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[3,4,6], 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

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 results which are physiologically more relevant than the previous models with uniform glycosylation. 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(Dennis et al., 1999; Dwek, 1996; Moremen et al., 2012). Additionally, envelope proteins from several high-risk viral pathogens which hijack the host protein production and glycosylation machineries, such as HIV (lentivirus)(Burton and Mascola, 2015), Coronavirus (Vankadari and Wilce, 2020; Walls et al., 2016), Lassa (arenavirus)(Sommerstein et al., 2015), Hepatitis C (flavivirus)(Zhang et al., 2017), Epstein Barr (herpesvirus)(Szakonyi et al., 2006), Ebola (filovirus)(Ilinykh et al., 2018; Lennemann et al., 2014), 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(Crispin et al., 2018; Doores, 2015; Karsten and Alter, 2017; Ringe et al., 2019; Stewart-Jones et al., 2016; Wagh et al., 2018). 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(Bradley et al., 2016; Haynes et al., 2014; Walker et al., 2010). 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; Julien et al., 2013;

65 Lyumkis et al., 2013; Pancera et al., 2014). Some bNAbs have evolved to include conserved
66 glycans as part of the epitope(Crispin and Doores, 2015; Falkowska et al., 2014; Mouquet et al.,
67 2012; Walker et al., 2009). Moreover, these surface glycans are also critical for Env folding, viral
68 assembly, and infectivity.

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

79 The extreme dynamics and conformational heterogeneity stemming from large variance of
80 possible constituent sugars, linkage, branching patterns, and rotamer flexibility, make the study of
81 these glycans immensely challenging(Imberty and Perez, 2000; Walsh, 2010). Traditional X-ray
82 crystallography methods are rendered ineffectual(Chang et al., 2007) since this heterogeneity
83 prevents proper crystallization of the glycoproteins. Inherent glycan flexibility also complicates
84 analysis by ensemble techniques like Nuclear Magnetic Resonance (NMR) and single-particle
85 electron cryomicroscopy (cryo-EM) where the signal from glycans is often too noisy for structural
86 interpretation at the resolutions necessary for building atomic structure. (Chang et al., 2007; Davis
87 and Crispin, 2010; Slynko et al., 2009; Woods et al., 1998). Within the ~300 experimental
88 structures of HIV-1 envelope glycoprotein in the Protein Databank (PDB) (Berman et al., 2000),
89 only a quarter of the total glycan content has been structurally resolved and are mostly in high
90 mannose form. Recently there have been a few cryo-EM and X-ray structures published of natively
91 glycosylated Env(Barnes et al., 2018; Gristick et al., 2016; Lee et al., 2016). Since most of these
92 structurally resolved glycans were stabilized by interaction with antibodies, they provide little
93 information about the structural and dynamic details of the individual and collective glycans.
94 Furthermore, all computational studies of HIV-1 Env glycosylation comprise of oligomannose

95 glycan moieties ranging from mannose-5 to mannose-9(Ferreira et al., 2018; Lemmin et al., 2017;
96 Liang et al., 2016; Stewart-Jones et al., 2016; Yang et al., 2017). The main drawback affecting the
97 quality of such MD simulations is the robustness of conformational sampling. Due to the
98 intrinsically dynamic nature of glycans, to effectively sample a biologically relevant energy
99 landscape of the glycoprotein, long and often multiple trajectories need to be run, preferably with
100 enhanced sampling techniques(Yang et al., 2017).

101 We had previously established a high throughput pipeline to robustly build atomistic
102 models of glycans, sample the glycan conformational space, construct the glycan network
103 topology, and extract molecular level description of the viral glycan shield (Berndsen et al., 2019).
104 We had also set up a protocol to experimentally validate the models by generating synthetic cryo-
105 EM datasets and comparing the resulting maps in a quantitative manner. Here, we have employed
106 a similar quantitative comparison using computational and experimental single-particle cryo-EM
107 data and show that the dynamics of high-mannose and native glycosylation remains practically
108 indistinguishable around the glycan stem β-mannose (BMA) residue. This provides further
109 rationale towards the need for computational modeling of complete glycans in a natively
110 glycosylated shield. Presently this seems to be the only way to elucidate the ramifications of such
111 complex sugars on the shield architecture. Therefore, we have built a glycosylated Env having
112 both oligomannose and complex glycans(Behrens et al., 2016; Cao et al., 2017) based on the
113 information provided by mass spectrometry (MS)(Behrens et al., 2016). To our knowledge, this is
114 the first time that a complete computational model of Env SOSIP was generated to include
115 processed glycoforms, thus obtaining results which are immunologically more relevant.

116 Using these atomistic models, we have employed graph theory to capture the glycan shield
117 topological network, pinpoint potential interaction pathways, and identify concerted behavior of
118 the glycans. The glycan-glycan and glycan-protein interactions influence the behavior of the shield
119 as a whole and can affect distant sites in the glycan network(Behrens et al., 2016; Ferreira et al.,
120 2018; Stewart-Jones et al., 2016). Analyses of various network attributes, such as relative centrality
121 of different glycans and critical subnetwork properties, have aided in detailed examination of the
122 native glycan shield in the context of bNAb interactions. Important global and local feature
123 differences due to native-like glycosylation pattern come to light as compared to all high-
124 oligomannose glycans, arising from the presence of charged sialic acids at the tips and fucose at

125 the base. Both the simulated models and cryo-EM maps identify a change in projection angle
126 relative to the peptide surface for fucosylated glycans. We have also been successful in quantifying
127 the glycan shield based on the number of sugar heavy atoms encountered over the antigen surface
128 that can be implemented to define regions of relative vulnerability and regions where the glycan
129 shield blocks access. Due to the rapid yield rate of this method, it can be applied to a large number
130 of diverse HIV-1 sequences or can be seamlessly extended to model glycan shields of other viruses
131 in a rapid manner.

132

133 **Results**

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

135 The soluble, recombinant BG505 SOSIP.664 (here on referred simply as BG505) trimer
136 has been well characterized as a native-like, Env-mimetic model, and serves as the prototypical
137 immunogen in several vaccine development programs(Sanders et al., 2013; Sanders et al.,
138 2015). For our study, we have used this system in the pre-fusion closed conformation, for which
139 several structures have been determined(Barnes et al., 2018; Gristick et al., 2016; Lee et al.,
140 2016; Stewart-Jones et al., 2016). A single protomer of this BG505 protein contains 28 N-linked
141 glycosylation sites. While all N-glycans share a common core sugar sequence, they are broadly
142 classified into three types **(Figure 1)**: (i) oligomannose, in which only mannose residues are
143 attached to the core; (ii) complex, in which differently branching "antennae" are attached to the
144 core; and (iii) hybrid, which is a mix of the first two types. Recent novel MS-based approaches
145 have been successful in identifying site-specific glycosylation profile in multiple HIV-1
146 trimers(Cao et al., 2017; Go et al., 2017). It is known that each glycosylation site at BG505 has
147 a distribution of multiple possible glycan species, depending on various factors, such as type of
148 cell lines, purification methods and sources, structural constraints, and so on. We used a site-
149 specific distribution of glycan type at each PNGS that was obtained by Behrens *et. al*(Behrens
150 et al., 2016) and Cao *et. al.*(Cao et al., 2017), to identify the most likely glycan species at each
151 site. The glycan species at each site was selected based on the highest relative abundance per
152 the MS studies (Figure 1 in reference (Behrens et al., 2016)), and is listed in Supplementary

153 Information (SI) **Table S1**. The structure schematic for each of the selected glycan species have
154 been shown in **Figure 1A**.

155 **High-throughput atomistic modeling of the native glycan shield**

156 In the static pictures as obtained from fully glycosylated structures(Gristick et al., 2016;
157 Stewart-Jones et al., 2016) and even single snapshots from computational models, a significant
158 fraction of the protein surface appears to be exposed, with each glycan taking up a particular
159 conformation (**Figure 1B**). However, due to their flexible and dynamic nature, these glycans
160 are not confined to a particular conformation, instead sampling a large volume in space. A
161 previous simulation study suggested(Guvench et al., 2011) that the root mean squared
162 deviations (RMSD) for the carbohydrate regions are more than 4 times larger than that of the
163 underlying protein loops. As a result of this conformational variability, the cumulative effect of
164 the glycans over time is like that of a cloud of glycan atoms that shield the underlying surface
165 from any approaching protein probe (an antibody for example) as illustrated in **Figure 1C**.

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

177 Due to the initial randomization of the glycan orientations, this integrated technique can
178 sufficiently sample a physiologically relevant conformational space accessible to carbohydrates in
179 a very short time, leading to the overall spatial shielding effect (**Figure 1C**). The torsion angle
180 distributions of 10 different inter-glycan linkages within the ensemble was compared with those

181 obtained from different glycan structures available in the PDB database using GlyTorsion
182 webserver(Lütteke et al., 2005) and is shown in **SI Figure S1**. These distributions match very well
183 between our generated ensemble and the PDB structures (see SI for details), indicating that the
184 modeled glycans sample a stereo-chemically relevant landscape. In order to understand the effects
185 just due to native glycans, we have also built a similar ensemble with mannose-9 glycans at all
186 glycosylation sites, (hereon referred as all-man9 model) for comparison.

Quantitative comparison of simulated models with cryo-EM data indicates glycoform-independent glycan stem dynamics

187 We generated simulated cryo-EM density maps for both the native and all-man9 ensembles
188 by utilizing the method that we established earlier to convert *in silico* atomistic models to synthetic
189 cryo-EM data(Berndsen et al., 2019). Then, we quantitatively compared them to the experimental
190 cryo-EM maps of BG505.SOSIP.664 from HEK 293F cells which produce a native-like glycan
191 shield including both oligomannose and complex glycans, and from HEK 293S cells which do not
192 produce complex glycans(Berndsen et al., 2019). The experimental and simulated maps show
193 similar features in the shield, with only the first few sugar residues at each site defined at high
194 resolution **(Figure 2A).** A scale-space analysis developed in our earlier study(Berndsen et al.,
195 2019) also shows a similar trend to the experimental data **(Figure 2B**). Progressive smoothing of
196 the map is accompanied by a large expansion in glycan volume that plateaus around 1.75-2 SD.
197 The effect of the glycans is best appreciated in comparison to a map generated from a non-
198 glycosylated ensemble **(Figure 2B - dashed line)**. This suggests the simulation is capturing native-
199 like dynamics in the glycan shield, at least globally. The 1.75 SD Gaussian filtered maps are
200 visualized at high and low thresholds in **Figure 2C**, revealing a similar overall shape and topology.
201 At the level of individual glycan dynamics, comparison of the normalized mean intensity at the
202 location of each BMA residue for the simulated map with that of the experimental BG505_293F
203 map reveals a strong positive correlation (**Figure 2D-E**), with a correlation coefficient of ~0.82
204 (p=6.82e-6). We also confirmed that local map intensity around BMA residues accurately captures
205 relative differences in Root Mean Squared Fluctuations (RMSF) at each glycan **(SI Figure S2)** for
206 the native ensemble.

207 An interesting outcome of the above comparison is that the correlation coefficients remain
208 similarly strong not only when we compare the all-man9 model to BG505_293S data, but also
209 when the native model is compared to high-oligomannose experiments and vice versa **(SI Figure**
210 **S2)**. Therefore, while these comparisons give credence to the validity of the model in accurately
211 capturing the relative position and flexibility of the glycan stems, they also indicate that the glycan
212 dynamics are practically identical within the observational limits of cryo-EM analysis between
213 complex and oligomannose glycans. The native glycosylation pattern on the shield topology only
214 manifest at the level of the flexible antennae beyond the stem and can therefore only be observed
215 in detail using computational models.

216 **Fucose-mediated changes in glycan orientation distinguish native glycan shield**

217 While the glycan stem shows indistinguishable dynamics at the BMA sugar, an important
218 difference between complete models of complex and all-man9 arise due to the presence of the
219 fucose ring at the core of the complex sugars (**Figure 3)**. By structurally aligning the first core
220 sugar and the three residues at the base of the glycan (n-1, n, n+1; where n is the glycosylated
221 asparagine residue) for all the structures in the ensemble, the oligomannose glycans are found to
222 be spaced symmetrically around the core, while complex sugars have a distinct bend away from
223 the side where the fucose ring is present **(Figure 3A)**. The change in projection angle of the
224 principal axis of the glycan relative to the protein surface is predicted to be ~10° in our models.
225 This difference in orientation preference, along with the presence of the negatively charged sialic
226 acid tips of the antennae, play notable roles that govern the differences between high oligomannose
227 and native glycosylation.

228 The change in projection angle from the fucose ring manifests as a reasonably large change
229 at the tips of the antennae, but the overall change in position at the core N-acetylglycosamine
230 (NAG) is too small to be captured in the high-resolution cryo-EM maps **(Figure 3B)**. However, a
231 close examination of the low-resolution difference map between the BG505_293F and
232 BG505_293S maps (Berndsen et al., 2019) does reveal some evidence for such changes in
233 projection angle at sites with complex glycans (**Figure 3C-E)**. For instance, the difference signal
234 around the glycans at N611 and N618 is asymmetric around the glycan stem and is therefore likely
235 not the result of differences in occupancy. The difference signal at N611 is located primarily on

236 the underside of the glycan stem, on the face opposite where we identified density likely belonging
237 to the fucose ring **(Figure 3D)**, which is suggestive of a change in projection angle similar to what
238 we observed in the simulations. In addition, examination of the glycan stem models built and
239 relaxed into the Gaussian filtered BG505_293F and BG505_293S maps used for assessing local
240 map intensity **(Figure 3E)** show a slight change in projection angle, further confirming this fucose-
241 mediated glycan reorientation.

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

243 Network analyses has historically been used to study protein allosteric frameworks and
244 evolutionary paths (Beleva Guthrie et al., 2018; Eargle and Luthey-Schulten, 2012; Huang et al.,
245 2017; Sethi et al., 2013; Skjaerven et al., 2014), and only recently begun to be applied to in
246 glycoprotein structural characterization (Lemmin et al., 2017). While each glycan exerts effects in
247 its immediate vicinity, due to the inter-glycan interactions, their influence can percolate over long
248 distances across the surface of protein(Berndsen et al., 2019). Long-range glycan interactions can
249 occur, as with perturbation of a glycan at one site can affect the processing and antibody
250 interactions of another glycan at a distant site(Behrens et al., 2016; Bricault et al., 2019; Doores,
251 2015; Stewart-Jones et al., 2016). We have previously shown that understanding long-range
252 interactions between glycans is important for characterizing global properties of the glycan
253 shield(Berndsen et al., 2019).

254 Now, we implement the graph-theory based approach to native glycan shield. Details of
255 modeling the network are recapitulated in the Methods section. The natively glycosylated BG505
256 trimer, like all HIV-1 Envs, has a highly dense glycosylation pattern and each glycan samples a
257 particular region in space (**SI Figure S3A-B**), such that neighboring glycans can occupy
258 overlapping regions (**Figure 1D**). The fraction of volume overlap gives a measure of the
259 interaction probability between the constituent glycans (**SI Figure S3C)**. Within a protomer there
260 are three main regions of overlap – the apex, the gp41 base, and the glycan-dense central high-
261 mannose patch (HMP). Inter-protomer overlaps are primarily due to V1 and V2 loop glycans near
262 the trimer apex. **Figure 4A** shows the obtained network of BG505 native glycosylation unfolded
263 and laid out in two dimensions for the ease of visualization. This is a force-directed layout, which
264 uses attractive forces between adjacent nodes and repulsive forces between distant nodes to reach

265 an optimum distribution of the node points in space(Fruchterman and Reingold, 1991). For
266 conceptualization, the network relative to the BG505 structure is provided in **SI Figure S3D.**

267 The nodes in the central region around the V4 loop, in the HMP, are very highly inter-
268 connected. Glycans in the V1, V2 apex are also reasonably well-connected, but the connections at
269 the base near gp41 are relatively sparse, both within the locality, and globally with the rest of the
270 network. Glycans at positions 234, 355, 462 and 276 connect the sparse base region of the network
271 with the dense apex and central regions. Glycan 156, 160, and those in the V2 loop, enable inter-
272 protomer glycan interactions at the apex, stabilizing the protein quaternary trimer
273 structure(Berndsen et al., 2019). Glycan 197, crucial in the binding of CD4bs and V3-specific
274 antibodies(Li et al., 2008; Liang et al., 2016), connects the central crowded region of the network
275 between neighboring protomers. gp41 glycans 611 and 637 also contribute to inter-protomer
276 pathways, with glycan 637 communicating with glycan 448 directly in the central mannose patch.

277 **Utilizing the glycosylation network to extract functionally relevant topological features**

278 The native glycosylation network is well-connected as indicated by the network degree, or
279 the number of connections a node has to other nodes (**SI Figure S3E**). Due to the high structural
280 fluctuations of the V2 loop, the glycans on and flanking V2 can interact with a number of other
281 neighboring nodes, having the highest degrees of connectivity. There is dense crowding of glycans
282 around the high mannose patch with a number of inter-connections, resulting in the glycans of this
283 patch being less processed because of reduced accessibility to enzymes(Coss et al., 2016; Doores
284 et al., 2010; Pritchard et al., 2015a). The network degree values are relatively lower around the
285 base of the protein on and around gp41, reflective of the sparse architecture of glycan topology in
286 this region. The site N332 is central in the glycan interaction network, and is known to have both
287 direct and subtle long-range influences over a number of bNAb sites(Behrens et al., 2016; Bricault
288 et al., 2019). We calculated the shortest paths to all glycans from the N332 glycan site using the
289 Floyd-Warshall algorithm(Floyd, 1962; Warshall, 1962), where the glycan distances were edge-
290 weighted (**Figure 4B)**. Our calculation shows the most probable pathway over which each glycan
291 feels the influence of the glycan at site N332, providing useful insights into how these inter-glycan
292 communications can occur.

293 To elucidate the relative influence of each of the glycans on the network, we calculated the
294 eigenvector centrality, or eigencentrality, of the nodes to measure its connectivity to the network.
295 A large relative value indicates that the node is well connected to the network and thus is
296 "centrally" located, whereas a low relative value indicates that the node is on the periphery of the
297 network. Operationally, the eigencentrality is the eigenvector associated with the largest
298 eigenvalue of the adjacency matrix, that in position (i,j) reports the overlap of the ensemble of
299 glycan at sites i and j. Value of that particular matrix element will be zero if the glycans in position
300 i and j do not physically interact (details given in Methods section). The normalized eigencentrality
301 of the glycans are projected on the network as a colormap in **Figure 4A**. The eigencentrality
302 increases towards the middle of the graph, with the glycans at the crowded HMP with a large
303 number of connections between each other having the highest centrality values. When we increase
304 the threshold of volume overlap needed to form an edge, those glycans with the least centrality
305 leave the network first. Glycan 611 is the first to be eliminated, succeeded by some of the inter-
306 protomer interactions and the other gp41 glycans, following the relative centrality scores. The
307 HMP with high eigencentrality values persists throughout even with high overlap thresholds. We
308 have previously confirmed for all-man9(Berndsen et al., 2019) that successive enzymatic digestion
309 of glycans from Env loosely follows a pattern that matches with the network centrality. Those
310 glycans which are sparsely connected to each other, having lower centrality, are eliminated earlier
311 during the process of endoH digestion. On the other hand, the glycans having higher network
312 centrality, such as those in the HMP, takes longer to be eliminated by the digestive enzymes. Thus,
313 the eigenvector centrality provides a measure of the crowding of the highly central glycans which
314 makes them difficult for endoglycosidase to access.

315 **<u>Influence of complex glycans on overall network topology</u>**

316 The degree of connectivity decreases almost throughout the network for the native
317 glycosylation (**SI Figure S4A**) compared to all-man9. The number of stable glycan-glycan
318 interactions at several glycan sites decreases because spatially proximal glycans avoid unfavorable
319 charge-charge interactions. For example, the uncharged glycans 234 and 355 takes up a position
320 central to all the neighboring charged sugars, increasing the distances between them, and the
321 charged N406 glycan buries itself in the middle of the surrounding high-mannose glycans (**Figure**
322 **4A**), clustering together with them. In fact, removal of this charged glycan N406 glycan can

323 increase the processing of the neighboring high-mannose moieties(Cao et al., 2018). Further
324 evidence for such alterations in glycan connectivity due to the presence of complex glycans can
325 be observed in the comparisons of the RMSF variations. A detailed analysis has been presented in
326 the SI and illustrated in **SI Figure S5**. Combination of the network node location differences and
327 RMSF variations suggest that the presence of the charged sialic acid at the tips of these complex
328 glycans can dictate the interactions between neighbors, increasing their structural variations if
329 surrounded by other charged glycans, or reducing them if a stable conformation buried between
330 uncharged high-mannose patches can be found. This overall decrease in connectivity slightly
331 increases the diameter of the native network, increasing the mean number of hops for the shortest
332 path from 2.4 hops (and distance 0.19) in the all-man9 model to 3 hops (and distance 0.27).

333 The distribution of centrality also shifts between the two glycosylation models (**SI Figure**
334 **S4B**). The centralities of the glycans at 137, 234, 355, 398, 406, 462 and those at gp41 increase
335 due to native glycosylation. On the other hand, the V1 and V2 loop glycans as well as 262, 295,
336 363 decrease from the all-man9 model. **Figure 5A** illuminates the difference in adjacency matrices
337 between the two models, with blue color indicating decrease in edge weight, and red indicating
338 increase in edge weight in native network, as compared to all-man9. Overall connectivity goes
339 down in the apex and the central patch for the native, and increases within gp41 glycans, as
340 compared to all-man9. The differences between the two networks (with at least 10% edge-weight
341 change) are shown in **Figure 5B** and **C**. The pathways around 197, as well as the inter-protomer
342 interactions involving 156, 185e and 637 become stronger in native glycosylation. (**Figure 5B**).
343 Some of these 197 connections, such as those with 276 and the V5 loop glycans occur across the
344 CD4 binding site. It was previously shown that in predominantly high-mannose Env structures,
345 CD4bs targeting antibody VRC01 has very little interaction with glycans 197 and 276(Stewart-
346 Jones et al., 2016). Conversely, in another Env structure with fully processed native glycans, a
347 VRC01-like antibody called IOMA interacted extensively with both 197 and 276(Gristick et al.,
348 2016). This matches our observation of increased orientation of glycan 197 over the CD4bs in
349 native glycosylation pattern. However, the paths connecting 137, 156, 301, and 197 (of
350 neighboring protomer) are significantly weakened in the native model (**Figure 5C**). Subnetworks
351 involving V4 loop glycans and some of the central mannose patch become weaker due to the
352 presence of native glycans.

353 **Community analysis to determine breaches in glycan shield**

354 In order to interact and neutralize the Env protein, antibodies look for breaches in the shield
355 where glycan densities are sparse. This information of relative sparsity is encoded within the
356 glycan network, in terms of edge clustering. We examine this clustering of glycan interactions
357 within the network by defining communities using a modularity maximization approach that
358 divides the network into sub-modules or groups (see Methods). Communities have high density of
359 edges between the nodes within a module and comparatively sparse edges connecting different
360 modules (**Figure 5D inset**). This analysis identifies five distinct communities within the BG505
361 native glycan network (**Figure 5D and E**). The apex glycans around loop V2 from the three
362 protomers together form a single community (1, green). Right below that, glycans 262, 295, 301,
363 332 and 448 (2, blue) forms a community around the V3 and alpha 2 helix. Glycans 133, 137,
364 197, 363, and 386 (3, yellow) and 339, 392, 398, 406 and 411 (4, red) form two distinct
365 communities involving the glycans on and surrounding V4 loop region. Rest of the glycans 88,
366 234, 276, 355 and 462 from gp120 and all four glycans from gp41 form the fifth community (5,
367 cyan), though the modularity value is lower due to sparser connections. The possibility of an
368 approaching probe reaching the protein surface through these strongly connected communities is
369 low. A similar study clustering glycans from microsecond simulation runs of BG505 SOSIP was
370 performed by Lemmin *et. al.*(Lemmin et al., 2017), identifying 4 glycan microdomains that
371 roughly correspond to our modules 1, 2, (3+4) and 5. However, in that study, mannose-5 glycoform
372 was used at all sites, and due to the smaller length of these glycans, some of the interactions,
373 including inter-protomer interactions were not observed. Remarkably, in that study, junctions
374 between microdomains were found to indicate regions of relative vulnerability. The communities
375 we identified, also demarcate the regions where the glycan shield can be penetrated. Broadly
376 neutralizing antibodies whose binding epitopes are known target these community boundaries
377 (**Figure 5D**). Thus, glycan community dynamics can help to determine susceptible regions of the
378 glycan shield, and can be further used for guiding immunological studies.

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

380 To capture the impact of the glycan shield acting as an immunological barrier, we
381 quantified the sugar barrier over the Env protein surface using the ensemble of structures that have

382 been generated. Towards this goal, we have defined the glycan encounter factor (GEF) as the
383 number of glycan heavy atoms encountered by an external probe approaching the protein surface.
384 This parameter is calculated at each residue on the surface of the protein within a probe of diameter
385 6Å calculated using our ensemble. The highest value obtained is 14 for the native model, and is
386 located in the HMP. Based on existing structures, the main interaction points between Env and
387 bNAbs are often hairpin-like loop regions. Even large-scale atomistic simulations suggest that the
388 first line of contact between Env and an antibody is through a loop region(Schmidt et al., 2013).
389 Accordingly, for our analysis, we have used a probe size of 6Å, which is the typical diameter of a
390 hairpin loop. At each residue present on the surface of the protein, the approaching probe was
391 considered in three directions (**Figure 6A**): perpendicular to the surface **z**, and then the **x** and **y**
392 directions spanning the plane parallel to the surface. The geometric mean of the **x**, **y** and **z** values
393 provide the GEF per residue on the Env surface. This value will go to zero when the glycan
394 encounter factor is zero from any one of the three cardinal directions. Thus, for any point on the
395 surface which has a dense glycan covering, such as D in **Figure 6B**, has a high glycan encounter
396 factor value, versus a point such as R where the glycan covering is sparse, which will have a low
397 GEF.

398 Previous evidence suggests that 70% of the Env ECD surface area is covered by
399 glycans(Pancera et al., 2014). Based on this, we determined the lower cut-off of GEF below which
400 we can define glycan holes. **Figure 6C** shows GEFs mapped onto the trimer protein surface. The
401 calculated solvent-exposed surface area of the protein part of our modeled BG505 structure
402 without considering glycans is 86,055 $Å^2$. Excluding the exposed region at the base of the soluble
403 SOSIP, this reduces to 79,672 $Å^2$. Varying the lower cut-off of GEF, we find that a cut-off of 1.5
404 GEF leads to 30% of the surface to be exposed. Regions of BG505 surface having a GEF less than
405 1.5 is colored in black in **Figure 6D**. From the figure, it is clear that the typical glycan holes
406 targeted by bNAbs in BG505, such as the CD4 binding site, the V3-loop epitope and the fusion
407 peptide binding region fall below this GEF cut-off. GEF tracks with epitopes that are relatively
408 generic to a broad range of Env strains(Burton et al., 2012; Xu et al., 2018).

409 Antibodies elicited by BG505.SOSIP.664 are mainly biased towards the missing 241 and
410 289 glycan holes (GH) and the cleft-of-trimer (COT) epitope regions as demonstrated by cryo-EM
411 and immunogenicity assays(Bianchi et al., 2018; McCoy et al., 2016) (see Figure 1 in reference

412 (Bianchi et al., 2018)). The GEF parameter identifies these BG505-specific epitopes as breaches
413 in the glycan shield (**Figure 6B**, "R" region; **Figure 6D** yellow dashed circle). On the other hand,
414 the densely glycosylated regions around V2, V4 loop and alpha 2 helix have high values of GEF.
415 At each point, the GEF value is given by a combination of all glycans in the vicinity that can come
416 in the way of the approaching probe. At the same time, we can analyze the extent of influence of
417 each glycan on the protein surface. Thus, we are now equipped with a tool to quantify the barrier
418 effect of the glycans individually or as a group for immunogen design, as further
419 demonstrated(Bianchi et al., 2018; McCoy et al., 2016) in the following section.

420 **Combining network topology and glycan encounter factor to inform on local and global**
421 **effects of neutralization**

422 The proposed network is useful in deciphering the impact of addition or deletion of glycan
423 on neutralization. Removal of the high conserved glycan at 197 by mutating the sequon leads to
424 enhanced neutralization sensitivity to a variety of CD4bs and V3-specific antibodies(Huang et al.,
425 2012; Li et al., 2008; Townsley et al., 2016). This glycan is situated proximal to the CD4 binding
426 site, and the tip of V3 loop, directly affecting the binding of these antibodies. However, past
427 experimental evidence also suggests that the deletion of this glycan at N197 increases the binding
428 affinity of two antibodies PG9 and PGT145 (Behrens et al., 2016) which target the trimer apex of
429 Env. The epitope of PG9 as determined from the PDB structure 5VJ6 include residues 160, 161,
430 167-173, 185, 305 and 307 (**Figure 7A**) and the epitope of PGT145 as determined from PDB
431 structure 5V8L include residues 123, 124, 127, 160-162, and 167-169 (**Figure 7A**). The footprint
432 of glycan 197 as per our GEF model is shown in **Figure 7B**. Residues of the antibody epitopes do
433 not overlap with those regions directly covered by glycan 197. Yet, removal of glycan 197 results
434 in significant reduction in the glycan encounter factor over the V2 antibody epitope as evidenced
435 in **Figure 7C** and **D**.

436 The removal of glycan 197 affects the glycans that were originally acting as barriers over
437 the epitopes to resample the available space in such a way that the barrier over the epitopes is now
438 reduced. Glycans 156, 160, 185e and 185h from the neighboring protomers directly shield the PG9
439 and PGT145 epitopes. Looking at the shortest paths of communication between residue 197 from
440 either of the protomers to all other glycans (**SI Figure S6A, B**) demonstrate that while 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 **SI Figure S7A**. 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 (**SI Figure S7B)**. 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 (**SI Figure S7C, D)**.

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. While 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 method 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 can add new perspective to current HIV vaccine design efforts.

Previous computational studies have generalized glycan moieties as simple oligo-mannoses such as mannose-5 and mannose-9 (Ferreira et al., 2018; Lemmin et al., 2017; Liang et al., 2016; Stewart-Jones et al., 2016; Yang et al., 2017) 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. While the overall aspects of the network are consistent between mannose-9 and native-like glycosylation, there are critical differences. 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

469 microenvironment, ultimately influencing the shield topology. Thus, our comprehensive
470 characterization of the shield is capable of capturing individual glycan effects which are
471 physiologically and immunologically more relevant.

472 Our network-based approach illuminates the collective behavior of glycans. We compare
473 the relative centrality of glycans, identify potential interaction pathways and detect stable
474 communities. The centrality or importance of glycans correlates well with experimental cryo-EM
475 data(Amanat et al., 2018; Berndsen et al., 2019). Glycans with lower centrality have lesser
476 influence on the graph, and are the first to be eliminated from the network, if the adjacency
477 threshold is increased. Similarly, the most central region of the network is the most resilient to
478 enzymatic action. Such centrality measures can help determine the ease of targeting glycans and
479 their modifications, guiding the process of immunogen development in the context of distinct
480 Envs. While the shield resilience is the main consideration for epitope exposure, some antibodies
481 have also evolved to capitalize on specific interactions with a number of the conserved glycans
482 during their molecular engagement with the epitope. Due to the stability of glycan interactions
483 within the detected communities, this help determine the antibody angle of approach that has been
484 shown to influence the breadth of bNAbs (Moore and Williamson, 2016).

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

495 The method we developed here for the structural modeling of the glycoprotein atomistic
496 ensemble and the subsequent development of the network is high-throughput compared to
497 traditional sampling by MD simulations. Due to the ease of this fast and efficient pipeline we are

498 now equipped with a tool to perform comparative structural studies due to glycan additions,
499 deletions or modifications, as well as other variations in the Env sequence. This is important to
500 understand the evolution of the glycan shield over longitudinal sampling of lineages. The structural
501 basis of addition or removal of glycans that are known to drive antibody maturation and
502 neutralization activities(LaBranche et al., 2018; Umotoy et al., 2019) can also be easily
503 investigated. Capitalizing on the relatively low computational overhead, our pipeline can be
504 integrated into a polyclonal epitope mapping assay to track the glycan shield as a function of
505 hierarchical antibody response, by modeling the shield network with the presence of different
506 antibodies and observing how the topology is altered.

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

521 In summary, HIV and the human immune system are at war constantly, and the virus uses
522 the Env glycan shield to mask the human immune surveillance. One battlefront in this war is the
523 glycan shield: while the virus evolves to develop resilience, the immune response counteracts by
524 looking for vulnerabilities. As efforts are underway to aid the immune system overcome this race
525 by conditioning it with engineered immunogens, there is a need to quantify the resistance and
526 vulnerability of glycan shield in a more quantitative manner. This is that first time that the native
527 glycan network and shielding has been spatially quantified. Our derivation of the Glycan

528 Encounter Factor measures the relative barrier over the Env protein surface, and can potentially
529 aid to distinguish subtle differences on the shield due to variations in the glycosylation or even
530 protein sequences. We are therefore armed with a set of validated *in silico* tools with which to help
531 the anti-HIV war efforts and guide immunogen design.

532

533 # STAR Methods

534 ## Contact for Reagent and Resource Sharing

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

538 ## Method Details

539 **High-throughput conformational modeling**

540 The ensemble of BG505 glycoprotein 3D conformations were built in atomistic detail by
541 implementing the ALLOSMOD(Guttman et al., 2013; Weinkam et al., 2012) package of
542 MODELLER(Eswar et al., 2006; Sali, 1995) [RRID: SCR_008395]. For ease of understanding,
543 the modeling protocol has been described as a flow schematic in **SI figure S8.** The protocol
544 consists of the following steps:

545 (a) ***Base template selection for protein scaffold:*** The BG505 protein scaffold was homology
546 modeled, by threading the protein sequence into available crystal structure templates.
547 Ligand-free SOSIP structure with PDB accession ID 4ZMJ(Kwon et al., 2015) was used
548 as the base template for gp120, that also guided the three-fold symmetry of the trimer.
549 5CEZ(Garces et al., 2015) was used as the base template for gp41 and since it has the least
550 number of missing residues among the available structures.
551 (b) ***Identifying flexible regions from available experimental PDB structures to guide***
552 ***conformational sampling:*** There are a large number of unstructured flexible regions in the

553 protein. In order to incorporate the fluctuations as dictated by the available structural data
554 we devised the following protocol. We aimed to construct 10 distinct protein models for
555 these flexible regions as starting scaffolds for glycan building. In order to model these, 34
556 different BG505 SOSIP experimental structures were selected below a resolution cut-off
557 of 4Å (see **SI Table S2**). Open-like, chimeric and synthetic models were not considered.
558 The set of 34 were structurally aligned using PyMOL(Schrödinger, 2015) and their residue-
559 wise backbone RMSF was calculated (**SI Figure S9A**). The base-level of 0.1 was used as
560 a threshold in the normalized RMSF measure. The unstructured residues (coils and turns)
561 were identified on 4ZMJ and 5CEZ using the STRIDE program(Frishman and Argos,
562 1995). All the identified regions with RMSF higher than the threshold (**SI Figure S9A**)
563 were considered for modeling based on multiple templates. These include residue ranges
564 64 – 82, 133 – 153, 184 – 189, 298 – 303, 312 – 314, 324 – 329, 396 – 413, 425 – 431, 457
565 – 464, and 609 – 618 (in HXB2 numbering).
566 (c) ***Template selection for identified flexible regions:*** Some of these identified regions have
567 missing residues in many of the PDBs (**SI Figure S9B**). Wherever enough models were
568 available, 10 distinct conformations were randomly selected as templates. If this number
569 10 could not be reached for some regions (for example, in loops V2 and V4) due to missing
570 residues (**SI Figure S9B**), those residues were modeled *ab initio*. All known disulphide
571 bonds were kept as additional restraints. We made sure these short-listed regions for *ab*
572 *initio* modeling were less than MODELLER-prescribed 14 residues in length. This is
573 because, modeling of such long loops are usually considered to be error-prone(Fiser et al.,
574 2000). The loop region 546-569 in gp41 is present only in the structure 5CEZ. We did not
575 model this region *ab initio*, and used 5CEZ as the only template for all models. The 10
576 models thus built were further tested for proper stereochemistry, based on MODELLER
577 optimization scores and PROCHECK scores(Laskowski et al., 1993). This results in
578 different starting orientations of the hypervariable loops, within the experimentally
579 determined limits of fluctuations.
580 (d) ***Modeling of glycans and their ensemble sampling:*** For each of the 10 selected protein
581 structures, glycans were initially added with random orientation, at the known
582 glycosylation sites, based on ideal geometries as dictated by CHARMM36(Best et al.,
583 2012; Huang and MacKerell Jr, 2013) force field internal coordinates, followed by a 1Å

584 randomization added to the overall atomic coordinates as described by Guttman et.
585 al.(Guttman et al., 2013). Once all the glycans were added, ensuing refinement steps of the
586 glycoprotein system optimized an energy function given by a combination of template-
587 based spatial restraints, CHARMM36 forcefield terms, and a soft sphere-like truncated
588 Gaussian term to prevent collisions. The structures were relaxed with 1000 steps of
589 conjugate gradient minimization followed by a short molecular dynamics equilibration of
590 500ps. Further refining with five rounds of simulated annealing was performed between
591 1,300K to 300K in 8 steps. The glycans and the loop regions were kept flexible during the
592 refinement steps. 100 fully glycosylated structures were modeled from each of the 10
593 selected protein models, resulting in the final ensemble containing 1000 different poses.
594 Template restraints were removed from the hypervariable regions with glycans in order for
595 them to sample a wider range of conformations during the glycoprotein relaxation phase.
596 These residues by HXB2 numbering are as follows: 143 to 152 (V1 loop), BG505-specific
597 insert residues 185A to 185I and 186 to 189 (V2 loop), 309 to 315 and 325 to 329 (V3
598 loop), 400 to 410 (V4 loop) and 458 to 464 (V5 loop) as determined from the LANL HIV
599 database (https://www.hiv.lanl.gov/).

600 **Glycan root mean square fluctuations and volume calculation**

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

604
$$RMSF_n = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{M}\sum_{m=1}^{M}|\vec{r}_{mnk} - \langle\vec{r}_{nk}\rangle|^2}$$

605 where $\vec{r}_{mnk}$ is the atomic position of heavy atom *k* of glycan *n* in snapshot *m*, $\langle\vec{r}_{nk}\rangle =$
606 $(\frac{1}{M})\sum_{m=1}^{M}\vec{r}_{mnk}$ is the average atomic position of heavy atom *k* in glycan *n*. K is the total number
607 of heavy atoms in the glycan. It is 127 for mannose-9, and varies depending on the type of glycan.
608 The ensemble for each model contains 1000 snapshots, making $M = 1000$ snapshots for each of

609 the two models. The standard deviations were obtained by dividing the 1000 snapshots into 4 sets
610 of M=250, and calculating the four sets of RMSF values **(SI Figure S5).**

611 The 1000 structures of the total ensemble are built from 10 initial starting protein
612 conformations, as described above. The main difference between these 10 conformations are the
613 variations in the loop regions, due to the missing residues in the templates. In order to remove the
614 effects of the loop fluctuations and consider the RMSF coming from the glycans alone, the reduced
615 RMSF values were also calculated in each of these 10 sub-models, and their average and s.d.
616 calculated (**SI Figure S5B**). The RMSF difference between the models **(SI Figure S5E)** were
617 obtained by subtracting the reduced RMSF values of all-man9 from native model (native minus
618 all-man9). Since, the average reduced RMSF value is ~4Å, only those RMSF difference values are
619 statistically significant which are above 0.2 Å, which corresponds to a p-value of 0.05, rejecting
620 those values within the null hypothesis. Ensemble of individual glycans were converted to atomic
621 densities at 1 Å gridsize in order to calculate the volume. Volume calculation was done using
622 CHIMERA package(Pettersen et al., 2004).

623 **<u>Glycan overlap and network analysis</u>**

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

638 and network analyses were performed using Python(Rossum, 1995) [RRID: SCR_008394] and
639 Matlab_R2018a packages(MATLAB, 2018) [RRID: SCR_001622].

640 **Eigenvector centrality calculation:**

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

644
$$x_v = \frac{1}{\lambda} \sum_{t \in N(v)} x_t$$

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

648
$$x_v = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

649 This has the form of the eigenvector equation **Ax**= λ**x.** With the added constraint of the
650 eigencentrality values needing to be non-negative, by the Perron-Frobenius theorem(Pillai et al.,
651 2005), the eigenvector corresponding to the largest eigenvalue gives the desired measure of
652 centrality. The eigenvector is a unit vector and therefore the centrality values add up to one. For
653 the purpose of this work, we have normalized the centrality values with respect to the node with
654 the highest centrality assigned at 1, to obtain the relative centrality values.

655 **Modularity maximization for community detection**

656 Community detection within the glycan network was performed using the modularity
657 maximization approach given by Newmann and Girvan(Newman, 2006; Newman and Girvan,
658 2004). Modularity Q is calculated as the difference between the fraction of edges that fall within
659 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].

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials, and further deposited in the library:
http://dx.doi.org/10.17632/yds8rtfjjy.1
Additional data related to this paper may be requested from the authors.

Supplemental Information

Supplemental Information includes details of torsion angle distributions, flexibility and volume sampling analysis of modeled glycan structures, along with 2 Tables and 9 Figures.

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

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

690

691 **Author Contributions**

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

696

697 **Competing Interests**

698 The authors declare no competing interests.

699

700 **References**

701 Amanat, F., Duehr, J., Oestereich, L., Hastie, K.M., Ollmann Saphire, E., and Krammer, F.
702 (2018). Antibodies to the Glycoprotein GP2 Subunit Cross-React between Old and New World
703 Arenaviruses. mSphere. 3(3). DOI: 10.1128/mSphere.00189-18.
704 An, H.J., Froehlich, J.W., and Lebrilla, C.B. (2009). Determination of glycosylation sites and
705 site-specific heterogeneity in glycoproteins. Curr Opin Chem Biol. 13(4), 421-426. DOI:
706 10.1016/j.cbpa.2009.07.022.
707 Barnes, C.O., Gristick, H.B., Freund, N.T., Escolano, A., Lyubimov, A.Y., Hartweger, H., West,
708 A.P., Jr., Cohen, A.E., Nussenzweig, M.C., and Bjorkman, P.J. (2018). Structural
709 characterization of a highly-potent V3-glycan broadly neutralizing antibody bound to natively-
710 glycosylated HIV-1 envelope. Nat Commun. 9(1), 1251. DOI: 10.1038/s41467-018-03632-y.

711 Behrens, A.J., Vasiljevic, S., Pritchard, L.K., Harvey, D.J., Andev, R.S., Krumm, S.A., Struwe,
712 W.B., Cupo, A., Kumar, A., Zitzmann, N., et al. (2016). Composition and Antigenic Effects of
713 Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. Cell Rep. 14(11), 2695-
714 2706. DOI: 10.1016/j.celrep.2016.02.058.
715 Beleva Guthrie, V., Masica, D.L., Fraser, A., Federico, J., Fan, Y., Camps, M., and Karchin, R.
716 (2018). Network analysis of protein adaptation: Modeling the functional impact of multiple
717 mutations. Molecular biology and evolution. 35(6), 1507-1519.
718 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N.,
719 and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28(1), 235-242. DOI:
720 10.1093/nar/28.1.235.
721 Berndsen, Z., Chakraborty, S., Wang, X., Cottrell, C., Torres, J., Diedrich, J.K., Lopez, C.A.,
722 Yates III, J.R., van-Gills, M.J., Paulson, J.C., et al. (2019). Visualization of the HIV-1 Env
723 Glycan Shield Across Scales. bioRxiv. 10.1101/839217.
724 Best, R.B., Zhu, X., Shim, J., Lopes, P.E., Mittal, J., Feig, M., and MacKerell Jr, A.D. (2012).
725 Optimization of the additive CHARMM all-atom protein force field targeting improved sampling
726 of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. Journal of chemical theory and
727 computation. 8(9), 3257-3273.
728 Bhatia, R., Gautam, S.K., Cannon, A., Thompson, C., Hall, B.R., Aithal, A., Banerjee, K., Jain,
729 M., Solheim, J.C., Kumar, S., et al. (2019). Cancer-associated mucins: role in immune
730 modulation and metastasis. Cancer Metastasis Rev. 38(1-2), 223-236. DOI: 10.1007/s10555-018-
731 09775-0.
732 Bianchi, M., Turner, H.L., Nogal, B., Cottrell, C.A., Oyen, D., Pauthner, M., Bastidas, R.,
733 Nedellec, R., McCoy, L.E., Wilson, I.A., et al. (2018). Electron-Microscopy-Based Epitope
734 Mapping Defines Specificities of Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope
735 Trimer Immunization. Immunity. 49(2), 288-300 e288. DOI: 10.1016/j.immuni.2018.07.009.
736 Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of
737 communities in large networks. Journal of statistical mechanics: theory and experiment.
738 2008(10), P10008.
739 Bradley, T., Fera, D., Bhiman, J., Eslamizar, L., Lu, X., Anasti, K., Zhang, R., Sutherland, L.L.,
740 Scearce, R.M., Bowman, C.M., et al. (2016). Structural Constraints of Vaccine-Induced Tier-2
741 Autologous HIV Neutralizing Antibodies Targeting the Receptor-Binding Site. Cell Rep. 14(1),
742 43-54. DOI: 10.1016/j.celrep.2015.12.017.
743 Bricault, C.A., Yusim, K., Seaman, M.S., Yoon, H., Theiler, J., Giorgi, E.E., Wagh, K., Theiler,
744 M., Hraber, P., Macke, J.P., et al. (2019). HIV-1 Neutralizing Antibody Signatures and
745 Application to Epitope-Targeted Vaccine Design. Cell Host Microbe. 25(1), 59-72 e58. DOI:
746 10.1016/j.chom.2018.12.001.
747 Burton, D.R., Ahmed, R., Barouch, D.H., Butera, S.T., Crotty, S., Godzik, A., Kaufmann, D.E.,
748 McElrath, M.J., Nussenzweig, M.C., Pulendran, B., et al. (2012). A Blueprint for HIV Vaccine
749 Discovery. Cell Host Microbe. 12(4), 396-407. DOI: 10.1016/j.chom.2012.09.008.
750 Burton, D.R., and Mascola, J.R. (2015). Antibody responses to envelope glycoproteins in HIV-1
751 infection. Nat Immunol. 16(6), 571-576. DOI: 10.1038/ni.3158.
752 Cao, L., Diedrich, J.K., Kulp, D.W., Pauthner, M., He, L., Park, S.R., Sok, D., Su, C.Y.,
753 Delahunty, C.M., Menis, S., et al. (2017). Global site-specific N-glycosylation analysis of HIV
754 envelope glycoprotein. Nat Commun. 8, 14954. DOI: 10.1038/ncomms14954.
755 Cao, L., Pauthner, M., Andrabi, R., Rantalainen, K., Berndsen, Z., Diedrich, J.K., Menis, S., Sok,
756 D., Bastidas, R., Park, S.R., et al. (2018). Differential processing of HIV envelope glycans on the

virus and soluble recombinant trimer. Nat Commun. 9(1), 3693. DOI: 10.1038/s41467-018-06121-4.
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., et al. (2007). Glycoprotein structural genomics: solving the glycosylation problem. Structure. 15(3), 267-273. DOI: 10.1016/j.str.2007.01.011.
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(24), 11132-11144. DOI: 10.1128/JVI.01542-16.
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. DOI: 10.1016/j.coviro.2015.02.002.
Crispin, M., Ward, A.B., and Wilson, I.A. (2018). Structure and Immune Recognition of the HIV Glycan Shield. Annu Rev Biophys. DOI: 10.1146/annurev-biophys-060414-034156.
Dalziel, M., Crispin, M., Scanlan, C.N., Zitzmann, N., and Dwek, R.A. (2014). Emerging principles for the therapeutic exploitation of glycosylation. Science. 343(6166), 1235681. DOI: 10.1126/science.1235681.
Davis, S.J., and Crispin, M. (2010). Solutions to the glycosylation problem for low-and high-throughput structural glycoproteomics. In: Functional and Structural Proteomics of Glycoproteins, (Springer), pp. 127-158.
Dennis, J.W., Granovsky, M., and Warren, C.E. (1999). Protein glycosylation in development and disease. Bioessays. 21(5), 412-421. DOI: 10.1002/(SICI)1521-1878(199905)21:5<412::AID-BIES8>3.0.CO;2-5.
Doores, K.J. (2015). The HIV glycan shield as a target for broadly neutralizing antibodies. The FEBS journal. 282(24), 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(31), 13800-13805. DOI: 10.1073/pnas.1006498107.
Dwek, R.A. (1996). Glycobiology: Toward Understanding the Function of Sugars. Chem Rev. 96(2), 683-720.
Eargle, J., and Luthey-Schulten, Z. (2012). NetworkView: 3D display and analysis of protein.RNA interaction networks. Bioinformatics. 28(22), 3000-3001. DOI: 10.1093/bioinformatics/bts546.
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. Current protocols in bioinformatics. 15(1), 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., et al. (2014). Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. Immunity. 40(5), 657-668. DOI: 10.1016/j.immuni.2014.04.009.
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(1), 15031. DOI: 10.1038/s41598-018-33390-2.
Fiser, A., Do, R.K., and Sali, A. (2000). Modeling of loops in protein structures. Protein Sci. 9(9), 1753-1773. Published online 2000/10/25 DOI: 10.1110/ps.9.9.1753.
Floyd, R.W. (1962). Algorithm 97: shortest path. Communications of the ACM. 5(6), 345.

803 Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment.
804 Proteins. 23(4), 566-579. DOI: DOI 10.1002/prot.340230412.
805 Fruchterman, T.M., and Reingold, E.M. (1991). Graph drawing by force-directed placement.
806 Software: Practice and experience. 21(11), 1129-1164.
807 Garces, F., Lee, J.H., de Val, N., de la Pena, A.T., Kong, L., Puchades, C., Hua, Y., Stanfield,
808 R.L., Burton, D.R., Moore, J.P., et al. (2015). Affinity Maturation of a Potent Family of HIV
809 Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. Immunity. 43(6),
810 1053-1063. DOI: 10.1016/j.immuni.2015.11.007.
811 Go, E.P., Ding, H., Zhang, S., Ringe, R.P., Nicely, N., Hua, D., Steinbock, R.T., Golabek, M.,
812 Alin, J., Alam, S.M., et al. (2017). Glycosylation Benchmark Profile for HIV-1 Envelope
813 Glycoprotein Production Based on Eleven Env Trimers. J Virol. 91(9). DOI: 10.1128/JVI.02428-
814 16.
815 Gristick, H.B., von Boehmer, L., West, A.P., Jr., Schamber, M., Gazumyan, A., Golijanin, J.,
816 Seaman, M.S., Fatkenheuer, G., Klein, F., Nussenzweig, M.C., et al. (2016). Natively
817 glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-
818 binding site. Nat Struct Mol Biol. 23(10), 906-915. DOI: 10.1038/nsmb.3291.
819 Guttman, M., Weinkam, P., Sali, A., and Lee, K.K. (2013). All-atom ensemble modeling to
820 analyze small-angle x-ray scattering of glycosylated proteins. Structure. 21(3), 321-331. DOI:
821 10.1016/j.str.2013.02.004.
822 Guvench, O., Mallajosyula, S.S., Raman, E.P., Hatcher, E., Vanommeslaeghe, K., Foster, T.J.,
823 Jamison, F.W., and MacKerell Jr, A.D. (2011). CHARMM additive all-atom force field for
824 carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling.
825 Journal of chemical theory and computation. 7(10), 3162-3180.
826 Haynes, B.F., Moody, M.A., Alam, M., Bonsignori, M., Verkoczy, L., Ferrari, G., Gao, F.,
827 Tomaras, G.D., Liao, H.X., and Kelsoe, G. (2014). Progress in HIV-1 vaccine development. J
828 Allergy Clin Immunol. 134(1), 3-10; quiz 11. DOI: 10.1016/j.jaci.2014.04.025.
829 Huang, J., Kang, B.H., Pancera, M., Lee, J.H., Tong, T., Feng, Y., Imamichi, H., Georgiev, I.S.,
830 Chuang, G.Y., Druz, A., et al. (2014). Broad and potent HIV-1 neutralization by a human
831 antibody that binds the gp41-gp120 interface. Nature. 515(7525), 138-142. DOI:
832 10.1038/nature13601.
833 Huang, J., and MacKerell Jr, A.D. (2013). CHARMM36 all-atom additive protein force field:
834 Validation based on comparison to NMR data. Journal of computational chemistry. 34(25),
835 2135-2145.
836 Huang, L., Liao, L., and Wu, C.H. (2017). Evolutionary analysis and interaction prediction for
837 protein-protein interaction network in geometric space. PloS one. 12(9), e0183495.
838 Huang, X., Jin, W., Hu, K., Luo, S., Du, T., Griffin, G.E., Shattock, R.J., and Hu, Q. (2012).
839 Highly conserved HIV-1 gp120 glycans proximal to CD4-binding region affect viral infectivity
840 and neutralizing antibody induction. Virology. 423(1), 97-106. DOI:
841 10.1016/j.virol.2011.11.023.
842 Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J Mol
843 Graph Model. 14(1), 33-38. DOI: Doi 10.1016/0263-7855(96)00018-5.
844 Ilinykh, P.A., Santos, R.I., Gunn, B.M., Kuzmina, N.A., Shen, X., Huang, K., Gilchuk, P., Flyak,
845 A.I., Younan, P., Alter, G., et al. (2018). Asymmetric antiviral effects of ebolavirus antibodies
846 targeting glycoprotein stem and glycan cap. PLoS Pathog. 14(8), e1007204. DOI:
847 10.1371/journal.ppat.1007204.

848 Imberty, A., and Perez, S. (2000). Structure, conformation, and dynamics of bioactive
849 oligosaccharides: theoretical approaches and experimental validations. Chem Rev. 100(12),
850 4567-4588.
851 Julien, J.P., Cupo, A., Sok, D., Stanfield, R.L., Lyumkis, D., Deller, M.C., Klasse, P.J., Burton,
852 D.R., Sanders, R.W., Moore, J.P., et al. (2013). Crystal structure of a soluble cleaved HIV-1
853 envelope trimer. Science. 342(6165), 1477-1483. DOI: 10.1126/science.1245625.
854 Karsten, C.B., and Alter, G. (2017). The HIV-1 Glycan Shield: Strategically Placed Kinks in the
855 Armor Improve Antigen Design. Cell Rep. 19(4), 669-670. DOI: 10.1016/j.celrep.2017.04.033.
856 Kwon, Y.D., Pancera, M., Acharya, P., Georgiev, I.S., Crooks, E.T., Gorman, J., Joyce, M.G.,
857 Guttman, M., Ma, X., Narpala, S., et al. (2015). Crystal structure, conformational fixation and
858 entry-related interactions of mature ligand-free HIV-1 Env. Nat Struct Mol Biol. 22(7), 522-531.
859 DOI: 10.1038/nsmb.3051.
860 LaBranche, C.C., McGuire, A.T., Gray, M.D., Behrens, S., Kwong, P.D.K., Chen, X., Zhou, T.,
861 Sattentau, Q.J., Peacock, J., Eaton, A., et al. (2018). HIV-1 envelope glycan modifications that
862 permit neutralization by germline-reverted VRC01-class broadly neutralizing antibodies. PLoS
863 Pathog. 14(11), e1007431. DOI: 10.1371/journal.ppat.1007431.
864 Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a
865 program to check the stereochemical quality of protein structures. Journal of applied
866 crystallography. 26(2), 283-291.
867 Lasky, L.A., Groopman, J.E., Fennie, C.W., Benz, P.M., Capon, D.J., Dowbenko, D.J.,
868 Nakamura, G.R., Nunes, W.M., Renz, M.E., and Berman, P.W. (1986). Neutralization of the
869 AIDS retrovirus by antibodies to a recombinant envelope glycoprotein. Science. 233(4760), 209-
870 212. DOI: 10.1126/science.3014647.
871 Lee, J.H., Ozorowski, G., and Ward, A.B. (2016). Cryo-EM structure of a native, fully
872 glycosylated, cleaved HIV-1 envelope trimer. Science. 351(6277), 1043-1048. DOI:
873 10.1126/science.aad2450.
874 Lemmin, T., Soto, C., Stuckey, J., and Kwong, P.D. (2017). Microsecond Dynamics and
875 Network Analysis of the HIV-1 SOSIP Env Trimer Reveal Collective Behavior and Conserved
876 Microdomains of the Glycan Shield. Structure. 25(10), 1631-1639 e1632. DOI:
877 10.1016/j.str.2017.07.018.
878 Lennemann, N.J., Rhein, B.A., Ndungo, E., Chandran, K., Qiu, X., and Maury, W. (2014).
879 Comprehensive functional analysis of N-linked glycans on Ebola virus GP1. MBio. 5(1),
880 e00862-00813. DOI: 10.1128/mBio.00862-13.
881 Li, Y., Cleveland, B., Klots, I., Travis, B., Richardson, B.A., Anderson, D., Montefiori, D.,
882 Polacino, P., and Hu, S.L. (2008). Removal of a single N-linked glycan in human
883 immunodeficiency virus type 1 gp120 results in an enhanced ability to induce neutralizing
884 antibody responses. J Virol. 82(2), 638-651. DOI: 10.1128/JVI.01691-07.
885 Liang, Y., Guttman, M., Williams, J.A., Verkerke, H., Alvarado, D., Hu, S.L., and Lee, K.K.
886 (2016). Changes in Structure and Antigenicity of HIV-1 Env Trimers Resulting from Removal of
887 a Conserved CD4 Binding Site-Proximal Glycan. J Virol. 90(20), 9224-9236. DOI:
888 10.1128/JVI.01116-16.
889 Lütteke, T., Frank, M., and von der Lieth, C.-W. (2005). Carbohydrate Structure Suite (CSS):
890 analysis of carbohydrate 3D structures derived from the PDB. Nucleic acids research.
891 33(suppl_1), D242-D246.

892 Lyumkis, D., Julien, J.P., de Val, N., Cupo, A., Potter, C.S., Klasse, P.J., Burton, D.R., Sanders,
893 R.W., Moore, J.P., Carragher, B., et al. (2013). Cryo-EM structure of a fully glycosylated soluble
894 cleaved HIV-1 envelope trimer. Science. 342(6165), 1484-1490. DOI: 10.1126/science.1245627.
895 MATLAB, 2018. MATLAB, Version R2018a. The MathWorks Inc Natick, MA.
896 McCoy, L.E., van Gils, M.J., Ozorowski, G., Messmer, T., Briney, B., Voss, J.E., Kulp, D.W.,
897 Macauley, M.S., Sok, D., Pauthner, M., et al. (2016). Holes in the Glycan Shield of the Native
898 HIV Envelope Are a Target of Trimer-Elicited Neutralizing Antibodies. Cell Rep. 16(9), 2327-
899 2338. DOI: 10.1016/j.celrep.2016.07.074.
900 Moore, P.L., and Williamson, C. (2016). Approaches to the induction of HIV broadly
901 neutralizing antibodies. Current Opinion in HIV and AIDS. 11(6), 569.
902 Moremen, K.W., Tiemeyer, M., and Nairn, A.V. (2012). Vertebrate protein glycosylation:
903 diversity, synthesis and function. Nat Rev Mol Cell Biol. 13(7), 448-462. DOI:
904 10.1038/nrm3383.
905 Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J.F., Halper-Stromberg, A.,
906 Gnanapragasam, P.N., Spencer, D.I., Seaman, M.S., et al. (2012). Complex-type N-glycan
907 recognition by potent broadly neutralizing HIV antibodies. Proc Natl Acad Sci U S A. 109(47),
908 E3268-3277. DOI: 10.1073/pnas.1217207109.
909 Nakagawa, H., Hato, M., Takegawa, Y., Deguchi, K., Ito, H., Takahata, M., Iwasaki, N.,
910 Minami, A., and Nishimura, S. (2007). Detection of altered N-glycan profiles in whole serum
911 from rheumatoid arthritis patients. J Chromatogr B Analyt Technol Biomed Life Sci. 853(1-2),
912 133-137. DOI: 10.1016/j.jchromb.2007.03.003.
913 Newman, M.E. (2006). Modularity and community structure in networks. Proceedings of the
914 national academy of sciences. 103(23), 8577-8582.
915 Newman, M.E., and Girvan, M. (2004). Finding and evaluating community structure in
916 networks. Physical review E. 69(2), 026113.
917 Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P.,
918 Chuang, G.Y., Ofek, G., et al. (2014). Structure and immune recognition of trimeric pre-fusion
919 HIV-1 Env. Nature. 514(7523), 455-461. DOI: 10.1038/nature13808.
920 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and
921 Ferrin, T.E. (2004). UCSF chimera - A visualization system for exploratory research and
922 analysis. Journal of Computational Chemistry. 25(13), 1605-1612. DOI: 10.1002/jcc.20084.
923 Pillai, S.U., Suel, T., and Cha, S. (2005). The Perron-Frobenius theorem: some of its
924 applications. IEEE Signal Processing Magazine. 22(2), 62-75.
925 Pritchard, L.K., Spencer, D.I., Royle, L., Bonomelli, C., Seabright, G.E., Behrens, A.J., Kulp,
926 D.W., Menis, S., Krumm, S.A., Dunlop, D.C., et al. (2015a). Glycan clustering stabilizes the
927 mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. Nat
928 Commun. 6, 7479. DOI: 10.1038/ncomms8479.
929 Pritchard, L.K., Vasiljevic, S., Ozorowski, G., Seabright, G.E., Cupo, A., Ringe, R., Kim, H.J.,
930 Sanders, R.W., Doores, K.J., Burton, D.R., et al. (2015b). Structural Constraints Determine the
931 Glycosylation of HIV-1 Envelope Trimers. Cell Rep. 11(10), 1604-1613. DOI:
932 10.1016/j.celrep.2015.05.017.
933 Ringe, R.P., Pugach, P., Cottrell, C.A., LaBranche, C.C., Seabright, G.E., Ketas, T.J.,
934 Ozorowski, G., Kumar, S., Schorcht, A., van Gils, M.J., et al. (2019). Closing and Opening
935 Holes in the Glycan Shield of HIV-1 Envelope Glycoprotein SOSIP Trimers Can Redirect the
936 Neutralizing Antibody Response to the Newly Unmasked Epitopes. J Virol. 93(4). DOI:
937 10.1128/JVI.01656-18.

938 Rossum, G. (1995). Python reference manual.
939 Sali, A. (1995). Comparative protein modeling by satisfaction of spatial restraints. Mol Med
940 Today. 1(6), 270-277.
941 Sanders, R.W., Derking, R., Cupo, A., Julien, J.P., Yasmeen, A., de Val, N., Kim, H.J., Blattner,
942 C., de la Pena, A.T., Korzun, J., et al. (2013). A next-generation cleaved, soluble HIV-1 Env
943 trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not
944 non-neutralizing antibodies. PLoS Pathog. 9(9), e1003618. DOI: 10.1371/journal.ppat.1003618.
945 Sanders, R.W., van Gils, M.J., Derking, R., Sok, D., Ketas, T.J., Burger, J.A., Ozorowski, G.,
946 Cupo, A., Simonich, C., Goo, L., et al. (2015). HIV-1 VACCINES. HIV-1 neutralizing
947 antibodies induced by native-like envelope trimers. Science. 349(6244), aac4223. DOI:
948 10.1126/science.aac4223.
949 Saphire, E.O., Schendel, S.L., Fusco, M.L., Gangavarapu, K., Gunn, B.M., Wec, A.Z.,
950 Halfmann, P.J., Brannan, J.M., Herbert, A.S., Qiu, X., et al. (2018). Systematic Analysis of
951 Monoclonal Antibodies against Ebola Virus GP Defines Features that Contribute to Protection.
952 Cell. 174(4), 938-952 e913. DOI: 10.1016/j.cell.2018.07.033.
953 Schmidt, A.G., Xu, H., Khan, A.R., O'Donnell, T., Khurana, S., King, L.R., Manischewitz, J.,
954 Golding, H., Suphaphiphat, P., Carfi, A., et al. (2013). Preconfiguration of the antigen-binding
955 site during affinity maturation of a broadly neutralizing influenza virus antibody. Proc Natl Acad
956 Sci U S A. 110(1), 264-269. DOI: 10.1073/pnas.1218256109.
957 Schrödinger, L.L.C. (2015). The PyMOL molecular graphics system, Version 1.8. Technical
958 Report.
959 Sethi, A., Tian, J., Derdeyn, C.A., Korber, B., and Gnanakaran, S. (2013). A mechanistic
960 understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein. PLoS
961 Comput Biol. 9(5), e1003046. DOI: 10.1371/journal.pcbi.1003046.
962 Skjaerven, L., Yao, X.Q., Scarabelli, G., and Grant, B.J. (2014). Integrating protein structural
963 dynamics and evolutionary analysis with Bio3D. BMC Bioinformatics. 15, 399. DOI:
964 10.1186/s12859-014-0399-6.
965 Slynko, V., Schubert, M., Numao, S., Kowarik, M., Aebi, M., and Allain, F.H. (2009). NMR
966 structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. J Am
967 Chem Soc. 131(3), 1274-1281. DOI: 10.1021/ja808682v.
968 Sommerstein, R., Flatz, L., Remy, M.M., Malinge, P., Magistrelli, G., Fischer, N., Sahin, M.,
969 Bergthaler, A., Igonet, S., Ter Meulen, J., et al. (2015). Arenavirus Glycan Shield Promotes
970 Neutralizing Antibody Evasion and Protracted Infection. PLoS Pathog. 11(11), e1005276. DOI:
971 10.1371/journal.ppat.1005276.
972 Stanley, P., Taniguchi, N., and Aebi, M. (2015). N-Glycans. In: Essentials of Glycobiology, rd,
973 A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T.
974 Kinoshita, N.H. Packer, et al. eds. (Cold Spring Harbor (NY)), pp. 99-111.
975 Stewart-Jones, G.B., Soto, C., Lemmin, T., Chuang, G.Y., Druz, A., Kong, R., Thomas, P.V.,
976 Wagh, K., Zhou, T., Behrens, A.J., et al. (2016). Trimeric HIV-1-Env Structures Define Glycan
977 Shields from Clades A, B, and G. Cell. 165(4), 813-826. DOI: 10.1016/j.cell.2016.04.010.
978 Szakonyi, G., Klein, M.G., Hannan, J.P., Young, K.A., Ma, R.Z., Asokan, R., Holers, V.M., and
979 Chen, X.S. (2006). Structure of the Epstein-Barr virus major envelope glycoprotein. Nat Struct
980 Mol Biol. 13(11), 996-1001. DOI: 10.1038/nsmb1161.
981 Taniguchi, N., and Kizuka, Y. (2015). Glycans and cancer: role of N-glycans in cancer
982 biomarker, progression and metastasis, and therapeutics. Adv Cancer Res. 126, 11-51. DOI:
983 10.1016/bs.acr.2014.11.001.

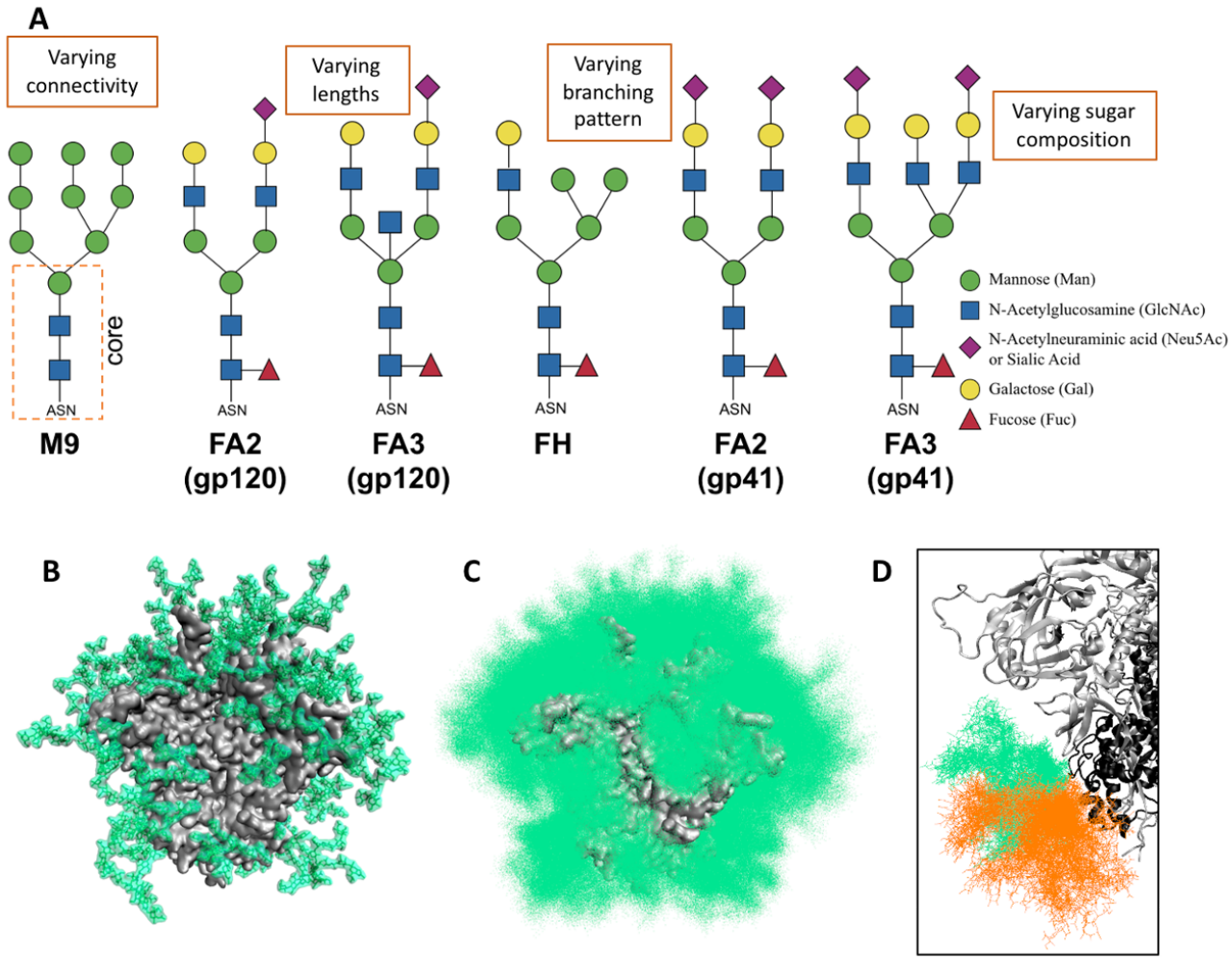
984 Townsley, S., Li, Y., Kozyrev, Y., Cleveland, B., and Hu, S.L. (2016). Conserved Role of an N-
985 Linked Glycan on the Surface Antigen of Human Immunodeficiency Virus Type 1 Modulating
986 Virus Sensitivity to Broadly Neutralizing Antibodies against the Receptor and Coreceptor
987 Binding Sites. J Virol. 90(2), 829-841. DOI: 10.1128/JVI.02321-15.
988 Umotoy, J., Bagaya, B.S., Joyce, C., Schiffner, T., Menis, S., Saye-Francisco, K.L., Biddle, T.,
989 Mohan, S., Vollbrecht, T., Kalyuzhniy, O., et al. (2019). Rapid and Focused Maturation of a
990 VRC01-Class HIV Broadly Neutralizing Antibody Lineage Involves Both Binding and
991 Accommodation of the N276-Glycan. Immunity. 51(1), 141-154 e146. DOI:
992 10.1016/j.immuni.2019.06.004.
993 Vankadari, N., and Wilce, J.A. (2020). Emerging WuHan (COVID-19) coronavirus: glycan
994 shield and structure prediction of spike glycoprotein and its interaction with human CD26.
995 Emerg Microbes Infect. 9(1), 601-604. Published online 2020/03/18 DOI:
996 10.1080/22221751.2020.1739565.
997 Wagh, K., Kreider, E.F., Li, Y., Barbian, H.J., Learn, G.H., Giorgi, E., Hraber, P.T., Decker,
998 T.G., Smith, A.G., Gondim, M.V., et al. (2018). Completeness of HIV-1 Envelope Glycan Shield
999 at Transmission Determines Neutralization Breadth. Cell Rep. 25(4), 893-908 e897. DOI:
1000 10.1016/j.celrep.2018.09.087.
1001 Walker, L.M., Phogat, S.K., Chan-Hui, P.Y., Wagner, D., Phung, P., Goss, J.L., Wrin, T., Simek,
1002 M.D., Fling, S., Mitcham, J.L., et al. (2009). Broad and potent neutralizing antibodies from an
1003 African donor reveal a new HIV-1 vaccine target. Science. 326(5950), 285-289. DOI:
1004 10.1126/science.1178746.
1005 Walker, L.M., Simek, M.D., Priddy, F., Gach, J.S., Wagner, D., Zwick, M.B., Phogat, S.K.,
1006 Poignard, P., and Burton, D.R. (2010). A limited number of antibody specificities mediate broad
1007 and potent serum neutralization in selected HIV-1 infected individuals. PLoS Pathog. 6(8),
1008 e1001028. DOI: 10.1371/journal.ppat.1001028.
1009 Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., DiMaio, F., Bosch, B.J.,
1010 and Veesler, D. (2016). Glycan shield and epitope masking of a coronavirus spike protein
1011 observed by cryo-electron microscopy. Nat Struct Mol Biol. 23(10), 899-905. Published online
1012 2016/09/13 DOI: 10.1038/nsmb.3293.
1013 Walsh, G. (2010). Biopharmaceutical benchmarks 2010. Nat Biotechnol. 28(9), 917-924. DOI:
1014 10.1038/nbt0910-917.
1015 Wang, C.C., Chen, J.R., Tseng, Y.C., Hsu, C.H., Hung, Y.F., Chen, S.W., Chen, C.M., Khoo,
1016 K.H., Cheng, T.J., Cheng, Y.S., et al. (2009). Glycans on influenza hemagglutinin affect receptor
1017 binding and immune response. Proc Natl Acad Sci U S A. 106(43), 18137-18142. DOI:
1018 10.1073/pnas.0909696106.
1019 Warshall, S., 1962. A theorem on boolean matrices, Journal of the ACM. Citeseer.
1020 Weinkam, P., Pons, J., and Sali, A. (2012). Structure-based model of allostery predicts coupling
1021 between distant sites. Proc Natl Acad Sci U S A. 109(13), 4875-4880. DOI:
1022 10.1073/pnas.1116274109.
1023 Woods, R.J., Pathiaseril, A., Wormald, M.R., Edge, C.J., and Dwek, R.A. (1998). The high
1024 degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the
1025 overall topology of the molecule. Eur J Biochem. 258(2), 372-386.
1026 Xu, K., Acharya, P., Kong, R., Cheng, C., Chuang, G.Y., Liu, K., Louder, M.K., O'Dell, S.,
1027 Rawi, R., Sastry, M., et al. (2018). Epitope-based vaccine design yields fusion peptide-directed
1028 antibodies that neutralize diverse strains of HIV-1. Nat Med. 24(6), 857-867. DOI:
1029 10.1038/s41591-018-0042-6.

1030 Yang, M., Huang, J., Simon, R., Wang, L.X., and MacKerell, A.D., Jr. (2017). Conformational
1031 Heterogeneity of the HIV Envelope Glycan Shield. Sci Rep. 7(1), 4435. DOI: 10.1038/s41598-
1032 017-04532-9.
1033 Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., and Korber, B. (2004).
1034 Tracking global patterns of N-linked glycosylation site variation in highly variable viral
1035 glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology.
1036 14(12), 1229-1246. DOI: 10.1093/glycob/cwh106.
1037 Zhang, X., Jia, R., Shen, H., Wang, M., Yin, Z., and Cheng, A. (2017). Structures and functions
1038 of the envelope glycoprotein in flavivirus infections. Viruses. 9(11), 338.
1039

1040 **Figures and Figure Legends**

1041

1042



1043

1044 **Figure1: Native glycosylation and shielding of HIV-1 Env.** (A). Schematic representation of N-
1045 glycan types. The common core of Man3GlcNAc2Asn is indicated. The glycan species selected
1046 for different sites of the BG505 native model (as given in **Table S1**). M9 ≡ mannose-9, FA2 ≡
1047 fucosylated two-antennae, FA3 ≡ fucosylated three-antennae, FH ≡ fucosylated hybrid. Complex
1048 glycans in gp41 are different from those in gp120, as per site-specific mass spectroscopy
1049 experiments. (B) Single snapshot of modeled natively glycosylated BG505 structure, with each
1050 glycan taking up a particular conformation. (C) Cumulative shielding effect over time due to the
1051 dynamic nature of the glycans. 100 randomly selected models from the 1000-structure ensemble
1052 is shown here. (D) Neighboring glycans within close proximity can sample overlapping volumes.
1053 Glycan ensembles N88 shown in green and N625 shown in orange.

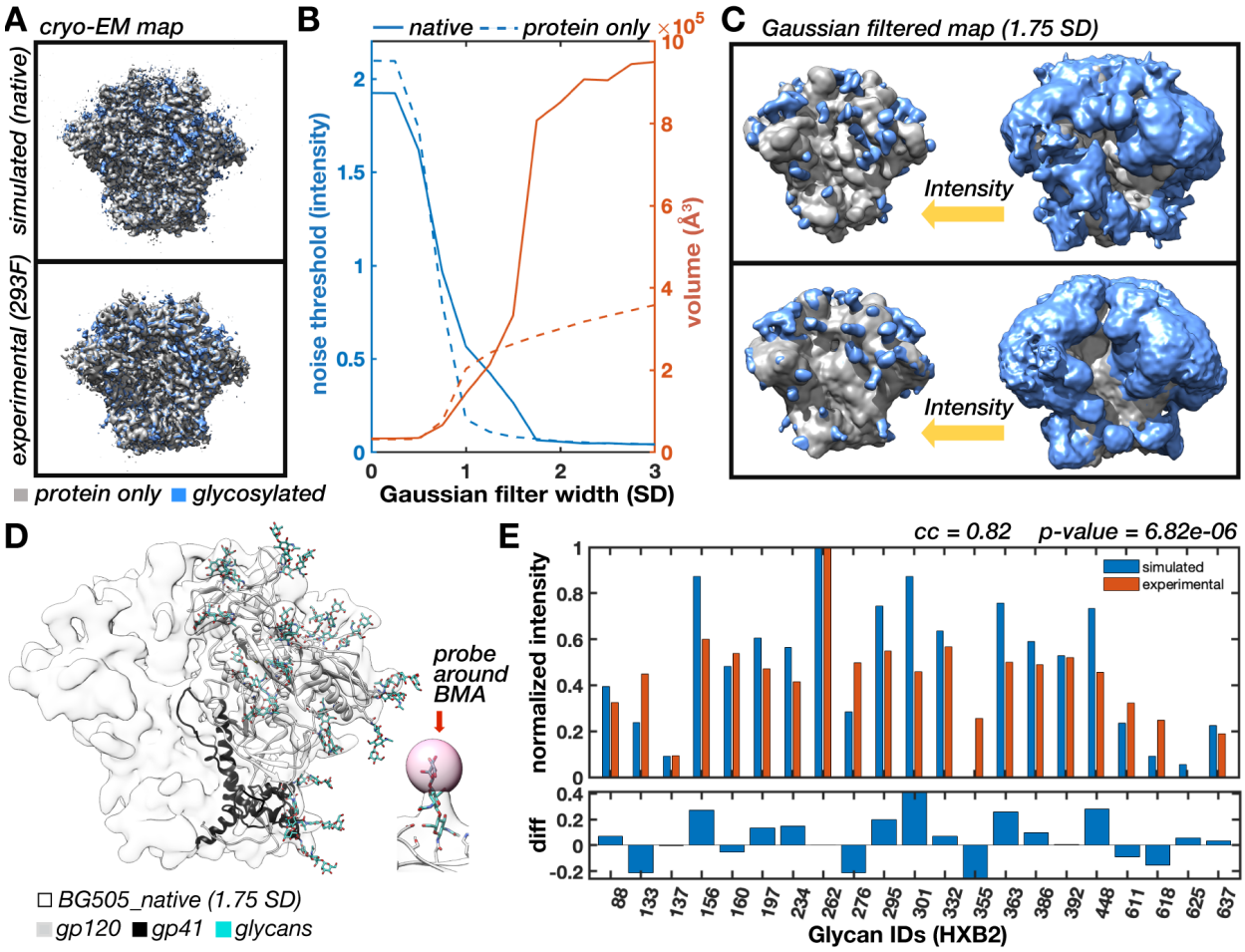


Figure 2: Comparison of glycan dynamics between experimental and simulated cryo-EM maps. (A) Experimental BG505_293F (top) and simulated BG505_native (bottom) cryo-EM maps sharpened and filtered at the global resolution (~3Å), with maps of de-glycosylated trimers displayed underneath to highlight the glycan density. The BG505_293F data is reproduced from (Berndsen et al., 2019) and density corresponding to the three RM20A3 Fabs bound to the base of the SOSIP were deleted for the sake of aesthetics (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 de-glycosylated maps showing the progressive emergence of glycan signal at lower resolution. (C) 1.75 SD Gaussian filtered maps from panel 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.

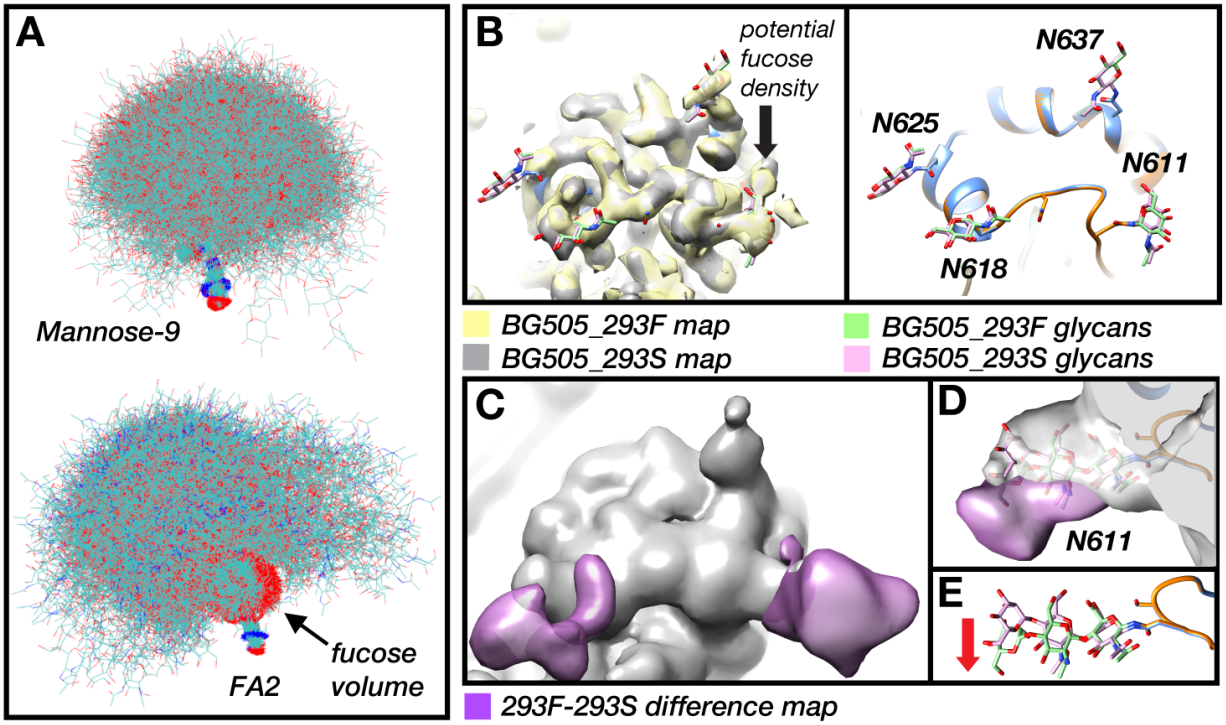


Figure 3: Model and Experimental cryo-EM data reveals subtle change in orientation due to core fucose. (A) Difference in overall glycan orientation between all-man9 and fucosylated complex glycan in simulated model. The first core sugar and three residues at the base of the glycan (n-1, n, n+1; where n is the glycosylated asparagine residue) were aligned for a single sugar ensemble. Oligomannose samples a space symmetrically around the core, while, 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 4 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.

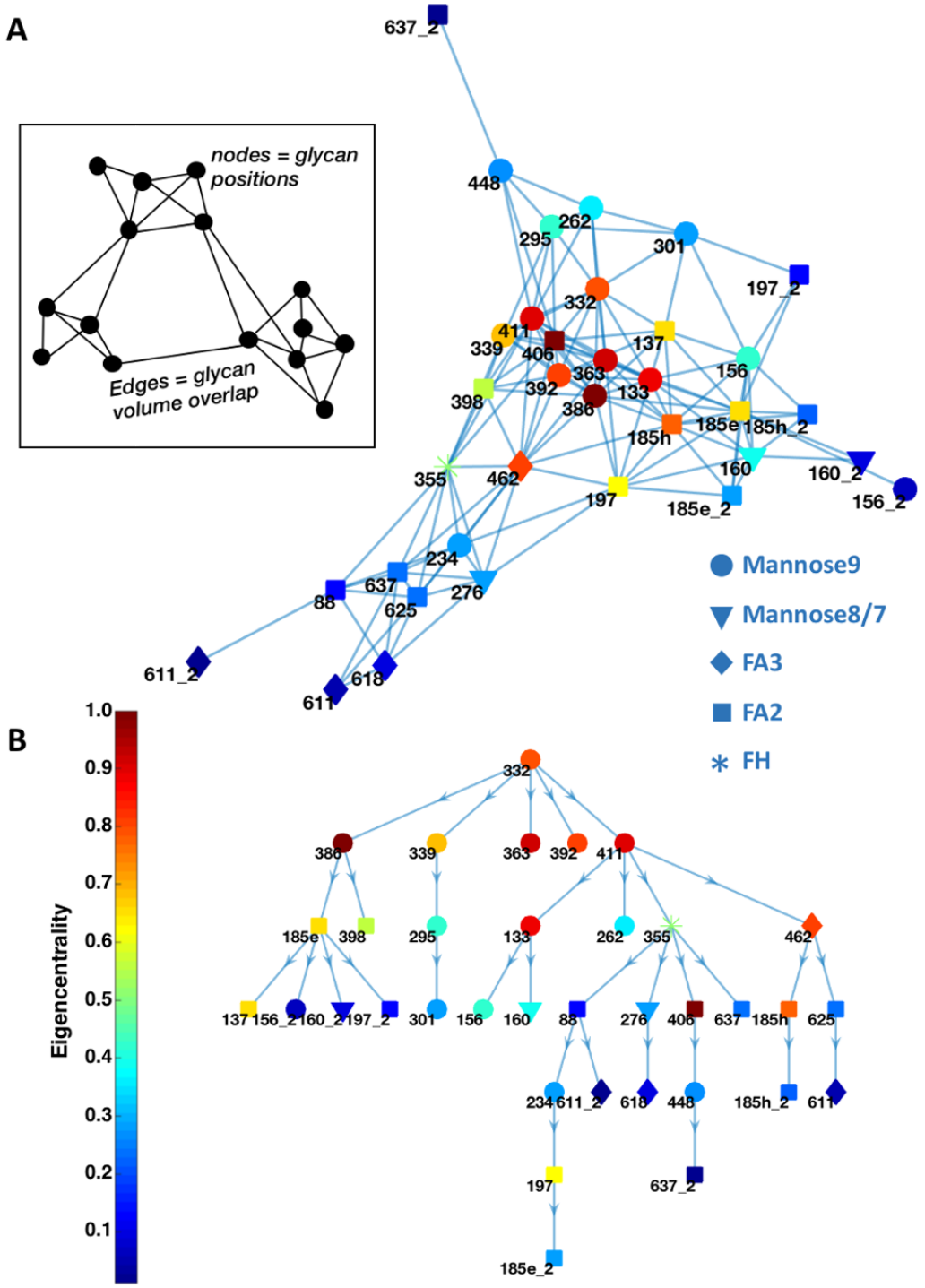


Figure 4: Spatial network of BG505 native glycosylation. (A) Visual representation of network in 2-dimensions, based on a force-directed layout. Each glycan forms a node-point on the graph, two nodes are connected by an edge (inset, from reference (Berndsen et al., 2019), Figure 8). The nodes are colored by the normalized eigenvector centrality of the glycans. (B) Shortest path of communication between glycan N332 and all other glycans in the network.

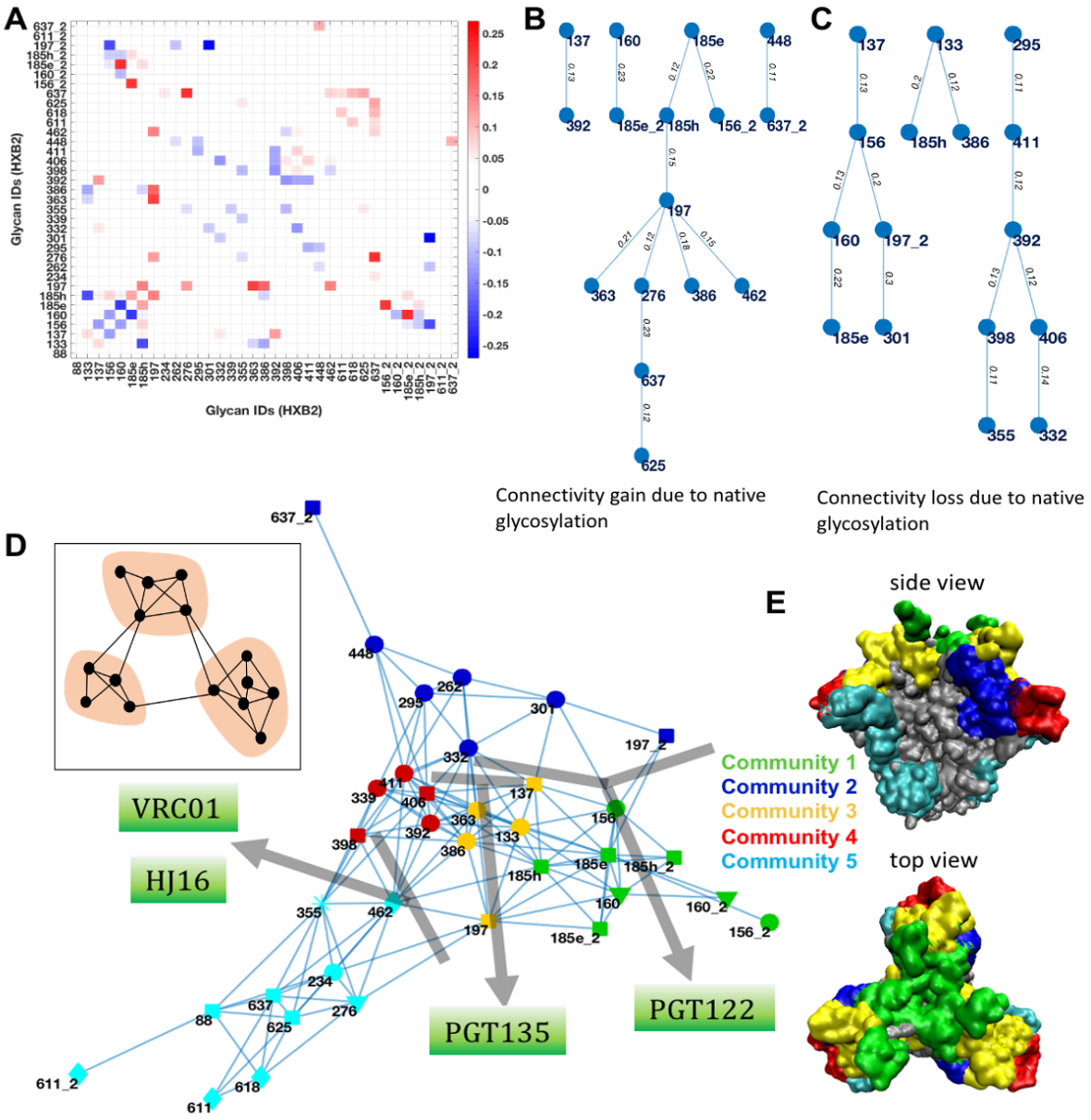


Figure 5: Network difference and community analysis. (A) Difference in adjacency matrices between native and all-man9 models (native minus all-man9). Blue color indicates decrease in edge weight, and red indicates increase in edge weight in native network, as compared to all-man9. (B) Increase and (C) 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) 5 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.

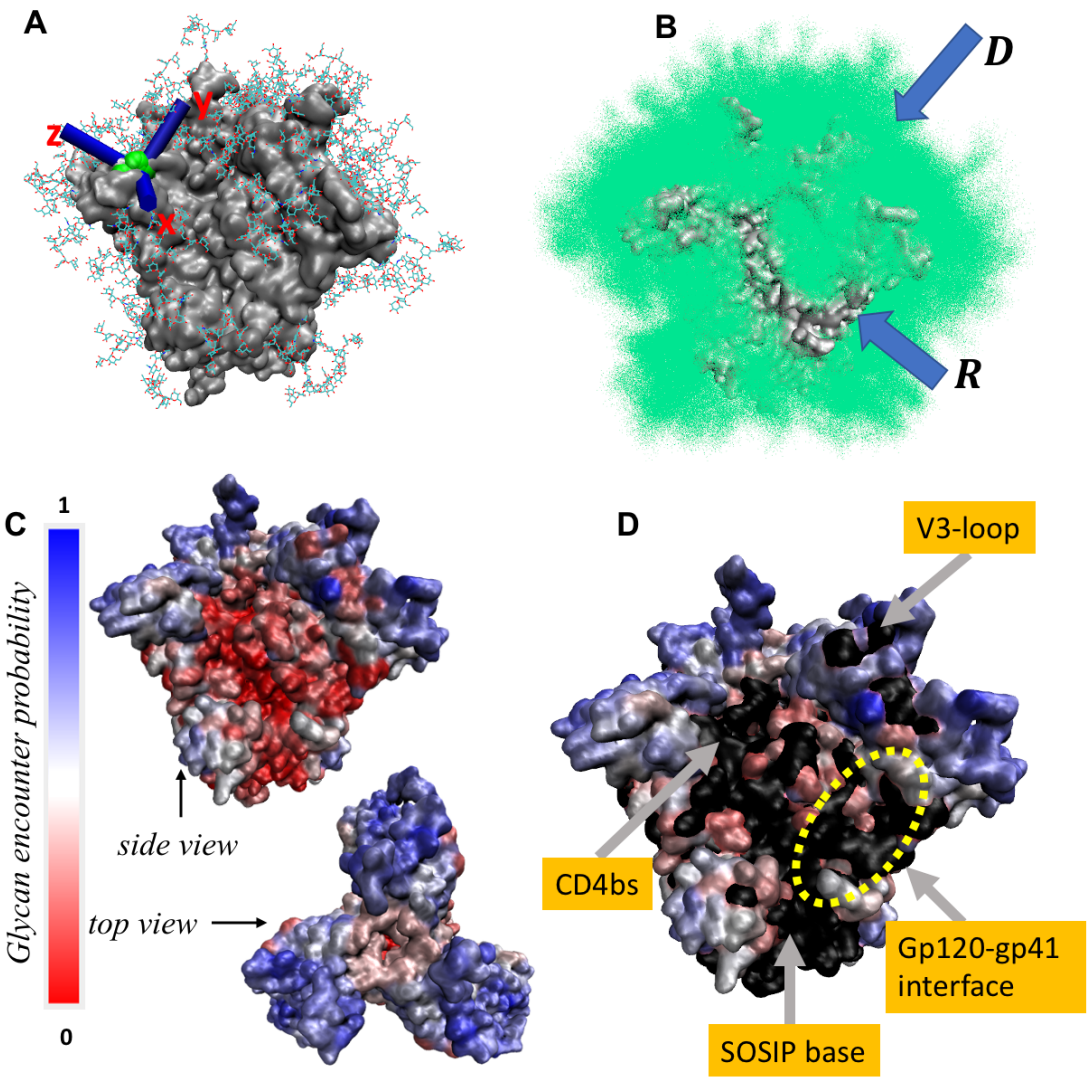
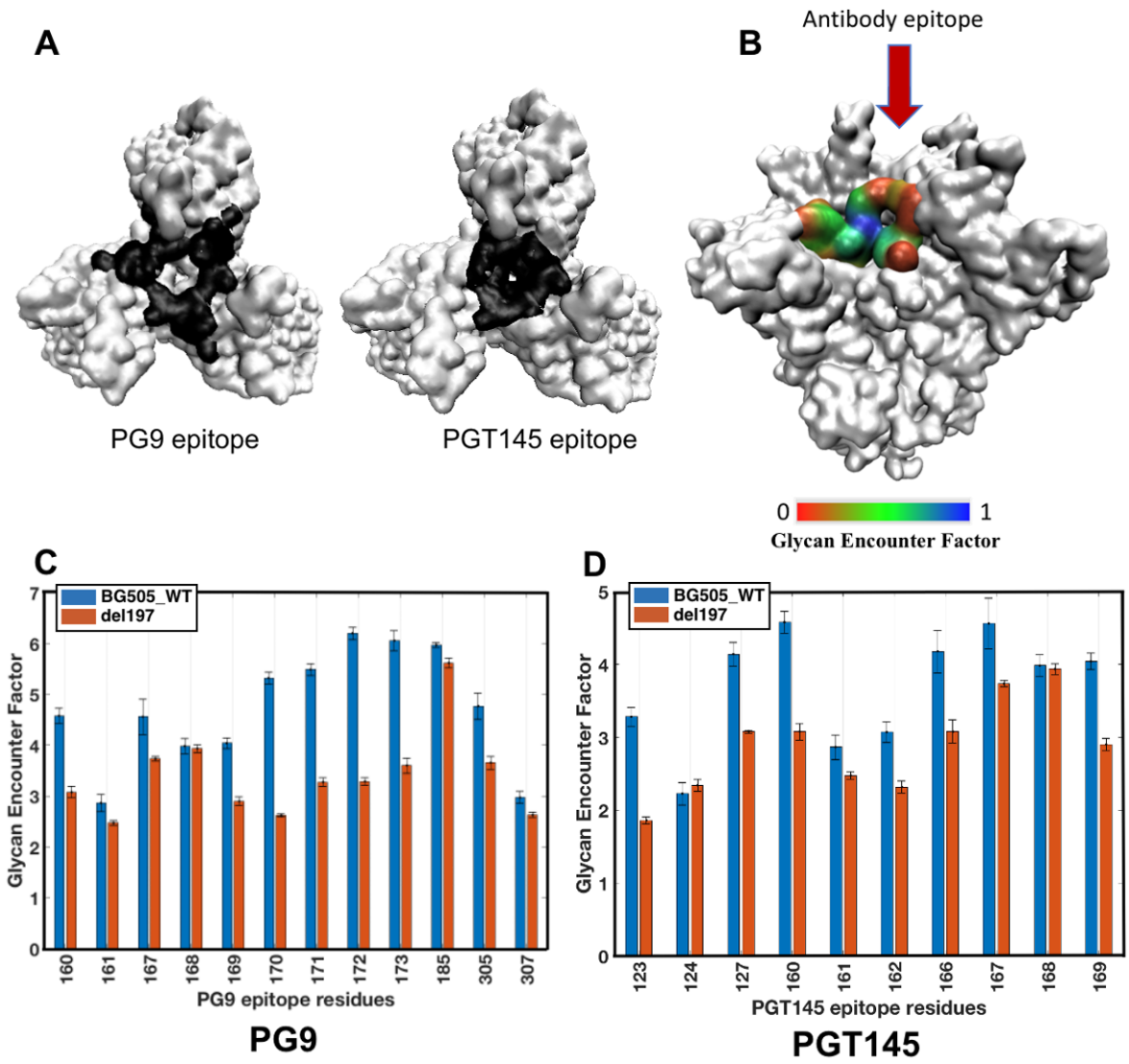


Figure 6: Glycan Encounter Factor (GEF) for quantifying shielding effect. (A) At each residue present on the surface of the protein, the approaching probe is considered in three directions **x, y** and **z.** (B) Any point on the surface which has a dense glycan covering, such as D, has a high glycan encounter factor value, versus a point such as R where the glycan covering is sparse, which will have a low GEF. (C) 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.



1107

1108 **Figure 7: Deletion of glycan N197 decreases glycan shielding at PG9 and PGT145 epitope.**
1109 (A) Top view of Env showing epitope regions of antibodies PG9 and PGT145 at the apex. (B)
1110 Footprint of glycan N197 on Env surface, colored by glycan encounter factor contributed by N197
1111 alone. (C) Glycan encounter factor over the PG9 and (D) PGT145 epitope residues. GEF decreases
1112 for both the epitopes due to deletion of N197 glycan.