# A Network-based approach for Quantifying the Resilience and Vulnerability of HIV-1 Native Glycan Shield

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

[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 Contacts: andrew@scripps.edu, gnana@lanl.gov

## Summary

The dense arrangement of N-glycans masking antigenic surfaces on the HIV-1 envelope (Env) protein acts as a shield from the adaptive immune system. The molecular complexity of glycan modifications and their inherent dynamic heterogeneity on a protein surface make experimental studies of glycoprotein structures a challenge. Here we have integrated a high-throughput atomistic modeling with graph-theory based method to capture the native glycan shield topological network and identify concerted behavior of these glycans. This is the first time that a complete computational model of an HIV-1 Env trimeric SOSIP structure has been generated with a native glycosylation pattern including both oligomannose and complex glycans, thus obtaining results which are immunologically more relevant. Important global and local feature differences due to the native-like glycosylation pattern have been identified, that stem from the charged sialic acid tips, fucose rings at the base, and different branching patterns of the complex glycans. Analyses of network attributes have aided in detailed description of the shield in a biological context. We have also derived a measure to quantify the shielding effect based on the number of glycan heavy atoms encountered over the antigenic protein surface that can define regions of relative vulnerability and resilience on the shield, and can be harnessed for potential immunogen design.

38 ## Introduction

39 Protein glycosylation is an essential aspect of post-translational modification, with 50-70%
40 of human proteins been estimated to be glycosylated to some degree[1]. These glycans play
41 significant roles in numerous biological processes, such as cellular signaling, recognition and
42 adhesion, protein folding, structural stability, and immune system interactions[2-4]. A detailed
43 comprehension of the three-dimensional structure and dynamics of these glycan moieties is
44 essential for a thorough understanding of the molecular basis of such functions. One of the most
45 frequently occurring protein glycosylation type found in nature is where the glycan moieties are
46 covalently attached to protein at asparagine (N) residues by an N-glycosidic bond. Investigation
47 of protein databases suggest that ~70% of proteins carry one or more motifs required for potential
48 N-glycosylation sites (PNGS, given by the sequon Asn-X-[Ser or Thr], where X is not proline)[5].
49 Such glycans usually have about 10-20 pyranose rings covalently connected in a tree-like structure
50 to the asparagine residue of the sequon. An important function of N-glycosylated proteins is their
51 role in viral and other pathogen-host relationships[6-9], making them a major focus for biomedical
52 research efforts.

53 Envelope proteins from several high-risk viral pathogens, such as HIV (lentivirus)[10],
54 Lassa (arenavirus)[11], Hepatitis C (flavivirus)[12], Epstein Barr (herpesvirus)[13], Ebola
55 (filovirus)[14, 15], and Influenza[8] are heavily glycosylated, where the host protein production
56 and glycosylation machineries are hijacked by these viruses. These surface-expressed viral
57 proteins are important immunological targets for neutralizing antibodies that can block viral
58 infection of cells, and form the primary focus of vaccine studies[16, 17]. However, the dynamic
59 and dense glycan coating can effectively act as a shield for the underlying envelope protein,
60 masking antigenic surfaces, barricading it from the adaptive immune system, and defending
61 against immune surveillance[18-23].

62 A deeper molecular level understanding of the glycan shields in these pathogenic viruses
63 may help inform vaccine design strategies that can overcome this protective barrier. However,
64 glycoproteins as a class remain recalcitrant to structural studies. The extreme dynamic
65 heterogeneity and conformational complexity stemming from large variance of possible
66 constituent sugars, linkage, branching patterns, and rotamer flexibility, make the study of these

glycans immensely challenging[24, 25]. High internal flexibility within each glycosidic bond and the compositional heterogeneity of the glycans prevent proper crystallization of these glycoproteins, rendering traditional X-ray crystallography methods ineffectual[26]. The inherent flexibility within glycan chains also makes structural determination by Nuclear Magnetic Resonance (NMR) and cryo-Electron Microscopy (cryo-EM) techniques challenging and incomplete[26-29]. Considering the case of HIV-1 alone, currently there are more than 300 experimental structures of envelope glycoprotein in the Protein Databank[30]. However, at best, only a quarter of the total envelope glycan content has been structurally resolved. Most of these were coordinated by antibodies and stabilized away from the true dynamic picture of the native trimer, providing little information about the structural and dynamic details of the surface glycans.

Computational modeling and molecular dynamics (MD) simulations have been particularly useful as a complementary approach towards the characterization of such systems[31-34], and have been harnessed to study the conformational dynamics of the HIV Env glycan shield[18, 35-37]. However, the main drawback affecting the quality of MD simulations is the robustness of conformational sampling. Due 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[36]. Such simulations for a system as large as the HIV-1, Ebola and other viral glycoproteins requires extensive computational power and time. We have previously established a high throughput pipeline to robustly build atomistic models of glycans, sample the glycan conformational space, construct the glycan network topology, and extract a molecular level description of the viral glycan shield[38]. In that study, we had also extensively validated the modeled ensembles against experimental cryo-EM data. This method drastically improves conventional sampling time, and yet retains the necessary accuracy to be physiologically relevant.

Here, we have applied our network modeling methodology on the HIV-1 Envelope Glycoprotein (Env) with a native glycosylation pattern. The 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 PNGS[39], and extremely dense glycosylation patterns

97 making up to approximately half the mass of the entire Env molecule[40]. As a result, there has
98 been only limited success in eliciting broadly neutralizing antibodies (bNAbs) to Env vaccine
99 immunogens[41-43]. These sugar moieties are highly dynamic[29], and have a median of 93 N-
100 linked glycans present per HIV-1 trimer[18]. Structures of bNAbs in complex with Env indicate
101 that these antibodies need to extend through the glycan shield in order to engage with the epitopes
102 at the protein surface[44-47]. Some bNAbs have evolved to include conserved glycans as part of
103 the epitope[48-51]. Moreover, these surface glycans are also critical for Env folding, viral
104 assembly, and infectivity. Given the overall importance of the Env glycan shield, understanding
105 its molecular architecture and dynamics as a whole is critical for Env-based vaccine design.

106 The glycosylated Env proteins further undergo additional processing in terms of α-
107 mannosidase and glycosyl transferases; the latter modifies the oligomannose into complex
108 sugars[51]. The level of processing of a particular glycan varies depending on its spatial location,
109 local protein content, local crowding by other glycans, and the producer cell type. A high degree
110 of processing is an indication of exposure or accessibility of sugars to enzymes. At regions with
111 dense crowding of glycans, steric constraints limit the activities of the carbohydrate processing
112 enzymes[52, 53] as the proteins fold and translocate through endoplasmic reticulum and Golgi
113 apparatus. This results in glycoform heterogeneity, and leads to additional problems for structural
114 studies, particularly X-ray crystallography. Thus, most structural studies have so far been
115 performed either by expressing all the glycans in the high-mannose form, or by removing majority
116 of the glycans. Recently there have been a few cryo-EM and X-ray structure determinations of
117 natively glycosylated Env[54-56], however in these structures most of the glycans are not visible
118 beyond the core glycan stem, unless stabilized by interactions with bound antibodies. Similarly,
119 while a number of structural studies have been performed on the HIV-1 Env glycosylation using
120 computational methods, all of these comprise of oligomannose glycan moieties ranging from
121 mannose-5 to mannose-9[18, 35-37, 57]. Here, we have modeled the natively glycosylated Env,
122 having both oligomannose and complex glycans[7, 58]. To our knowledge, this is the first time
123 that a complete computational model of Env SOSIP was generated having native glycosylation
124 pattern including processed glycoforms, thus obtaining results which are physiologically and
125 immunologically more relevant.

126 Using the natively glycosylated Env trimer of BG505 SOSIP, we have employed graph
127 theory to capture the glycan shield topological network, pinpoint potential interaction pathways,
128 and identify concerted behavior of the glycans. The glycan-glycan and glycan-protein interactions
129 influence the behavior of the shield as a whole and can affect distant sites in the glycan network[7,
130 18, 35]. Such effects can be captured by this graph-based approach. Analyses of various network
131 attributes, such as relative centrality of different glycan positions and critical subnetwork
132 properties, have aided in detailed examination of the native glycan shield in the context of bNAb
133 interactions. Important global and local feature differences due to native-like glycosylation pattern
134 come to light as compared to all high-oligomannose glycans. We have also been successful in
135 quantifying the glycan shield based on the number of sugar heavy atoms encountered over the
136 antigen surface that can be implemented to define regions of relative vulnerability and regions
137 where the glycan shield blocks access. This information can be incorporated into immunogen
138 design strategies. Due to the rapid yield rate of this method, it can be carried out for a large number
139 of diverse HIV-1 sequences, or can be seamlessly extended to model glycan shields of other
140 viruses.

141

142 **Results**

143 **Selection of site-specific glycans for native glycosylation of HIV-1 Env**

144 The soluble, recombinant BG505 SOSIP.664 (here on referred simply as BG505) trimer
145 has been well characterized as a native-like, Env-mimetic model, and serves as the prototypical
146 immunogen in several vaccine development programs[59, 60]. For our study, we have used this
147 system in the pre-fusion closed conformation, for which several structures have been
148 determined[18, 54-56]. A single protomer of this BG505 protein contains 28 N-linked
149 glycosylation sites. While all N-glycans share a common core sugar sequence, they are broadly
150 classified into three types (**Figure 1A**): (i) oligomannose, in which only mannose residues are
151 attached to the core; (ii) complex, in which differently branching "antennae" are attached to the
152 core; and (iii) hybrid, which is a mix of the first two types. Recent novel mass spectrometry
153 (MS)-based approaches have been successful in identifying site-specific glycosylation profile

154 in multiple HIV-1 trimers[58, 61]. It is known that each glycosylation site at BG505 has a
155 distribution of multiple possible glycan species, depending on various factors, such as type of
156 cell lines, purification methods and sources, structural constraints, and so on. We used a site-
157 specific distribution of glycan type at each PNGS that was obtained by Behrens *et. al*[7] and
158 Cao *et. al.*[58], to identify the most likely glycan species at each site. The glycan chosen at each
159 site is given in **Supplementary Table S1**. Here we selected the particular species with the
160 highest relative abundance per the MS studies (Figure 1 in reference 7). The structure schematic
161 for each of the selected glycan species have been shown in **Figure 1B**.

162 **Atomistic modeling of the native glycan shield**

163 Due to the highly dense glycosylation on Env surface, an approaching antibody
164 generally requires to navigate through the barrier of sugars to form interactions with the
165 underlying protein epitope. In the static pictures as obtained from fully glycosylated
166 structures[18, 54] and even single snapshots from computational models, a significant fraction
167 of the protein surface appears to be exposed, with each glycan taking up a particular
168 conformation (**Figure 2A**). However, due to their flexible and dynamic nature, these glycans
169 are not confined to a particular conformation, instead sampling a large volume in space. A
170 previous simulation study suggested[62] that the root mean squared deviations (RMSD) for the
171 carbohydrate regions are more than 4 times larger than that of the underlying protein loops. As
172 a result of this conformational variability, the cumulative effect of the glycans over time is like
173 that of a cloud of glycan atoms that shield the underlying surface from any approaching protein
174 probe (an antibody for example) as illustrated in **Figure 2B**.

175 We have modeled this entire glycan shield with its extensive conformational sampling,
176 over the Env protein. Details of the modeling protocol has been established previously[38] and is
177 reiterated briefly here. A robust ensemble of BG505 glycoprotein 3D conformations in atomistic
178 detail was generated by utilizing a template-free glycan modeling pipeline including a sequence
179 of refinement steps with restraints to enforce proper stereochemistry. The underlying protein
180 scaffold was built by homology modeling from available structures. The glycans were modeled *ab*
181 *initio* by implementing the ALLOSMOD[63, 64] package of MODELLER[65, 66] in a streamlined
182 pipeline. **Figure 2A** shows a single such pose of the BG505 glycoprotein. Due to the initial

183 randomization of the glycan orientations, this integrated technique can sufficiently sample a
184 physiologically relevant conformational space accessible to carbohydrates, in a very short time.
185 The complete ensemble has been depicted in **Figure 2B**, which shows that the extensive landscape
186 sampling of the all glycans leads to the overall spatial shielding effect.

187 The individual glycans modeled by this method covers a biologically relevant landscape.
188 The PNGS asparagine chi1 dihedral, as well as the phi and psi dihedral distributions of 9 different
189 inter-glycan linkages within the ensemble was compared with those obtained from different glycan
190 structures available in the PDB database, as described in **Supplementary Figure S1**. The torsion
191 angle distributions from the PDB were obtained using GlyTorsion webserver[67]. These
192 distributions match very well between our generated ensemble and the PDB structures, as
193 demonstrated in details in the Supplementary Information. We had previously built a similar
194 ensemble with mannose-9 glycans at all glycosylation sites[38], and had validated this method for
195 ensemble generation by quantitatively comparing the mannose-9 models to cryo-EM data from
196 oligomannose-predominant BG505.SOSIP.664 Env structures utilizing progressive low-pass
197 filtering[38]. That study confirmed that our methodology for characterizing the structural and
198 dynamical properties of the glycan shield accurately captures the properties of the physiological
199 ensemble. In this current study, we used the same ensemble (hereon referred as all-man9 model)
200 for comparison to understand the effects of native glycosylation.

201 **Influence of complex glycans on glycan shield**

202 Glycan dynamics, type, and inter-glycan interactions determine the local shielding effect
203 over the Env protein. Here, we looked at the dynamics of individual glycans and how they change
204 due to native-like glycosylation, ultimately governing the glycan shield properties. Each glycan
205 samples a large region in space as shown in **Figures 2C** and **2D**. These fluctuations of course
206 become much more extensive, when the glycans are present on the variable loops, due to the
207 dynamic movement of the protein backbone of the loops at these regions. The root mean squared
208 fluctuations (RMSF) measure and the sampled volume for those glycans located on the loops are
209 generally much higher, as seen in **Figure 3A** and **Supplementary Figure S2A**. However, aligning
210 the protein backbone and considering the reduced RMSF contribution coming only from the
211 glycans (see Methods), we see that the fluctuations between different glycans are comparable, with

212 a maximum difference of ~1.5Å (**Supplementary Figure S2B**). As for the glycan specific sampled
213 volume, after removing the flexibilities brought on by protein loop motions, the glycans on and
214 flanking loop V2 and those in gp41 cover the largest regions in space (**Figure 3B**).

215 Considering the differences between native-like glycosylation and the all-man9 model,
216 some interesting consistent patterns emerge. Comparing the reduced RMSF between native and
217 all-man9 glycosylation in **Figure 3C**, we see that while glycans 88, 355, 398 and 406 have
218 decreased fluctuations in the native model, those of glycans 234, 462, 611, 618, 625 and 637 have
219 increased significantly. With the exception of 234, all of these glycans are complex sugars. The
220 presence of the charged sialic acid at the tips of almost all of these complex glycans can dictate
221 the interactions between neighboring glycans, increasing their structural variations if surrounded
222 by other charged glycans, or reducing them if a stable conformation buried between uncharged
223 high-mannose patches can be found. The man-9 glycan at N234 itself is centrally located between
224 a number charged glycans, those at 88, 462, 625 and 637. The high fluctuations of this glycan can
225 stem from its attempt to screen these neighboring negatively charged sugars. The complex sugars
226 in gp41 (611, 618, 625 and 637) have also been shown to have high variations in cryo-EM
227 maps[68]. The total volume sampled is generally larger for native glycosylation, compared to the
228 all-man9 model (**Figure 3D)**. This is not unexpected, since complex glycans generally have larger
229 number of sugars, including the bulky fucose ring at the base. The centrally located high-mannose
230 patch has almost similar volumes in both the models.

231 Another important difference in models that include complex glycans versus the all-man9
232 model results from the presence of the fucose ring at the core of the complex glycans
233 (**Supplementary Figure S2C)**. By structurally aligning the first core sugar and the three residues
234 at the base of the glycan (n-1, n, n+1; where n is the glycosylated asparagine residue) for all the
235 structures in the ensemble, the oligomannose glycans are found to be spaced symmetrically around
236 the core, while complex sugars have a distinct bend away from the side where the fucose ring is
237 present. This difference in orientation preference, along with the presence of the negatively
238 charged sialic acid tips of the antennae, play notable roles that govern the differences between high
239 oligomannose and native glycosylation of the Env. A close examination of the high-threshold
240 difference maps generated in our previous study[38] also reveals such changes in dynamics around
241 single glycans. For instance, the cryoEM map intensity difference around the complex glycan at

242 N618 is located primarily on the underside of the glycan stalk, facing the body of the RM20A3
243 Fab, suggesting that complex type glycans can interact weakly with the Fab in a way high-mannose
244 type glycans cannot. A change in projection angle relative to the peptide surface can also lead to a
245 difference signal; for example the difference signal around the glycan at N611 is strongest below
246 the glycan stalk supporting the hypothesis that a change in glycan projection angle can add
247 asymmetry to the glycan stalk and could affect its dynamics.

248 **Network analysis of glycan topology explicates the shield connectivity**

249 The BG505 trimer, like all HIV-1 Envs, has a highly dense glycosylation pattern and each
250 glycan sample a particular region in space (**Figure 2D**), such that neighboring glycans can occupy
251 overlapping regions (**Figure 4A**). The fraction of volume overlap gives a measure of the
252 interaction probability between the constituent glycans. **Figure 4B** shows a heat map of the glycan-
253 glycan volume overlap within a protomer, while **Figure 4C** represents inter-protomer overlap.
254 Within a protomer there are three main regions of overlap – the apex, the gp41 base, and the
255 glycan-dense central high-mannose patch. Inter-protomer overlaps are mainly due to V1 and V2
256 loop glycans near the trimer apex.

257 While each glycan exerts effects in its immediate vicinity, due to the inter-glycan
258 interactions, their influence can percolate over long distances across the surface of protein. Long-
259 range glycan interactions can occur, as with perturbation of a glycan at one site can affect the
260 processing and antibody interactions of another glycan at a distant site[7, 18, 23, 69].
261 Understanding long-range interactions between glycans is important for characterizing global
262 properties of the glycan shield. We employed graph theory to describe the glycan shield
263 topological network and to examine if this network can successfully capture the long-range
264 interactions.

265 Network analyses has historically been used to study protein allosteric frameworks and
266 evolutionary paths [70-74], and only recently has begun to be applied to in glycoprotein structural
267 characterization [37]. We have previously established this graph-based method, where the network
268 from all-man9 model was verified against cryo-EM data[38]. High network degree correlation
269 with per-glycan map intensity, and centrality correlation with progressive enzymatic digestion of

270 glycans affirms the strength of this network-based approach to capture the shield topology
271 accurately. Detailed methods of modeling the network are recapitulated in the Methods section.
272 Our network is based on a combination of intra and inter-protomer glycan volume overlap maps
273 (**Figures 4B** and **4C**). **Figure 5A** shows the obtained network of BG505 native glycosylation
274 unfolded and laid out in 2 dimensions for the ease of visualization. This is a force-directed layout,
275 which uses attractive forces between adjacent nodes and repulsive forces between distant nodes to
276 reach an optimum distribution of the node points in space[75]. **Supplementary Figure S3** gives
277 another representation of the network with respect to the BG505 structure.

278 The overall topological features remain the same as the all-man9 network described in our
279 previous study[38]. The nodes in the central region around the V4 loop, in the high-mannose patch,
280 are very highly inter-connected. Glycans in the V1, V2 apex are also reasonably well-connected,
281 but the connections at the base near gp41 are relatively sparse, both within the locality, and globally
282 with the rest of the network. Glycans at positions 355, 462 and 276 connect the sparse base region
283 of the network with the dense apex and central regions. Glycan 160, and those in the V2 loop,
284 enable inter-protomer glycan interactions at the apex, stabilizing the protein quaternary trimer
285 structure[38]. Glycan 197, crucial in the binding of CD4bs and V3-specific antibodies[57, 76],
286 connects the central crowded region of the network between neighboring protomers. gp41 glycans
287 611 and 637 also contribute to inter-protomer pathways, with glycan 637 communicating with
288 glycan 448 directly in the central mannose patch.

289 **Utilizing the glycan graph to extract functionally relevant topological features**

290 The degree of a node in a network is the number of connections it has to other nodes
291 (**Figure 5B**). The native glycosylation network is well-connected. Due to the high structural
292 fluctuations of the V2 loop, the glycans on and flanking V2 can interact with a number of other
293 neighboring nodes, having the highest degrees of connectivity. There is dense crowding of glycans
294 around the high mannose patch with a number of inter-connections, resulting in the glycans of this
295 patch being less processed because of reduced accessibility to enzymes[52, 77, 78]. The network
296 degree values are the lowest around the base of the protein on and around gp41, reflective of the
297 sparse architecture of glycan topology in this region.

298 While the shield is a result of the cumulative effect of all the glycans, the importance of
299 each of these glycans is not uniform. To elucidate the relative influence of each of the glycans on
300 the network, we calculated the eigenvector centrality, or eigencentrality, of the nodes to measure
301 its connectivity to the network. A large relative value indicates that the node is well connected to
302 the network and thus is "centrally" located, whereas a low relative value indicates that the node is
303 on the periphery of the network. Operationally, the eigencentrality is the eigenvector associated to
304 the largest eigenvalue of the adjacency matrix, that in position (i,j) reports the overlap of the
305 ensemble of glycan at sites i and j. That value will be zero if the glycans in position i and j do not
306 physically interact (details given in Methods section). The normalized eigencentrality of the
307 glycans are projected on the network as a colormap in **Figure 6A**. The eigencentrality increases
308 towards the middle of the graph, with the glycans at the crowded central patch with a large number
309 of connections between each other having the highest centrality values. When we increase the
310 threshold of volume overlap needed to form an edge, those glycans with the least centrality leave
311 the network first. Glycan 611 is the first to be eliminated, succeeded by some of the inter-protomer
312 interactions and the other gp41 glycans, following the relative centrality scores. The core patch of
313 glycans with high eigencentrality values, mainly consisting of high-mannose glycans, persist
314 throughout the subnetwork.

315 We have previously shown[38] that successive enzymatic digestion of glycans from Env
316 follows a pattern that matches with the network centrality. Those glycans which are sparsely
317 connected to each other, having lower centrality, are eliminated earlier during the process of endoH
318 digestion. On the other hand, the glycans having higher network centrality, such as those in the
319 high-mannose patch, takes longer to be eliminated by the digestive enzymes. The eigencentrality
320 calculated from the all-man9 network and localized intensities after two hours of endoH digestion
321 (Figure 8, reference [38]) have a Pearson's correlation coefficient of ~0.8, and a p-value of 1.14e-
322 05. Thus the eigenvector centrality provides a quantitative measure of the crowding of the highly
323 central glycans that makes them difficult to access for stripping off by endoH action. This
324 comparison with experiments gives compelling evidence of the validity of the network.

325 Next, we calculate the shortest paths between any two sites on the glycan interaction
326 network. As an example, **Figure 6B** shows the shortest paths to all glycans from the glycan at site
327 N332. The site N332 is central in the glycan interaction network, and is known to have both direct

328 and subtle long-range influences over a number of bNAb sites[7, 69]. Our calculation shows the
329 most probable pathway over which each glycan feels the influence of the glycan at site N332. The
330 shortest paths were calculated using the Floyd-Warshall algorithm[79, 80], where the inter-glycan
331 distances were edge-weighted.

332 **Influence of complex glycans on overall network topology**

333 The degree of connectivity decreases almost throughout the network for the native
334 glycosylation (**Supplementary Figure S4A**). The number of stable glycan-glycan interactions at
335 several glycan sites decreases because spatially proximal glycans avoid unfavorable charge-charge
336 interactions. For example, the uncharged glycan 234 takes up a position central to all the
337 neighboring charged sugars, increasing the distances between them, and the charged N406 glycan
338 buries itself in the middle of the surrounding high-mannose glycans (**Figure 6A**), clustering
339 together with them. In fact, removal of this charged glycan N406 glycan can increase the
340 processing of the neighboring high-mannose moieties[81]. This can result from breaking down of
341 this clustering of glycans, leading to increased accessibility by glycosyltransferases and other
342 glycan-processive enzymes. A decrease in connectivity slightly increases the overall diameter of
343 the native network, increasing the mean number of hops for the shortest path from 2.4 hops (and
344 distance 0.19) in the all-man9 model to 3 hops (and distance 0.27)

345 The distribution of centrality also shifts between the two models (**Supplementary Figure**
346 **S4B**). The centralities of the V1 and V2 glycans, along with those at 197, 234, 276 and 462 increase
347 due to native glycosylation. On the other hand, the high-mannose glycans present in and around
348 V3 region decrease from the all-man9 model. **Figure 7A** illuminates the difference in adjacency
349 matrices between the two models, with blue color indicating at least 5% decrease in edge weight,
350 and red indicating at least 5% increase in edge weight in native network, as compared to all-man9.
351 Overall connectivity goes down in the apex and the central patch for the native, and increases
352 within gp41 glycans, as compared to all-man9. The differences between the two networks are
353 shown in **Figure 7B** and **C**. The pathways 137 to loop V2 glycans and 332 as well as those between
354 355 through 625 to 618 become stronger in native glycosylation. A new connectivity comes up
355 between 197 and 276, and a number of paths starting from 197 become shorter (**Figure 7B**). Some
356 of these 197 connections, such as those with 276 and the V5 loop glycans occur across the CD4

357 binding site. It was previously shown that in predominantly high-mannose Env structures, CD4bs
358 targeting antibody VRC01 has very little interaction with glycans 197 and 276[18]. Conversely, in
359 another Env structure with fully processed native glycans, a VRC01-like antibody called IOMA
360 interacted extensively with both 197 and 276[54]. This matches our observation of increased
361 orientation of glycan 197 over the CD4bs in native glycosylation pattern. However, the paths
362 connecting 137, 156, 301, and 197 (of neighboring protomer) are significantly weakened in the
363 native model (**Figure 7C**). Two other subnetworks involving V1,V2 loop glycans and the central
364 mannose patch become weaker due to the presence of native glycans.

365 **Quantifying the vulnerability of glycan shield for antibody-based neutralization**

366 To capture the impact of the glycan shield acting as an immunological barrier, we
367 quantified the sugar barrier over the Env protein surface using the ensemble of structures that have
368 been generated. Towards this goal, we have defined the glycan encounter factor (GEF) to be the
369 number of glycan heavy atoms encountered by an external probe approaching the protein surface.
370 This factor is calculated at each residue on the surface of the protein within a probe of diameter
371 6Å calculated using our ensemble. The highest value obtained is 12 for the native model, and is
372 located in the HMP. Based on existing structures, the main interaction points between Env and
373 bNAbs are often hairpin-like loop regions. Even large-scale atomistic simulations suggest that the
374 first line of contact between Env and an antibody is through a peptide loop[82]. Accordingly, for
375 our analysis, we have used a probe size of 6Å, which is the typical diameter of a hairpin loop. At
376 each residue present on the surface of the protein, the approaching probe was considered in three
377 directions (**Figure 8A**): perpendicular to the surface **z**, and then the **x** and **y** directions spanning
378 the plane parallel to the surface. The geometric mean of the three values were taken to get the final
379 GEF per residue on the Env surface. This value will go to zero when the glycan encounter factor
380 is zero from any one of the three cardinal directions. Thus, for any point on the surface which has
381 a dense glycan covering, such as D in **Figure 8B**, has a high glycan encounter factor value, versus
382 a point such as R where the glycan covering is sparse, which will have a low GEF.

383 **Figure 8C** shows normalized GEFs mapped onto the trimer protein surface (**Figure 8C**).
384 Previous evidence suggests that 70% of the Env ECD surface area is covered by glycans[45].
385 Based on this, we determined the lower cut-off of GEF below which we can define glycan holes.

386 The calculated solvent-exposed surface area of the protein part of our modeled BG505 structure
387 without considering glycans is 86,055 $Å^2$. Excluding the exposed region at the base of the soluble
388 SOSIP, this reduces to 79,672 $Å^2$. Varying the lower cut-off of GEF, we find that a cut-off of 1.5
389 GEF leads to 30% of the surface to be exposed. Regions of BG505 surface having a GEF less than
390 1.5 is colored in black in **Figure 8D**. From the figure, it is clear that the typical glycan holes
391 targeted by bNAbs in BG505, such as the CD4 binding site, the V3-loop epitope and the fusion
392 peptide binding region fall below this GEF cut-off. GEF tracks with epitopes that are relatively
393 generic to a broad range of Env strains[83, 84].

394 Antibodies elicited by BG505.SOSIP.664 are mainly biased towards the missing 241 and
395 289 glycan hole (GH) and the cleft-of-trimer (COT) epitope regions as demonstrated by cryo-EM
396 and immunogenicity assays[85, 86] (see Figure 1 in reference [86]). The GEF parameter identifies
397 these BG505-specific eptiopes as breaches in the glycan shield (**Figure 8B**, "R" region; **Figure**
398 **8D** yellow dashed circle). On the other hand, the densely glycosylated regions around V2, V4 loop
399 and alpha 2 helix have high values of GEF. At each point, the GEF value is given by a combination
400 of all glycans in the vicinity that can come in the way of the approaching probe. At the same time,
401 we can analyze the extent of influence of each glycan on the protein surface. Thus, GEF
402 calculations could aid in interpreting Env sensitivity profiles to bNAbs, and for estimating the
403 impact of interactions of glycans with in the network on epitope exposure for immunogen design.
404 And we are now equipped with a tool to quantify the barrier effect of the glycans individually or
405 as a group, as further demonstrated[85, 86] in the following section.

406 **<u>Combining network topology and glycan encounter factor to inform on local and global</u>**
407 **<u>effects of neutralization</u>**

408 To better understand the dynamics of the glycan shield in the context of neutralization by
409 bNAbs, we examine the clustering of glycans within the network. We define communities using a
410 modularity maximization approach that divides the network into sub-modules or groups (see
411 Methods). Communities have dense connections between the nodes within a module and
412 comparatively sparse connections between nodes in different modules (**Figure 9A inset**). This
413 analysis identifies five distinct communities within the BG505 native glycan network (**Figure 9A**
414 **and B**). The apex glycans around loop V2 from the three protomers together form a single

415 community (1, green). Right below that, glycans 137, 262, 295, 301, and 448 (2, blue) forms a
416 community around the V3 and alpha 2 helix. Glycans 133, 197, 363, 386 and 392 (3, yellow) and
417 339, 398, 406 and 411 (4, red) form two distinct communities involving the glycans on and
418 surrounding V4 loop region. The rest of the glycans with 88, 234, 276, 355 and 462 from gp120
419 and all four glycans from gp41 form the fifth community (5, cyan), though the modularity value
420 is lower due to sparser connections. While the glycan shield is well connected even between the
421 communities, the glycan-glycan interactions within each community is much higher. The
422 possibility of an approaching probe reaching the protein surface through these strongly connected
423 communities is low. A similar study clustering glycans from microsecond simulation runs of
424 BG505 SOSIP was performed by Lemmin *et. al.*[37], identifying 4 glycan microdomains that
425 roughly correspond to our modules 1, 2, (3+4) and 5. However, in that study, mannose-5 glycoform
426 was used at all sites, and due to the smaller length of these glycans, some of the interactions,
427 including inter-protomer interactions were not observed. Remarkably, in that study, junctions
428 between microdomains were found to indicate regions of relative vulnerability. The communities
429 we identified, also demarcate the regions where the glycan shield can be penetrated. Broadly
430 neutralizing antibodies whose binding epitopes are known target these community boundaries
431 (**Figure 9A**). Thus, glycan community dynamics can help to determine susceptible regions of the
432 glycan shield, and can be further used for guiding immunological studies.

433 The proposed network is useful in deciphering the impact of addition or deletion of glycan
434 on neutralization. Removal of the high conserved glycan at 197 by mutating the sequon leads to
435 enhanced neutralization sensitivity to a variety of CD4bs and V3-specific antibodies[76, 87, 88].
436 This glycan is situated proximal to the CD4 binding site, and the tip of V3 loop, directly affecting
437 the binding of these antibodies. However, past experimental evidence also suggests that the
438 deletion of this glycan at N197 increases the binding affinity of two antibodies PG9 and PGT145
439 [7] which target the trimer apex of Env. The epitope of PG9 as determined from the PDB structure
440 5VJ6 include residues 160, 161, 167-173, 185, 305 and 307 (**Figure 10A**) and the epitope of
441 PGT145 as determined from PDB structure 5V8L include residues 123, 124, 127, 160-162, and
442 167-169 (**Figure 10A**). The footprint of glycan 197 as per our GEF model is shown in **Figure 10B**.
443 Residues of the antibody epitopes do not overlap with those regions directly covered by glycan

444 197. Yet, removal of glycan 197 results in significant reduction in the glycan encounter factor over
445 the V2 antibody epitope as evidenced in **Figure 10C** and **D**.

446 The removal of glycan 197 affects the glycans that were originally acting as barriers over
447 the epitopes to resample the available space in such a way that the barrier over the epitopes is now
448 reduced. Glycans 156, 160, 185e and 185h from the neighboring protomers directly shield the PG9
449 and PGT145 epitopes. Looking at the shortest paths of communication between residue 197 from
450 either of the protomers to all other glycans (**Supplementary Figure S5A and S5B**) demonstrate
451 that while glycan 185e interacts with glycan 197, other glycans covering the epitope regions
452 communicate with 197 via a series of inter-glycan interactions. Deletion of 197 does not only affect
453 the V1/V2 loop region glycans. The difference in adjacency matrix due to this perturbation is
454 illustrated in **Supplementary Figure S6**. Interactions such as those between glycans N133 - N386,
455 N295 – N406, N398 – N339, and N411 – N448 are reduced more than 10%. On the other hand,
456 glycan 276, which was originally interacting strongly with 197 now forms new interactions with
457 N295, N332 and N411 and N448 glycans. Therefore, deletion of the glycan at 197 also causes
458 significant changes to the shield topology up to the V4 loop region, and our methodology
459 describing the glycan shield network corroborates the experimental findings, and sheds light on
460 the most probable pathways through which the glycan-glycan communications occur.

461

462 **Discussion**

463 The dense arrangement of N-glycans masks antigenic surfaces on Env, acting as a dynamic
464 shield protecting the protein from the adaptive immune system. Moreover, since host proteins have
465 similar glycosylation pathways, these self-glycans are generally immunologically inert.
466 Understanding the structure and dynamics of the glycan shield as a whole is therefore important
467 for Env-based vaccine design. While X-ray, NMR and cryo-EM structures have supplied a number
468 of important molecular details about the Env glycoprotein, they do not account for the high
469 flexibility and dynamics of these glycans that leads to the glycan shield. Detailed characterization
470 of the glycan shield and the quantification of the resilience and vulnerability of this barrier over
471 the Env surface can add new perspective and depth to current HIV vaccine design efforts.

472 All previous computational studies have generalized glycan moieties as simple oligo-
473 mannoses such as mannose-5 and mannose-9 [18, 35-37, 57] for the purpose of modeling. For the
474 first time, we have incorporated native glycosylation, and included complex sugars based on site-
475 specific mass spectrometry results. While the overall glycan aspects of the network are consistent
476 between mannose-9 and native-like glycosylation, there are critical differences. The most
477 noteworthy are a consequence of the presence of the bulky fucose ring at the base or the negatively
478 charged sialic acids at the antennae tips. These lead to overall rearrangement of glycan orientations
479 affecting its microenvironment, and ultimately influencing the shield topology. Thus, our
480 comprehensive view of the shield is capable of capturing individual glycan effects which are
481 physiologically and immunologically more relevant.

482 Our network-based approach enables in understanding the collective behavior of the
483 glycans. We can compare the relative centrality of glycans, and identify potential interaction
484 pathways. Importantly, the centrality or importance of glycans correlates well with experimental
485 cryo-EM data[38]. Glycans with lower centrality have lesser influence on the graph, and are the
486 first to be eliminated from the network, if the adjacency threshold is increased. At the same time,
487 the most central region of the network is the most resilient to enzymatic action. Such centrality
488 measures can help determine the ease of targeting glycans and their modifications, guiding the
489 process of immunogen development in the context of distinct Envs.

490 Complex networks of glycan interactions that give rise to long range effects over the shield
491 topology are evident[7]. We have also identified specific communities of glycans in the shield that
492 have high degrees of interaction within each community. The antigen surface under each such
493 community is therefore better shielded, and the regions under community junctions are more
494 susceptible to antibody binding. The boundaries of such glycan domains have been previously
495 identified to be regions of vulnerability[37]. While the shield resilience is the main consideration
496 for epitope exposure, some antibodies have also evolved to capitalize on specific interactions with
497 a number of the conserved glycans during their molecular engagement with the epitope. Due to
498 the stability of glycan interactions within the communities, the community detection can also aid
499 in the determination of antibody angle of approach that has been shown to influence the breadth
500 of bNAbs [89].

There are known common and uncommon ‘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[22]. Previous advances have been made in identifying breaches in the shield based on the area of influence of each glycan[22]. In this study, 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 epitope regions[69]. Therefore, we have derived a measure to quantify the shielding effect based on the encounter factor of glycan 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 and the subsequent development of the network is high-throughput compared to traditional sampling by MD simulation methods. Due to the ease of this fast and efficient pipeline 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 protein. 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[90, 91] can also be easily investigated utilizing this pipeline. Capitalizing on the relatively low computational overhead of this approach, 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.

While this approach captures the equilibrium ensemble of glycan spatial distribution, it lacks temporal information regarding the glycoprotein dynamics and any topological transitions that might occur on shorter timescales. Computationally costly MD simulations could potentially address the dynamics of hierarchical glycan topology. These simulations can also provide insights on how antibodies navigate the transient accessibility of epitope through highly flexible glycans and to refine the 6Å probe considered here.

529 Beyond HIV-1 and other viral envelope proteins, the significance of glycoproteins in a vast
530 array of biological processes from protein folding, cellular communication to immune-regulation,
531 make them a fast-emerging field of interest in biomedical research. Changes in these glycosylation
532 patterns have been associated with various diseases, including rheumatoid arthritis[92], and
533 cancers[93]. Additionally, many of the current therapeutic antibodies in the market are N-linked
534 glycoproteins, and the significance of N-glycans is becoming increasingly evident [94]. The ease
535 of modeling of glycan network utilizing our approach makes it translatable to other systems, and
536 can assist in determining the role of these glycans in conjugation with the underlying proteins at a
537 molecular level. In comparison to N-glycans, our understanding of O-glycosylation is limited.
538 This class of glycans are relatively more complex in terms of structural variations. It would be an
539 interesting challenge to generalize our methodology to encompass O-glycosylated systems as well,
540 that are present to some extent on HIV-1 Env. This can also help gain molecular insights on such
541 refractory systems, like Ebola that have both N- and O-linked glycosylation, and in understanding
542 the pathological implications of dense O-glycans in mucin associated cancers[95].

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

554

555

556

557 # STAR Methods

558 ## Contact for Reagent and Resource Sharing

559 Further information and requests for resources and reagents should be directed to and will be
560 fulfilled by the Lead Contacts Andrew B. Ward (andrew@scripps.edu) and S. Gnanakaran
561 (gnana@lanl.gov).

562 ## Method Details

563 **High-throughput conformational modeling**

564 The ensemble of BG505 glycoprotein 3D conformations were built in atomistic detail by
565 implementing the ALLOSMOD[63, 64] package of MODELLER[65, 66]. The BG505 protein
566 scaffold was homology modeled, by threading the protein sequence into available crystal structure
567 templates. SOSIP structure with PDB accession ID 4ZMJ[96] was used as the template for gp120,
568 that also guided the three-fold symmetry of the trimer. 5CEZ[97] was used as the template for
569 gp41 and since it has the least number of missing residues among the available structures. The
570 missing residues in the loops were modeled *ab initio*, using known disulphide bonds as additional
571 restraints. 100 protein models were generated, and the best 10 were selected as starting scaffolds
572 for glycan building, based on MODELLER optimization scores, and stereochemistry scores as
573 determined by PROCHECK[98]. This results in different starting orientations of the hypervariable
574 loops. For each of the 10 selected protein structures, glycans were initially added with random
575 orientation, at the known glycosylation sites, based on ideal geometries as dictated by
576 CHARMM36[99, 100] force field internal coordinates, followed by a 1Å randomization added to
577 the overall atomic coordinates as described by Guttman et. al.[64]. Once all the glycans were
578 added, ensuing refinement steps of the glycoprotein system optimized an energy function given by
579 a combination of template-based spatial restraints, CHARMM36 forcefield terms, and a soft
580 sphere-like truncated Gaussian term to prevent collisions. The structures were relaxed with 1000
581 steps of conjugate gradient minimization followed by a short molecular dynamics equilibration of
582 500ps. Further refining with five rounds of simulated annealing was performed between 1,300K
583 to 300K in 8 steps. The glycans and the loop regions were kept flexible during the refinement
584 steps. 100 fully glycosylated structures were modeled from each of the 10 selected protein models,

585 resulting in the final ensemble containing 1000 different poses. In order to let these unstructured
586 loops sample a wider range of conformations, we removed the template restraints from these loop
587 regions during the protein homology modeling phase. These residues by HXB2 numbering are as
588 follows: 143 to 152 (V1 loop), BG505-specific insert residues 185A to 185I and 186 to 189 (V2
589 loop), 309 to 315 and 325 to 329 (V3 loop), 400 to 410 (V4 loop) and 458 to 464 (V5 loop) as
590 determined from the LANL HIV database (https://www.hiv.lanl.gov/).

591 **Glycan root mean square fluctuations**

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

595 $$RMSF_n = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{M}\sum_{m=1}^{M}|\vec{r}_{mnk} - \langle\vec{r}_{nk}\rangle|^2}$$

596 where $\vec{r}_{mnk}$ is the atomic position of heavy atom $k$ of glycan $n$ in snapshot $m$, $\langle\vec{r}_{nk}\rangle =$
597 $(\frac{1}{M})\sum_{m=1}^{M}\vec{r}_{mnk}$ is the average atomic position of heavy atom $k$ in glycan $n$. K is the total number
598 of heavy atoms in the glycan. It is 127 for man-9, and varies depending on the type of glycan. The
599 ensemble for each model contains 1000 snapshots, making $M = 1000$ snapshots for each of the two
600 models. The standard deviations (s.d.) were obtained by dividing the 1000 snapshots into 4 sets of
601 M=250, and calculating the four sets of RMSF values (Figure 3A).

602 The 1000 structures of the total ensemble are built from 10 initial starting protein
603 conformations, as described above. The main difference between these 10 conformations are the
604 variations in the loop regions, due to the missing residues in the templates. In order to reduce the
605 effects of the loop fluctuations and consider the RMSF coming from the glycans alone, the reduced
606 RMSF values were also calculated in each of these 10 sub-models, and their average and s.d.
607 calculated (**Supplementary Figure S2B**). The RMSF difference between the models (Figure 3D)
608 were obtained by subtracting the reduced RMSF values of all-man9 from native model (native
609 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.

**Glycan volume 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 (Figure 5A 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[101] and Matlab_R2018a packages[102].

**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[103], 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.

**<u>Modularity maximization for community detection</u>**

Community detection within the glycan network was performed using the modularity maximization approach given by Newmann and Girvan[104, 105]. 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[106] in Matlab R2018a.

## Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

## Supplemental Information

Supplemental Information includes 6 Figures, 1 Table, and details of torsion angle distributions of modeled glycan structure.


## Acknowledgements

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.W 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.

## Competing Interests

The authors declare no competing interests.

685 **References**


686 1. An, H.J., J.W. Froehlich, and C.B. Lebrilla, *Determination of glycosylation sites and site-*
687 *specific heterogeneity in glycoproteins.* Curr Opin Chem Biol, 2009. **13**(4): p. 421-6.
688 2. Dwek, R.A., *Glycobiology: Toward Understanding the Function of Sugars.* Chem Rev,
689 1996. **96**(2): p. 683-720.
690 3. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation:*
691 *diversity, synthesis and function.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 448-62.
692 4. Dennis, J.W., M. Granovsky, and C.E. Warren, *Protein glycosylation in development and*
693 *disease.* Bioessays, 1999. **21**(5): p. 412-21.
694 5. Stanley, P., N. Taniguchi, and M. Aebi, *N-Glycans*, in *Essentials of Glycobiology*, rd, et
695 al., Editors. 2015: Cold Spring Harbor (NY). p. 99-111.
696 6. Wang, B., et al., *Mechanistic understanding of N-glycosylation in Ebola virus*
697 *glycoprotein maturation and function.* J Biol Chem, 2017. **292**(14): p. 5860-5870.
698 7. Behrens, A.J., et al., *Composition and Antigenic Effects of Individual Glycan Sites of a*
699 *Trimeric HIV-1 Envelope Glycoprotein.* Cell Rep, 2016. **14**(11): p. 2695-706.
700 8. Wang, C.C., et al., *Glycans on influenza hemagglutinin affect receptor binding and*
701 *immune response.* Proc Natl Acad Sci U S A, 2009. **106**(43): p. 18137-42.
702 9. Nothaft, H. and C.M. Szymanski, *Bacterial protein N-glycosylation: new perspectives*
703 *and applications.* J Biol Chem, 2013. **288**(10): p. 6912-20.
704 10. Burton, D.R. and J.R. Mascola, *Antibody responses to envelope glycoproteins in HIV-1*
705 *infection.* Nat Immunol, 2015. **16**(6): p. 571-6.
706 11. Sommerstein, R., et al., *Arenavirus Glycan Shield Promotes Neutralizing Antibody*
707 *Evasion and Protracted Infection.* PLoS Pathog, 2015. **11**(11): p. e1005276.
708 12. Zhang, X., et al., *Structures and functions of the envelope glycoprotein in flavivirus*
709 *infections.* Viruses, 2017. **9**(11): p. 338.
710 13. Szakonyi, G., et al., *Structure of the Epstein-Barr virus major envelope glycoprotein.* Nat
711 Struct Mol Biol, 2006. **13**(11): p. 996-1001.
712 14. Lennemann, N.J., et al., *Comprehensive functional analysis of N-linked glycans on Ebola*
713 *virus GP1.* MBio, 2014. **5**(1): p. e00862-13.
714 15. Ilinykh, P.A., et al., *Asymmetric antiviral effects of ebolavirus antibodies targeting*
715 *glycoprotein stem and glycan cap.* PLoS Pathog, 2018. **14**(8): p. e1007204.
716 16. Amanat, F., et al., *Antibodies to the Glycoprotein GP2 Subunit Cross-React between Old*
717 *and New World Arenaviruses.* mSphere, 2018. **3**(3).
718 17. Saphire, E.O., et al., *Systematic Analysis of Monoclonal Antibodies against Ebola Virus*
719 *GP Defines Features that Contribute to Protection.* Cell, 2018. **174**(4): p. 938-952 e13.
720 18. Stewart-Jones, G.B., et al., *Trimeric HIV-1-Env Structures Define Glycan Shields from*
721 *Clades A, B, and G.* Cell, 2016. **165**(4): p. 813-26.
722 19. Karsten, C.B. and G. Alter, *The HIV-1 Glycan Shield: Strategically Placed Kinks in the*
723 *Armor Improve Antigen Design.* Cell Rep, 2017. **19**(4): p. 669-670.
724 20. Ringe, R.P., et al., *Closing and Opening Holes in the Glycan Shield of HIV-1 Envelope*
725 *Glycoprotein SOSIP Trimers Can Redirect the Neutralizing Antibody Response to the*
726 *Newly Unmasked Epitopes.* J Virol, 2019. **93**(4).
727 21. Crispin, M., A.B. Ward, and I.A. Wilson, *Structure and Immune Recognition of the HIV*
728 *Glycan Shield.* Annu Rev Biophys, 2018.

729 22. Wagh, K., et al., *Completeness of HIV-1 Envelope Glycan Shield at Transmission*
730 *Determines Neutralization Breadth.* Cell Rep, 2018. **25**(4): p. 893-908 e7.
731 23. Doores, K.J., *The HIV glycan shield as a target for broadly neutralizing antibodies.* The
732 FEBS journal, 2015. **282**(24): p. 4679-4691.
733 24. Walsh, G., *Biopharmaceutical benchmarks 2010.* Nat Biotechnol, 2010. **28**(9): p. 917-24.
734 25. Imberty, A. and S. Perez, *Structure, conformation, and dynamics of bioactive*
735 *oligosaccharides: theoretical approaches and experimental validations.* Chem Rev,
736 2000. **100**(12): p. 4567-88.
737 26. Chang, V.T., et al., *Glycoprotein structural genomics: solving the glycosylation problem.*
738 Structure, 2007. **15**(3): p. 267-73.
739 27. Davis, S.J. and M. Crispin, *Solutions to the glycosylation problem for low-and high-*
740 *throughput structural glycoproteomics*, in *Functional and Structural Proteomics of*
741 *Glycoproteins*. 2010, Springer. p. 127-158.
742 28. Slynko, V., et al., *NMR structure determination of a segmentally labeled glycoprotein*
743 *using in vitro glycosylation.* J Am Chem Soc, 2009. **131**(3): p. 1274-81.
744 29. Woods, R.J., et al., *The high degree of internal flexibility observed for an oligomannose*
745 *oligosaccharide does not alter the overall topology of the molecule.* Eur J Biochem,
746 1998. **258**(2): p. 372-86.
747 30. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
748 31. Kumar, S. and P. Cieplak, *Role of N-glycosylation in activation of proMMP-9. A*
749 *molecular dynamics simulations study.* PLoS One, 2018. **13**(1): p. e0191157.
750 32. Lee, H.S., Y. Qi, and W. Im, *Effects of N-glycosylation on protein conformation and*
751 *dynamics: Protein Data Bank analysis and molecular dynamics simulation study.*
752 Scientific reports, 2015. **5**: p. 8926.
753 33. Dong, C., et al., *Long-ranged Protein-glycan Interactions Stabilize von Willebrand*
754 *Factor A2 Domain from Mechanical Unfolding.* Sci Rep, 2018. **8**(1): p. 16017.
755 34. Hang, I., et al., *Analysis of site-specific N-glycan remodeling in the endoplasmic*
756 *reticulum and the Golgi.* Glycobiology, 2015. **25**(12): p. 1335-49.
757 35. Ferreira, R.C., et al., *Structural Rearrangements Maintain the Glycan Shield of an HIV-1*
758 *Envelope Trimer After the Loss of a Glycan.* Sci Rep, 2018. **8**(1): p. 15031.
759 36. Yang, M., et al., *Conformational Heterogeneity of the HIV Envelope Glycan Shield.* Sci
760 Rep, 2017. **7**(1): p. 4435.
761 37. Lemmin, T., et al., *Microsecond Dynamics and Network Analysis of the HIV-1 SOSIP*
762 *Env Trimer Reveal Collective Behavior and Conserved Microdomains of the Glycan*
763 *Shield.* Structure, 2017. **25**(10): p. 1631-1639 e2.
764 38. Berndsen, Z., Chakraborty, S., Wang, X., Cottrell, C., Torres, J., Lopez, C., van-Gills,
765 M., Paulson, J., Gnanakaran, S., Ward, A. B., *Visualization of HIV-1 Env Glycan Shield*
766 *Across Scales* (submitted for peer review), bioRxiv doi: 10.1101/839217.
767 39. Zhang, M., et al., *Tracking global patterns of N-linked glycosylation site variation in*
768 *highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza*
769 *hemagglutinin.* Glycobiology, 2004. **14**(12): p. 1229-46.
770 40. Lasky, L.A., et al., *Neutralization of the AIDS retrovirus by antibodies to a recombinant*
771 *envelope glycoprotein.* Science, 1986. **233**(4760): p. 209-12.
772 41. Walker, L.M., et al., *A limited number of antibody specificities mediate broad and potent*
773 *serum neutralization in selected HIV-1 infected individuals.* PLoS Pathog, 2010. **6**(8): p.
774 e1001028.

775 42. Bradley, T., et al., *Structural Constraints of Vaccine-Induced Tier-2 Autologous HIV*
776 *Neutralizing Antibodies Targeting the Receptor-Binding Site.* Cell Rep, 2016. **14**(1): p.
777 43-54.
778 43. Haynes, B.F., et al., *Progress in HIV-1 vaccine development.* J Allergy Clin Immunol,
779 2014. **134**(1): p. 3-10; quiz 11.
780 44. Huang, J., et al., *Broad and potent HIV-1 neutralization by a human antibody that binds*
781 *the gp41-gp120 interface.* Nature, 2014. **515**(7525): p. 138-42.
782 45. Pancera, M., et al., *Structure and immune recognition of trimeric pre-fusion HIV-1 Env.*
783 Nature, 2014. **514**(7523): p. 455-61.
784 46. Julien, J.P., et al., *Crystal structure of a soluble cleaved HIV-1 envelope trimer.* Science,
785 2013. **342**(6165): p. 1477-83.
786 47. Lyumkis, D., et al., *Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1*
787 *envelope trimer.* Science, 2013. **342**(6165): p. 1484-90.
788 48. Mouquet, H., et al., *Complex-type N-glycan recognition by potent broadly neutralizing*
789 *HIV antibodies.* Proc Natl Acad Sci U S A, 2012. **109**(47): p. E3268-77.
790 49. Walker, L.M., et al., *Broad and potent neutralizing antibodies from an African donor*
791 *reveal a new HIV-1 vaccine target.* Science, 2009. **326**(5950): p. 285-9.
792 50. Falkowska, E., et al., *Broadly neutralizing HIV antibodies define a glycan-dependent*
793 *epitope on the prefusion conformation of gp41 on cleaved envelope trimers.* Immunity,
794 2014. **40**(5): p. 657-68.
795 51. Crispin, M. and K.J. Doores, *Targeting host-derived glycans on enveloped viruses for*
796 *antibody-based vaccine design.* Curr Opin Virol, 2015. **11**: p. 63-9.
797 52. Pritchard, L.K., et al., *Glycan clustering stabilizes the mannose patch of HIV-1 and*
798 *preserves vulnerability to broadly neutralizing antibodies.* Nat Commun, 2015. **6**: p.
799 7479.
800 53. Pritchard, L.K., et al., *Structural Constraints Determine the Glycosylation of HIV-1*
801 *Envelope Trimers.* Cell Rep, 2015. **11**(10): p. 1604-13.
802 54. Gristick, H.B., et al., *Natively glycosylated HIV-1 Env structure reveals new mode for*
803 *antibody recognition of the CD4-binding site.* Nat Struct Mol Biol, 2016. **23**(10): p. 906-
804 915.
805 55. Lee, J.H., G. Ozorowski, and A.B. Ward, *Cryo-EM structure of a native, fully*
806 *glycosylated, cleaved HIV-1 envelope trimer.* Science, 2016. **351**(6277): p. 1043-8.
807 56. Barnes, C.O., et al., *Structural characterization of a highly-potent V3-glycan broadly*
808 *neutralizing antibody bound to natively-glycosylated HIV-1 envelope.* Nat Commun,
809 2018. **9**(1): p. 1251.
810 57. Liang, Y., et al., *Changes in Structure and Antigenicity of HIV-1 Env Trimers Resulting*
811 *from Removal of a Conserved CD4 Binding Site-Proximal Glycan.* J Virol, 2016. **90**(20):
812 p. 9224-36.
813 58. Cao, L., et al., *Global site-specific N-glycosylation analysis of HIV envelope*
814 *glycoprotein.* Nat Commun, 2017. **8**: p. 14954.
815 59. Sanders, R.W., et al., *HIV-1 VACCINES. HIV-1 neutralizing antibodies induced by*
816 *native-like envelope trimers.* Science, 2015. **349**(6244): p. aac4223.
817 60. Sanders, R.W., et al., *A next-generation cleaved, soluble HIV-1 Env trimer, BG505*
818 *SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-*
819 *neutralizing antibodies.* PLoS Pathog, 2013. **9**(9): p. e1003618.

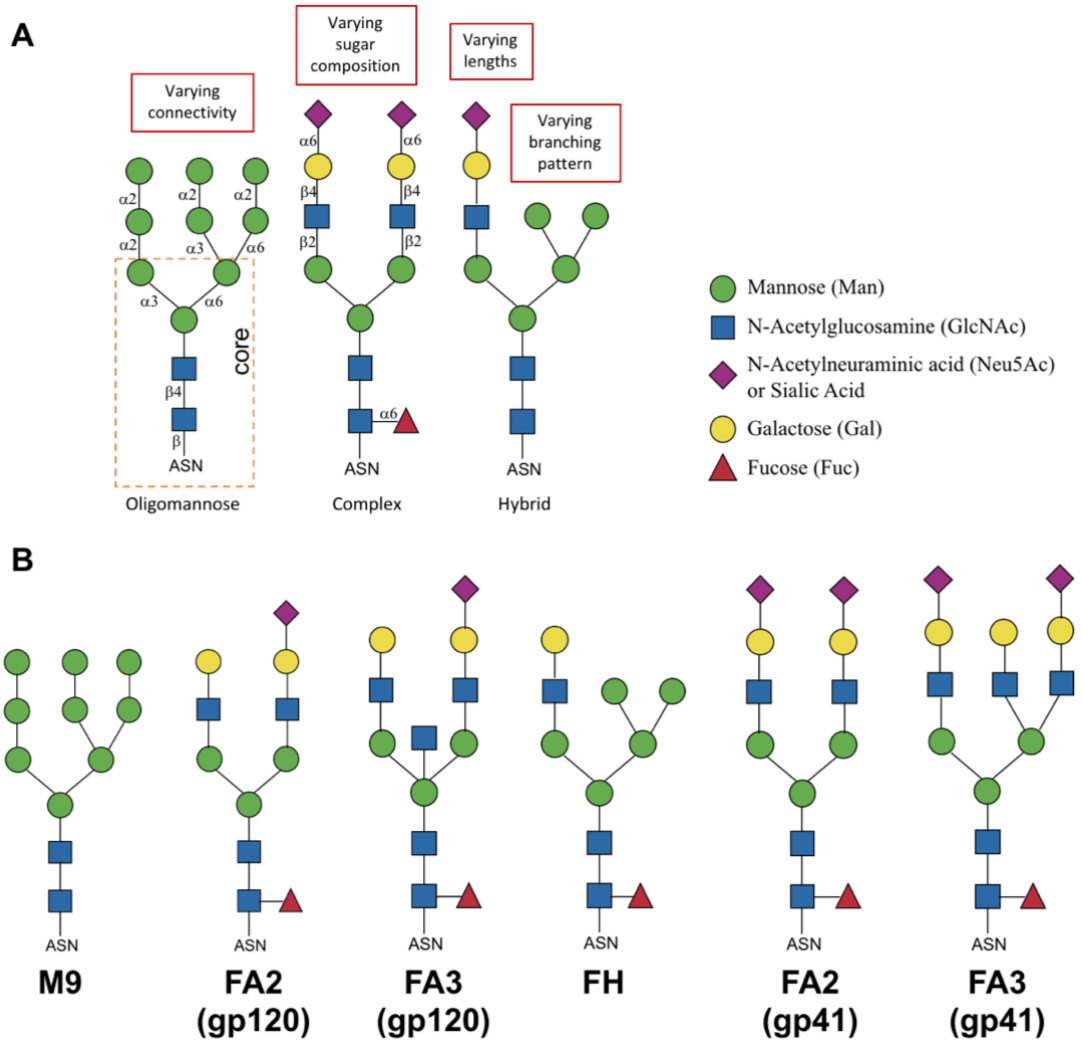
61. Go, E.P., et al., *Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein Production Based on Eleven Env Trimers.* J Virol, 2017. **91**(9).
62. Guvench, O., et al., *CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling.* Journal of chemical theory and computation, 2011. **7**(10): p. 3162-3180.
63. Weinkam, P., J. Pons, and A. Sali, *Structure-based model of allostery predicts coupling between distant sites.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 4875-80.
64. Guttman, M., et al., *All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins.* Structure, 2013. **21**(3): p. 321-31.
65. Eswar, N., et al., *Comparative protein structure modeling using Modeller.* Current protocols in bioinformatics, 2006. **15**(1): p. 5.6. 1-5.6. 30.
66. Sali, A., *Comparative protein modeling by satisfaction of spatial restraints.* Mol Med Today, 1995. **1**(6): p. 270-7.
67. Lütteke, T., M. Frank, and C.-W. von der Lieth, *Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB.* Nucleic acids research, 2005. **33**(suppl_1): p. D242-D246.
68. Ward, A.B. and I.A. Wilson, *The HIV-1 envelope glycoprotein structure: Nailing down a moving target.* Immunological reviews, 2017. **275**(1): p. 21-32.
69. Bricault, C.A., et al., *HIV-1 Neutralizing Antibody Signatures and Application to Epitope-Targeted Vaccine Design.* Cell Host Microbe, 2019. **25**(1): p. 59-72 e8.
70. Beleva Guthrie, V., et al., *Network analysis of protein adaptation: Modeling the functional impact of multiple mutations.* Molecular biology and evolution, 2018. **35**(6): p. 1507-1519.
71. Huang, L., L. Liao, and C.H. Wu, *Evolutionary analysis and interaction prediction for protein-protein interaction network in geometric space.* PloS one, 2017. **12**(9): p. e0183495.
72. Eargle, J. and Z. Luthey-Schulten, *NetworkView: 3D display and analysis of protein.RNA interaction networks.* Bioinformatics, 2012. **28**(22): p. 3000-1.
73. Skjaerven, L., et al., *Integrating protein structural dynamics and evolutionary analysis with Bio3D.* BMC Bioinformatics, 2014. **15**: p. 399.
74. Sethi, A., et al., *A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein.* PLoS Comput Biol, 2013. **9**(5): p. e1003046.
75. Fruchterman, T.M. and E.M. Reingold, *Graph drawing by force-directed placement.* Software: Practice and experience, 1991. **21**(11): p. 1129-1164.
76. Li, Y., et al., *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, 2008. **82**(2): p. 638-51.
77. Coss, K.P., et al., *HIV-1 Glycan Density Drives the Persistence of the Mannose Patch within an Infected Individual.* J Virol, 2016. **90**(24): p. 11132-11144.
78. Doores, K.J., et al., *Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens.* Proc Natl Acad Sci U S A, 2010. **107**(31): p. 13800-5.
79. Warshall, S. *A theorem on boolean matrices*. in *Journal of the ACM*. 1962. Citeseer.
80. Floyd, R.W., *Algorithm 97: shortest path.* Communications of the ACM, 1962. **5**(6): p. 345.
81. Cao, L., et al., *Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer.* Nat Commun, 2018. **9**(1): p. 3693.

866 82. Schmidt, A.G., et al., *Preconfiguration of the antigen-binding site during affinity*
867 *maturation of a broadly neutralizing influenza virus antibody.* Proc Natl Acad Sci U S A,
868 2013. **110**(1): p. 264-9.
869 83. Xu, K., et al., *Epitope-based vaccine design yields fusion peptide-directed antibodies that*
870 *neutralize diverse strains of HIV-1.* Nat Med, 2018. **24**(6): p. 857-867.
871 84. Burton, D.R., et al., *A Blueprint for HIV Vaccine Discovery.* Cell Host Microbe, 2012.
872 **12**(4): p. 396-407.
873 85. McCoy, L.E., et al., *Holes in the Glycan Shield of the Native HIV Envelope Are a Target*
874 *of Trimer-Elicited Neutralizing Antibodies.* Cell Rep, 2016. **16**(9): p. 2327-38.
875 86. Bianchi, M., et al., *Electron-Microscopy-Based Epitope Mapping Defines Specificities of*
876 *Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization.*
877 Immunity, 2018. **49**(2): p. 288-300 e8.
878 87. Townsley, S., et al., *Conserved Role of an N-Linked Glycan on the Surface Antigen of*
879 *Human Immunodeficiency Virus Type 1 Modulating Virus Sensitivity to Broadly*
880 *Neutralizing Antibodies against the Receptor and Coreceptor Binding Sites.* J Virol,
881 2016. **90**(2): p. 829-41.
882 88. Huang, X., et al., *Highly conserved HIV-1 gp120 glycans proximal to CD4-binding*
883 *region affect viral infectivity and neutralizing antibody induction.* Virology, 2012.
884 **423**(1): p. 97-106.
885 89. Moore, P.L. and C. Williamson, *Approaches to the induction of HIV broadly neutralizing*
886 *antibodies.* Current Opinion in HIV and AIDS, 2016. **11**(6): p. 569.
887 90. LaBranche, C.C., et al., *HIV-1 envelope glycan modifications that permit neutralization*
888 *by germline-reverted VRC01-class broadly neutralizing antibodies.* PLoS Pathog, 2018.
889 **14**(11): p. e1007431.
890 91. Umotoy, J., et al., *Rapid and Focused Maturation of a VRC01-Class HIV Broadly*
891 *Neutralizing Antibody Lineage Involves Both Binding and Accommodation of the N276-*
892 *Glycan.* Immunity, 2019. **51**(1): p. 141-154 e6.
893 92. Nakagawa, H., et al., *Detection of altered N-glycan profiles in whole serum from*
894 *rheumatoid arthritis patients.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007.
895 **853**(1-2): p. 133-7.
896 93. Taniguchi, N. and Y. Kizuka, *Glycans and cancer: role of N-glycans in cancer*
897 *biomarker, progression and metastasis, and therapeutics.* Adv Cancer Res, 2015. **126**: p.
898 11-51.
899 94. Dalziel, M., et al., *Emerging principles for the therapeutic exploitation of glycosylation.*
900 Science, 2014. **343**(6166): p. 1235681.
901 95. Bhatia, R., et al., *Cancer-associated mucins: role in immune modulation and metastasis.*
902 Cancer Metastasis Rev, 2019. **38**(1-2): p. 223-236.
903 96. Kwon, Y.D., et al., *Crystal structure, conformational fixation and entry-related*
904 *interactions of mature ligand-free HIV-1 Env.* Nat Struct Mol Biol, 2015. **22**(7): p. 522-
905 31.
906 97. Garces, F., et al., *Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily*
907 *Focused on Accommodating or Avoiding Glycans.* Immunity, 2015. **43**(6): p. 1053-63.
908 98. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of*
909 *protein structures.* Journal of applied crystallography, 1993. **26**(2): p. 283-291.

99. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone $\phi$, $\psi$ and side-chain $\chi 1$ and $\chi 2$ dihedral angles.* Journal of chemical theory and computation, 2012. **8**(9): p. 3257-3273.
100. Huang, J. and A.D. MacKerell Jr, *CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data.* Journal of computational chemistry, 2013. **34**(25): p. 2135-2145.
101. Rossum, G., *Python reference manual.* 1995.
102. MATLAB, *MATLAB, Version R2018a*. 2018, The MathWorks Inc Natick, MA.
103. Pillai, S.U., T. Suel, and S. Cha, *The Perron-Frobenius theorem: some of its applications.* IEEE Signal Processing Magazine, 2005. **22**(2): p. 62-75.
104. Newman, M.E., *Modularity and community structure in networks.* Proceedings of the national academy of sciences, 2006. **103**(23): p. 8577-8582.
105. Newman, M.E. and M. Girvan, *Finding and evaluating community structure in networks.* Physical review E, 2004. **69**(2): p. 026113.
106. Blondel, V.D., et al., *Fast unfolding of communities in large networks.* Journal of statistical mechanics: theory and experiment, 2008. **2008**(10): p. P10008.

927 **Figures and Figure Legends**

928




929

930

931 **Figure1: Schematic representation of N-glycan types.** (a). The three generalized N-glycan types
932 commonly found in glycoproteins are high oligomannose, complex, and hybrid (having a mix of
933 oligomannose and complex branches). The common core of Man3GlcNAc2Asn is indicated.
934 Variability of these glycans stem from varying connectivity, sugar composition, lengths and
935 branching patterns. (b). The glycan species selected for different sites of the BG505 native model
936 (as given in Table S1). M9 ≡ mannose-9, FA2 ≡ fucosylated two-antennae, FA3 ≡ fucosylated
937 three-antennae, FH ≡ fucosylated hybrid. It must be noted that the complex glycans in gp41 (N611,
938 N618, N625, N637) are different from those in gp120, as per site-specific mass spectroscopy
939 experiments.

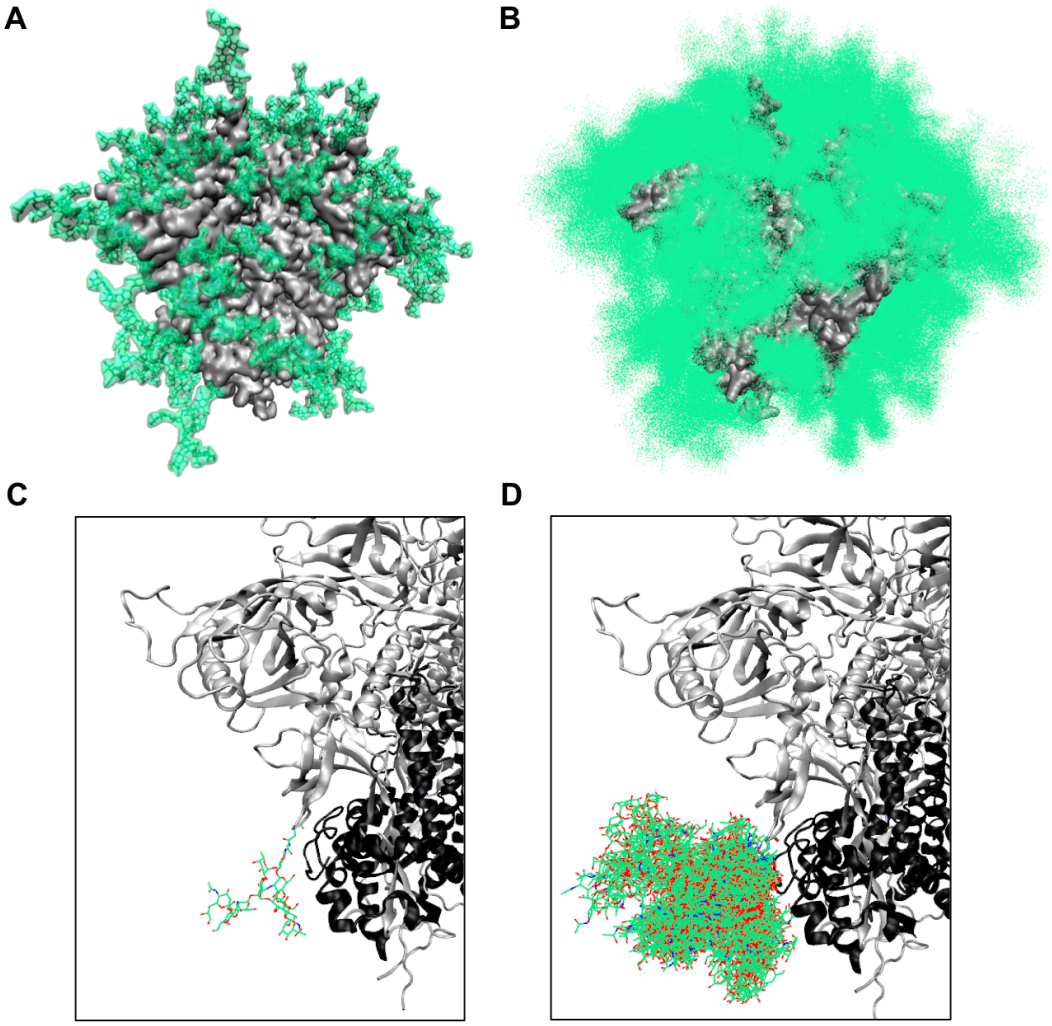



**Figure 2: Spatial shielding of Env protein by glycans.** (a) A single snapshot of modeled natively glycosylated BG505 structure, with each glycan taking up a particular conformation. Protein surface in grey, glycans in green. (b) Cumulative shielding effect over time due to the flexible and dynamic nature of the glycans. 100 randomly selected models from the 1000-structure ensemble is shown here. (c) Glycan N88 in one particular pose. gp120 (grey), gp41 (black), glycan carbon atoms (green), glycan oxygen atoms (red), glycan nitrogen atoms (blue). (d) Spatial sampling by glycan N88. Each glycan can take a variety of different conformations and orientations, sampling a large volume in space.

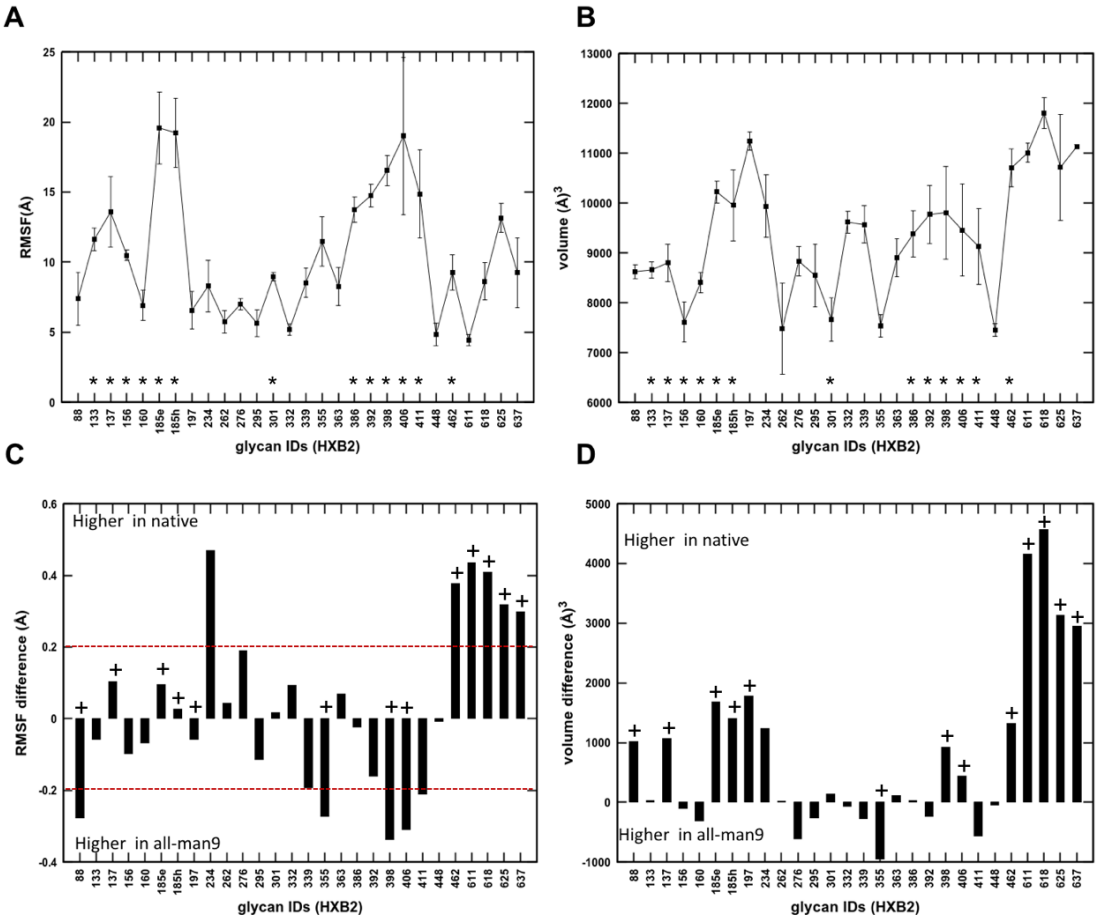
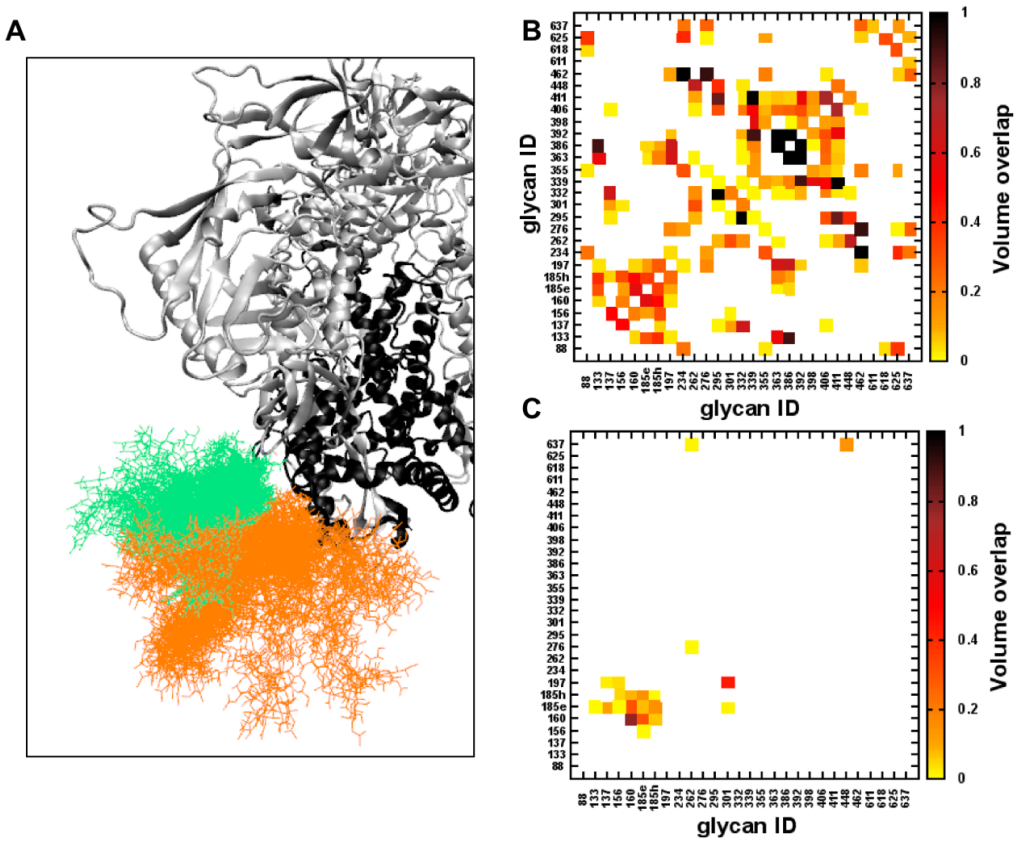



**Figure 3: Structural fluctuations of individual glycans in BG505 native model.** Glycans present in the hyper-variable loop regions of gp120 are indicated by *. Glycans modeled as fucosylated complex or hybrid glycoforms are indicated by +. (a) Site-specific Root Mean Squared fluctuations (RMSF). (b) Sampled volume per glycan at each PNGS. (c) Difference in RMSF between native and all-man9 model. Native minus all-man9 values are plotted. (d) Difference in sampled volume between native and all-man9 model. Native minus all-man9 values are plotted.




958

959

960 **Figure 4: Glycan-glycan volume overlap.** (a) Neighboring glycans within close proximity can
961 sample overlapping volumes. Glycan N88 shown in green and glycan N625 shown in orange.
962 gp120 (grey) and gp41 (black) are shown in cartoon representation. (b) Probability distribution
963 map of inter-glycan fractional overlap within a single protomer, and (c) between inter-protomer
964 glycans.

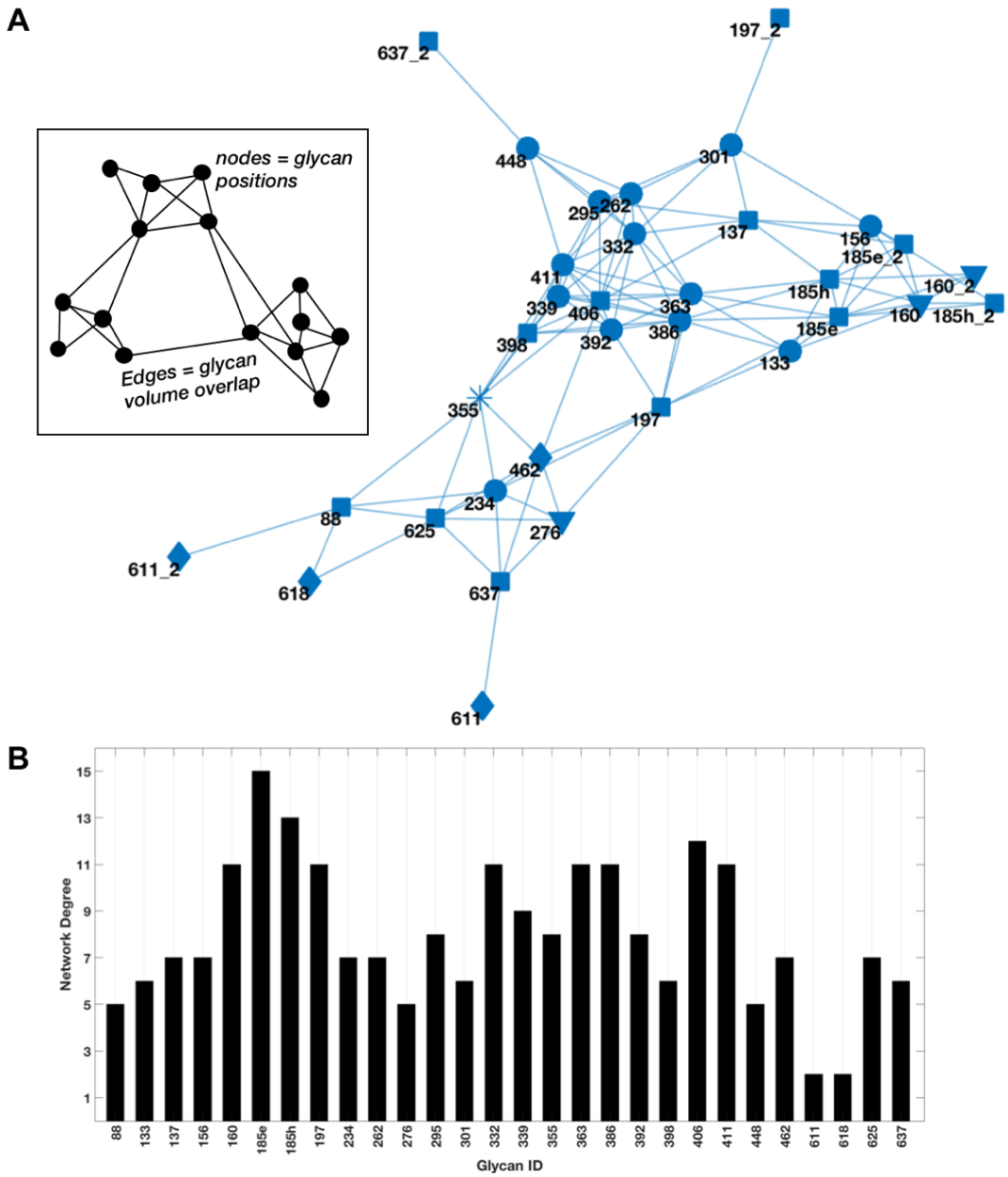



**Figure 5: Spatial network of BG505 native glycosylation.** (a) Visual representation of network in 2-dimensions, based on a force-directed layout. Each glycan forms a node-point on the graph, two nodes are connected by an edge (inset, from reference [38], Fig.8) if there is interaction as per Figure 4B and C, scaled by the fraction of overlap. (b) The network degree of each node or glycan, given by the number of other nodes it is connected to.

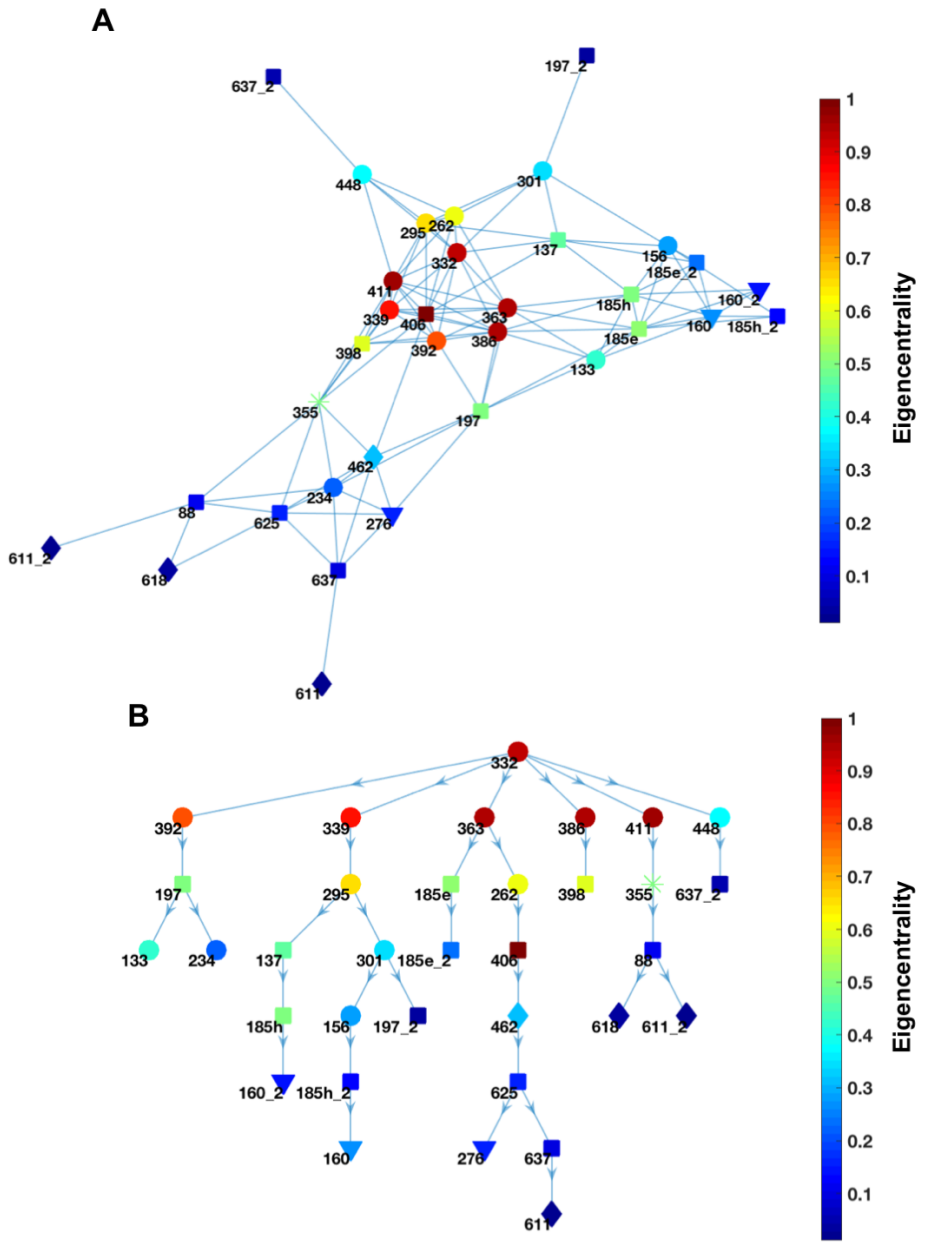



**Figure 6: BG505 native glycan network properties**. (a) Normalized eigenvector centrality of the glycans projected on the network. (b) Shortest path of communication between glycan N332 and all other glycans in the network.

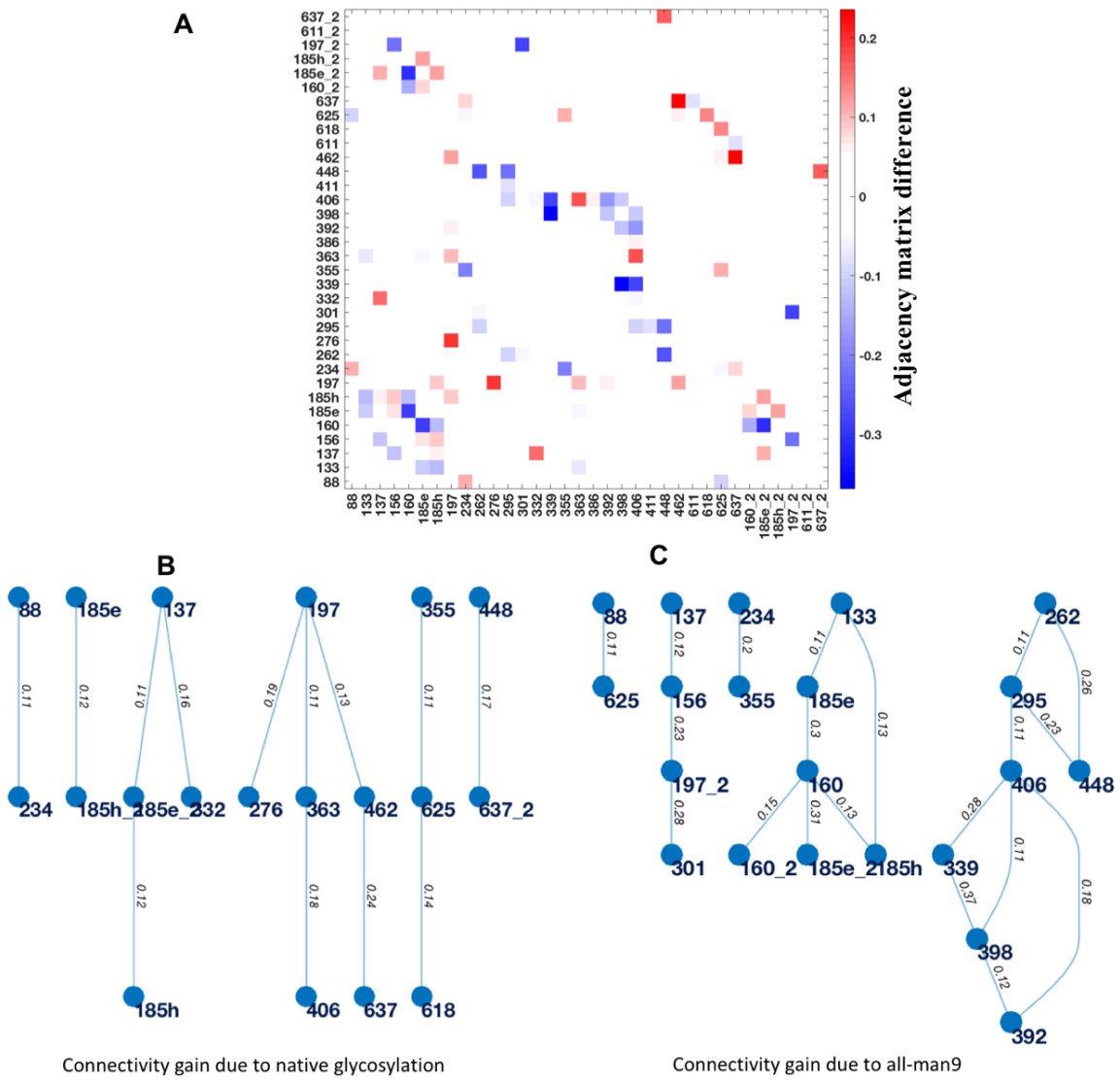
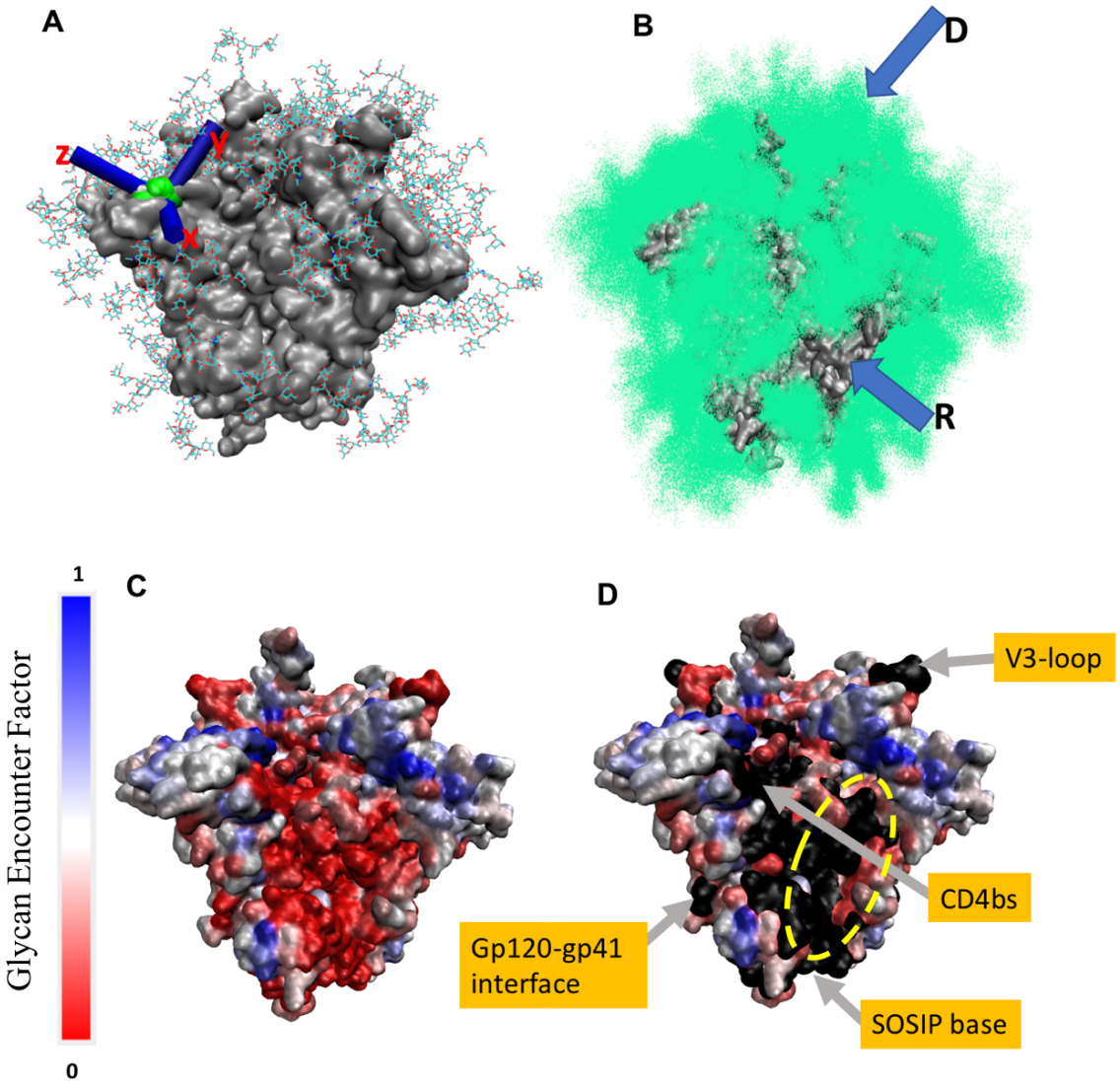
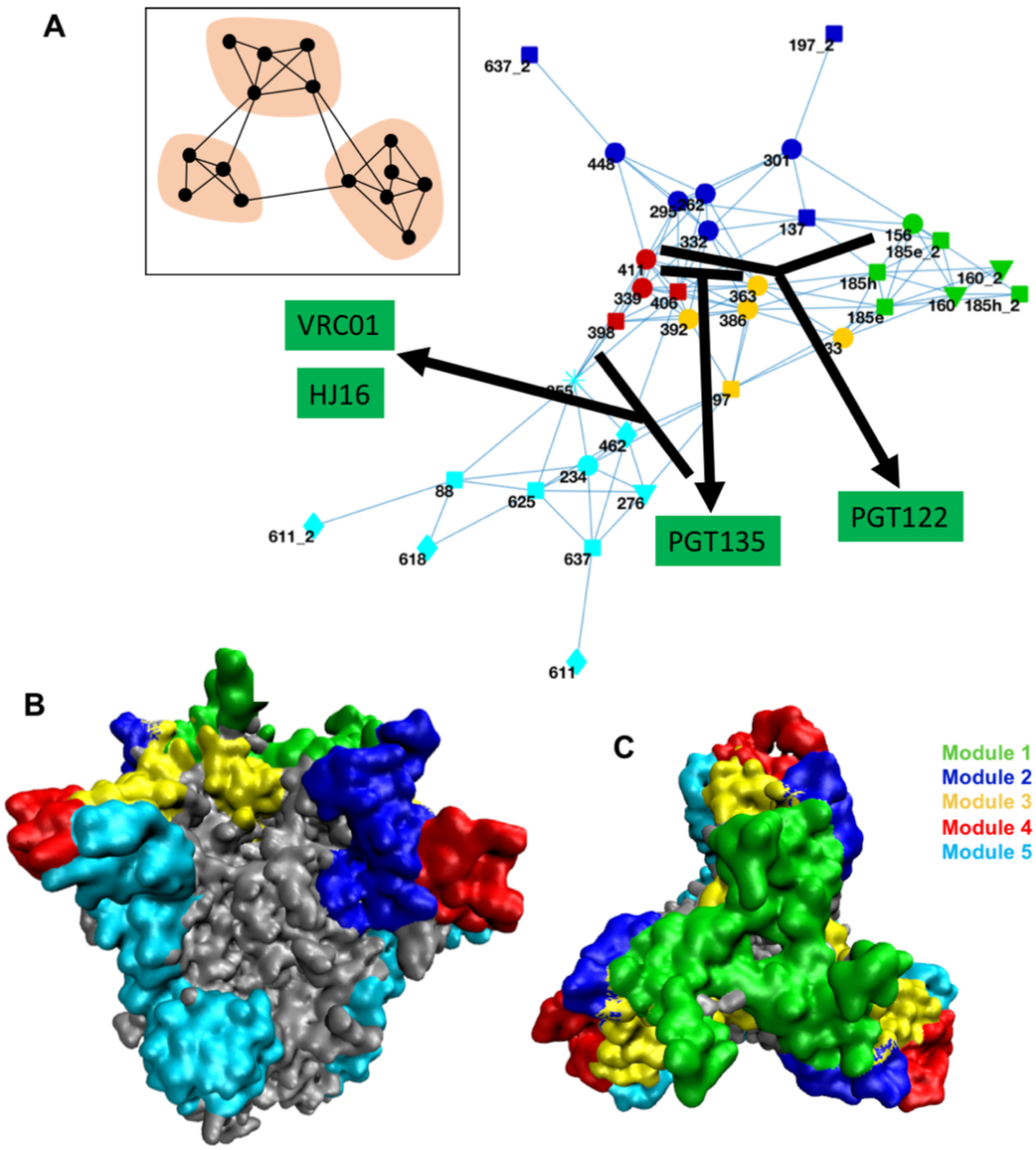



**Figure 7: Network difference between native and all-man9 glycosylation.** (a) Difference in adjacency matrices between the two models, - native minus man9. Blue color indicates at least 5% decrease in edge weight, and red indicates at least 5% increase in edge weight in native network, as compared to all-man9. (b) Increase and (c) decrease in connectivity due to native glycosylation in comparison with all-man9 model.




983

984 **Figure 8: Glycan Encounter Factor (GEF) for quantifying shielding effect.** (a) At each residue
985 present on the surface of the protein, the approaching probe is considered in three directions **x, y**
986 and **z.** (b) Any point on the surface which has a dense glycan covering, such as D, has a high glycan
987 encounter factor value, versus a point such as R where the glycan covering is sparse, which will
988 have a low GEF. (c) Representation of normalized GEF on Env surface, given by a colormap. (d)
989 Regions of BG505 surface having a GEF less than 1.5 (normalized GEF <0.14) is colored in black.
990 Typical known antibody epitopes are indicated by arrows. BG505-specific GH and COT epitope
991 region demarcated by yellow dashed circle.




992

993

994 **Figure 9: Subcommunities within network.** (a) 5 different subcommunities were identified
995 based on modularity maximization. The sub-community junctions identify susceptible regions in
996 the shield where antibodies tend to bind. (b) Location of each sub-community projected on the
997 Env surface, including the (c) top view.

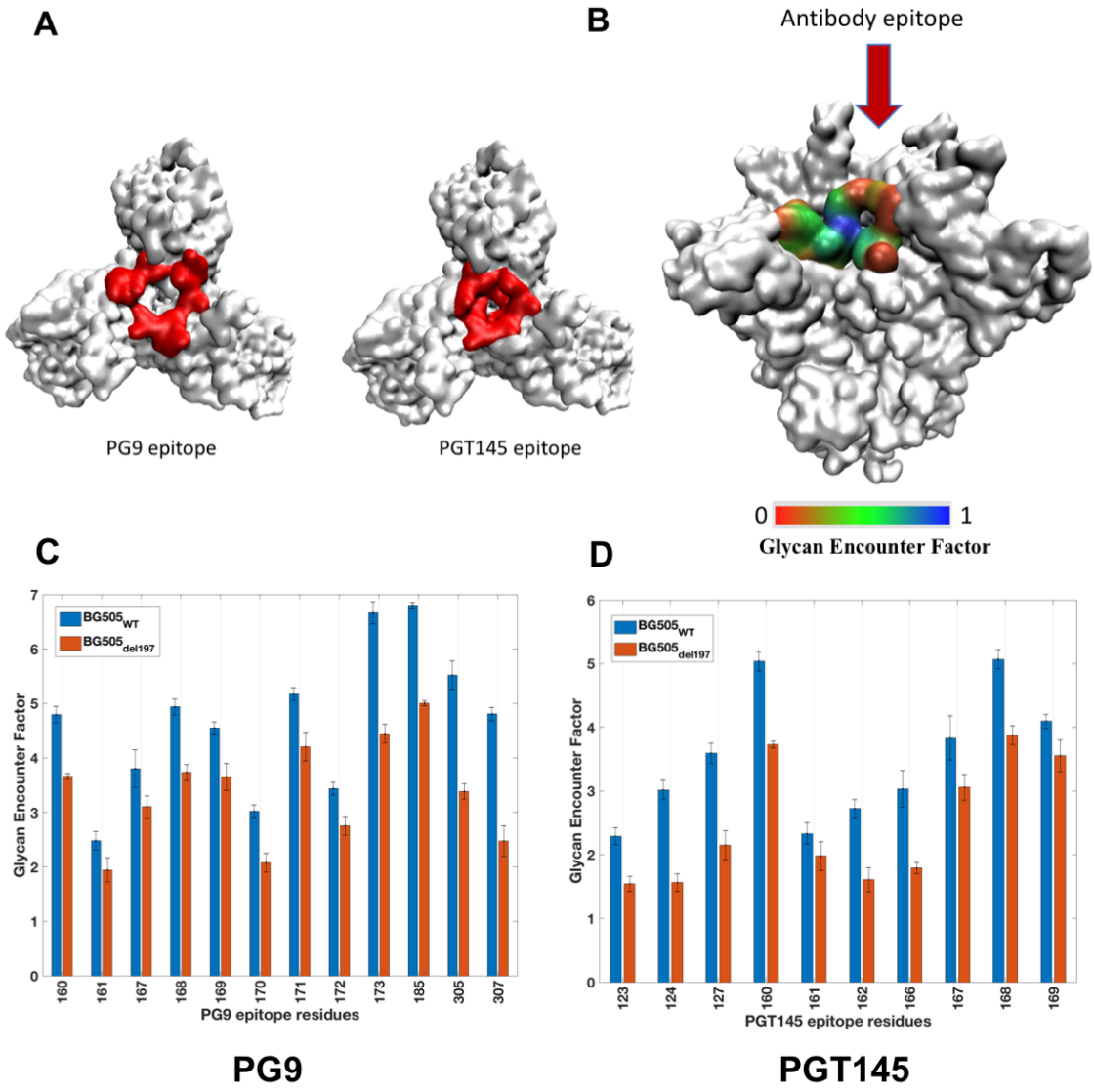



**Figure 10: Deletion of glycan N197 decreases glycan shielding at PG9 and PGT145 epitope.** (a) Top view of Env showing epitope regions of antibodies PG9 and PGT145 at the apex. (b) Footprint of glycan N197 on Env surface, colored by glycan encounter factor contributed by N197 alone. (c) Normalized glycan encounter factor over the PG9 and (d) PGT145 epitope residues. GEF decreases for both the epitopes due to deletion of N197 glycan.