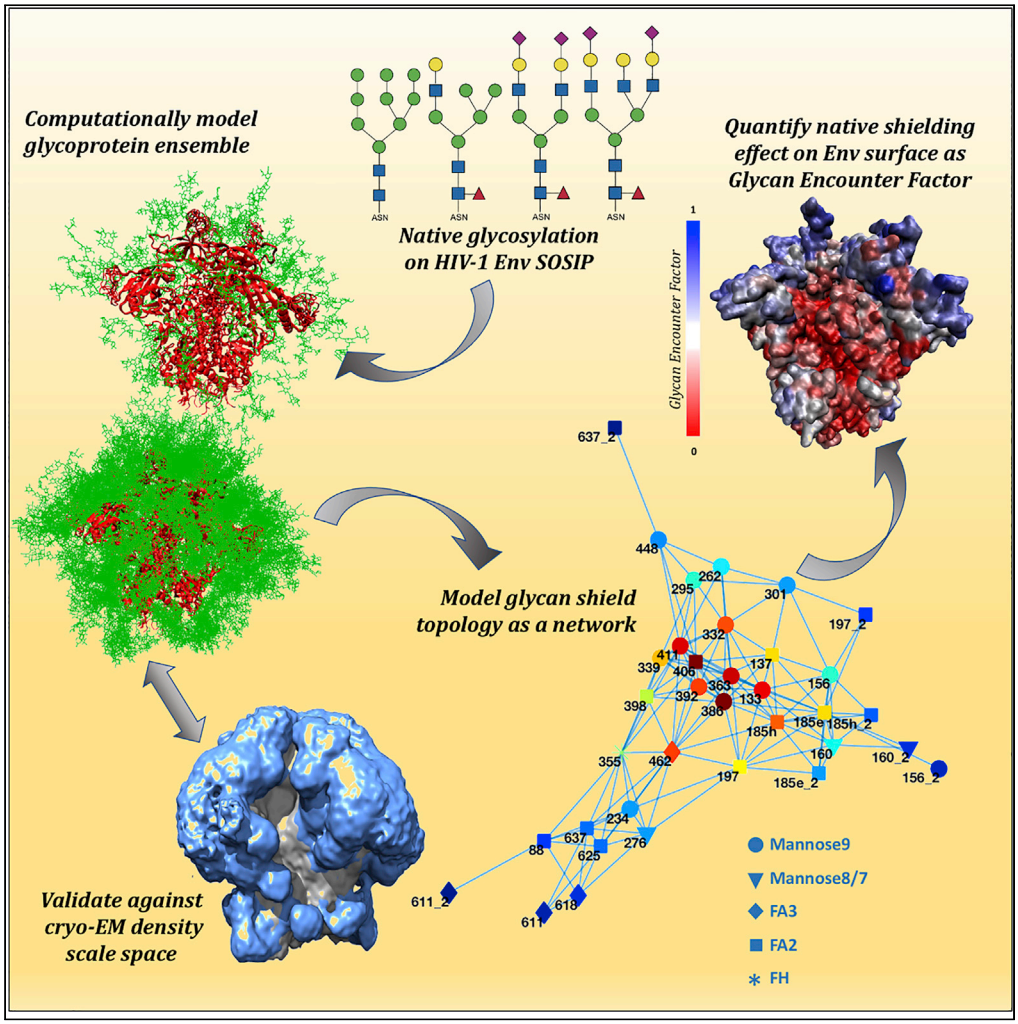
iScience

CellPress
OPEN ACCESS

Article

Quantification of the Resilience and Vulnerability of HIV-1 Native Glycan Shield at Atomistic Detail



Srirupa Chakraborty, Zachary T. Berndsen, Nicolas W. Hengartner, Bette T. Korber, Andrew B. Ward, S. Gnanakaran

andrew@scripps.edu (A.B.W.)
gnana@lanl.gov (S.G.)

HIGHLIGHTS

HIV-1 Env soluble BG505 SOSIP glycoprotein ensemble modeled with native glycosylation

Cryo-EM comparison suggests core fucose may lead to subtle glycan orientation changes

Native glycan shield topology modeled as a network of glycan-glycan volume overlaps

Glycan Encounter Factor developed to quantify the shielding effect on Env surface

Chakraborty et al., iScience 23, 101836
December 18, 2020 © 2020
https://doi.org/10.1016/j.isci.2020.101836

iScience

CellPress
OPEN ACCESS

Article

Quantification of the Resilience and Vulnerability of HIV-1 Native Glycan Shield at Atomistic Detail

Srirupa Chakraborty,[1,2,5] Zachary T. Berndsen,[3,4,5] Nicolas W. Hengartner,[1] Bette T. Korber,[1] Andrew B. Ward,[3,4,*] and S. Gnanakaran[1,6,*]

SUMMARY

Dense surface glycosylation on the HIV-1 envelope (Env) protein acts as a shield from the adaptive immune system. However, the molecular complexity and flexibility of glycans make experimental studies a challenge. Here we have integrated high-throughput atomistic modeling of fully glycosylated HIV-1 Env with graph theory to capture immunologically important features of the shield topology. This is the first complete all-atom model of HIV-1 Env SOSIP glycan shield that includes both oligomannose and complex glycans, providing physiologically relevant insights of the glycan shield. This integrated approach including quantitative comparison with cryo-electron microscopy data provides hitherto unexplored details of the native shield architecture and its difference from the high-mannose glycoform. We have also derived a measure to quantify the shielding effect over the antigenic protein surface that defines regions of relative vulnerability and resilience of the shield and can be harnessed for rational immunogen design.

INTRODUCTION

Protein glycosylation is an essential aspect of post-translational modification, with 50%–70% of human proteins been estimated to be glycosylated to some degree (An et al., 2009). These glycans play significant roles in numerous biological processes, such as cellular signaling, recognition and adhesion, protein folding, structural stability, and immune system interactions (Dwek, 1996; Moremen et al., 2012; Dennis et al., 1999). Additionally, envelope proteins from several high-risk viral pathogens that hijack the host protein production and glycosylation machineries, such as HIV (lentivirus) (Burton and Mascola, 2015), Coronavirus (Walls et al., 2016; Vankadari and Wilce, 2020), Lassa (arenavirus) (Sommerstein et al., 2015), Hepatitis C (flavivirus) (Zhang et al., 2017), Epstein Barr (herpesvirus) (Szakonyi et al., 2006), Ebola (filovirus) (Lennemann et al., 2014; Ilinykh et al., 2018), and Influenza (Wang et al., 2009) are also heavily glycosylated. These surface-expressed viral proteins are important immunological targets for neutralizing antibodies that can block viral infection of cells and form the primary focus of vaccine studies (Amanat et al., 2018; Saphire et al., 2018). However, the dynamic and dense glycan coating can effectively act as a shield for the underlying envelope protein, masking antigenic surfaces, barricading it from the adaptive immune system, and defending against immune surveillance (Stewart-Jones et al., 2016; Karsten and Alter, 2017; Ringe et al., 2019; Crispin et al., 2018; Wagh et al., 2018; Doores, 2015). A deeper molecular-level understanding of the glycan shields in these pathogenic viruses may help inform vaccine design strategies that can overcome this protective barrier.

The HIV-1 Envelope Glycoprotein (Env) is a heterodimeric trimer composed of proteins gp120 and gp41, and is responsible for the molecular recognition of the host receptor and fusion into host target cells. A number of obstacles hinder traditional vaccine design methods in case of HIV-1, namely, its remarkable sequence diversity, conformational plasticity, dramatic shifts in position and number of glycans (Zhang et al., 2004), and extremely dense glycosylation making up to approximately half the mass of the entire Env molecule (Lasky et al., 1986). As a result, there has been only limited success in eliciting broadly neutralizing antibodies (bNAbs) to Env vaccine immunogens (Walker et al., 2010; Bradley et al., 2016; Haynes et al., 2014). Structures of bNAbs in complex with Env indicate that these antibodies need to extend through the glycan shield in order to engage with the epitopes at the protein surface (Huang et al., 2014; Pancera et al., 2014; Julien et al., 2013; Lyumkis et al., 2013). Some bNAbs have evolved to include conserved glycans as

[1]Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[2]Center for Non-Linear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[4]IAVI Neutralizing Antibody Center and Collaboration of AIDS Vaccine Discovery, The Scripps Research Institute, La Jolla, CA 92037, USA

[5]These authors contributed equally

[6]Lead Contact

*Correspondence: andrew@scripps.edu (A.B.W.), gnana@lanl.gov (S.G.)

https://doi.org/10.1016/j.isci.2020.101836

part of the epitope (Mouquet et al., 2012; Walker et al., 2009; Falkowska et al., 2014; Crispin and Doores, 2015). Moreover, these surface glycans are also critical for Env folding, viral assembly, and infectivity.

Almost all the glycans on the Env are N-linked and occur at potential N-glycosylation sites (PNGS, given by the sequon Asn-X-[Ser or Thr], where X is not proline) (Stanley et al., 2017). Such glycans usually have about 10–20 pyranose rings covalently connected in a tree-like structure to the asparagine (N) residue of the sequon. These can further undergo additional processing (Crispin and Doores, 2015) into complex sugars from simple oligomannose, depending on their spatial location, local protein content, crowding, and producer cell type. A high degree of processing is indicative of exposure or accessibility of sugars to enzymes. At regions with dense crowding of glycans, steric constraints limit the activities of the carbohydrate processing enzymes (Pritchard et al., 2015a, 2015b) as the proteins fold and translocate through endoplasmic reticulum and Golgi apparatus.

The extreme dynamics and conformational heterogeneity stemming from large variance of possible constituent sugars, linkage, branching patterns, and rotamer flexibility make the study of these glycans immensely challenging (Walsh, 2010; Imberty and Perez, 2000). Traditional X-ray crystallography methods are rendered ineffectual (Chang et al., 2007) since this heterogeneity prevents proper crystallization of the glycoproteins. Inherent glycan flexibility also complicates analysis by ensemble techniques like nuclear magnetic resonance (NMR) and single-particle electron cryomicroscopy (cryo-EM) where the signal from glycans is often too noisy for structural interpretation at the resolutions necessary for building atomic structure (Chang et al., 2007; Davis and Crispin, 2010; Slynko et al., 2009; Woods et al., 1998). Within the ~300 experimental structures of HIV-1 envelope glycoprotein in the Protein Data Bank (PDB) (Berman et al., 2000), only a quarter of the total glycan content has been structurally resolved and are mostly in high mannose form. Recently there have been a few cryo-EM and X-ray structures published of natively glycosylated Env (Gristick et al., 2016; Lee et al., 2016; Barnes et al., 2018). Since most of these structurally resolved glycans were stabilized by interaction with antibodies, they provide little information about the structural and dynamic details of the individual and collective glycans. Furthermore, all computational studies of HIV-1 Env glycosylation comprise oligomannose glycan moieties ranging from mannose-5 to mannose-9 (Ferreira et al., 2018; Yang et al., 2017; Lemmin et al., 2017; Stewart-Jones et al., 2016; Liang et al., 2016). The main drawback affecting the quality of such MD simulations is the robustness of conformational sampling. Owing to the intrinsically dynamic nature of glycans, to effectively sample a biologically relevant energy landscape of the glycoprotein, long and often multiple trajectories need to be run, preferably with enhanced sampling techniques (Yang et al., 2017).

Here, we have performed a quantitative comparison between computational and experimental single-particle cryo-EM data demonstrating that the dynamics of high-mannose and native glycosylation remains practically indistinguishable around the glycan stem β-mannose (BMA) residue. This provides further rationale toward the need for computational modeling of complete glycans in a natively glycosylated shield. Presently such computational methods seem to be the only viable way to elucidate structural ramifications of such complex sugars on the shield architecture. Therefore, we have built a glycosylated Env having both oligomannose and complex glycans (Behrens et al., 2016; Cao et al., 2017) based on the information provided by mass spectrometry (MS) (Behrens et al., 2016). To our knowledge, this is the first time that a complete computational model of Env SOSIP was generated to include processed glycoforms, thus obtaining results that are immunologically more relevant.

Using these atomistic models, we have employed graph theory to capture the glycan shield topological network, pinpoint potential interaction pathways, and identify concerted behavior of the glycans. The glycan-glycan and glycan-protein interactions influence the behavior of the shield as a whole and can affect distant sites in the glycan network (Behrens et al., 2016; Stewart-Jones et al., 2016; Ferreira et al., 2018). Analyses of various network attributes, such as relative centrality of different glycans and critical subnetwork properties, have aided in detailed examination of the native glycan shield in the context of bNAb interactions. Important global and local feature differences due to native-like glycosylation pattern come to light as compared with all high-oligomannose glycans, arising from the presence of charged sialic acids at the tips and fucose at the base. Both the simulated models and cryo-EM maps identify a slight change in projection angle relative to the peptide surface for fucosylated glycans. We have also been successful in quantifying the glycan shield based on the number of sugar heavy atoms encountered over the antigen surface that can be implemented to define regions of relative vulnerability and regions where the glycan shield

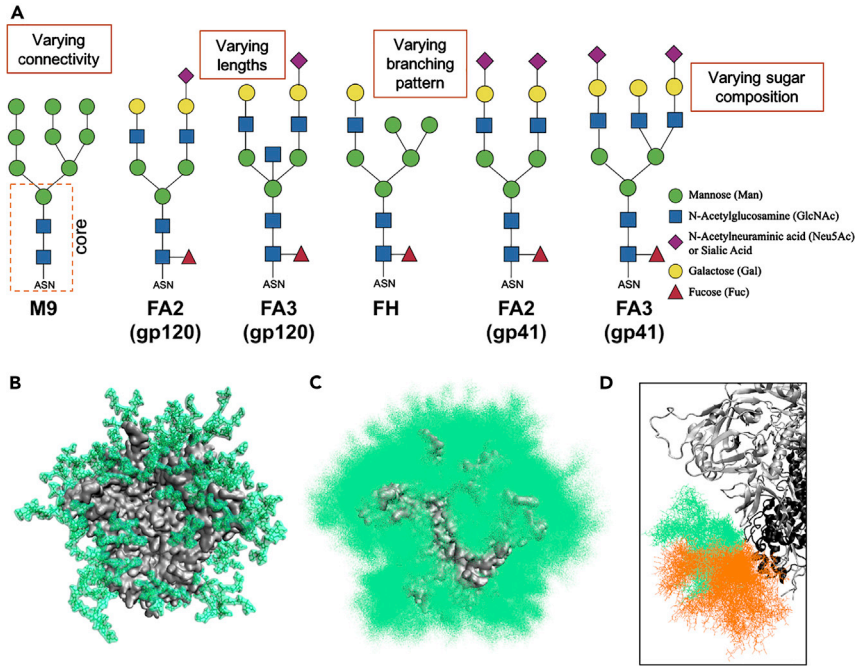


Figure 1. Native Glycosylation and Shielding of HIV-1 Env

(A) Schematic representation of N-glycan types. The common core of Man3GlcNAc2Asn is indicated. The glycan species selected for different sites of the BG505 native model (as given in Table S1). M9 ≡ mannose-9, FA2 ≡ fucosylated two-antennae, FA3 ≡ fucosylated three-antennae, FH ≡ fucosylated hybrid. Complex glycans in gp41 are different from those in gp120, as per site-specific mass spectroscopy experiments.

(B) Single snapshot of modeled natively glycosylated BG505 structure, with each glycan taking up a particular conformation.

(C) Cumulative shielding effect over time due to the dynamic nature of the glycans. One hundred randomly selected models from the 1000-structure ensemble is shown here.

(D) Neighboring glycans within close proximity can sample overlapping volumes. Glycan ensembles N88 shown in green and N625 shown in orange.

blocks access. Owing to the rapid yield rate of this method (see Supplementary Information), it can be applied to a large number of diverse HIV-1 sequences or can be seamlessly extended to model glycan shields of other viruses in a rapid manner.

RESULTS

Selection of Site-Specific Glycans for Native Glycosylation of HIV-1 Env

The soluble, recombinant BG505 SOSIP.664 (here on referred simply as BG505) trimer has been well characterized as a native-like, Env-mimetic model and serves as the prototypical immunogen in several vaccine development programs (Sanders et al., 2013, 2015). For our study, we have used this system in the pre-fusion closed conformation, for which several structures have been determined (Gristick et al., 2016; Stewart-Jones et al., 2016; Barnes et al., 2018; Lee et al., 2016). A single protomer of this BG505 protein contains 28 N-linked glycosylation sites. Although all N-glycans share a common core sugar sequence, they are broadly classified into three types (Figure 1): (1) oligomannose, in which only mannose residues are attached to the core; (2) complex, in which differently branching "antennae" are attached to the core; and (3) hybrid, which is a mix of the first two types. Recent novel MS-based approaches have been successful in identifying site-specific glycosylation profile in multiple HIV-1 trimers (Cao et al., 2017; Go et al., 2017). It is known that each glycosylation site at BG505 has a distribution of multiple possible glycan species, depending on various factors, such as type of cell lines, purification methods and sources, and structural constraints. We used a site-specific distribution of glycan type at each PNGS that was obtained by Behrens et. al (2016) and Cao et. al. (2017), to identify the most likely glycan species at each site. The glycan species at each site was selected based on the highest relative abundance per the MS studies (Figure 1 in Behrens et al., 2016) and is listed in Table S1. The structure schematic for each of the selected glycan species has been shown in Figure 1A.

High-Throughput Atomistic Modeling of the Native Glycan Shield

In the static pictures as obtained from fully glycosylated structures (Stewart-Jones et al., 2016; Gristick et al., 2016) and even single snapshots from computational models, a significant fraction of the protein surface appears to be exposed, with each glycan taking up a particular conformation (Figure 1B). However, owing to their flexible and dynamic nature, these glycans are not confined to a particular conformation, instead sampling a large volume in space. A previous simulation study suggested (Guvench et al., 2011) that the root-mean-square deviations (RMSDs) for the carbohydrate regions are more than four times larger than that of the underlying protein loops. As a result of this conformational variability, the cumulative effect of the glycans over time is like that of a cloud of glycan atoms that shield the underlying surface from any approaching protein probe (an antibody, for example) as illustrated in Figure 1C.

Here we have incorporated experimental information-driven conformational sampling in the flexible unstructured protein regions to model the entire glycan shield, improving upon a previously established protocol (Berndsen et al., 2020). Details of the modeling procedure are given in the Methods section, and are briefly reiterated here. A robust ensemble of BG505 glycoprotein atomic models was generated by utilizing a template-free glycan modeling pipeline including a sequence of refinement steps with restraints to enforce proper stereochemistry. The underlying protein scaffold was built by utilizing templates from a number of experimentally determined Env structures deposited in the PDB (Table S2) by assimilating the differences in local structural regions including loops that arise owing to flexibility. The glycans were modeled *ab initio* on several distinct protein scaffolds by implementing the ALLOSMOD (Weinkam et al., 2012; Guttman et al., 2013) package of MODELLER (Eswar et al., 2006; Sali, 1995) in a streamlined pipeline.

Owing to the initial randomization of the glycan orientations, this integrated technique can sufficiently sample a physiologically relevant conformational space accessible to carbohydrates in a very short time, leading to the overall spatial shielding effect (Figure 1C). The torsion angle distributions of 10 different inter-glycan linkages within the ensemble was compared with those obtained from different glycan structures available in the PDB database using GlyTorsion webserver (Lütteke et al., 2005) and is shown in Figures S1–S4. These distributions match very well between our generated ensemble and the PDB structures (see Supplemental Information for details), indicating that the modeled glycans sample a stereochemically correct landscape. In order to understand the effects just due to native glycans, we have also built a similar ensemble with mannose-9 glycans at all glycosylation sites (hereon referred as all-man9 model) using the same protocol for comparison.

Quantitative Comparison of Simulated Models with Cryo-EM Data Indicates Glycoform-Independent Glycan Stem Dynamics

We generated simulated cryo-EM density maps for both the native and all-man9 ensembles by utilizing the method that we established earlier to convert *in silico* atomistic models to synthetic cryo-EM data (Berndsen et al., 2020). Then, we quantitatively compared them with the experimental cryo-EM maps of BG505.SOSIP.664 from HEK 293F cells, which produce a native-like glycan shield including both oligomannose and complex glycans, and from HEK 293S cells, which do not produce complex glycans (Berndsen et al., 2020). The experimental and simulated maps show similar features in the shield, with only the first few sugar residues at each site defined at high resolution (Figure 2A). A scale-space analysis developed in our earlier study (Berndsen et al., 2020) also shows a similar trend to the experimental data (Figure 2B). Progressive smoothing of the map is accompanied by a large expansion in glycan volume that plateaus around 1.75–2 SD. The effect of the glycan shield on the scale-space profile is best appreciated in comparison with a map generated from a non-glycosylated ensemble (Figure 2B, dashed line). The 1.75 SD Gaussian filtered maps are visualized at high and low thresholds in Figure 2C, revealing a similar overall shape and topology.

Another way to assess heterogeneity in cryo-EM maps is with local resolution. So, we compared the local resolution estimates for the simulated and experimental maps (Figure S5) and observed similar trends between the two. The simulated map is of higher resolution in the protein core, which is to be expected, but displays a similar distribution of resolutions within the glycan shield (Figures S5A and S5B), with resolution falling with distance from the protein surface. The radial average local resolution plot also captures the similar falloff in resolution with distance from the center of the protein, although the simulated map maintains a slightly higher resolution until just beyond the outer radius of the glycan shield (~80Å, Figure S5C). This could indicate that our models underestimate the glycan sampling slightly; however, some of the

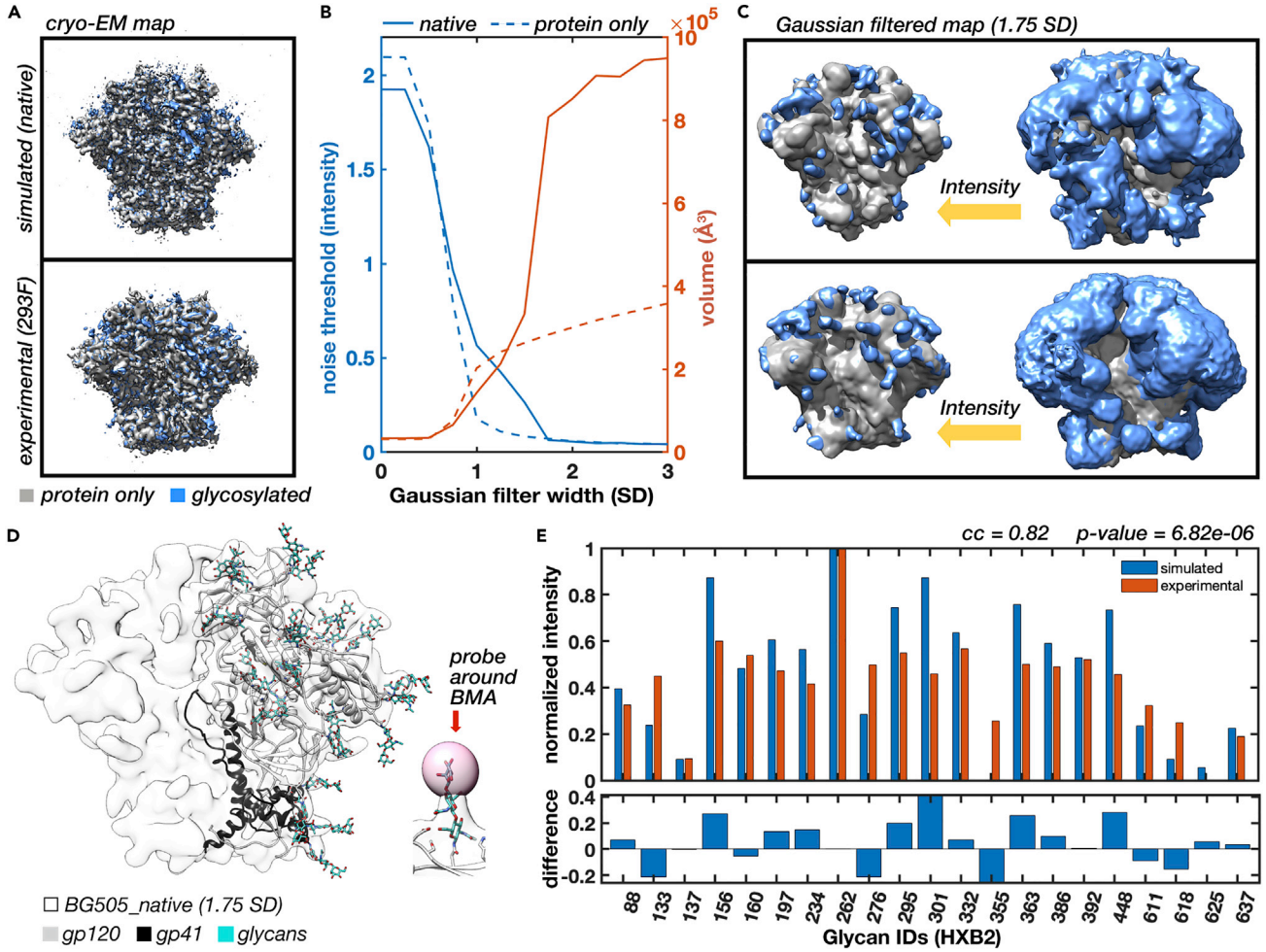


Figure 2. Comparison of Glycan Dynamics between Experimental and Simulated Cryo-EM Maps

(A) Experimental BG505_293F (top) and simulated BG505_native (bottom) cryo-EM maps sharpened and filtered at the global resolution (~3Å), with maps of de-glycosylated trimers displayed underneath to highlight the glycan density. The BG505_293F data are reproduced from Berndsen et al. (2020) and density corresponding to the three RM20A3 Fabs bound to the base of the SOSIP was deleted for better comparison.

(B) Noise threshold and map volume at the noise threshold as a function of Gaussian filter width (standard deviation, SD) of the BG505_native and protein-only maps showing the progressive emergence of glycan signal at lower resolution.

(C) The 1.75 SD Gaussian filtered maps from (A) displayed at high- and low-isosurface thresholds.

(D) Gaussian filtered BG505_native map with relaxed atomic model used for measuring local glycan dynamics along with a close up of a single glycan with a spherical probe around the BMA residue.

(E) Bar plot showing relationship between the normalized site-specific mean local intensity from the BG505_293F (experimental) and BG505_native (simulated) maps along with the absolute difference at each site along with the Pearson correlation coefficient.

discrepancy can be attributed to the overly simplified methods used to generate the simulated data (i.e., unrealistic noise model, inability to simulate data abnormalities from experimental/environmental errors, CTF not included). Together, these comparisons suggest the ALLOSMOD method is capturing realistic levels heterogeneity in the glycan shield, at least globally.

At the level of individual glycan dynamics, comparison of the normalized mean intensity at the location of each BMA residue for the simulated map with that of the experimental BG505_293F map reveals a strong positive correlation (Figures 2D–2E), with a correlation coefficient of ~0.82 ($p = 6.82 \times 10^{-6}$). For both simulated and experimental maps, the highest intensities emanate from highly conserved glycan N262, suggesting it is the most structured. Previous evidence suggests that this glycan is crucial in terms of Env function and stability owing to its extensive interactions with protein core and other neighboring glycans (Mathys et al., 2014; Kong et al., 2015). We also confirmed that local map intensity around BMA

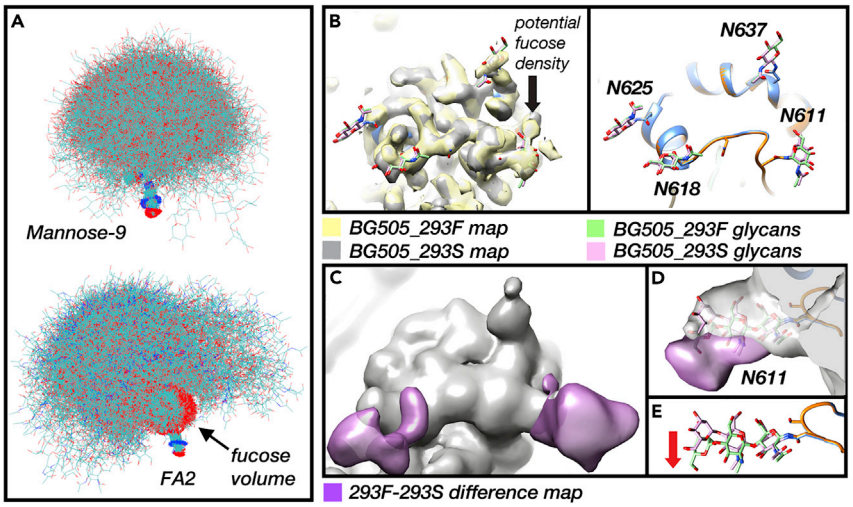


Figure 3. Model and Experimental Cryo-EM Data Reveal Subtle Change in Orientation due to Core Fucose

(A) Difference in overall glycan orientation between all-man9 and fucosylated complex glycan in simulated model. First core sugar and three residues at the base of the glycan (n-1, n, n+1; where n is the glycosylated asparagine residue) were aligned for a single sugar ensemble. Oligomannose samples a space symmetrically around the core, whereas complex sugars have a distinct bend away from the side where the fucose ring is present.

(B) Cryo-EM maps and refined atomic models for BG505_293F and BG505_293S (add citation) showing the four glycans from gp41. Only the core NAG residues can be confidently built into the sharpened maps.

(C) Gaussian filtered BG505_293S map (gray) and BG505_293F-BG505_293S difference map (purple).

(D) Close up of the glycan at N611 showing asymmetric difference signal around the glycan stem.

(E) Full glycan stalks built and relaxed into both Gaussian filtered maps showing more pronounced change in projection angle.

residues accurately captures relative differences in root-mean-square fluctuations (RMSFs) at each glycan (Figure S6) for the native ensemble. An interesting outcome of the above comparison is that the correlation coefficients remain similarly strong not only when we compare the all-man9 model to BG505_293S data but also when the native model is compared with high-oligomannose experiments and vice versa (Figure S6). Therefore, although these comparisons give credence to the validity of the model in accurately capturing the relative position and flexibility of the glycan stems, they also indicate that the glycan dynamics are practically identical within the observational limits of cryo-EM analysis between complex and oligomannose. The native glycosylation pattern on the shield topology only manifest at the level of the flexible antennae beyond the stem and can therefore only be observed in detail using computational models.

Fucose-Mediated Changes in Glycan Orientation Distinguish Native Glycan Shield

Although the glycan stem shows indistinguishable dynamics at the BMA sugar, an important difference between complete models of complex and all-man9 arise owing to the presence of the fucose ring at the core of the complex sugars (Figure 3). By structurally aligning the first core sugar and the three residues at the base of the glycan (n-1, n, n+1; where n is the glycosylated asparagine residue) for all the structures in the ensemble, the oligomannose glycans are found to be spaced symmetrically around the core, whereas complex sugars have a distinct bend away from the side where the fucose ring is present (Figure 3A). This is not evident in overall ensemble orientations of the complex glycan, and get picked up by the alignment, indicating that it is likely arising from the internal torsion space difference due to fucosylation. The flexibility of loops and glycan-glycan interactions can also mask this effect in the overall ensemble. For example, this observation is more pronounced in the sparsely situated gp41 glycans than those in the hypervariable loops and HMP. The change in projection angle of the principal axis of the glycan relative to the protein surface is predicted to be ~10° in our models. This difference in orientation preference, along with the presence of the negatively charged sialic acid tips of the antennae, play notable roles that govern the differences between high oligomannose and native glycosylation.

The change in projection angle from the fucose ring manifests as a reasonably large change at the tips of the antennae, but the overall average change in position at the core N-acetylglycosamine (NAG) is too

small to be captured in the high-resolution cryo-EM maps (Figure 3B). However, a close examination of the low-resolution difference map between the BG505_293F and BG505_293S maps (Berndsen et al., 2020) does reveal some evidence for such changes in projection angle at sites with complex glycans (Figures 3C–3E). For instance, the difference signal around the glycans at N611 and N618 is asymmetric around the glycan stem and is therefore likely not the result of differences in occupancy. The difference signal at N611 is located primarily on the underside of the glycan stem, on the face opposite where we identified density likely belonging to the fucose ring (Figure 3D), which is suggestive of a change in projection angle similar to what we observed in the simulations. In addition, examination of the glycan stem models built and relaxed into the Gaussian filtered BG505_293F and BG505_293S maps used for assessing local map intensity (Figure 3E) show a slight change in projection angle, further confirming this fucose-mediated glycan reorientation.

Network Analysis of Glycan Topology Explicates the Shield Connectivity

Network analyses have historically been used to study protein allosteric frameworks and evolutionary paths (Beleva Guthrie et al., 2018; Huang et al., 2017; Eargle and Luthey-Schulten, 2012; Skjaerven et al., 2014; Sethi et al., 2013) and only recently have begun to be applied to in glycoprotein structural characterization (Lemmin et al., 2017). Although each glycan exerts effects in its immediate vicinity, owing to the inter-glycan interactions, their influence can percolate over long distances across the surface of protein (Berndsen et al., 2020). Long-range glycan interactions can occur, as with perturbation of a glycan at one site can affect the processing and antibody interactions of another glycan at a distant site (Behrens et al., 2016; Stewart-Jones et al., 2016; Doores, 2015; Bricault et al., 2019; Hargett et al., 2019).

Here we have applied such a graph-theory based approach to model the native glycan shield network. Details of network modeling are recapitulated in the Methods section. The natively glycosylated BG505 trimer, like all HIV-1 Envs, has a highly dense glycosylation pattern and each glycan samples a particular region in space (Figures S7A and S7B), such that neighboring glycans can occupy overlapping regions (Figure 1D). The fraction of volume overlap gives a measure of the interaction probability between the constituent glycans (Figure S7C). Within a protomer there are three main regions of overlap—the apex, the gp41 base, and the glycan-dense central high-mannose patch (HMP). Inter-protomer overlaps are primarily due to V1 and V2 loop glycans near the trimer apex. Figure 4A shows the obtained network of BG505 native glycosylation unfolded and laid out in two dimensions for the ease of visualization. This is a force-directed layout, which uses attractive forces between adjacent nodes and repulsive forces between distant nodes to reach an optimum distribution of the node points in space (Fruchterman and Reingold, 1991). For conceptualization, the network relative to the BG505 structure is provided in Figure S7D.

The nodes in the central region around the V4 loop, in the HMP, are very highly inter-connected. Glycans in the V1, V2 apex are also reasonably well connected, but the connections at the base near gp41 are relatively sparse, both within the locality and globally with the rest of the network. Glycans at positions 234, 355, 462, and 276 connect the sparse base region of the network with the dense apex and central regions. Glycan 156, 160, and those in the V2 loop enable inter-protomer glycan interactions at the apex, stabilizing the protein quaternary trimer structure (Berndsen et al., 2020). Glycan 197, crucial in the binding of CD4bs and V3-specific antibodies (Liang et al., 2016; Li et al., 2008), connects the central crowded region of the network between neighboring protomers. gp41 glycans 611 and 637 also contribute to inter-protomer pathways, with glycan 637 communicating with glycan 448 in the neighboring protomer in the central mannose patch. Both glycans 448 and 637 together form part of PGT151 epitope, evincing their structural proximity in experiments (Lee et al., 2016).

Utilizing the Glycosylation Network to Extract Functionally Relevant Topological Features

The native glycosylation network is well connected as indicated by the network degree, or the number of connections a node has to other nodes (Figure S7E). Owing to the high structural fluctuations of the V2 loop, the glycans on and flanking V2 can interact with a number of other neighboring nodes, having the highest degrees of connectivity. There is dense crowding of glycans around the high mannose patch with a number of inter-connections, resulting in the glycans of this patch being less processed because of reduced accessibility to enzymes (Coss et al., 2016; Pritchard et al., 2015a; Doores et al., 2010). The network degree values are relatively lower around the base of the protein on and around gp41, reflective of the sparse architecture of glycan topology in this region. The site N332 is central in the glycan interaction network and is known to have both direct and subtle long-range influences over a number of bNAb sites

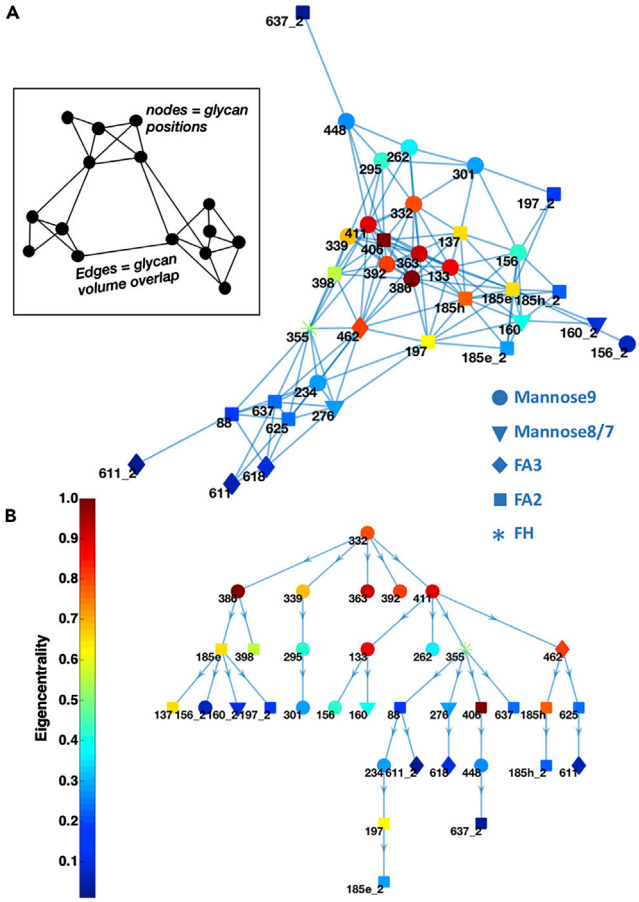


Figure 4. Spatial Network of BG505 Native Glycosylation

(A) Visual representation of network in two dimensions, based on a force-directed layout. Each glycan forms a node-point on the graph, two nodes are connected by an edge (inset, from reference Berndsen et al., 2020, Figure 8). The nodes are colored by the normalized eigenvector centrality of the glycans.

(B) Shortest path of communication between glycan N332 and all other glycans in the network.

(Behrens et al., 2016; Bricault et al., 2019). We calculated the shortest paths to all glycans from the N332 glycan site using the Floyd-Warshall algorithm (Warshall, 1962; Floyd, 1962), where the glycan distances were edge-weighted (Figure 4B). Our calculation shows the most probable pathway over which each glycan feels the influence of the glycan at site N332, providing useful insights into how these inter-glycan communications can occur.

To elucidate the relative influence of each of the glycans on the network, we calculated the eigenvector centrality, or eigencentrality, of the nodes to measure its connectivity to the network. A large relative value indicates that the node is well connected to the network and thus is "centrally" located, whereas a low relative value indicates that the node is on the periphery of the network. Operationally, the eigencentrality is the eigenvector associated with the largest eigenvalue of the adjacency matrix, that in position (i,j) reports the overlap of the ensemble of glycan at sites i and j. Value of that particular matrix element will be zero if the glycans in position i and j do not physically interact (details given in Methods section). The advantage of this centrality measure over some of the more commonly used quantifiers such as "betweenness. Centrality" is that it intrinsically considers the effects of glycans not immediately neighboring to the node of interest, thus involving higher-order interactions. Also, it is independent of the information flow pathway, which can deviate from the shortest paths for biological systems. The normalized eigencentrality of the glycans are projected on the network as a colormap in Figure 4A. The eigencentrality increases toward the middle of the graph, with the glycans at the crowded HMP with a large number of connections between each other having the highest centrality values. When we increase the threshold of volume overlap needed to form an

edge, those glycans with the least centrality leave the network first. Glycan 611 is the first to be eliminated, succeeded by some of the inter-protomer interactions and the other gp41 glycans, following the relative centrality scores. The HMP with high eigencentrality values persists throughout even with high overlap thresholds. We have previously confirmed for mannose-9 (Berndsen et al., 2020) that successive enzymatic digestion of glycans from Env loosely follows a pattern that matches with the network centrality. Those glycans that are sparsely connected to each other, having lower centrality, are eliminated earlier during the process of endoH digestion. On the other hand, the glycans having higher network centrality, such as those in the HMP, takes longer to be eliminated by the digestive enzymes. Thus, the eigenvector centrality provides a measure of the crowding of the highly central glycans, which makes them difficult for endoglycosidase to access.

Influence of Complex Glycans on Overall Network Topology

The degree of connectivity decreases almost throughout the network for the native glycosylation (Figure S8A) compared with all-man9. The number of stable glycan-glycan interactions at several glycan sites decreases because spatially proximal glycans avoid unfavorable charge-charge interactions. For example, the uncharged glycans 234 and 355 takes up a position central to all the neighboring charged sugars, increasing the distances between them, and the charged N406 glycan buries itself in the middle of the surrounding high-mannose glycans (Figure 5A), clustering together with them. In fact, removal of this charged glycan N406 glycan can increase the processing of the neighboring high-mannose moieties (Cao et al., 2018b). Further evidence for such alterations in glycan connectivity due to the presence of complex glycans can be observed in the comparisons of the RMSF variations. A detailed analysis has been presented in the Supplemental Information and illustrated in Figure S9. Combination of the network node location differences and RMSF variations suggest that the presence of the charged sialic acid at the tips of these complex glycans can dictate the interactions between neighbors, increasing their structural variations if surrounded by other charged glycans or reducing them if a stable conformation buried between uncharged high-mannose patches can be found. This overall decrease in connectivity slightly increases the diameter of the native network, increasing the mean number of hops for the shortest path from 2.4 hops (and distance 0.19) in the all-man9 model to 3 hops (and distance 0.27).

The distribution of centrality also shifts between the two glycosylation models (Figure S8B). The centralities of the glycans at 137, 234, 355, 398, 406, 462 and those at gp41 increase owing to native glycosylation. On the other hand, the V1 and V2 loop glycans as well as 262, 295, 363 decrease from the all-man9 model. Figure 5A illuminates the difference in adjacency matrices between the two models, with blue color indicating decrease in edge weight and red indicating increase in edge weight in native network, as compared with all-man9. Overall connectivity goes down in the apex and the central patch for the native and increases within gp41 glycans, as compared with all-man9. The differences between the two networks (with at least 10% edge-weight change) are shown in Figures 5B and 5C. The pathways around 197, as well as the inter-protomer interactions involving 156, 185e, and 637, become stronger in native glycosylation (Figure 5B). Some of these 197 connections, such as those with 276 and the V5 loop glycans occur across the CD4-binding site. It was previously shown that in predominantly high-mannose Env structures, CD4bs targeting antibody VRC01 has very little interaction with glycans 197 and 276 (Stewart-Jones et al., 2016). Conversely, in another Env structure with fully processed native glycans, a VRC01-like antibody called IOMA interacted extensively with both 197 and 276 (Gristick et al., 2016). This matches our observation of increased orientation of glycan 197 over the CD4bs in native glycosylation pattern. However, the paths connecting 137, 156, 301, and 197 (of neighboring protomer) are significantly weakened in the native model (Figure 5C). Subnetworks involving V4 loop glycans and some of the central mannose patch become weaker owing to the presence of native glycans.

Community Analysis to Determine Breaches in Glycan Shield

In order to interact and neutralize the Env protein, antibodies look for breaches in the shield where glycan densities are sparse. This information of relative sparsity is encoded within the glycan network, in terms of edge clustering. We examine this clustering of glycan interactions within the network by defining communities using a modularity maximization approach that divides the network into sub-modules or groups (see Methods). Communities have high density of edges between the nodes within a module and comparatively sparse edges connecting different modules (Figure 5D inset). This analysis identifies five distinct communities within the BG505 native glycan network (Figures 5D and 5E). The apex glycans around loop V2 from the three protomers together form a single community (1, green). Right below that, glycans 262, 295, 301, 332, and 448 (2, blue) forms a community around the V3 and alpha 2 helix. The same cluster of glycans has been independently identified as a

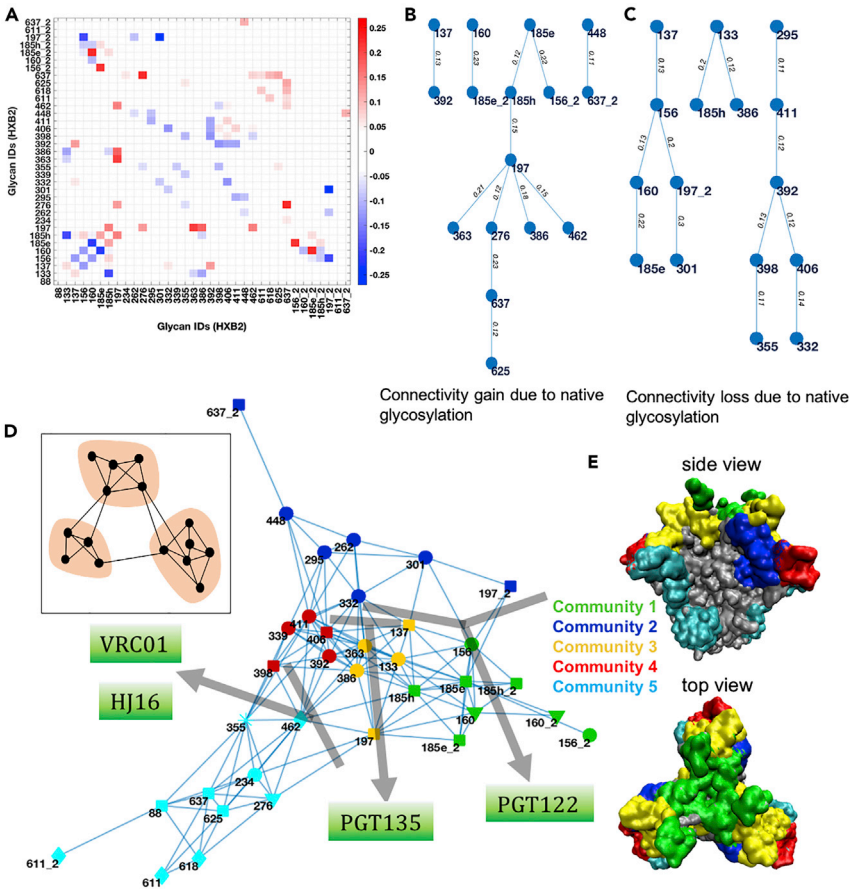


Figure 5. Network Difference and Community Analysis

(A) Difference in adjacency matrices between native and all-man9 models (native minus all-man9). Blue color indicates decrease in edge weight, and red indicates increase in edge weight in native network, as compared with all-man9.

(B and C) (B) Increase and (C) decrease in connectivity due to native glycosylation in comparison with all-man9 model (only those edges that change by 10% or more are shown).

(D) Five different communities were identified based on modularity maximization. The sub-community junctions identify susceptible regions in the shield where the shown HIV antibodies tend to bind.

(E) Location of each community projected on the Env surface.

correlated microdomain based on interdependent sequon frequencies (Hargett et al., 2019). Glycans 133, 137, 197, 363, and 386 (3, yellow) and 339, 392, 398, 406, and 411 (4, red) form two distinct communities involving the glycans on and surrounding V4 loop region. Rest of the glycans 88, 234, 276, 355, and 462 from gp120 and all four glycans from gp41 form the fifth community (5, cyan), although the modularity value is lower due to sparser connections. The possibility of an approaching probe reaching the protein surface through these strongly connected communities is low. A similar study clustering glycans from microsecond simulation runs of BG505 SOSIP was performed by Lemmin et al. (2017), identifying four glycan microdomains that roughly correspond to our modules 1, 2, (3 + 4), and 5. However, in that study, mannose-5 glycoform was used at all sites, and owing to the smaller length of these glycans, some of the interactions, including inter-protomer interactions, were not observed. Remarkably, in that study, junctions between microdomains were found to indicate regions of relative vulnerability. The communities we identified also demarcate the regions where the glycan shield can be penetrated. Broadly neutralizing antibodies (bNAbs) whose binding epitopes are known, target these community boundaries (Figure 5D). Thus, glycan community dynamics can help to determine susceptible regions of the glycan shield and can be further used for guiding immunological studies.

Quantifying the Vulnerability of Glycan Shield for Antibody-Based Neutralization

To capture the impact of the glycan shield acting as an immunological barrier, we quantified the sugar barrier over the Env protein surface using the ensemble of structures that have been generated. Toward this

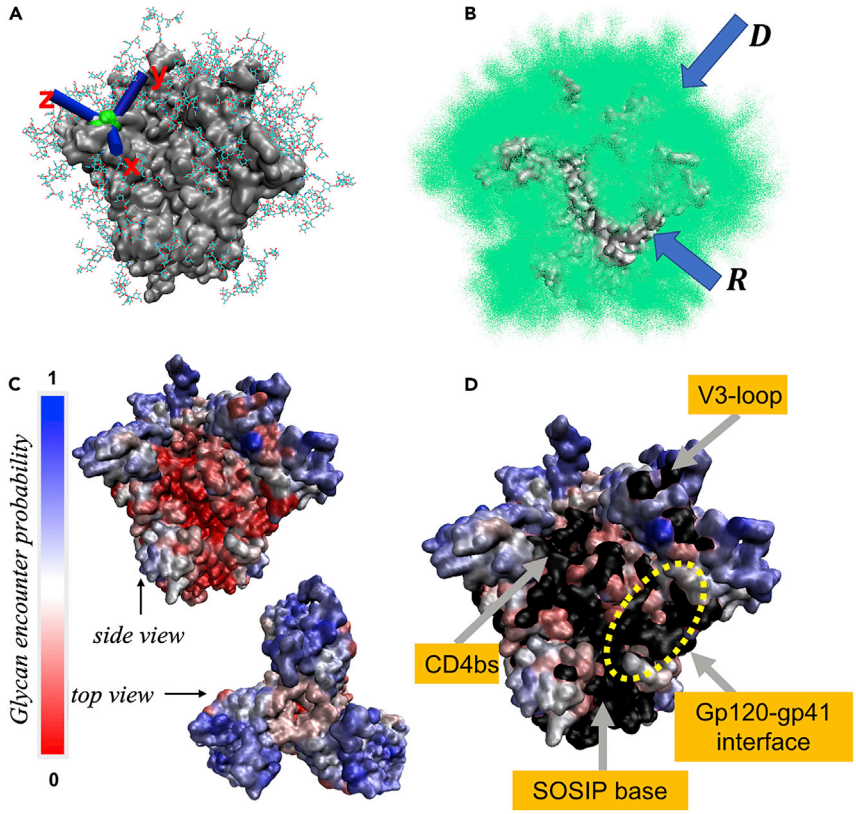


Figure 6. Glycan Encounter Factor (GEF) for Quantifying Shielding Effect

(A) At each residue present on the surface of the protein, the approaching probe is considered in three directions **x**, **y**, and **z**.

(B) Any point on the surface that has a dense glycan covering, such as D, has a high glycan encounter factor value, versus a point such as R where the glycan covering is sparse, which will have a low GEF.

(C and D) (C) Representation of GEF on Env surface, given by a colormap. Regions of BG505 surface having a GEF less than 0.14 is colored in black (D). Typical known antibody epitopes are indicated by arrows. BG505-specific GH and COT epitope region demarcated by yellow dashed circle.

goal, we have defined the glycan encounter factor (GEF) as the number of glycan heavy atoms encountered by an external probe approaching the protein surface. This parameter is calculated at each residue on the surface of the protein within a probe of diameter 6Å calculated using our ensemble. The highest value obtained is 14 for the native model and is located in the HMP. Based on existing structures, the main interaction points between Env and bNAbs are often hairpin-like loop regions. Even large-scale atomistic simulations suggest that the first line of contact between Env and an antibody is through a loop region (Schmidt et al., 2013). Accordingly, for our analysis, we have used a probe size of 6Å, which is the typical diameter of a hairpin loop. At each residue present on the surface of the protein, the approaching probe was considered in three directions (Figure 6A): perpendicular to the surface **z**, and then the **x** and **y** directions spanning the plane parallel to the surface. The geometric mean of the **x**, **y**, and **z** values provide the GEF per residue on the Env surface. This value will go to zero when the glycan encounter factor is zero from any one of the three cardinal directions. Thus, for any point on the surface that has a dense glycan covering, such as D in Figure 6B, it has a high glycan encounter factor value, versus a point such as R where the glycan covering is sparse, which will have a low GEF.

Previous evidence suggests that 70% of the Env ECD surface area is covered by glycans (Pancera et al., 2014). Based on this, we determined the lower cutoff of GEF below which we can define glycan holes. Figure 6C shows GEFs mapped onto the trimer protein surface. The calculated solvent-exposed surface area of the protein part of our modeled BG505 structure without considering glycans is 86,055 $Å^2$. Excluding the exposed region at the base of the soluble SOSIP, this reduces to 79,672 $Å^2$. Varying the lower cutoff of GEF,

we find that a cutoff of 1.5 GEF leads to 30% of the surface to be exposed. Regions of BG505 surface having a GEF less than 1.5 is colored in black in Figure 6D. From the figure, it is clear that the typical glycan holes targeted by bNAbs in BG505, such as the CD4-binding site, the V3-loop epitope, and the fusion peptide-binding region, fall below this GEF cutoff. GEF tracks with epitopes that are relatively generic to a broad range of Env strains (Xu et al., 2018; Burton et al., 2012).

Antibodies elicited by BG505.SOSIP.664 are mainly biased toward the missing 241 and 289 glycan holes (GH) and the cleft-of-trimer (COT) epitope regions as demonstrated by cryo-EM and immunogenicity assays (Mccoy et al., 2016; Bianchi et al., 2018) (see Figure 1 in reference Bianchi et al., 2018). The GEF parameter identifies these BG505-specific epitopes as breaches in the glycan shield (Figure 6B, "R" region; Figure 6D yellow dashed circle). On the other hand, the densely glycosylated regions around V2, V4 loop and alpha 2 helix have high values of GEF. At each point, the GEF value is given by a combination of all glycans in the vicinity that can come in the way of the approaching probe. At the same time, we can analyze the extent of influence of each glycan on the protein surface. Thus, we are now equipped with a tool to quantify the barrier effect of the glycans individually or as a group for immunogen design, as further demonstrated (Mccoy et al., 2016; Bianchi et al., 2018) in the following section.

Combining Network Topology and Glycan Encounter Factor to Inform on Local and Global Effects of Neutralization

The proposed network is useful in deciphering the impact of addition or deletion of glycan on neutralization. Removal of the highly conserved glycan at 197 by mutating the sequon leads to enhanced neutralization sensitivity to a variety of CD4bs and V3-specific antibodies (Townsley et al., 2016; Li et al., 2008; Huang et al., 2012). This glycan is situated proximal to the CD4-binding site and the tip of V3 loop, directly affecting the binding of these antibodies. However, past experimental evidence also suggests that the deletion of this glycan at N197 increases the binding affinity of two antibodies PG9 and PGT145 (Behrens et al., 2016) that target the trimer apex of Env. The epitope of PG9 as determined from the PDB structure 5VJ6 includes residues 160, 161, 167–173, 185, 305, and 307 (Figure 7A), and the epitope of PGT145 as determined from PDB structure 5V8L includes residues 123, 124, 127, 160–162, and 167–169 (Figure 7A). The footprint of glycan 197 as per our GEF model is shown in Figure 7B. Residues of the antibody epitopes do not overlap with those regions directly covered by glycan 197. Yet, removal of glycan 197 results in significant reduction in the glycan encounter factor over the V2 antibody epitope as evidenced in Figures 7C and 7D.

The removal of glycan 197 affects the glycans that were originally acting as barriers over the epitopes to resample the available space in such a way that the barrier over the epitopes is now reduced. Glycans 156, 160, 185e, and 185h from the neighboring protomers directly shield the PG9 and PGT145 epitopes. Looking at the shortest paths of communication between residue 197 from either of the protomers to all other glycans (Figures S10A and S10B) demonstrates that, although the shortest path between glycans 160 and 197 is direct, other glycans covering the epitope regions communicate with 197 across neighboring protomers or via a series of inter-glycan interactions. Deletion of 197 does not only affect the V1/V2 loop region glycans. The difference in adjacency matrix due to this perturbation is illustrated in Figure S11A. There is a slight modification of the community structure within the network as well, with glycan 392 leaving module 4 (red) and entering module 3 (yellow) instead (Figure S11B). Interactions such as those between glycans N88–N625, N133–N363, and interprotomer N156–N185e are reduced more than 10%. On the other hand, glycan 276, which was originally interacting strongly with 197 now forms new interactions with N295, N332, and N411 glycans (Figures S11C and S11D).

DISCUSSION

Dense arrangement of N-glycans masking antigenic surfaces on Env acts as a dynamic shield from the adaptive immune system. Understanding the structure and dynamics of the glycan shield as a whole is therefore important for characterizing the immunologically inert regions of Env. Although high-resolution X-ray, NMR, and cryo-EM structures have supplied a number of important molecular details about the Env glycoprotein, they do not account for the high flexibility and dynamics of these glycans that leads to the glycan shield. Our integrated approach for capturing dynamics of the glycans beyond the core β-mannose and quantification of the barrier resilience and vulnerability as a function of glycosylation pattern provides an improved methodology for obtaining new perspective to current HIV vaccine design efforts.

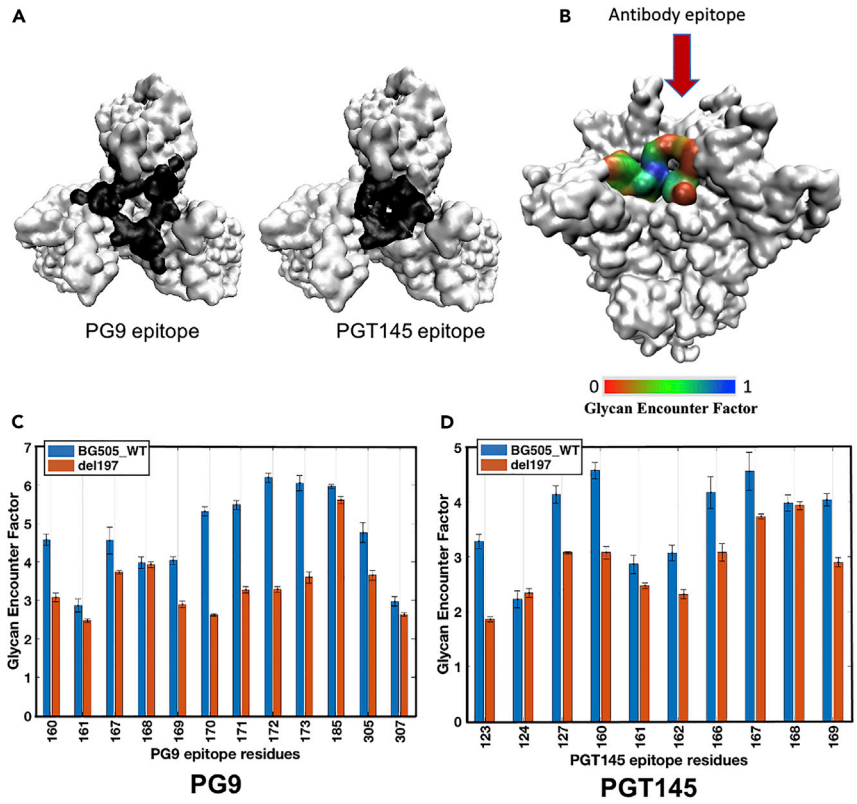


Figure 7. Deletion of Glycan N197 Decreases Glycan Shielding at PG9 and PGT145 Epitope

(A) Top view of Env showing epitope regions of antibodies PG9 and PGT145 at the apex.

(B) Footprint of glycan N197 on Env surface, colored by glycan encounter factor contributed by N197 alone.

(C and D) (C) Glycan encounter factor over the PG9 and (D) PGT145 epitope residues. GEF decreases for both the epitopes owing to deletion of N197 glycan. Standard deviations are shown as +/- error bars in the bar plots.

Previous computational studies have generalized glycan moieties as simple oligo-mannoses such as mannose-5 and mannose-9 (Ferreira et al., 2018; Yang et al., 2017; Lemmin et al., 2017; Stewart-Jones et al., 2016; Liang et al., 2016) for the purpose of modeling. For the first time, we have incorporated native glycosylation and included complex sugars based on site-specific mass spectrometry results. Our study suggests that the overall topology of the glycan shield remains similar globally between mannose-9 and native glycosylation. This gives credence to the use of oligomannose systems that have been extensively used in the literature for studying shield properties. However, there are some critical and somewhat subtle differences that arise. The most noteworthy are a consequence of the presence of the bulky fucose ring at the base or the negatively charged sialic acids at the antennae tips. These lead to overall rearrangement of glycan orientations affecting its microenvironment, ultimately influencing the shield topology. Thus, our comprehensive characterization of the shield is capable of capturing individual glycan effects that are physiologically and immunologically more relevant. Although the surface potential is more positive for Env protein, especially in gp120 (Kwon et al., 2015), the neutral MAN9 glycans and negatively charged sialated complex glycans modify the exposed electrostatic surface (see Figure S12) and can therefore influence antibody binding.

Our network-based approach illuminates the collective behavior of glycans. We compare the relative centrality of glycans, identify potential interaction pathways, and detect stable communities. The centrality or importance of glycans correlates well with experimental cryo-EM data (Berndsen et al., 2020; Amanat et al., 2018). Glycans with lower centrality have lesser influence on the graph and are the first to be eliminated from the network, if the adjacency threshold is increased. Similarly, the most central region of the network is the most resilient to enzymatic action. Such centrality measures can help determine the ease of targeting glycans and their modifications, guiding the process of immunogen development in the context of distinct Envs. Although the shield resilience is the main consideration for epitope exposure, some antibodies have

also evolved to capitalize on specific interactions with a number of the conserved glycans during their molecular engagement with the epitope. The stable glycan interactions within the detected communities help determine the antibody angle of approach that has been shown to influence the breadth of bNAbs (Moore and Williamson, 2016).

There are known common and uncommon "glycan holes" that can open up in the shield to make the virus more vulnerable, or conversely get covered resulting in immune escape, as a result of evolutionary addition/deletion or shift of glycans (Wagh et al., 2018). Although some advances have been made previously in identifying these breaches based on the area of influence of each glycan (Wagh et al., 2018), here we have found that this area can vary depending on the glycan type, charge, neighbors, etc. Amino acid signature analyses suggest that even minor perturbations such as single site mutations could potentially change the shielding effect over certain epitopes (Bricault et al., 2019). Therefore, we have derived a measure to quantify the shielding effect based on the encounter of glycans over the antigenic protein surface. This tool allows us to define regions of relative vulnerability and resilience in the glycan shield.

The method we developed here for the structural modeling of the glycoprotein atomistic ensemble (Figures S13 and S14) and the subsequent development of the network is high-throughput compared with traditional sampling by MD simulations (see Supplementary Information). Here we have tested it against BG505, which has been extremely well characterized as a model system. Comparison of glycan profile between recombinant SOSIPs and virion-derived Envs has suggested that the resultant glycosylation differences can cause hotspots of elevated glycan processing in the latter (Struwe et al., 2018) and lead to altered antigenicity (Cao et al., 2018a). Thus, knowledge of how the shield topology can change based on glycoforms is critical for guiding effective vaccine design and manufacturing protocols. The computational toolset we have presented here, along with insights into native glycosylation effects, provide a good handle to obtain glycan-variation-based structural description of the shield. We are now equipped with a tool to perform comparative structural studies due to glycan additions, deletions, or modifications, as well as other variations in the Env sequence. This is important to understand the evolution of the glycan shield over longitudinal sampling of lineages. The structural basis of addition or removal of glycans that are known to drive antibody maturation and neutralization activities (Labranche et al., 2018; Umotoy et al., 2019) can also be easily investigated. Capitalizing on the relatively low computational overhead, our pipeline can be integrated into a polyclonal epitope mapping assay to track the glycan shield as a function of hierarchical antibody response, by modeling the shield network with the presence of different antibodies and observing how the topology is altered.

Beyond HIV-1 Env and other viral envelope proteins, the significance of glycoproteins in a vast array of biological processes from protein folding, cellular communication to immune-regulation makes them a fast-emerging field of interest in biomedical research. Changes in these glycosylation patterns have been associated with various diseases, including rheumatoid arthritis (Nakagawa et al., 2007) and cancers (Taniguchi and Kizuka, 2015). Additionally, many of the current therapeutic antibodies in the market are N-linked glycoproteins, and the significance of N-glycans is becoming increasingly evident (Dalziel et al., 2014). The ease of modeling of glycan network utilizing our approach makes it translatable to other systems and can assist in determining the role of these glycans in conjugation with the underlying proteins at a molecular level. In comparison with N-glycans, O-linked glycosylations are relatively more complex in terms of structural variations, and our understanding is limited. Generalization of our methodology to encompass O-glycosylated systems can provide molecular insights on such refractory systems, like Ebola that has both N- and O-linked glycosylations, and explore the pathological implications of dense O-glycans in mucin-associated cancers (Bhatia et al., 2019).

In summary, HIV and the human immune system are at war constantly, and the virus uses the Env glycan shield to mask the human immune surveillance. One battlefront in this war is the glycan shield: the virus evolves to develop resilience, whereas the immune response counteracts by looking for vulnerabilities. As efforts are underway to aid the immune system overcome this race by conditioning it with engineered immunogens, there is a need to quantify the resistance and vulnerability of glycan shield in a more quantitative manner. This is the first time that the native glycan network and shielding has been spatially quantified. Our derivation of the Glycan Encounter Factor measures the relative barrier over the Env protein surface and can potentially aid to distinguish subtle differences on the shield due to variations in the glycosylation or even protein sequences. We are therefore armed with a set of validated *in silico* tools with which to help the anti-HIV war efforts and guide immunogen design.

Limitations of the Study

Although the high-throughput modeling method used in this study provides extensive conformational sampling, we cannot form any conjecture on the extent to which it can emulate the population distributions as obtained from large-scale or enhanced sampling-based MD simulations. Extensive comparison with MD simulations is required to see the performance differences and will form the basis of an important future exploration. Nevertheless, published MD studies also suffer from sampling limitations associated with strongly persisting glycan-glycan interactions and extended glycan torsional landscape (Yang et al., 2017; Woods, 2018). For the purpose of this study, reasonably small standard deviation values seen in the per-glycan RMSF and volume plots (Figure S9) suggest that there is convergence within the modeled glycan ensemble. Moreover, this method accurately captures the torsional space of glycans as seen in Figures S1–S4. Another limitation here is our utilization of a 6Å probe, while being close to a hairpin loop which is usually the first line of contact by antibodies, may not encompass all possible interaction poses between the antibodies and the epitopes. Also, much of the theoretical network topologies predicted in this study lack wet laboratory experimental evidence. Orthogonal experiments, perhaps through further SPECM or mass spectrometry analysis along with site-directed glycan removal would help to directly validate these networks.

Resource Availability

Lead Contact

Further requests for resources and materials should be directed to the lead contact, S. Gnanakaran (gnana@lanl.gov).

Materials Availability

This study did not yield new unique reagents.

Data and Code Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Further original data have been deposited to Mendeley Data: https://doi.org/10.17632/yds8rtfjjy.1.

Additional data and algorithms related to this paper may be requested from the authors.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101836.

ACKNOWLEDGMENTS

S.C., B.T.K., and S.G. were funded by grants from the NIH (Center for HIV/AIDS Vaccine Immunology and immunogen Discovery, UM1 AI100645; Consortia for HIV/AIDS Vaccine Development, UM1 AI144371). Z.T.B. and A.B.W. were supported by the National Institute of Allergy and Infectious Diseases grants: UM1 AI100663 and UM1 AI144462, and Collaboration for AIDS Vaccine Discovery (CAVD) grant OPP1115782. S.C. was also partially supported by the Center for Nonlinear Studies (CNLS) at Los Alamos National Laboratory (LANL). We acknowledge Timothy Travers and Cesar Lopez for suggestions on glycan modeling, Kshitij Wagh for insights on glycosylation holes in HIV, Kevin Weihe and Rory Henderson from Duke University for helpful discussions on Env immune response. S.C. and S.G. acknowledge the LANL High Performance Computing Division for providing computational facilities.

AUTHOR CONTRIBUTIONS

Conceptualization, S.C., Z.T.B., N.W.H., B.T.K., A.B.W., and S.G; Methodology, S.C., Z.T.B., N.W.H., A.B.W., and S.G.; Modeling and Simulations, S.C.; Cryo-EM experiments, Z.T.B.; Formal Analysis, S.C. and Z.T.B.; Visualization, S.C. and Z.T.B.; Manuscript Preparation, S.C., Z.T.B., N.W.H., B.T.K., A.B.W., and S.G.; Supervision, A.B.W. and S.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 10, 2020
Revised: October 22, 2020
Accepted: November 17, 2020
Published: December 18, 2020

REFERENCES

Amanat, F., Duehr, J., Oestereich, L., Hastie, K.M., Ollmann Saphire, E., and Krammer, F. (2018). Antibodies to the glycoprotein GP2 subunit cross-react between old and new world arenaviruses. mSphere *3*, e00189–e00218.

An, H.J., Froehlich, J.W., and Lebrilla, C.B. (2009). Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. Curr. Opin. Chem. Biol. *13*, 421–426.

Barnes, C.O., Gristick, H.B., Freund, N.T., Escolano, A., Lyubimov, A.Y., Hartweger, H., West, A.P.,, JR., Cohen, A.E., Nussenzweig, M.C., and Bjorkman, P.J. (2018). Structural characterization of a highly-potent V3-glycan broadly neutralizing antibody bound to natively-glycosylated HIV-1 envelope. Nat. Commun. *9*, 1251.

Behrens, A.J., Vasiljevic, S., Pritchard, L.K., Harvey, D.J., Andev, R.S., Krumm, S.A., Struwe, W.B., Cupo, A., Kumar, A., Zitzmann, N., and Crispin, M. (2016). Composition and antigenic effects of individual glycan sites of a trimeric HIV-1 envelope glycoprotein. Cell Rep. *14*, 2695–2706.

Beleva Guthrie, V., Masica, D.L., Fraser, A., Federico, J., Fan, Y., Camps, M., and Karchin, R. (2018). Network analysis of protein adaptation: modeling the functional impact of multiple mutations. Mol. Biol. Evol. *35*, 1507–1519.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. Nucleic Acids Res. *28*, 235–242.

Berndsen, Z., Chakraborty, S., Wang, X., Cottrell, C., Torres, J., Diedrich, J.K., Lopez, C.A., Yates, J.R., III, Van-Gills, M.J., Paulson, J.C., et al. (2020). Visualization of the HIV-1 env glycan shield across scales. Proc. Natl. Acad. Sci. U S A *117*, 28014–28025.

Bhatia, R., Gautam, S.K., Cannon, A., Thompson, C., Hall, B.R., Aithal, A., Banerjee, K., Jain, M., Solheim, J.C., Kumar, S., and Batra, S.K. (2019). Cancer-associated mucins: role in immune modulation and metastasis. Cancer Metastasis Rev. *38*, 223–236.

Bianchi, M., Turner, H.L., Nogal, B., Cottrell, C.A., Oyen, D., Pauthner, M., Bastidas, R., Nedellec, R., Mccoy, L.E., Wilson, I.A., and Burton, D.R. (2018). Electron-microscopy-based epitope mapping defines specificities of polyclonal antibodies elicited during HIV-1 BG505 envelope trimer immunization. Immunity *49*, 288–300.e8.

Bradley, T., Fera, D., Bhiman, J., Eslamizar, L., Lu, X., Anasti, K., Zhang, R., Sutherland, L.L., Scearce, R.M., Bowman, C.M., and Stolarchuk, C. (2016). Structural constraints of vaccine-induced tier-2 autologous HIV neutralizing antibodies targeting the receptor-binding site. Cell Rep. *14*, 43–54.

Bricault, C.A., Yusim, K., Seaman, M.S., Yoon, H., Theiler, J., Giorgi, E.E., Wagh, K., Theiler, M., Hraber, P., Macke, J.P., and Kreider, E.F. (2019). HIV-1 neutralizing antibody signatures and application to epitope-targeted vaccine design. Cell Host Microbe *25*, 59–72.e8.

Burton, D.R., Ahmed, R., Barouch, D.H., Butera, S.T., Crotty, S., Godzik, A., Kaufmann, D.E., Mcelrath, M.J., Nussenzweig, M.C., Pulendran, B., and Scanlan, C.N. (2012). A Blueprint for HIV vaccine discovery. Cell Host Microbe *12*, 396–407.

Burton, D.R., and Mascola, J.R. (2015). Antibody responses to envelope glycoproteins in HIV-1 infection. Nat. Immunol. *16*, 571–576.

Cao, L., Diedrich, J.K., Kulp, D.W., Pauthner, M., He, L., Park, S.R., Sok, D., Su, C.Y., Delahunty, C.M., Menis, S., and Andrabi, R. (2017). Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. Nat. Commun. *8*, 14954.

Cao, L., Pauthner, M., Andrabi, R., Rantalainen, K., Berndsen, Z., Diedrich, J.K., Menis, S., Sok, D., Bastidas, R., and Park, S.-K.R. (2018a). Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer. Nat. Commun. *9*, 1–14.

Cao, L., Pauthner, M., Andrabi, R., Rantalainen, K., Berndsen, Z., Diedrich, J.K., Menis, S., Sok, D., Bastidas, R., Park, S.R., and Delahunty, C.M. (2018b). Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer. Nat. Commun. *9*, 3693.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., and Dwek, R.A. (2007). Glycoprotein structural genomics: solving the glycosylation problem. Structure *15*, 267–273.

Coss, K.P., Vasiljevic, S., Pritchard, L.K., Krumm, S.A., Glaze, M., Madzorera, S., Moore, P.L., Crispin, M., and Doores, K.J. (2016). HIV-1 glycan density drives the persistence of the mannose patch within an infected individual. J. Virol. *90*, 11132–11144.

Crispin, M., and Doores, K.J. (2015). Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. Curr. Opin. Virol. *11*, 63–69.

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

Dalziel, M., Crispin, M., Scanlan, C.N., Zitzmann, N., and Dwek, R.A. (2014). Emerging principles for the therapeutic exploitation of glycosylation. Science *343*, 1235681.

Davis, S.J., and Crispin, M. (2010). Solutions to the Glycosylation Problem for Low-and High-Throughput Structural Glycoproteomics. Functional and Structural Proteomics of Glycoproteins (Springer).

Dennis, J.W., Granovsky, M., and Warren, C.E. (1999). Protein glycosylation in development and disease. Bioessays *21*, 412–421.

Doores, K.J. (2015). The HIV glycan shield as a target for broadly neutralizing antibodies. FEBS J. *282*, 4679–4691.

Doores, K.J., Bonomelli, C., Harvey, D.J., Vasiljevic, S., Dwek, R.A., Burton, D.R., Crispin, M., and Scanlan, C.N. (2010). Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. Proc. Natl. Acad. Sci. U S A *107*, 13800–13805.

Dwek, R.A. (1996). Glycobiology: toward understanding the function of sugars. Chem. Rev. *96*, 683–720.

Eargle, J., and Luthey-Schulten, Z. (2012). NetworkView: 3D display and analysis of protein.RNA interaction networks. Bioinformatics *28*, 3000–3001.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics *15*, 5.6.1–5.6.30.

Falkowska, E., Le, K.M., Ramos, A., Doores, K.J., Lee, J.H., Blattner, C., Ramirez, A., Derking, R., Van Gils, M.J., Liang, C.H., and Mcbride, R. (2014). Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. Immunity *40*, 657–668.

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.

Floyd, R.W. (1962). Algorithm 97: shortest path. Commun. ACM *5*, 345.

Fruchterman, T.M., and Reingold, E.M. (1991). Graph drawing by force-directed placement. Software Pract. Ex. *21*, 1129–1164.

Go, E.P., Ding, H., Zhang, S., Ringe, R.P., Nicely, N., Hua, D., Steinbock, R.T., Golabek, M., Alin, J., Alam, S.M., and Cupo, A. (2017). Glycosylation benchmark profile for HIV-1 envelope

glycoprotein production based on eleven env trimers. J. Virol. *91*, e02428-16.

Gristick, H.B., Von Boehmer, L., West, A.P., Jr., Schamber, M., Gazumyan, A., Golijanin, J., Seaman, M.S., Fatkenheuer, G., Klein, F., Nussenzweig, M.C., and Bjorkman, P.J. (2016). Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. Nat. Struct. Mol. Biol. *23*, 906–915.

Guttman, M., Weinkam, P., Sali, A., and Lee, K.K. (2013). All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. Structure *21*, 321–331.

Guvench, O., Mallajosyula, S.S., Raman, E.P., Hatcher, E., Vanommeslaeghe, K., Foster, T.J., Jamison, F.W., and Mackerell, A.D., Jr. (2011). CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. J. Chem. Theory Comput. *7*, 3162–3180.

Hargett, A.A., Wei, Q., Knoppova, B., Hall, S., Huang, Z.Q., Prakash, A., Green, T.J., Moldoveanu, Z., Raska, M., Novak, J., and Renfrow, M.B. (2019). Defining HIV-1 envelope N-glycan microdomains through site-specific heterogeneity profiles. J. Virol. *93*, e01177-18.

Haynes, B.F., Moody, M.A., Alam, M., Bonsignori, M., Verkoczy, L., Ferrari, G., Gao, F., Tomaras, G.D., Liao, H.X., and Kelsoe, G. (2014). Progress in HIV-1 vaccine development. J. Allergy Clin. Immunol. *134*, 3–10, quiz 11.

Huang, J., Kang, B.H., Pancera, M., Lee, J.H., Tong, T., Feng, Y., Imamichi, H., Georgiev, I.S., Chuang, G.Y., Druz, A., and Doria-Rose, N.A. (2014). Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. Nature *515*, 138–142.

Huang, L., Liao, L., and Wu, C.H. (2017). Evolutionary analysis and interaction prediction for protein-protein interaction network in geometric space. PLoS One *12*, e0183495.

Huang, X., Jin, W., Hu, K., Luo, S., Du, T., Griffin, G.E., Shattock, R.J., and Hu, Q. (2012). Highly conserved HIV-1 gp120 glycans proximal to CD4-binding region affect viral infectivity and neutralizing antibody induction. Virology *423*, 97–106.

Ilinykh, P.A., Santos, R.I., Gunn, B.M., Kuzmina, N.A., Shen, X., Huang, K., Gilchuk, P., Flyak, A.I., Younan, P., Alter, G., et al. (2018). Asymmetric antiviral effects of ebolavirus antibodies targeting glycoprotein stem and glycan cap. PLoS Pathog. *14*, e1007204.

Imberty, A., and Perez, S. (2000). Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. Chem. Rev. *100*, 4567–4588.

Julien, J.P., Cupo, A., Sok, D., Stanfield, R.L., Lyumkis, D., Deller, M.C., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., et al. (2013). Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science *342*, 1477–1483.

Karsten, C.B., and Alter, G. (2017). The HIV-1 glycan shield: strategically placed kinks in the armor improve antigen design. Cell Rep. *19*, 669–670.

Kong, L., Torrents De La Pena, A., Deller, M.C., Garces, F., Sliepen, K., Hua, Y., Stanfield, R.L., Sanders, R.W., and Wilson, I.A. (2015). Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. Acta Crystallogr. D Biol. Crystallogr. *71*, 2099–2108.

Kwon, Y.D., Pancera, M., Acharya, P., Georgiev, I.S., Crooks, E.T., Gorman, J., Joyce, M.G., Guttman, M., Ma, X., Narpala, S., and Soto, C. (2015). Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. Nat. Struct. Mol. Biol. *22*, 522–531.

Labranche, C.C., Mcguire, A.T., Gray, M.D., Behrens, S., Kwong, P.D.K., Chen, X., Zhou, T., Sattentau, Q.J., Peacock, J., Eaton, A., and Greene, K. (2018). HIV-1 envelope glycan modifications that permit neutralization by germline-reverted VRC01-class broadly neutralizing antibodies. PLoS Pathog. *14*, e1007431.

Lasky, L.A., Groopman, J.E., Fennie, C.W., Benz, P.M., Capon, D.J., Dowbenko, D.J., Nakamura, G.R., Nunes, W.M., Renz, M.E., and Berman, P.W. (1986). Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein. Science *233*, 209–212.

Lee, J.H., Ozorowski, G., and Ward, A.B. (2016). Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. Science *351*, 1043–1048.

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

Lennemann, N.J., Rhein, B.A., Ndungo, E., Chandran, K., Qiu, X., and Maury, W. (2014). Comprehensive functional analysis of N-linked glycans on Ebola virus GP1. MBio *5*, e00862-13.

Li, Y., Cleveland, B., Klots, I., Travis, B., Richardson, B.A., Anderson, D., Montefiori, D., Polacino, P., and Hu, S.L. (2008). Removal of a single N-linked glycan in human immunodeficiency virus type 1 gp120 results in an enhanced ability to induce neutralizing antibody responses. J. Virol. *82*, 638–651.

Liang, Y., Guttman, M., Williams, J.A., Verkerke, H., Alvarado, D., Hu, S.L., and Lee, K.K. (2016). Changes in structure and antigenicity of HIV-1 env trimers resulting from removal of a conserved CD4 binding site-proximal glycan. J. Virol. *90*, 9224–9236.

Lütteke, T., Frank, M., and Von Der Lieth, C.-W. (2005). Carbohydrate structure suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res. *33*, D242–D246.

Lyumkis, D., Julien, J.P., De Val, N., Cupo, A., Potter, C.S., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., Carragher, B., et al. (2013). Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. Science *342*, 1484–1490.

Mathys, L., Francois, K.O., Quandte, M., Braakman, I., and Balzarini, J. (2014). Deletion of the highly conserved N-glycan at Asn260 of HIV-1 gp120 affects folding and lysosomal degradation of gp120, and results in loss of viral infectivity. PLoS One *9*, e101181.

Mccoy, L.E., Van Gils, M.J., Ozorowski, G., Messmer, T., Briney, B., Voss, J.E., Kulp, D.W., Macauley, M.S., Sok, D., Pauthner, M., and Menis, S. (2016). Holes in the glycan shield of the native HIV envelope are a target of trimer-elicited neutralizing antibodies. Cell Rep. *16*, 2327–2338.

Moore, P.L., and Williamson, C. (2016). Approaches to the induction of HIV broadly neutralizing antibodies. Curr. Opin. HIV AIDS *11*, 569.

Moremen, K.W., Tiemeyer, M., and Nairn, A.V. (2012). Vertebrate protein glycosylation: diversity, synthesis and function. Nat. Rev. Mol. Cell Biol. *13*, 448–462.

Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J.F., Halper-Stromberg, A., Gnanapragasam, P.N., Spencer, D.I., Seaman, M.S., and Schuitemaker, H. (2012). Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. Proc. Natl. Acad. Sci. U S A *109*, E3268–E3277.

Nakagawa, H., Hato, M., Takegawa, Y., Deguchi, K., Ito, H., Takahata, M., Iwasaki, N., Minami, A., and Nishimura, S. (2007). Detection of altered N-glycan profiles in whole serum from rheumatoid arthritis patients. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. *853*, 133–137.

Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.Y., Ofek, G., and Stewart-Jones, G.B. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature *514*, 455–461.

Pritchard, L.K., Spencer, D.I., Royle, L., Bonomelli, C., Seabright, G.E., Behrens, A.J., Kulp, D.W., Menis, S., Krumm, S.A., Dunlop, D.C., and Crispin, D.J. (2015a). Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. Nat. Commun. *6*, 7479.

Pritchard, L.K., Vasiljevic, S., Ozorowski, G., Seabright, G.E., Cupo, A., Ringe, R., Kim, H.J., Sanders, R.W., Doores, K.J., Burton, D.R., et al. (2015b). Structural constraints determine the glycosylation of HIV-1 envelope trimers. Cell Rep. *11*, 1604–1613.

Ringe, R.P., Pugach, P., Cottrell, C.A., Labranche, C.C., Seabright, G.E., Ketas, T.J., Ozorowski, G., Kumar, S., Schorcht, A., Van Gils, M.J., and Crispin, M. (2019). Closing and opening holes in the glycan shield of HIV-1 envelope glycoprotein SOSIP trimers can redirect the neutralizing antibody response to the newly unmasked epitopes. J. Virol. *93*, e01656-18.

Sali, A. (1995). Comparative protein modeling by satisfaction of spatial restraints. Mol. Med. Today *1*, 270–277.

Sanders, R.W., Derking, R., Cupo, A., Julien, J.P., Yasmeen, A., De Val, N., Kim, H.J., Blattner, C., De La Pena, A.T., Korzun, J., and Golabek, M. (2013). A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses

multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. PLoS Pathog. *9*, e1003618.

Sanders, R.W., Van Gils, M.J., Derking, R., Sok, D., Ketas, T.J., Burger, J.A., Ozorowski, G., Cupo, A., Simonich, C., Goo, L., et al. (2015). HIV-1 VACCINES. HIV-1 neutralizing antibodies induced by native-like envelope trimers. Science *349*, aac4223.

Saphire, E.O., Schendel, S.L., Fusco, M.L., Gangavarapu, K., Gunn, B.M., Wec, A.Z., Halfmann, P.J., Brannan, J.M., Herbert, A.S., Qiu, X., and Wagh, K.; Viral Hemorrhagic Fever Immunotherapeutic Consortium (2018). Systematic analysis of monoclonal antibodies against ebola virus GP defines features that contribute to protection. Cell *174*, 938–952.e13.

Schmidt, A.G., Xu, H., Khan, A.R., O'Donnell, T., Khurana, S., King, L.R., Manischewitz, J., Golding, H., Suphaphiphat, P., Carfi, A., and Settembre, E.C. (2013). Preconfiguration of the antigen-binding site during affinity maturation of a broadly neutralizing influenza virus antibody. Proc. Natl. Acad. Sci. U S A *110*, 264–269.

Sethi, A., Tian, J., Derdeyn, C.A., Korber, B., and Gnanakaran, S. (2013). A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein. PLoS Comput. Biol. *9*, e1003046.

Skjaerven, L., Yao, X.Q., Scarabelli, G., and Grant, B.J. (2014). Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC Bioinformatics *15*, 399.

Slynko, V., Schubert, M., Numao, S., Kowarik, M., Aebi, M., and Allain, F.H. (2009). NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. J. Am. Chem. Soc. *131*, 1274–1281.

Sommerstein, R., Flatz, L., Remy, M.M., Malinge, P., Magistrelli, G., Fischer, N., Sahin, M., Bergthaler, A., Igonet, S., Ter Meulen, J., et al. (2015). Arenavirus glycan shield promotes neutralizing antibody evasion and protracted infection. PLoS Pathog. *11*, e1005276.

Stanley, P., Taniguchi, N., and Aebi, M. (2017). N-Glycans. In Essentials of Glycobiology, Ajit Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, and N.H. Packer, et al., eds. (Cold Spring Harbor Laboratory Press), p. Chapter 9.

Stewart-Jones, G.B., Soto, C., Lemmin, T., Chuang, G.Y., Druz, A., Kong, R., Thomas, P.V., Wagh, K., Zhou, T., Behrens, A.J., et al. (2016). Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. Cell *165*, 813–826.

Struwe, W.B., Chertova, E., Allen, J.D., Seabright, G.E., Watanabe, Y., Harvey, D.J., Medina-Ramirez, M., Roser, J.D., Smith, R., and Westcott, D. (2018). Site-specific glycosylation of virion-derived HIV-1 Env is mimicked by a soluble trimeric immunogen. Cell Rep. *24*, 1958–1966.e5.

Szakonyi, G., Klein, M.G., Hannan, J.P., Young, K.A., Ma, R.Z., Asokan, R., Holers, V.M., and Chen, X.S. (2006). Structure of the Epstein-Barr virus major envelope glycoprotein. Nat. Struct. Mol. Biol. *13*, 996–1001.

Taniguchi, N., and Kizuka, Y. (2015). Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. Adv. Cancer Res. *126*, 11–51.

Townsley, S., Li, Y., Kozyrev, Y., Cleveland, B., and Hu, S.L. (2016). Conserved role of an N-linked glycan on the surface antigen of human immunodeficiency virus type 1 modulating virus sensitivity to broadly neutralizing antibodies against the receptor and coreceptor binding sites. J. Virol. *90*, 829–841.

Umotoy, J., Bagaya, B.S., Joyce, C., Schiffner, T., Menis, S., Saye-Francisco, K.L., Biddle, T., Mohan, S., Vollbrecht, T., Kalyuzhniy, O., et al. (2019). Rapid and focused maturation of a VRC01-class HIV broadly neutralizing antibody lineage involves both binding and accommodation of the N276-Glycan. Immunity *51*, 141–154.e6.

Vankadari, N., and Wilce, J.A. (2020). Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. Emerg. Microbes Infect. *9*, 601–604.

Wagh, K., Kreider, E.F., Li, Y., Barbian, H.J., Learn, G.H., Giorgi, E., Hraber, P.T., Decker, T.G., Smith, A.G., Gondim, M.V., et al. (2018). Completeness of HIV-1 envelope glycan shield at transmission determines neutralization breadth. Cell Rep. *25*, 893–908.e7.

Walker, L.M., Phogat, S.K., Chan-Hui, P.Y., Wagner, D., Phung, P., Goss, J.L., Wrin, T., Simek, M.D., Fling, S., Mitcham, J.L., et al. (2009). Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science *326*, 285–289.

Walker, L.M., Simek, M.D., Priddy, F., Gach, J.S., Wagner, D., Zwick, M.B., Phogat, S.K., Poignard, P., and Burton, D.R. (2010). A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. PLoS Pathog. *6*, e1001028.

Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., Dimaio, F., Bosch, B.J., and Veesler, D. (2016). Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat. Struct. Mol. Biol. *23*, 899–905.

Walsh, G. (2010). Biopharmaceutical benchmarks 2010. Nat. Biotechnol. *28*, 917–924.

Wang, C.C., Chen, J.R., Tseng, Y.C., Hsu, C.H., Hung, Y.F., Chen, S.W., Chen, C.M., Khoo, K.H., Cheng, T.J., Cheng, Y.S., et al. (2009). Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc. Natl. Acad. Sci. U S A *106*, 18137–18142.

Warshall, S. (1962). A theorem on boolean matrices. J. ACM *9*, 11–12.

Weinkam, P., Pons, J., and Sali, A. (2012). Structure-based model of allostery predicts coupling between distant sites. Proc. Natl. Acad. Sci. U S A *109*, 4875–4880.

Woods, R.J. (2018). Predicting the structures of glycans, glycoproteins, and their complexes. Chem. Rev. *118*, 8005–8024.

Woods, R.J., Pathiaseril, A., Wormald, M.R., Edge, C.J., and Dwek, R.A. (1998). The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. Eur. J. Biochem. *258*, 372–386.

Xu, K., Acharya, P., Kong, R., Cheng, C., Chuang, G.Y., Liu, K., Louder, M.K., O'Dell, S., Rawi, R., Sastry, M., et al. (2018). Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. Nat. Med. *24*, 857–867.

Yang, M., Huang, J., Simon, R., Wang, L.X., and Mackerell, A.D.,, JR. (2017). Conformational heterogeneity of the HIV envelope glycan shield. Sci. Rep. *7*, 4435.

Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., and Korber, B. (2004). Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology *14*, 1229–1246.

Zhang, X., Jia, R., Shen, H., Wang, M., Yin, Z., and Cheng, A. (2017). Structures and functions of the envelope glycoprotein in flavivirus infections. Viruses *9*, 338.

iScience, Volume *23*

Supplemental Information

Quantification of the Resilience and Vulnerability of HIV-1 Native Glycan Shield at Atomistic Detail

Srirupa Chakraborty, Zachary T. Berndsen, Nicolas W. Hengartner, Bette T. Korber, Andrew B. Ward, and S. Gnanakaran

Supplemental Information

Quantification of the Resilience and Vulnerability of HIV-1 Native Glycan Shield at Atomistic Detail

Srirupa Chakraborty[1,2,*], Zachary T. Berndsen[3,4,*], Nicolas W. Hengartner[1], Bette T. Korber[1], Andrew B. Ward[3,4,5], and S. Gnanakaran[1,5]

[1]Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2]Center for Non-Linear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[4]IAVI Neutralizing Antibody Center and Collaboration of AIDS Vaccine Discovery, The Scripps Research Institute, La Jolla, CA 92037, USA

*These authors contributed equally

[5]Corresponding. authors: andrew@scripps.edu, gnana@lanl.gov

Modeled glycan ensembles sample a biologically relevant landscape:

The PNGS asparagine chi1 dihedral, and the phi and psi dihedral distributions of 9 different inter-glycan linkages within the ensemble was compared with those obtained from different glycan structures available in the PDB database, as described in **Figure S1-S4**. Torsion angle distributions from the PDB were obtained using GlyTorsion (Lütteke et al., 2005). These distributions match well between our generated ensemble and the PDB structures, Four of the psi angles, namely, the (i)1-3 linkage between two α-mannose sugars; (ii) 1-3 linkage between α-mannose and β-mannose; (iii) 1-2 linkage between two α-mannoses; and (iv) N-acetylglucosamine and α-mannose sample a wider distribution of angles in our models. It must be kept in mind that while the PDB distributions included ~17,000 structures for the ASN and ~7,000 structures for 1-4 linkage between two N-acetylglucosamines at the glycan core, there were less than 2000 structures for all other distributions. Also, these sugar linkages in the PDB come from both N-glycan and O-glycan types, have conformational restraints stemming from experimental structure determination,

and are not nearly as densely packed as in our HIV model. Our ensemble, on the other hand, has 84 glycans, with 1000 conformations each. The larger number of structures per linkage, as well as the dense interactions between neighboring glycans can result in some of the distributions sampling wider range of torsion angles.

Glycan fluctuations and their dependence on complex sugars

Glycan dynamics, type, and inter-glycan interactions determine the local shielding effect over the Env protein. We have looked at the dynamics of individual glycans and how they change due to native-like glycosylation, ultimately governing the glycan shield properties. Each glycan samples a large region in space as shown in **Figure S6A and S6B**. These fluctuations of course become much more extensive, when the glycans are present on the variable loops, due to the dynamic movement of the protein backbone of the loops at these regions. The root mean squared fluctuations (RMSF) measure and the sampled volume for those glycans located on the loops are generally much higher, as seen in **Figure S8A** and **Figure S8C**. However, aligning the protein backbone and considering the reduced RMSF contribution coming only from the glycans (see Methods), we see that the fluctuations between different glycans are comparable, with a maximum difference of ~1.5Å (**Figure S8B**). As for the glycan specific sampled volume, after removing the flexibilities brought on by protein loop motions, the glycans on loop V2 and V4 and those in gp41 cover the largest regions in space (**Figure S8D**).

Considering the differences between native-like glycosylation and the all-man9 model, some interesting consistent patterns emerge. Comparing the reduced RMSF between native and all-man9 glycosylation in **Figure S8E,** we see that while glycans 185h, 197, 355, 398 and 406 have decreased fluctuations in the native model, those of glycans 88, 462, 611, 618, 625 and 637 have increased significantly. With the exception of 355, which is a fucosylated hybrid (FH), all of these glycans are complex sugars. The presence of the charged sialic acid at the tips of these complex glycans can dictate the interactions between neighbors, increasing their structural variations if surrounded by other charged glycans, or reducing them if a stable conformation buried between uncharged high-mannose patches can be found. The FH glycan at N355 itself is centrally located between a number charged glycans, those at 88, 398, 462, and some of the gp41 sugars. The reduced fluctuations of this glycan can stem from its attempt to screen these neighboring

negatively charged sugars, thus forming stabilizing interactions. The complex sugars in gp41 (611, 618, 625 and 637) have also been shown to have high variations in cryo-EM maps(Ward and Wilson, 2017). The total volume sampled is generally larger for native glycosylation, compared to the all-man9 model (**Figure S8F)**. This is not unexpected, since complex glycans generally have larger number of sugars, including the bulky fucose ring at the base. The centrally located high-mannose patch has almost similar volumes in both the models.

Performance comparison with classical MD:

Classical MD requires initial modeling of glycans based on available tools (allosmod/glycosciences/toucan/glycam). This is followed by equilibration and production run. For a similar system size (~500,000 atoms), the throughput obtained for NAMD based multi-CPU runs in 2500 parallel Intel Xeon Broadwell processors is ~70 ns per day, or two weeks of continuous runs for 1μs. Extensive sampling will require several microseconds of runs, either in classical or through enhanced sampling. Though at this point there is no direct method to compare the performance of the two since the ALLOSMOD method does not give any time evidence, this high-throughput method performs at 100 model generation in ~22 hours. 10 sets were run in parallel to obtain ensemble of 1000 models.

TRANSPARENT METHODS

High-throughput conformational modeling

The ensemble of BG505 glycoprotein 3D conformations were built in atomistic detail by implementing the ALLOSMOD(Weinkam et al., 2012, Guttman et al., 2013) package of MODELLER(Eswar et al., 2006, Sali, 1995). For ease of understanding, the modeling protocol has been described as a flow schematic in **SI figure S13.** The protocol consists of these steps:

(a) ***Base template selection for protein scaffold:*** The BG505 protein scaffold was homology modeled, by threading the protein sequence into available crystal structure templates. Ligand-free SOSIP structure with PDB accession ID 4ZMJ(Kwon et al., 2015) was used as the base template for gp120, that also guided the three-fold symmetry of the trimer. 5CEZ(Garces et al., 2015) was used as the base template for gp41 and since it has the least number of missing residues among the available structures.

(b) ***Identifying flexible regions from available experimental PDB structures to guide conformational sampling:*** There are a large number of unstructured flexible regions in the protein. In order to incorporate the fluctuations as dictated by the available structural data we devised the following protocol. We aimed to construct 10 distinct protein models for these flexible regions as starting scaffolds for glycan building. In order to model these, 34 different BG505 SOSIP experimental structures were selected below a resolution cut-off of 4Å (see **SI Table S2**). Open-like, chimeric and synthetic models were not considered. The set of 34 were structurally aligned using PyMOL(Schrödinger, 2015) and their residue-wise backbone RMSF was calculated (**SI Figure S14A**). The base-level of 0.1 was used as a threshold in the normalized RMSF measure. The unstructured residues (coils and turns) were identified on 4ZMJ and 5CEZ using the STRIDE program(Frishman and Argos, 1995). All the identified regions with RMSF higher than the threshold (**SI Figure S14A**) were considered for modeling based on multiple templates. These include residue ranges 64 – 82, 133 – 153, 184 – 189, 298 – 303, 312 – 314, 324 – 329, 396 – 413, 425 – 431, 457 – 464, and 609 – 618 (in HXB2 numbering).

(c) ***Template selection for identified flexible regions:*** Some of these identified regions have missing residues in many of the PDBs (**SI Figure S14B**). Wherever enough models were

available, 10 distinct conformations were randomly selected as templates. If this number 10 could not be reached for some regions (for example, in loops V2 and V4) due to missing residues (**SI Figure S14B**), those residues were modeled *ab initio*. All known disulphide bonds were kept as additional restraints. We made sure these short-listed regions for *ab initio* modeling were less than MODELLER-prescribed 14 residues in length. This is because, modeling of such long loops are usually considered to be error-prone(Fiser et al., 2000). The loop region 546-569 in gp41 is present only in the structure 5CEZ. We did not model this region *ab initio*, and used 5CEZ as the only template for all models. The 10 models thus built were further tested for proper stereochemistry, based on MODELLER optimization scores and PROCHECK scores(Laskowski et al., 1993). This results in different starting orientations of the hypervariable loops, within the experimentally determined limits of fluctuations.

(d) ***Modeling of glycans and their ensemble sampling:*** For each of the 10 selected protein structures, glycans were initially added with random orientation, at the known glycosylation sites, based on ideal geometries as dictated by CHARMM36(Best et al., 2012, Huang and MacKerell Jr, 2013) force field internal coordinates, followed by a 1Å randomization added to the overall atomic coordinates as described by Guttman et. al.(Guttman et al., 2013). Once all the glycans were added, ensuing refinement steps of the glycoprotein system optimized an energy function given by a combination of template-based spatial restraints, CHARMM36 forcefield terms, and a soft sphere-like truncated Gaussian term to prevent collisions. The structures were relaxed with 1000 steps of conjugate gradient minimization followed by a short molecular dynamics equilibration of 500ps. Further refining with five rounds of simulated annealing was performed between 1,300K to 300K in 8 steps. The glycans, the PNGS asparagine residues (as default by ALLOSMOD glycan modeling) and the loop regions were kept unrestrained during the refinement steps to allow for the flexibility required to accommodate a physiologically correct glycan ensemble. 100 fully glycosylated structures were modeled from each of the 10 selected protein models, resulting in the final ensemble containing 1000 different poses. Template restraints were removed from the hypervariable regions with glycans in order for them to sample a wider range of conformations during the glycoprotein relaxation phase. These residues by HXB2 numbering are as follows: 143 to 152 (V1 loop), BG505-specific

insert residues 185A to 185I and 186 to 189 (V2 loop), 309 to 315 and 325 to 329 (V3 loop), 400 to 410 (V4 loop) and 458 to 464 (V5 loop) as determined from the LANL HIV database (https://www.hiv.lanl.gov/).

This improved method of data-driven modeling of the flexible protein regions significantly ameliorates the overall cross-correlation observed between simulated and experimental SPECM intensities at the glycan cores (main text **Figure 2E**) as compared to our previous study(Berndsen et al., 2020), such as those at N137, N262, N301, and N611.

Glycan root mean square fluctuations and volume calculation

For both the native and all-man9 glycosylated models, the root mean square fluctuations (RMSF) of each glycan (with index *n*) was calculated as an average over all its heavy atoms, by the following equation:

$$RMSF_n = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{M}\sum_{m=1}^{M}|\vec{r}_{mnk} - \langle\vec{r}_{nk}\rangle|^2}$$

where $\vec{r}_{mnk}$ is the atomic position of heavy atom k of glycan n in snapshot m, $\langle\vec{r}_{nk}\rangle = (\frac{1}{M})\sum_{m=1}^{M}\vec{r}_{mnk}$ is the average atomic position of heavy atom k in glycan n. K is the total number of heavy atoms in the glycan. It is 127 for mannose-9, and varies depending on the type of glycan. The ensemble for each model contains 1000 snapshots, making M = 1000 snapshots for each of the two models. The standard deviations were obtained by dividing the 1000 snapshots into 4 sets of M=250, and calculating the four sets of RMSF values for each of the three protomers **(SI Figure S9).**

The 1000 structures of the total ensemble are built from 10 initial starting protein conformations, as described above. The main difference between these 10 conformations are the variations in the loop regions, due to the missing residues in the templates. In order to remove the effects of the loop fluctuations and consider the RMSF coming from the glycans alone, the reduced RMSF values were also calculated in each of these 10 sub-models, and their average and s.d. calculated (**SI Figure S9B**). The RMSF difference between the models **(SI Figure S9E)** were

obtained by subtracting the reduced RMSF values of all-man9 from native model (native minus all-man9). Since, the average reduced RMSF value is ~4Å, only those RMSF difference values are statistically significant which are above 0.2 Å, which corresponds to a p-value of 0.05, rejecting those values within the null hypothesis. Ensemble of individual glycans were converted to atomic densities at 1 Å grid size in order to calculate the volume. Volume calculation was done using CHIMERA package(Pettersen et al., 2004).

Glycan overlap and network analysis

The inter-glycan overlap is calculated as the total fraction of heavy atoms from the two glycans that come within 5Å of each other. Let us consider the example of mannose-9 to illustrate the parameter of overlap. A single mannose-9 glycan has 127 heavy atoms. Since our ensemble is composed of 1000 possible structures, there are effectively 127,000 heavy atoms per ensemble of mannose-9 at one position. The fraction of the total number of heavy atoms from two neighboring ensembles that come within contact distance defines the overlap fraction. Since mannose-9 is the most commonly occurring glycoform in our system, we have used it as our reference for normalization of the overlap probability. An overlap greater than or equal to 50% of heavy atoms from two neighboring mannose-9 glycans is assigned as 1. This overlap matrix is used to define the adjacency matrix for our network analysis. Each glycan functions as a node of the graph (main text **Figure 4A inset)**, and two nodes are connected by an edge if there is at least 5% overlap as per our overlap definition given above. The edge length is inversely proportional to the overlap value, i.e., the larger the overlap, the closer two nodes (glycans) are in the graph. Only those glycans from the neighboring protomers are considered, that have an inter-protomer edge. All graph theory and network analyses were performed using Python(Rossum, 1995) and Matlab_R2018a packages(MATLAB, 2018).

Eigenvector centrality calculation:

For a given graph, G, with adjacency matrix $\mathbf{A}=(a_{v,t})$ where $a_{v,t}$ is the edge weight connecting nodes v and t ($a_{v,t}=0$ when there is no connection), the relative centrality score x, of node v can be defined as:

$$x_v = \frac{1}{\lambda} \sum_{t \in N(v)} x_t$$

where $N(v)$ is the set of neighbors directly connected to *v,* and λ is a constant. From the definition of the adjacency matrix where the elements go to zero if two nodes are not connected, the above equation can be expressed as:

$$x_v = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

This has the form of the eigenvector equation **Ax**= λ**x.** With the added constraint of the eigencentrality values needing to be non-negative, by the Perron-Frobenius theorem(Pillai et al., 2005), the eigenvector corresponding to the largest eigenvalue gives the desired measure of centrality. The eigenvector is a unit vector and therefore the centrality values add up to one. For the purpose of this work, we have normalized the centrality values with respect to the node with the highest centrality assigned at 1, to obtain the relative centrality values.

Modularity maximization for community detection

Community detection within the glycan network was performed using the modularity maximization approach given by Newmann and Girvan(Newman, 2006, Newman and Girvan, 2004). Modularity Q is calculated as the difference between the fraction of edges that fall within a module and the expected fraction if the edges were distributed in random.

$$Q = \sum_{i=1}^{k} (e_{ii} - a_i^2)$$

Where e_{ii} is the fraction of edges in module *i*; a_i is the fraction of edges with at least one end in module *i.* It is calculated by a greedy heuristic, beginning with the trivial system of each node being a cluster, and merging two clusters that will increase modularity by the largest value, stopping when any further merge would decrease the modularity. This approach is known to work well for small networks similar to our system. The calculations were implemented through a standard algorithm(Blondel et al., 2008) in Matlab R2018a.

Visualization of Structures: All visualizations of structures for figures were generated using the Visual Molecular Dynamics package(Humphrey et al., 1996) [RRID: SCR_001820].

Table S1: Different glycan species at BG505 SOSIP PNGS selected based on past site-specific mass spectroscopy data, related to Figure 1. Structures corresponding to each species is given in Figure 1A (main text).

Glycan ID	88	133	137	156	160	185e	185h	197	234	262	276	295	301	332
Glycan type	FA2	M9	FA2	M9	M8	FA2	FA2	FA2	M9	M9	M7	M9	M9	M9

Glycan ID	339	355	363	386	392	398	406	411	448	462	611	618	625	637
Glycan type	M9	FH	M9	M9	M9	FA2	FA2	M9	M9	FA3	FA3	FA3	FA2	FA2

Table S2: List of experimental BG505 SOSIP structures used for flexibility analysis, related to Figure 1. PDB accession IDs are given. BG505_293S and 293F are cryo-EM structures with high oligomannose and native-like glycosylated structures obtained in our previous work(Berndsen et al., 2020). 5CJX has three protomers having different loop orientations. Therefore, they were considered separately.

Structure	Resolution (Å)	Method
6MTJ	2.34	X-Ray
6MTN	2.5	X-Ray
6MU7	2.5	X-Ray
6MU6	2.55	X-Ray
6NNJ	2.6	X-Ray
6NM6	2.74	X-Ray
6NNF	2.76	X-Ray
5V7J	2.91	X-Ray
6MU8	2.99	X-Ray
5CEZ	3.03	X-Ray
5U7M	3.03	X-Ray
5U7O	3.03	X-Ray
4TVP	3.1	X-Ray
5FYL	3.1	X-Ray
BG505_293S	3.1	Cryo-EM
BG505_293F	3.1	Cryo-EM
6MPG	3.2	Cryo-EM
4ZMJ	3.31	X-Ray
6NFC	3.43	Cryo-EM
5T3Z	3.5	X-Ray
5UTF	3.5	X-Ray
6OT1	3.5	Cryo-EM
5CJX_1	3.58	X-Ray
5CJX_2	3.58	X-Ray
5CJX_3	3.58	X-Ray
6CDI	3.6	Cryo-EM
6NF2	3.7	Cryo-EM
6NF5	3.71	Cryo-EM
6CDE	3.8	Cryo-EM
6CH7	3.8	X-Ray
6MPH	3.8	Cryo-EM
6DFH	3.85	Cryo-EM
5T3X	3.9	X-Ray
6ORO	3.9	Cryo-EM

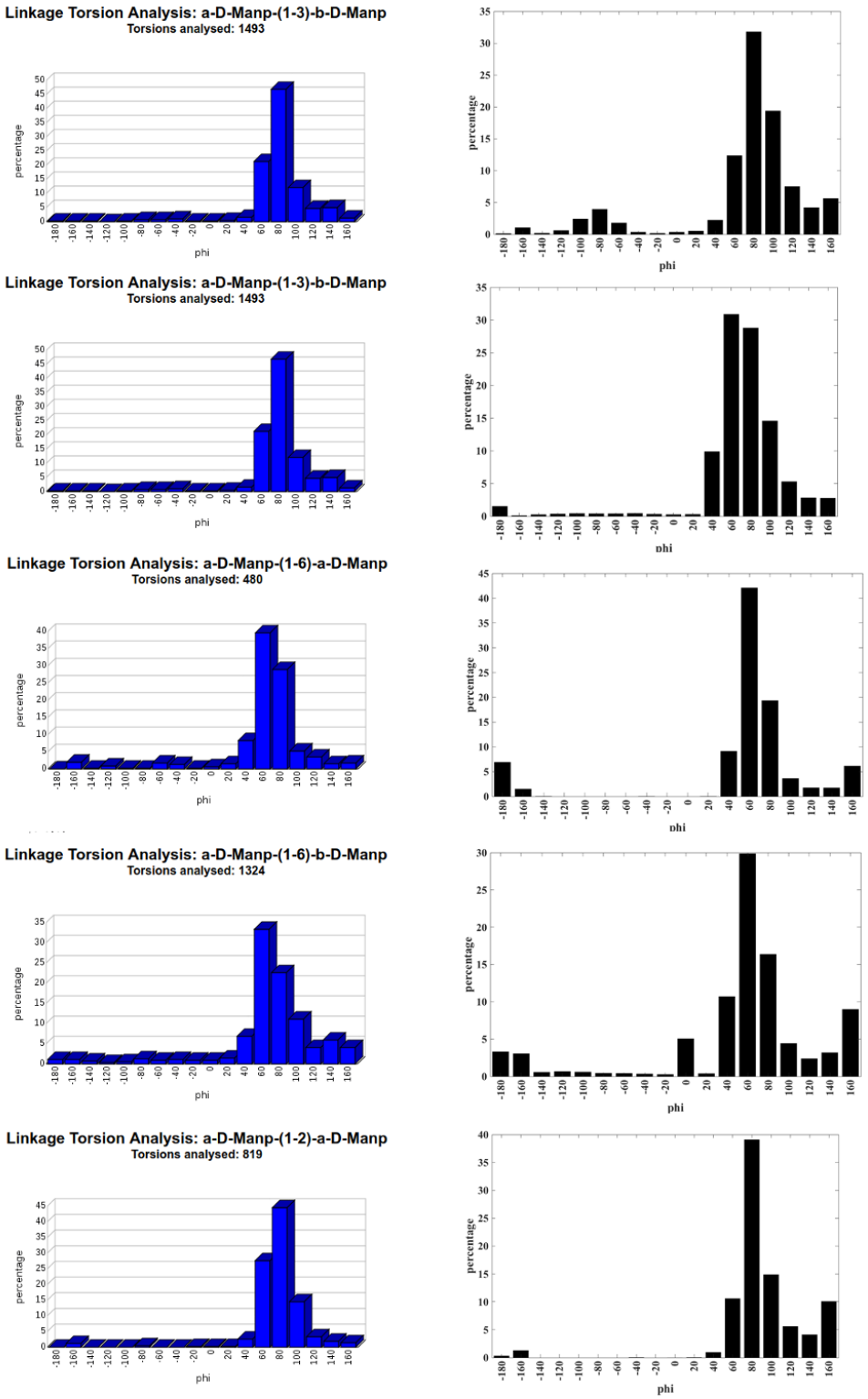


Figure S1: Phi (ϕ) dihedral distributions of different inter-glycan linkages involving mannose rings related to Figure 1. Comparison of the modeled ensemble with those obtained from different glycan structures available in the PDB is shown. Distributions on the left are from the PDB, as obtained by GlyTorsion. Distributions on the right are from the models. Number of torsions analyzed for each dihedral type from PDB is given.

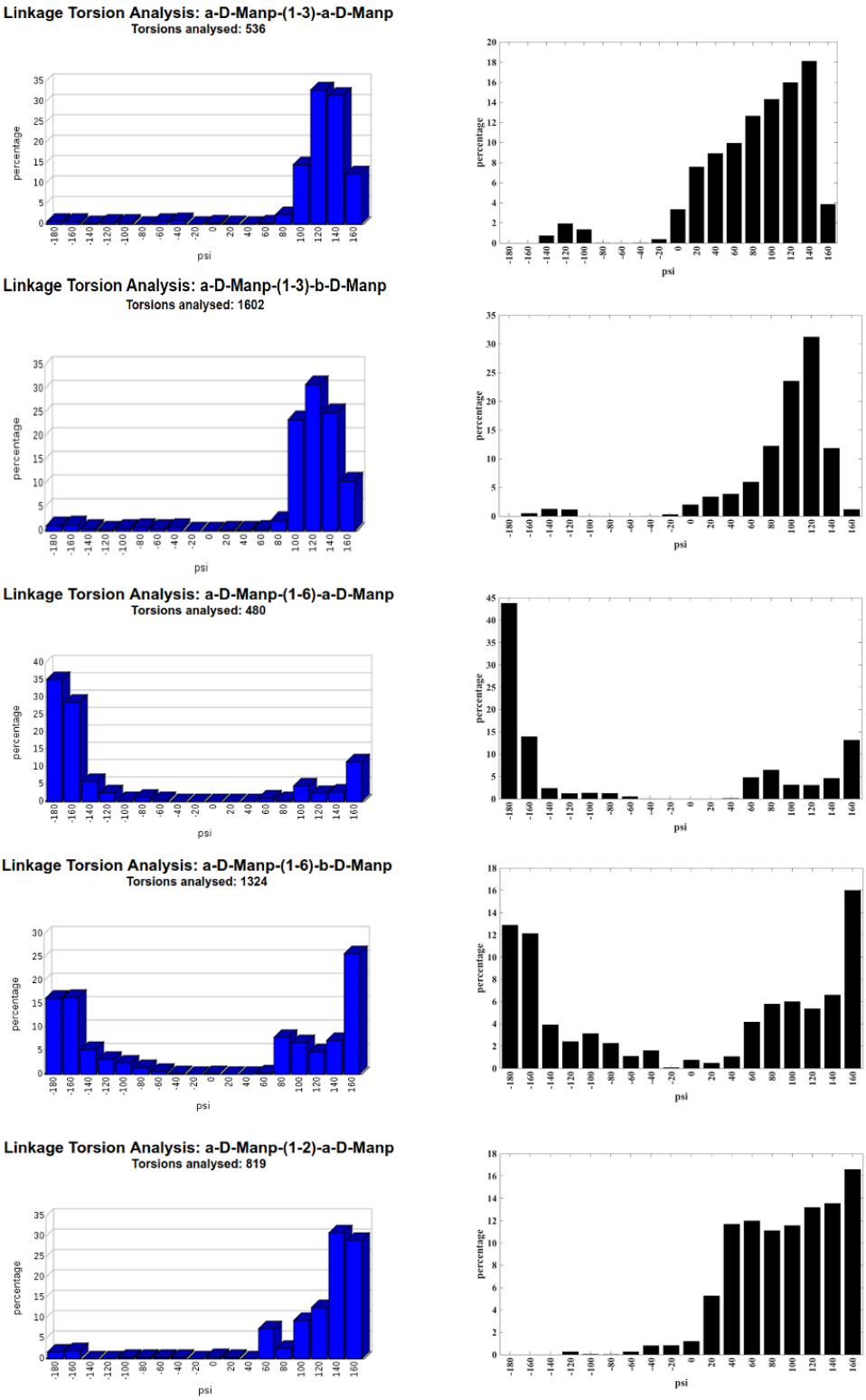


Figure S2: Psi (ψ) dihedral distributions of different inter-glycan linkages involving mannose rings related to Figure 1. Comparison of the modeled ensemble with those obtained from different glycan structures available in the PDB is shown. Distributions on the left are from the PDB, as obtained by GlyTorsion. Distributions on the right are from the models. Number of torsions analyzed for each dihedral type from PDB is given.

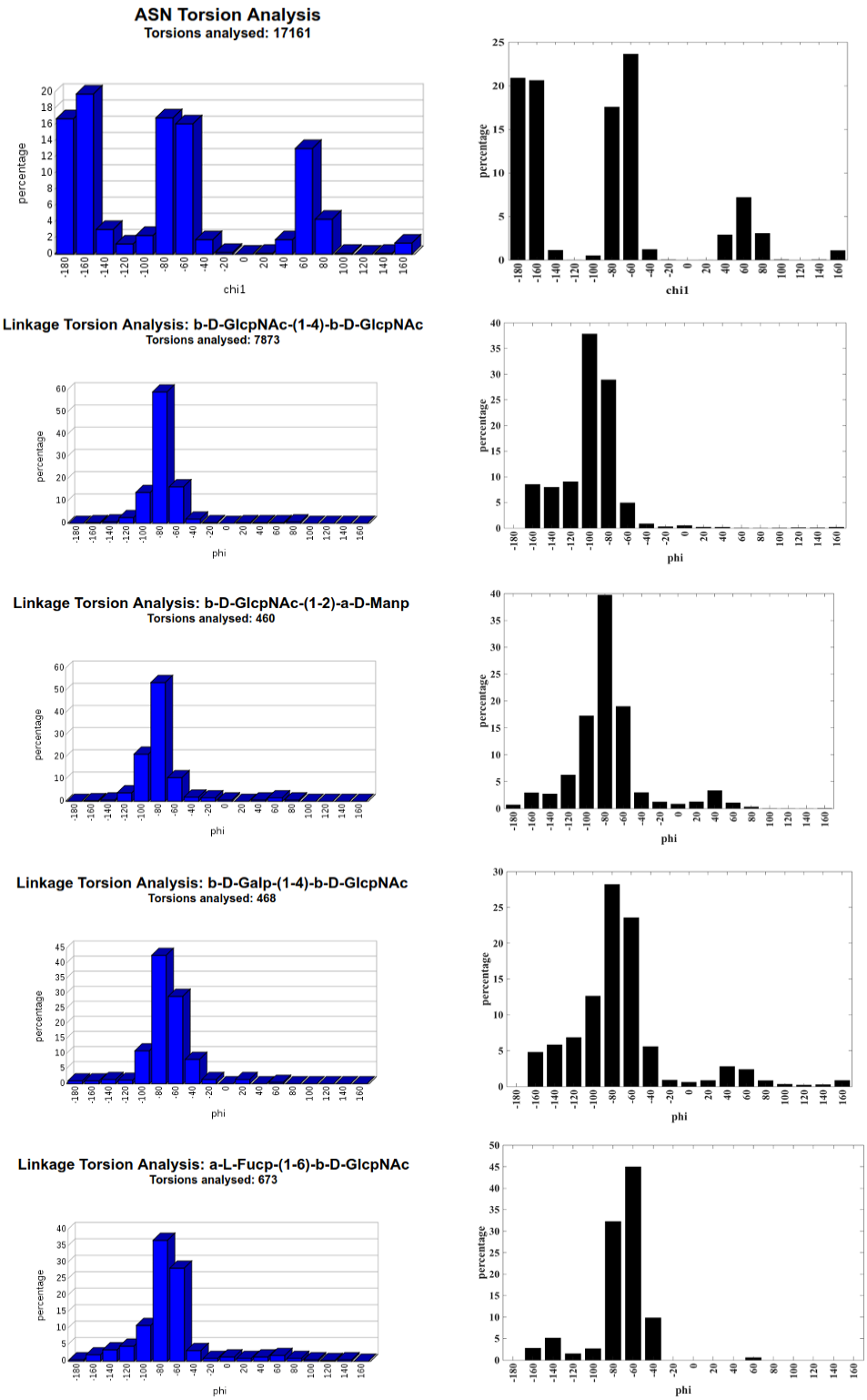


Figure S3: PNGS asparagine chi1 (top row), and non-mannose sugar phi (ϕ) dihedral distributions of different inter-glycan linkages related to Figure 1. Comparison of the modeled ensemble with those obtained from different glycan structures available in the PDB is shown. Distributions on the left are from the PDB, as obtained by GlyTorsion. Distributions on the right are from the models. Number of torsions analyzed for each dihedral type from PDB is given.

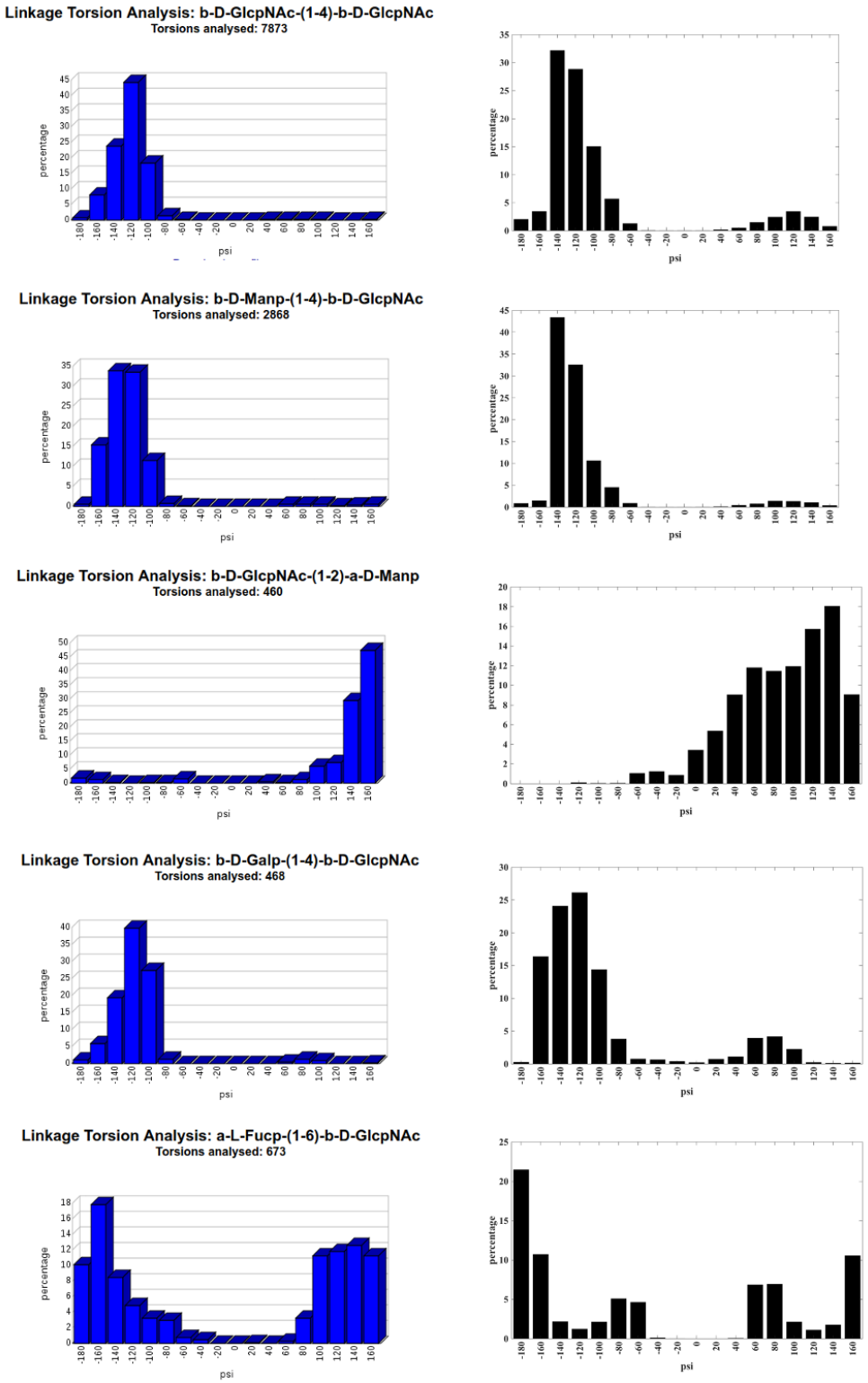


Figure S4: Psi (ψ) dihedral distributions of different inter-glycan linkages involving non-mannose sugars related to Figure 1. Comparison of the modeled ensemble with those obtained from different glycan structures available in the PDB is shown. Distributions on the left are from the PDB, as obtained by GlyTorsion. Distributions on the right are from the models. Number of torsions analyzed for each dihedral type from PDB is given.

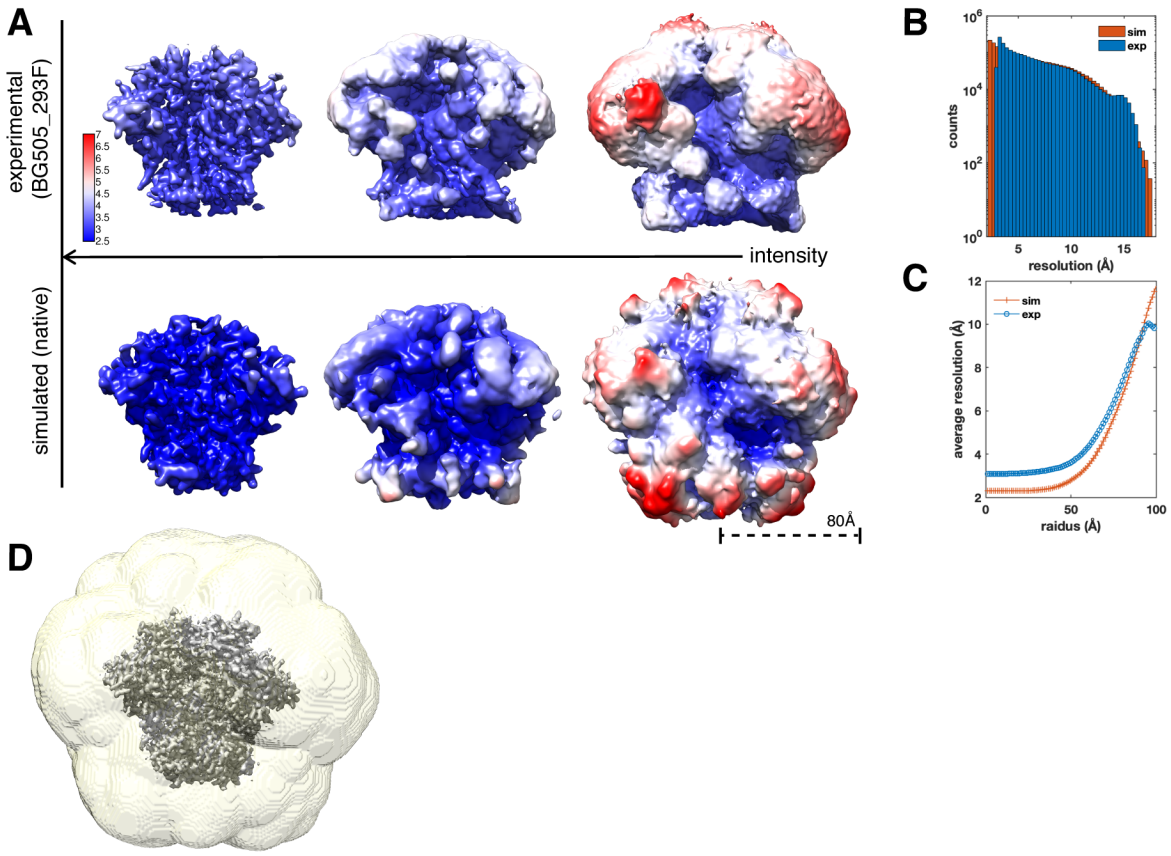
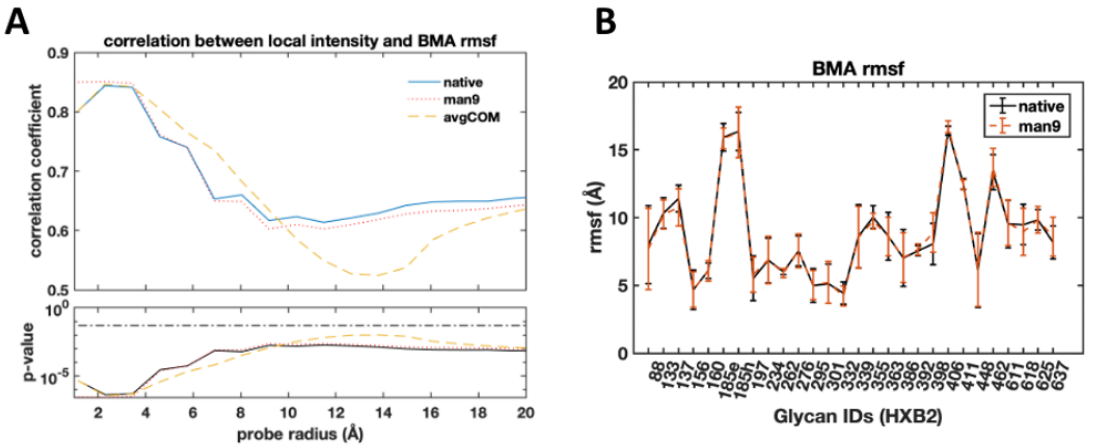


Figure S5 | Local resolution analysis of experimental and simulated cryo-EM maps.
(A) Gaussian filtered (σ = 1.5) experimental cryo-EM map of BG505 SOSIP.664 expressed in HEK293F cells and simulated cryo-EM map generated from the natively glycosylated ALLOSMOD ensemble shown at three isosurface thresholds and colored by local resolution estimates, related to Figure 2. Experimental data was previously published (Berndsen et al., 2020). The density corresponding to the base binding Fabs was removed via masking to make this figure. (B) Histograms of masked local resolution maps. (C) Average local resolution as a function of radius from the center of the map. (D) Mask used for local resolution maps.



C

	BG505_293F	BG505_293S	Sim_native	Sim_man9
BG505_293F	1 (0)	0.90 (1.60E-08)	0.82 (6.80E-06)	0.83 (2.30E-06)
BG505_293S	0.90 (1.60E-08)	1 (0)	0.70 (1.80E-04)	0.73 (2.30E-04)
Sim_native	0.82 (6.80E-06)	0.70 (1.80E-04)	1 (0)	0.95 (1.40E-10)
Sim_man9	0.83 (2.30E-06)	0.73 (2.30E-04)	0.95 (1.40E-10)	1 (0)

Figure S6: Glycan dynamics are effectively indistinguishable at the level of core BMA between native and all-man9 model, related to Figure 2. (A) Pearson correlation coefficient and p-value between mean local map intensity around BMA residues and inverse average rmsf of each BMA residue as a function of probe radius. Also plotted is the correlation when using the true C3 average BMA centers of mass (avgCOM). (B) Comparison of BMA root mean squared fluctuations between native and all-man9 models. (C) Table highlighting the correlation coefficients of local BMA intensities between experimental and simulated native and all-man9 maps, p-values given in parentheses. The correlation remains similarly high between all maps at the level of the glycan stem.

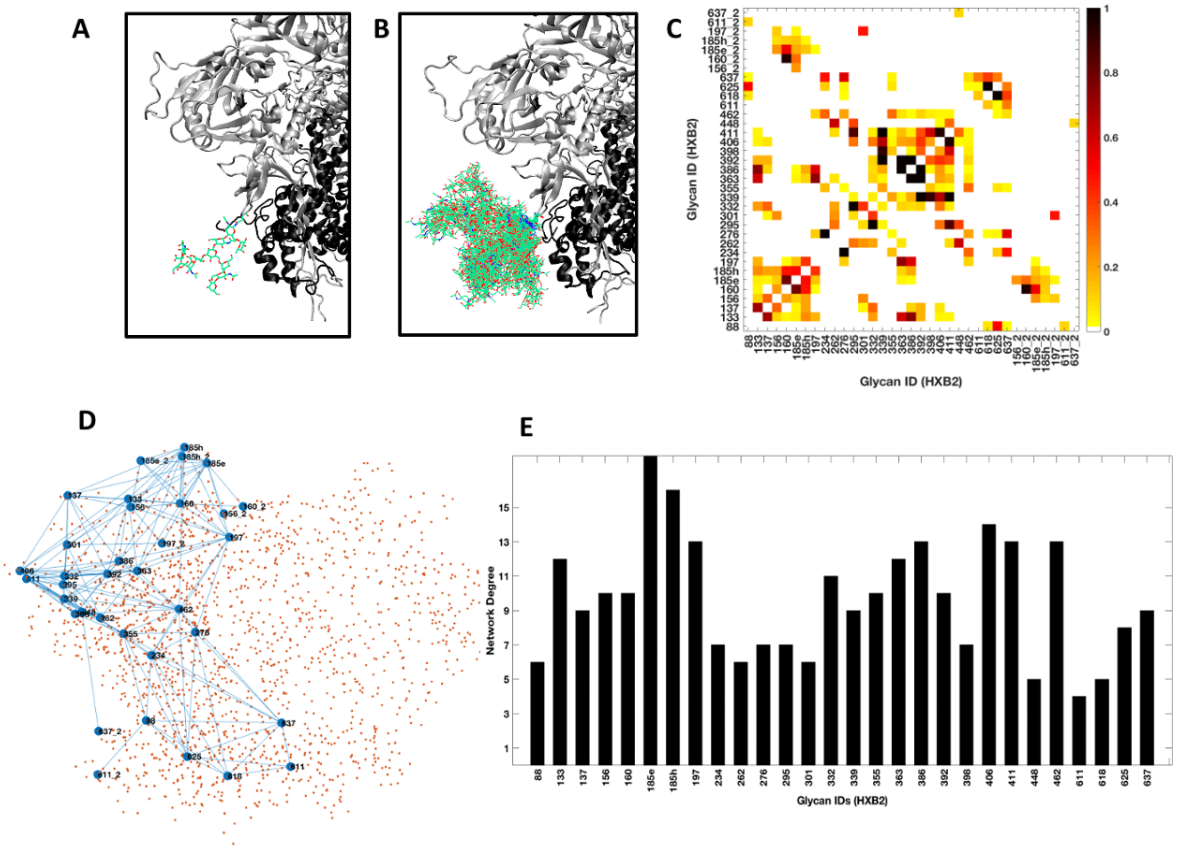


Figure S7: Topological native glycosylation shield network from glycan volume overlap, related to Figure 4. (A) Glycan N88 in one particular pose. (B) Spatial sampling by glycan N88. Each glycan can take a variety of different conformations and orientations, sampling a large volume in space. (C) Probability distribution map of inter-glycan fractional overlap. Neighboring protomer glycans are indicated by suffix '_2'. (D) Network projected on the Env structure in 2-dimensions. The orange dots indicate the projection of Env protein structure in 2D. Glycan nodes are given by circles, connected by edge lines, colored in blue. (E) The network degree of each node or glycan, given by the number of other nodes it is connected to.

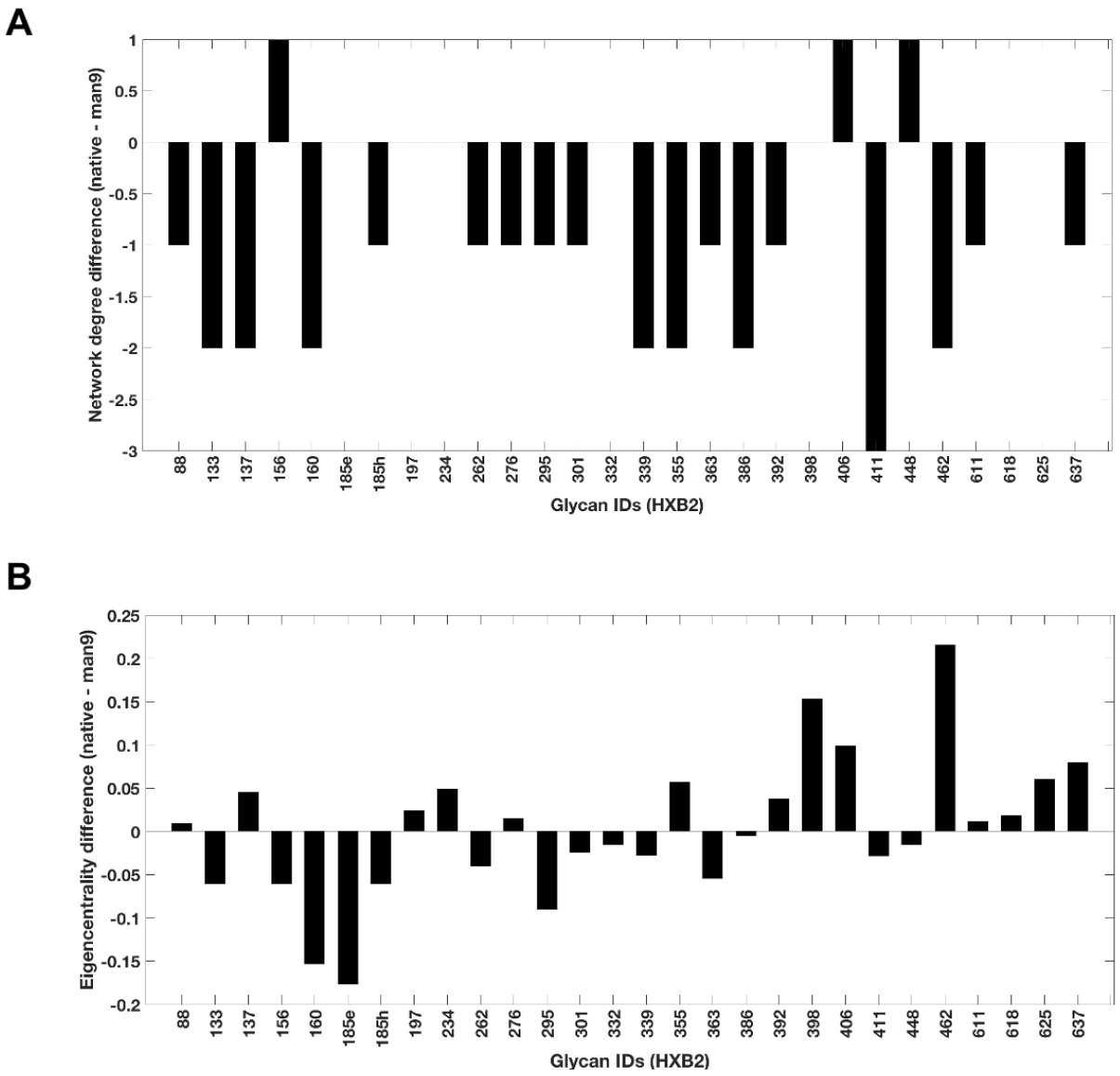


Figure S8:

Network difference between native and all-man9 glycosylation, related to Figure 5.

Differences are calculated as native minus all-man9. (A) Network degree difference. (B) Eigenvector centrality difference.

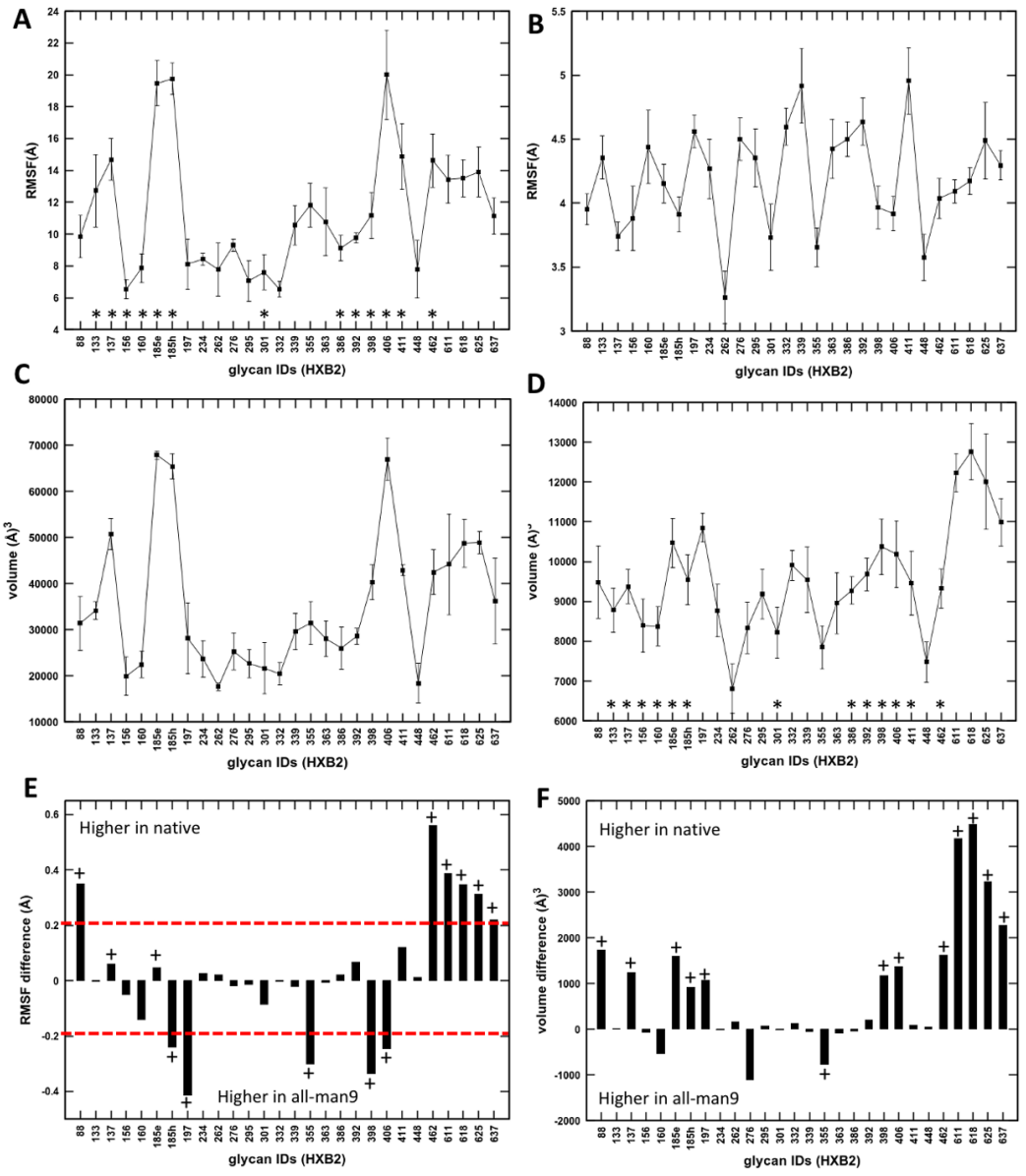


Figure S9: Structural fluctuations and volume sampling of individual glycans in BG505 native model. Glycans present in the hyper-variable loop regions of gp120 are indicated by *. Glycans modeled as fucosylated complex or hybrid glycoforms are indicated by +, related to Figure 5. (A) Site-specific Root Mean Squared fluctuations (RMSF). (B) Reduced RMSF per glycan, where underlying protein backbone are locally aligned to remove the contribution from protein fluctuations. (C) Sampled volume per glycan at each PNGS. (D) Reduced sampled volume per glycan, removing the contribution from protein fluctuations. (E) Difference in RMSF between native and all-man9 model. Native minus all-man9 values are plotted here. (F) Difference in sampled volume between native and all-man9 model. Native minus all-man9 values are plotted here.

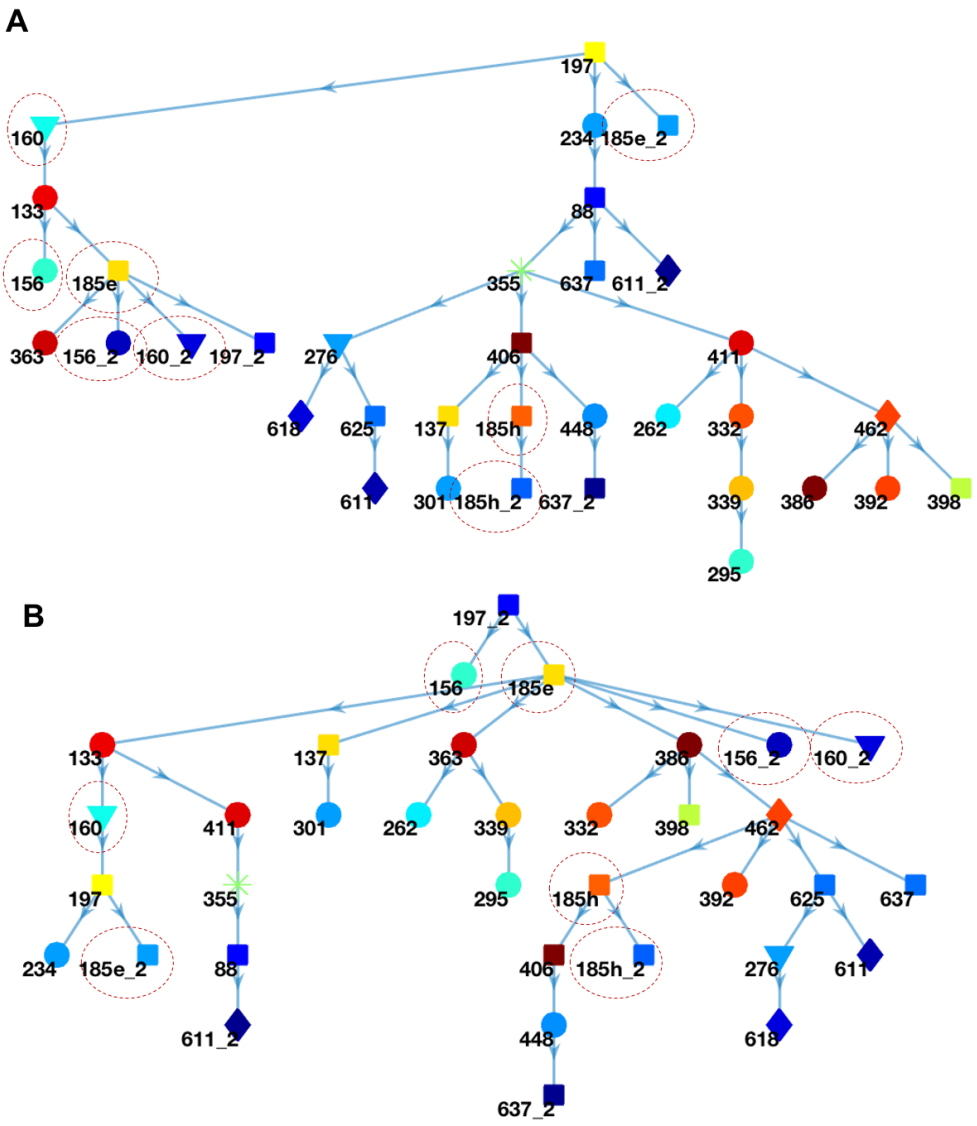


Figure S10:

Glycan 197 directly and indirectly affects glycans in the immediate vicinity of PG9 and PGT145 epitopes, related to Figure 7. (A) Shortest paths of communication from glycan N197 to all other glycans in the network. (B) Shortest path of communication from glycan N197 in the neighboring protomer. In both the graphs, those glycans that directly contribute to shielding of PG9 and PGT145 epitopes are circled by red dotted lines.

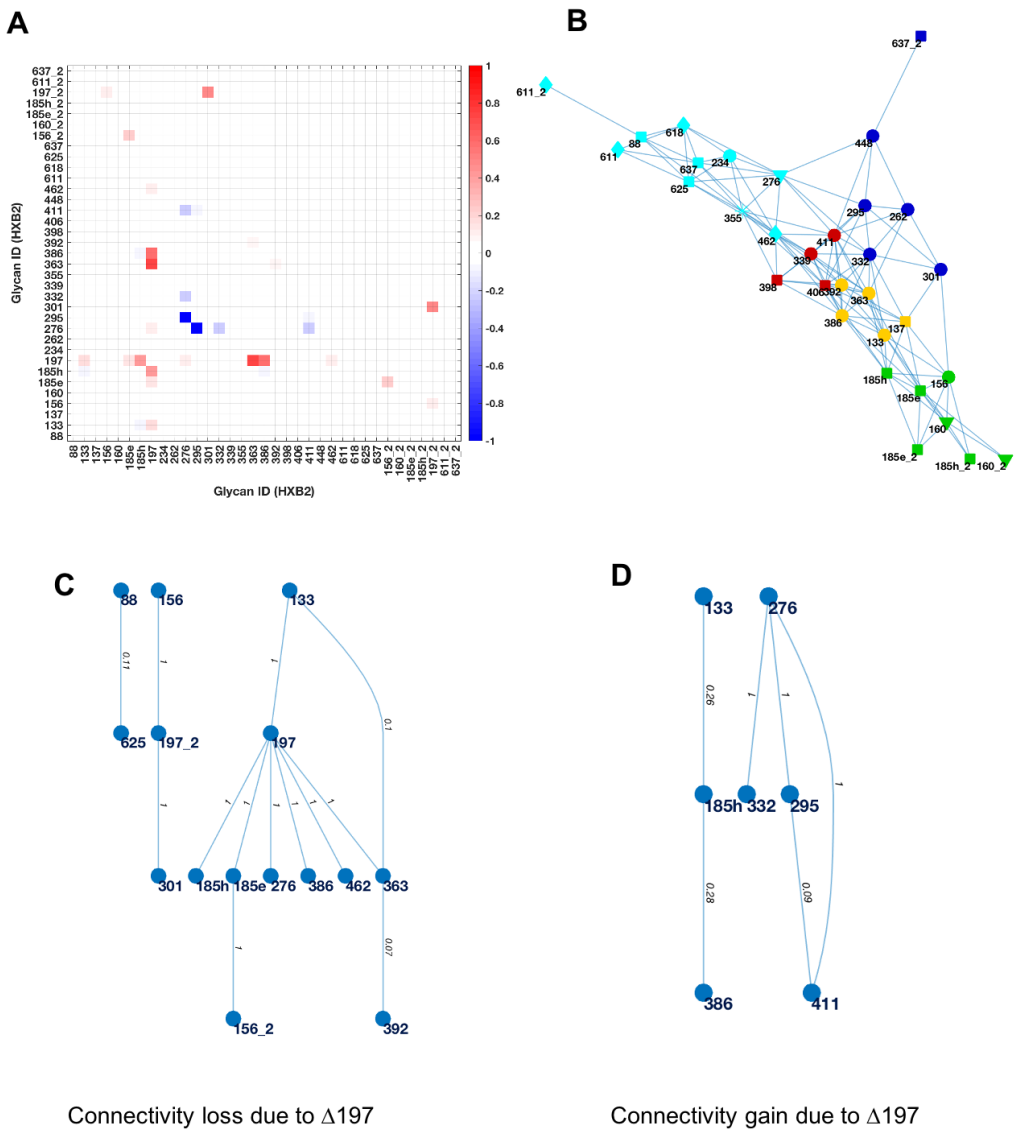


Figure S11:

Network difference between native glycosylation and deletion of N197 glycan, related to Figure 7. (A) Difference in adjacency matrices between the two models, - native minus del197. Red color indicates at least 5% decrease in edge weight, and blue indicates at least 5% increase in edge weight in Δ197 network, as compared to native glycosylation pattern. (B) Network communities in the Δ197 model. (C) Decrease and (D) increase in connectivity due to del197 model in comparison with native glycosylation.

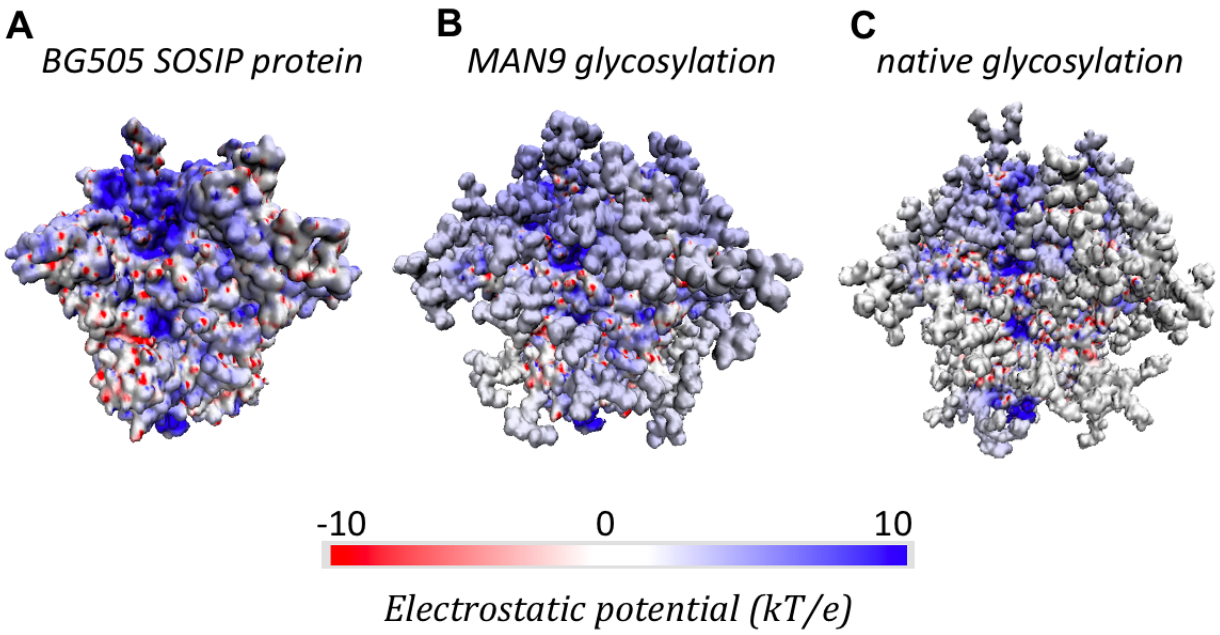


Figure S12:

Electrostatic potential surface of BG505 SOSIP. (A) Protein surface without glycans, (B) surface with MAN9 glycosylation, (C) surface with native glycosylation. All electrostatic potential calculations are shown on representative structures, and calculations were done using APBS(Baker et al., 2001). The electropositivity of the exposed surface goes down from unglycosylated to all-man9, becoming even more neural with native glycosylation, related to Discussion section and Figure 6.

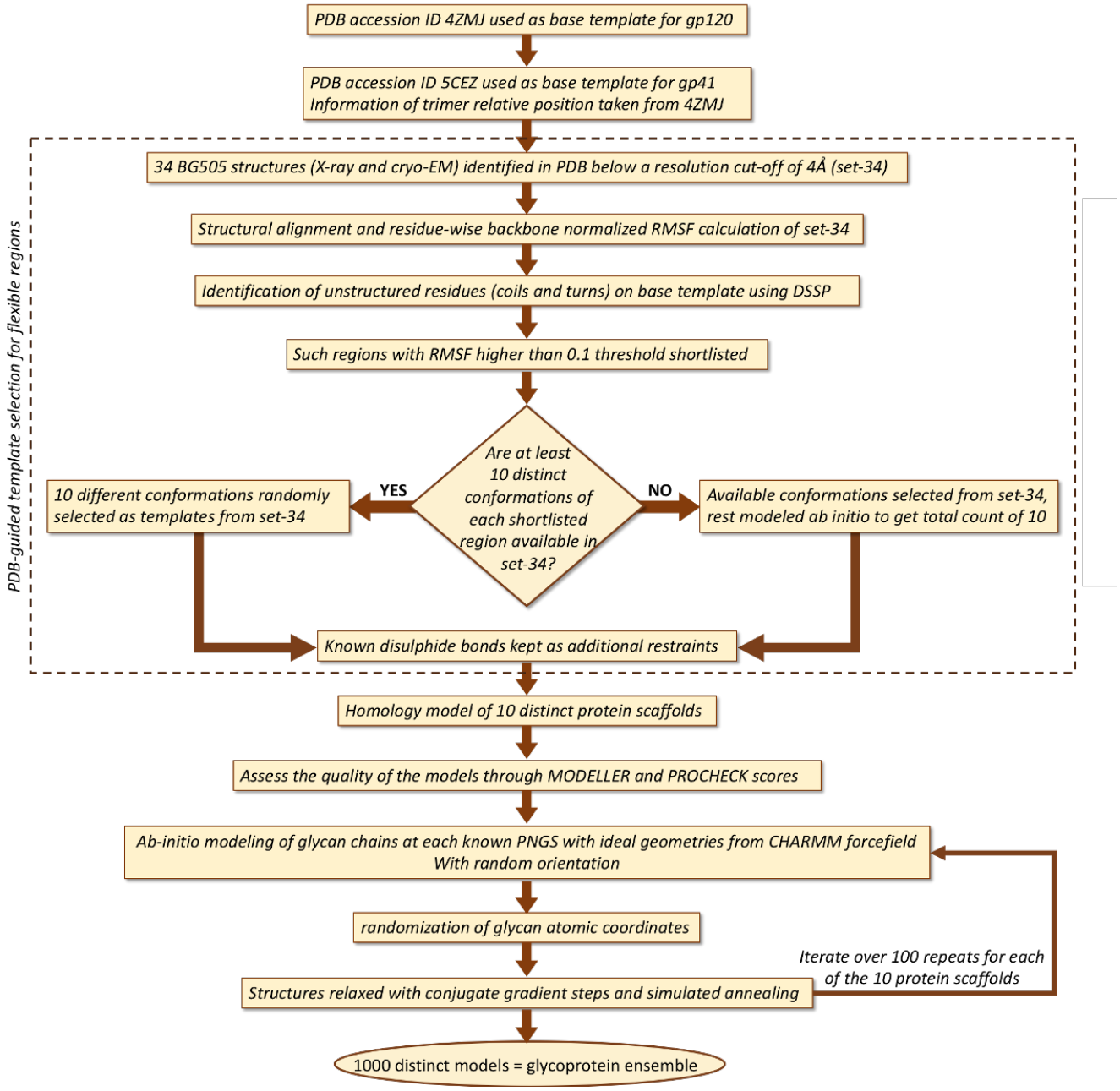


Figure S13:

Workflow describing modeling pipeline for BG505-SOSIP Env glycoprotein ensemble.
Related to Modeling Methods and Figures 1-2.

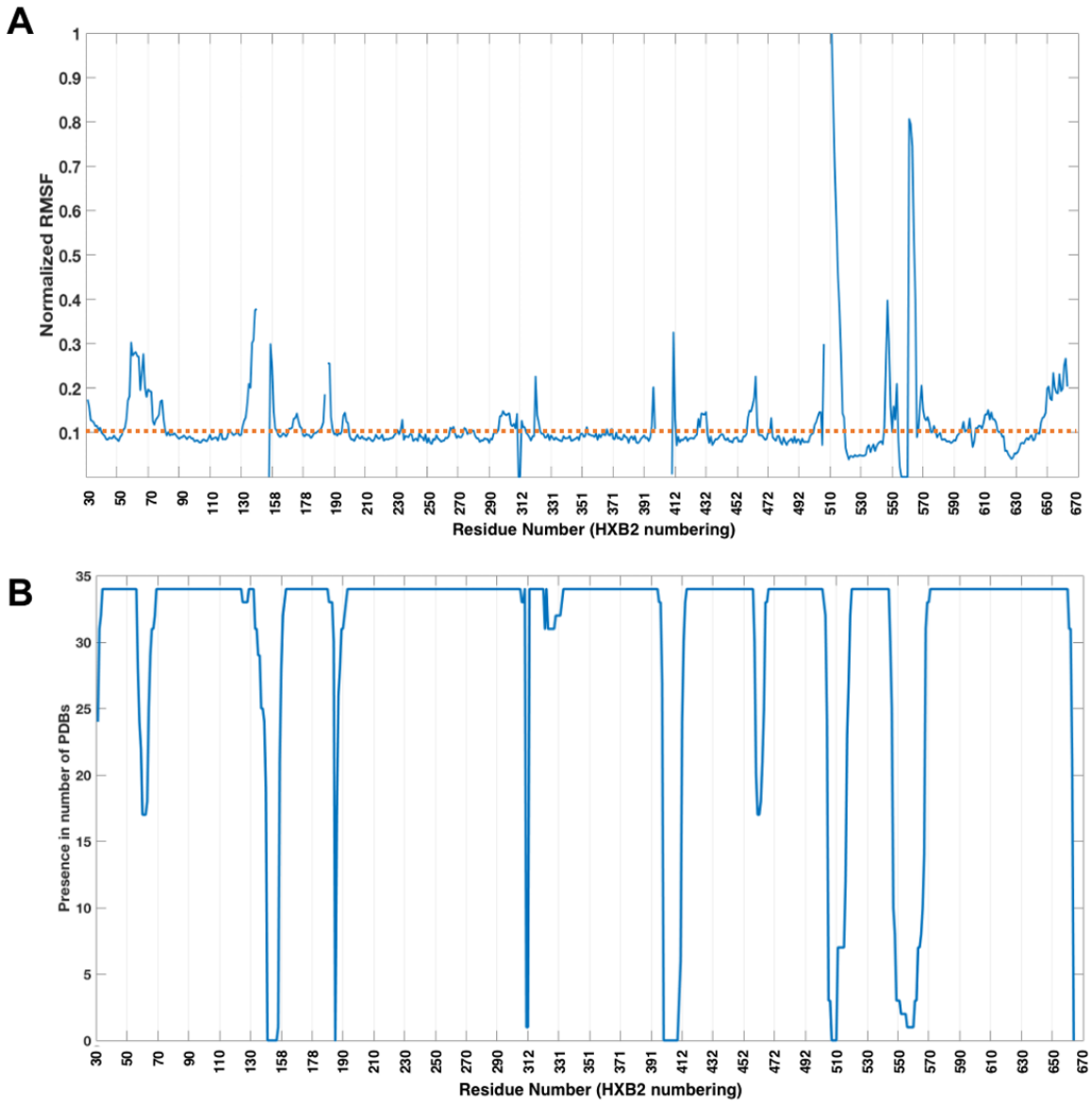


Figure S14:

Homology modeling of unstructured loop regions based on variations in available PDB structures. (A) Root Mean Squared Fluctuations in protein backbone position based on 34 different PDB structures. Those loop regions having fluctuations higher than the threshold line (dotted, red) were modeled based on 10 randomly selected available datapoints where residues are present. (B) Residue-wise presence in the 34 BG505 PDB candidates. Some of loop regions such as V2 and V4 have missing residues in all available structures. Related to Modeling Methods and Figures 1-2.

REFERENCES

BAKER, N. A., SEPT, D., JOSEPH, S., HOLST, M. J. & MCCAMMON, J. A. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A,* 98**,** 10037-41.

BERNDSEN, Z., CHAKRABORTY, S., WANG, X., COTTRELL, C., TORRES, J., DIEDRICH, J. K., LOPEZ, C. A., YATES III, J. R., VAN-GILLS, M. J., PAULSON, J. C., GNANAKARAN, S. & WARD, A. B. 2020. Visualization of the HIV-1 Env Glycan Shield Across Scales. *Proc Natl Acad Sci U S A***,** 202000260.

BEST, R. B., ZHU, X., SHIM, J., LOPES, P. E., MITTAL, J., FEIG, M. & MACKERELL JR, A. D. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of chemical theory and computation,* 8**,** 3257-3273.

BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. & LEFEBVRE, E. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment,* 2008**,** P10008.

ESWAR, N., WEBB, B., MARTI-RENOM, M. A., MADHUSUDHAN, M., ERAMIAN, D., SHEN, M. Y., PIEPER, U. & SALI, A. 2006. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics,* 15**,** 5.6. 1-5.6. 30.

FISER, A., DO, R. K. & SALI, A. 2000. Modeling of loops in protein structures. *Protein Sci,* 9**,** 1753-73.

FRISHMAN, D. & ARGOS, P. 1995. Knowledge-based protein secondary structure assignment. *Proteins-Structure Function and Genetics,* 23**,** 566-579.

GARCES, F., LEE, J. H., DE VAL, N., DE LA PENA, A. T., KONG, L., PUCHADES, C., HUA, Y., STANFIELD, R. L., BURTON, D. R., MOORE, J. P., SANDERS, R. W., WARD, A. B. & WILSON, I. A. 2015. Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. *Immunity,* 43**,** 1053-63.

GUTTMAN, M., WEINKAM, P., SALI, A. & LEE, K. K. 2013. All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. *Structure,* 21**,** 321-31.

HUANG, J. & MACKERELL JR, A. D. 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry,* 34**,** 2135-2145.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling,* 14**,** 33-38.

KWON, Y. D., PANCERA, M., ACHARYA, P., GEORGIEV, I. S., CROOKS, E. T., GORMAN, J., JOYCE, M. G., GUTTMAN, M., MA, X., NARPALA, S., SOTO, C., TERRY, D. S., YANG, Y., ZHOU, T., AHLSEN, G., BAILER, R. T., CHAMBERS, M., CHUANG, G. Y., DORIA-ROSE, N. A., DRUZ, A., HALLEN, M. A., HARNED, A., KIRYS, T., LOUDER, M. K., O'DELL, S., OFEK, G., OSAWA, K., PRABHAKARAN, M., SASTRY, M., STEWART-JONES, G. B., STUCKEY, J., THOMAS, P. V., TITTLEY, T., WILLIAMS, C., ZHANG, B., ZHAO, H., ZHOU, Z., DONALD, B. R.,

LEE, L. K., ZOLLA-PAZNER, S., BAXA, U., SCHON, A., FREIRE, E., SHAPIRO, L., LEE, K. K., ARTHOS, J., MUNRO, J. B., BLANCHARD, S. C., MOTHES, W., BINLEY, J. M., MCDERMOTT, A. B., MASCOLA, J. R. & KWONG, P. D. 2015. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat Struct Mol Biol,* 22**,** 522-31.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography,* 26**,** 283-291.

LÜTTEKE, T., FRANK, M. & VON DER LIETH, C.-W. 2005. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic acids research,* 33**,** D242-D246.

MATLAB 2018. MATLAB, Version R2018a. The MathWorks Inc Natick, MA.

NEWMAN, M. E. 2006. Modularity and community structure in networks. *Proceedings of the national academy of sciences,* 103**,** 8577-8582.

NEWMAN, M. E. & GIRVAN, M. 2004. Finding and evaluating community structure in networks. *Physical review E,* 69**,** 026113.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry,* 25**,** 1605-1612.

PILLAI, S. U., SUEL, T. & CHA, S. 2005. The Perron-Frobenius theorem: some of its applications. *IEEE Signal Processing Magazine,* 22**,** 62-75.

ROSSUM, G. 1995. Python reference manual.

SALI, A. 1995. Comparative protein modeling by satisfaction of spatial restraints. *Mol Med Today,* 1**,** 270-7.

SCHRÖDINGER, L. L. C. 2015. The PyMOL molecular graphics system, Version 1.8. *Technical Report*.

WARD, A. B. & WILSON, I. A. 2017. The HIV-1 envelope glycoprotein structure: Nailing down a moving target. *Immunological reviews,* 275**,** 21-32.

WEINKAM, P., PONS, J. & SALI, A. 2012. Structure-based model of allostery predicts coupling between distant sites. *Proc Natl Acad Sci U S A,* 109**,** 4875-80.