Visualization of the HIV-1 Env Glycan Shield Across Scales

Zachary T. Berndsen[1,2,#], Srirupa Chakraborty[5,6,#], Xiaoning Wang[2,3] , Christopher A. Cottrell[1,2], Jonathan L. Torres[1], Jolene K. Diedrich[3], Cesar A. López[5], John R. Yates III[3] , Marit J. van Gils[7], James C. Paulson[2,3,4], S Gnanakaran[5] *, Andrew B. Ward[1,2] *.

[1]Department of Integrative Structural and Computational Biology, The [2]IAVI Neutralizing Antibody Center and Collaboration of AIDS Vaccine Discovery, Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID), [3]Department of Molecular Medicine, and [4]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. [5]Theoretical Biology and Biophysics Group and [6]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States. [7]Department of Medical Microbiology, Amsterdam UMC, location AMC, University of Amsterdam, 1105 AZ Amsterdam, Netherlands

[#]These authors contributed equally

*Correspondence: andrew@scripps.edu, gnana@lanl.gov

Abstract

The dense array of N-linked glycans on the HIV-1 Envelope Glycoprotein (Env), known as the "glycan shield", is a key determinant of immunogenicity, yet intrinsic heterogeneity confounds typical structure-function analysis. Here we present an integrated approach of single-particle electron cryomicroscopy (cryo-EM), computational modeling, and site-specific mass-spectrometry (MS) to probe glycan shield structure and behavior at multiple levels. We found that dynamics lead to an extensive network of inter-glycan interactions that drive the formation of higher-order structure within the glycan shield. This structure defines diffuse boundaries between buried and exposed protein surface and creates a mapping of potentially immunogenic sites on Env. Analysis of Env expressed in different cell lines revealed how cryo-EM can detect subtle changes in glycan occupancy, composition, and dynamics that impact glycan shield structure and epitope accessibility. Importantly, this identified unforeseen changes in the glycan shield of Env obtained from expression in the same CHO cell line used for GMP production. Finally, by capturing the enzymatic deglycosylation of Env in a time-resolved manner we found that highly connected glycan clusters are resistant to digestion and help stabilize the pre-fusion trimer, suggesting the glycan shield may function beyond immune evasion.

Significance Statement

The HIV-1 Env "glycan shield" masks the surface of the protein from immune recognition, yet intrinsic heterogeneity defies a typical structure-function description. Using a complementary approach of cryo-EM, computational modeling, and mass-spectrometry we show how heterogeneity and dynamics affect glycan shield structure across scales. Our combined approach facilitated the development of new cryo-EM data analysis methods and allowed for validation of

models against experiment. Comparison of Env across a range of glycosylation states revealed how subtle differences in composition impact glycan shield structure and affect the accessibility of epitopes on the surface. Finally, time-resolved cryo-EM experiments uncovered how highly connected glycan clusters help stabilize the pre-fusion trimer, suggesting the glycan shield may function beyond immune evasion.

Introduction

The Human Immunodeficiency Virus Type 1 (HIV-1) Envelope Glycoprotein (Env) is the sole antigen on the surface of the virion and has evolved several tactics for evading the adaptive immune system, chief among which is extensive surface glycosylation (1-3). Env has one of the highest densities of N-linked glycosylation sites known, with glycans accounting for ~1/2 the mass of the molecule (4-8). This sugar coat, or "glycan shield," is common among viral fusion proteins and is believed to be a primary hurdle in the development of neutralizing antibodies against Env during infection and vaccination (9, 10). Therefore, arriving at a consistent and general description of its structure and dynamics may prove necessary in designing an effective Env-based immunogen and be of broader importance towards understanding the structure and function of other densely glycosylated proteins.

N-linked glycans can be highly dynamic, however, and can vary substantially in length, connectivity, and chemical composition (11, 12). On the Env ectodomain surface there are ~95 glycosylation sites on average (6), and mass-spectrometry analysis (MS) has revealed that compositional heterogeneity exists between sites as well as at the same site between different copies of Env (13). While heterogeneity and dynamics are critical for immune evasion, they are incompatible with a typical structure-function description. When the first structures of the fully glycosylated Env were solved with single-particle electron cryomicroscopy (cryo-EM) (14, 15) most glycans were not resolved beyond the first or second sugar ring except when stabilized by antibodies. This suggested that the glycan shield is highly dynamic, however, due to the missing high-resolution details, structural analysis remained focused on Env-antibody interactions, while experimental investigations into the glycan shield shifted primarily towards chemical analysis via MS (11). One crystallography study specifically addressing glycan shield structure was recently published and the data shows evidence for highly stabilized glycans engaging in a wide range of inter-glycan contacts (6), which is at odds with the cryo-EM results. Such a disagreement between the two techniques highlights the difficulty in studying the glycan shield and the need for more focused experiments.

Computational modeling, in particular molecular dynamics (MD) simulations, has proven useful as a complementary approach for characterizing glycoprotein structure and dynamics (16-19), including Env (6, 20-22). Three sets of atomistic dynamics simulations of fully glycosylated Env trimers were recently reported and these predicted that dynamics would lead to extensive shielding of antigenic surface and the formation of complicated networks among interacting glycans (6, 21, 22). In one study, glycans were observed clustering into distinct micro-domains over long timescales, and it was found that neutralizing antibodies preferentially target the interfaces

between these domains, suggesting the glycan shield may possess biologically relevant large-scale structure. Although MD is a widely accepted method for modeling glycoproteins, atomistic simulations for systems as large as Env require substantial computational resources to access biologically relevant spatial and temporal scales and can suffer from initial model bias. Therefore, more computationally efficient methods for modeling fully glycosylated Env are needed. Furthermore, there are still no established methods for experimentally validating the results of these simulations.

Here, we describe an integrated approach combining cryo-EM, computational modeling, and site-specific MS aimed at illuminating glycan shield structure and behavior at multiple levels. First, we introduce a model system for our cryo-EM experiments and describe the complete structure of the native Env glycan shield for the first time using scale-space and 3-D variability analysis. In parallel, we developed a high-throughput modeling pipeline for rapidly generating diverse ensembles of fully glycosylated Env at atomistic resolution, enabling quantification of glycan-specific geometric properties and concerted behavior within the glycan shield as a whole. We then used these ensembles as ground-truth for the creation of simulated cryo-EM data, which facilitated the development of new analysis tools and enabled the validation of theoretical models against cryo-EM experiments.

With our integrated approach, we show that the glycan shield is highly dynamic but exhibits variation in dynamics between glycans due in part to crowding and other geometric and energetic constraints. We found that dynamics give rise to a network of inter-glycan interactions that drive the formation of higher-order structure within the glycan shield. This blurry low-resolution structure creates diffuse boundaries between buried and exposed protein surface that define potential sites of vulnerability. Using Env expressed in three common cell lines, we show how differences in glycan composition and occupancy can be detected by cryo-EM and result in changes to glycan shield structure and dynamics that affect the accessibility of epitopes on the surface. To better inform our results we present site-specific mass-spectrometry data for the same samples and demonstrate that cryo-EM can complement such studies. Finally, by exposing Env to endoglycosidase digestion and capturing reaction intermediates with cryo-EM in a time-resolved manner, we found that highly connected glycan clusters are resistant to digestion and act to stabilize the pre-fusion trimer structure, implying the glycan shield may function beyond immune evasion.

Results

Env in complex with a base-specific antibody as a model system for cryo-EM analysis of the native glycan shield

We choose BG505 SOSIP.664 as a model system to demonstrate our approach because it is the first and most widely studied of the soluble pre-fusion stabilized Envs (23, 24). Native Env is a transmembrane protein that forms a trimer of dimers composed of two gene products, gp120 and gp41, which are responsible for receptor binding and membrane fusion respectively [add citation]. The acronym *SOSIP* is in reference to the engineered disulfide bonds between gp120 and gp41 and the isoleucine to proline mutation in the HR1 helix, which act to stabilize the pre-fusion conformation, while 664 is the final residue in truncated ectodomain (23). BG505 SOSIP.664 is also currently in the

102 first ever human clinical trials of a trimeric subunit-based HIV vaccine
103 (https://clinicaltrials.gov/ct2/show/NCT03699241). In Fig. 1 we present the cryo-EM reconstruction and refined
104 atomic model of BG505 SOSIP.664 purified from HEK293F cells in complex with three copies of the fragment antigen
105 binding domain (Fab) of the non-neutralizing antibody, RM20A3. This complex will be referred to as BG505_293F from
106 here on. The RM20A3 Fab was included because it improves orientation bias, as illustrated by the complete and near-
107 uniform angular distribution (Fig. 1*C*), and it binds to a non-native epitope at the base of the SOSIP trimer that is
108 devoid of glycans (Fig. 1*D*). This leaves the native glycan shield undisturbed and means that RM2OA3 Fab binding is
109 independent of glycosylation, making the SOSIP/base-Fab complex ideal for this study.
110 The global resolution of this C3-symmetric reconstruction determined by Fourier shell correlation (FSC) is
111 ~3.1Å (*SI Appendix*, Fig. S1*A*), and the bulk of the protein is at or near this resolution (Fig. 1*E*). ~80% of Env residues
112 could be confidently built into this map, while most glycans are ill-defined beyond the core N-acetylglucosamine
113 (NAG), accounting for only ~15% of the total predicted glycan mass (Fig. 1). Glycans located on the poorly resolved
114 flexible loops (N185e, N185h, N398, N406, N411, N462) could not be identified at all, however, site-specific MS data
115 confirmed they are indeed glycosylated (25) (*SI Appendix*, Fig. S2*A*). N-linked glycans on Env can be roughly classified
116 into two types, high-mannose and complex, based on the extent of intracellular processing, and they differ in
117 composition and structure (one possible configuration of each type is depicted in Fig. S 1F). Even if the glycans were
118 better resolved, averaging of individual projections during 3-D reconstruction means we cannot uniquely identify the
119 specific type of glycan at most sites because they possess a mixture of several different types (26) (*SI Appendix*, Fig.
120 S2*A*). Of those we could identify, a few were found to be near-uniformly processed (designated as >75% complex), for
121 example the glycans at N611 and N618 on gp41, and it should be possible to confirm this by the presence of a core
122 fucose residue. However, we were unable to identify clear density for fucose at any of these sites (Fig. 1*F*). Together,
123 these observations show that the glycan shield is highly dynamic with respect to the protein core but exhibits some
124 differences in ordering between glycans, and that it is not possible to distinguish between the different types of
125 glycans from a single cryo-EM map when the glycans are not stabilized by antibodies.
126
127 **Scale-space and 3-D variance analysis reveal interconnectivity and higher-order structure within the glycan shield**
128 Although we lack the resolution to build precise atomic models, there is still a great deal of information about glycan
129 shield structure and dynamics that can be gleaned from the cryo-EM data. At low isosurface thresholds noise appears
130 surrounding Env's well resolved protein core where the missing glycan mass should be (Fig. 1*E*). To look for meaningful
131 structure within this noisy signal we analyzed properties of the map across a range of scales by progressively
132 smoothing it with a Gaussian filter of increasing standard deviation (SD). Resolution in the glycan shield falls off with
133 distance from the protein surface, therefore as the map is smoothed more and more glycan signal should become
134 visible until the map encompasses the majority of the space occupied by the ensemble of configurations. This process
135 could be captured by measuring the volume of the map across the scale-space. To calculate the volume, however,
136 one must first choose an appropriate threshold. For our analysis we implemented an auto-thresholding method based

137 a simple measure of topological connectivity described in the *SI Appendix* (*SI Appendix*, Fig. S3*A-F*). We refer to the
138 resulting intensity as the *noise threshold* since it represents the lowest threshold before the appearance of noise. Fig.
139 2*A* shows the noise threshold and volume of the map at the noise threshold as a function of Gaussian filter width.
140 Both curves have a sigmoidal shape; once the map is sufficiently smoothed it undergoes a period of rapid expansion
141 in volume and reduction in noise threshold that is presumably capturing the appearance of signal from the poorly
142 resolved regions of the map. The curves then plateau around ~1.5-2 SD, suggesting there is negligible gain in signal
143 with further smoothing.

144 Examination of these filtered maps confirms the progressive emergence of more structure in the glycan
145 shield (Fig. 2*B*). By taking 2-D slices through six representative maps and displaying them on top of one another we
146 see how glycosylated surfaces undergo substantially more dilation (red arrows) than the non-glycosylated surfaces.
147 Once the isosurface has expanded to include signal from glycan antennae, neighboring glycans begin to merge,
148 creating a canopy-like effect that occludes underlying protein surfaces (black arrows). This implies that neighboring
149 glycans sample overlapping volumes and have the potential to interact. Also apparent is the negligible gain in signal
150 moving from 1.5 to 2 SD (blue and pink). We therefore concluded that the 1.5 SD Gaussian filtered map is the ideal
151 scale for interpreting glycan shield structure in this reconstruction. Visualization of this filtered map across a range of
152 thresholds reveals a hierarchy of structural features (Fig. 2*C*). At high thresholds, the glycan shield is composed
153 primarily of isolated and well-defined glycans stems, while the majority of the protein surface is exposed (gray map).
154 Moving to lower thresholds, it progresses through a series of higher-order structural states defined by isolated clusters
155 of multiple overlapping glycans, until finally becoming completely interconnected near the noise threshold. This low
156 threshold structure occludes large regions of protein surface and represents the full extent of the glycan shield.

157 As a complementary approach for visualizing the dynamic glycan shield, we performed 3-D variability analysis
158 in the SPARX software package (27, 28) (Fig. 2*D*). 3-D variability, which is closely related to 3-D variance, should be
159 high anywhere there is significant heterogeneity in the map. Indeed, we see high variability around known flexible
160 regions such as the constant domains of the Fabs (clipped from view) as well as variable loops and the exterior of the
161 protein surface at the sites of N-linked glycans. Variability in the glycans is highest at the distal ends of the glycan
162 stems and expands outward as the threshold is reduced, revealing extensive interconnectivity (Fig. 2*D*). 2-D slices
163 through the variability map highlight the canopy effect between neighboring glycans, which acts to occlude underlying
164 protein surfaces (red arrows - Fig. 2*D*). Overall the shape and topology are similar to the Gaussian filtered map. A
165 more complete description of the 3-D variability can be found in the *SI Appendix*.
166

167 **A high-throughput atomistic modeling (HT-AM) pipeline for generating large ensembles of glycosylated Env**

168 Cryo-EM does not capture atomic details of individual molecules, so to better understand the results presented above,
169 we performed atomistic simulations of fully glycosylated BG505 SOSIP.664. To overcome the sampling limitations of
170 MD simulations, we employed a high throughput atomistic modeling (HT-AM) pipeline based around the ALLOSMOD
171 package of the MODELLER software suite (29), which has been used previously to generate small ensembles of

172 glycosylated Env (30). This methodology accounts for both energetic and spatial constraints on glycan sampling by a
173 combination of empirical energy minimization-based structural relaxation and simulated annealing (described in detail
174 by Guttman et al. (29), along with initial randomization of glycan orientations. However, it does not provide temporal
175 information about dynamics. Here, we have built on this by incorporating 10 different protein scaffolds and additional
176 variability in the V2 and V4 loops (Fig. 3*A*) and utilized it to generate much larger ensembles of 1000 fully glycosylated
177 Env structures with uniform mannose-9 (Man9) glycans. Ten such models are shown in Fig. 3*C*, one from each protein
178 scaffold, while the full set of models at a single glycosylation site is shown in Fig. 3*D*. We also repeated the simulation
179 with uniform mannose-5 (Man5) glycosylation for comparison.
180
181 **The HT-AM pipeline capture spatial and energetic constraints on glycan flexibility**
182 To assess how spatial and energetic constraints affect the flexibility of individual glycans in our simulations we
183 calculated the root mean-squared fluctuation (rmsf) for each glycan across all 1000 models after aligning the protein
184 scaffold. Our method captures variability between glycans (Fig. 3*E*), and the additional flexibility imparted by the 10
185 starting protein scaffolds can be appreciated by comparison to the glycan only rmsf (Fig. 3*F*). For example, glycans
186 located on the flexible loops have a much higher rmsf than all other glycans (185e, 185h, 406 and 411) whereas
187 glycans at the N262, N301, N332, N448, and N611 sites have lower rmsf. This leads to a large difference in sampled
188 volumes between the most and least dynamic glycans (*SI Appendix*, Fig. S4*A*). We also see an increase in average rmsf
189 of the individual glycan residues starting from the core NAG and moving outward to the tips of each antenna (*SI*
190 *Appendix*, Fig. S5*A-C*), which is in line with the cryo-EM results showing reduced resolution beyond the core NAG (Fig.
191 1*D*).
192 We observed a similar trend for the Man5 ensemble (*SI Appendix*, Fig. S5*G-H*), however, by comparing the
193 average rmsf values of the 5th mannose residue alone, we found a slight increase in rmsf and sampled volume at most
194 sites compared to the Man9 ensemble (Fig. 3*G* and *SI Appendix*, Fig. S5*I-J*). We attribute this effect to increased
195 crowding in the glycan canopy from the more massive Man9 glycans. In support of this, there is a significant positive
196 correlation between glycan flexibility glycan crowding (see *SI Appendix*), however, only when considering relatively
197 large neighborhoods (*SI Appendix*, Fig. S6*A*). A similar trend was observed by Stewart-Jones et al. which they attributed
198 to different "shells" of influencing glycans (6). Such higher-order dynamic effects could be possible in light of our
199 structural observations.
200 In addition to crowding effects, the local protein structure will also influence glycan dynamics. In our
201 modeling pipeline, the protein backbone was kept harmonically restrained close to the template to allow for extensive
202 sampling of glycan conformations using simulated annealing. Thus, we see that the Asn sidechains of residues 88, 160,
203 197, 234 and 262 all have very low rmsf (*SI Appendix*, Fig. S5*F*), possibly stemming from limited torsional space
204 available during modeling. The glycosylated Asn residues in gp41 have relatively low rmsf as well (N611, N618, N625,
205 and N637), being situated on stable helical bundles (*SI Appendix*, Fig. S4*B*). This ultimately results in a relative reduction
206 of the glycan dynamics at some of these sites (Fig. 3*C*). Correcting for the contribution to fluctuations coming from

207 the underlying protein, we observed that the rmsf between the different glycans are comparable, ranging from 3 Å to
208 5 Å, with similar scale of SD (Fig. 3*F*).
209
210 **Simulated cryo-EM maps reproduce defining features of the experimental data**
211 Given the single-molecule nature of cryo-EM datasets, the ensembles generated by the HT-AM pipeline can be seen
212 as representing the individual particles that go into a 3-D reconstruction. With that in mind, we established a protocol
213 for transforming the ensembles into simulated cryo-EM datasets (Fig. 4*A*). For comparison, we replicated the process
214 for the Man5 ensemble as well as a protein-only ensemble (BG505_Man9, BG505_Man5 and BG505_PO respectively).
215 The simulated maps reproduced some of the defining features of the experimental data. For instance, refinement was
216 dominated by the stable protein core and only the first few sugar residues at each site are defined when the map is
217 filtered to the global FSC resolution (Fig. 4*A*). This leads to a similar scale-space structure (Fig. 4*B*), with a plateau again
218 appearing around 1.5-2 SD when the maps are filtered to the same initial resolution, suggesting that the HT-AM
219 pipeline is capturing physiologically relevant sampling in the glycan shield, at least globally. Importantly, we found that
220 the volume of the 1.5 SD Gaussian filtered map at the noise threshold closely approximates the total volume sampled
221 by the ensemble (Fig. 4*B* - dashed lines). Comparing the curves for all three simulated reconstructions (Fig. 4*B*) we see
222 how differences in glycosylation manifest as differences in volume at low resolutions, suggesting cryo-EM can be used
223 to capture global changes in glycan shield composition between two reconstructions of the same Env. Visualization of
224 the 1.5 SD Gaussian filtered BG505_Man9 map at high and low thresholds reveals a similar evolution towards a more
225 connected topology as the threshold is reduced and extensive shielding of protein surface (Fig. 4*C*). We also calculated
226 the 3-D variability map for the BG505_Man9 reconstruction and observed a similar result to the experimental data
227 (Fig. 4*D*). In addition, we compared the SPARX variability map to the true 3-D variance (*SI Appendix*, Fig. S7*B*) and
228 found negligible differences between the two (See the *SI Appendix* for details).
229
230 **Measuring glycan dynamics in cryo-EM maps**
231 Measuring intensity around individual glycans in the cryo-EM maps should allow us to assess their relative dynamics.
232 To confirm this, we used the simulated BG505_Man9 map since we could make direct comparison to the rmsf values.
233 However, only the first 1-2 glycan residues are defined at most sites in the high-resolution map (Figure5A) and their
234 dynamics deviate significantly from the rest of the glycan (*SI Appendix*, Fig. S8*A*). The third glycan residue, β-mannose
235 (BMA), more closely approximates the average rmsf (mean deviation ~2Å) and the relative differences between sites
236 (*SI Appendix*, Fig. S8*A*-*B*). In addition, the glycan stems are well defined in the 1.5 SD Gaussian filtered map, which
237 allowed us to identify their locations in the map with reasonable accuracy. So, we built and relaxed glycan stems into
238 the 1.5 SD Gaussian filtered BG505_Man9 simulated map (Fig. 5*A*) at every glycosylation site we could confidently
239 identify (23/28 per protomer). We could not identify the V2 and V4 loop glycans at N185e, N185h, N406, and N411,
240 which had the highest rmsf values in the simulation, nor the glycan at N339, which has a lower rmsf but projects
241 directly towards the heterogeneous V4 loop. The positions of the BMA residues were then used to analyze local map

242 statistics within a spherical probe around each glycan (Fig. 5*A*). We observe a strong positive correlation between the
243 inverse rmsf and normalized mean intensity (Fig. 5*B*-*C*), with the peak correlation coefficient of ~0.88 (p=2.1e-8)
244 occurring when using the unfiltered map and a probe radius of 2.3Å. Thus, local intensity around BMA residues
245 accurately captures ground-truth differences in relative dynamics between glycans and can be used to make direct
246 comparisons between the simulated and experimental cryo-EM maps.
247
248 **The HT-AM pipeline reproduces physiologically relevant trends in glycan dynamics measured by cryo-EM**
249 With a method in place for measuring glycan dynamics from cryo-EM maps, we then made direct comparisons with
250 the experimental data. We built and relaxed glycan stems into the 1.5 SD Gaussian filtered BG505_293F map as
251 described above and could identify clear density at 21/28 glycosylation sites per monomer, two less than from the
252 simulated map (N398 and N426). The other five missing glycans were common to both maps, meaning the HT-AM
253 pipeline captured physiologically relevant dynamics at these sites, at least up to the detection limits of this method.
254 Overall, we found that the HT-AM pipeline captured a similar trend in glycan dynamics with a correlation coefficient
255 between the two of ~0.46 (p = 0.03) (Fig. 5*D*).
256 In addition to the V4 and V5 loop glycans at N398 and N462, we also saw a large deviation at the N137 glycan
257 on the V1 loop, as well as the N262 and N301 glycans. In the BG505_293F map, the N262 glycan was the most ordered
258 due to stabilizing contacts with the gp120 core, and these interactions may not have been accurately captured by the
259 simulation given the restricted protein dynamics. In gp41, a large deviation also occurred at the N611 glycan, which
260 we attribute to restrictive sampling of the protein backbone as previously discussed. To complicate the comparison,
261 the N618 and N625 sites are under-occupied as revealed by MS (*SI Appendix*, Fig. S2*A*). Although sub-occupancy will
262 cause reduced signal intensity due to averaging, the severity of this effect is diminished when dynamics at that site
263 are large relative to other glycans, which is the case for the glycan at N625.
264
265 **Detecting site-specific changes in glycan dynamics, occupancy and chemical composition from cryo-EM maps.**
266 To test whether these methods could be used to detect site-specific changes in dynamics and occupancy between
267 cryo-EM maps of differentially glycosylated Env, we performed a comparative analysis between the BG505_Man9 and
268 BG505_Man5 simulated maps. Given that stems of a Man9 and Man5 glycan are identical, the only changes in
269 intensity around the BMA residue should arise from differences in dynamics alone. On average, we see a ~17%
270 reduction in intensity indicative of increased dynamics, which is in line the rmsf data (Fig. 6*A*). We also accurately
271 detect the largest increase and decrease in dynamics at the N262 and N234 sites respectively. To verify we could
272 detect changes in occupancy, we removed the glycan at the N625 site from half of the models and re-refined the data
273 (referred to as BG505_Man9HO for "half occupancy"). Not surprisingly, we see a ~50% reduction in mean intensity
274 from the fully occupied reconstruction around this site (Fig. 6*A*).
275 Another technique that should be sensitive to subtle changes between two cryo-EM maps is difference
276 mapping. Indeed, the change in occupancy at the N625 site is apparent in the BG505_Man9 - BG505_Man9HO

277 difference map (Fig. 6*B*). At high threshold, the signal is localized around the glycan stem and extends to the protein
278 surface. In the BG505_Man9 - BG505_Man5 difference map (Fig. 6*C*), however, the difference signal is strongest
279 where the distal tips of the Man9 glycans would be and does not extend to the protein surface. These results establish
280 cryo-EM as a tool for measuring glycan dynamics as well as changes in chemical composition and occupancy between
281 differentially glycosylated Envs.
282
283 **Insights gained from analysis of simulated data allows improved characterization of cell-type specific differences**
284 **in glycan shield**
285 To test our methods experimentally, we collected cryo-EM data on complexes of RM20A3 Fab and BG505 SOSIP.664
286 expressed in two additional common cell lines that produce major and minor changes in glycosylation; HEK293S cells,
287 and a stable CHO cell line (referred to as BG505_293S and BG505_CHO respectively). The stable CHO cell line
288 expressed Env sample was provided to us by IAVI as part of the demo run conducted prior to the currently on-going
289 human clinical trials, and should be identical to the final vaccine product (31). Both new datasets refined to ~3Å-
290 resolution (*SI Appendix*, Fig. S1*A*) and the three atomic structures are nearly identical (Cα RMSDs ~ 1Å - *SI Appendix*,
291 Fig. S1*B*). Scale-space (*SI Appendix*, Fig. S3*G-I*) and 3-D variability analysis (*SI Appendix*, Fig. S9*A-C*) also revealed similar
292 low-resolution structural features to BG505_293F, however, comparison of local map intensity around BMA residues
293 uncovered some significant difference in the gp41 glycans, particularly at the N611 and N625 sites, potentially from
294 changes in occupancy (Fig. 7*A*). This is further supported by site-specific MS data collected for the two additional
295 samples (*SI Appendix*, Fig. S2*A*).
296 The 293S sample contains only high-mannose type glycans and should therefore allow detection of complex
297 glycans via difference mapping, and to a lesser extent, different distributions of high-mannose glycans. Indeed, we
298 see strong difference signal around the primarily complex gp41 glycans in both the difference maps (Fig. 7*B-C*). We
299 also observe signal around the primarily complex glycan at N88 (Fig. 7*F*). In addition, clear difference signal appeared
300 around the V2, V4, and V5 loops, specifically near the glycans at N185e and h, N398, and N462, all of which are shown
301 to be complex by MS. The MS data also shows a subtle difference in both occupancy and percentage of complex
302 glycans at the N398 site between the 293F and CHO sample, which could explain why there is strong difference signal
303 in one map and not the other. According to the MS data, the only significant differences in occupancy between the
304 293F and 293S samples occurred at the N137, N133, and N611 sites, however the difference maps do not contain any
305 signal around the N611 glycan, suggesting there may be discrepancies between the two methods.
306 The MS analysis also detected differences in occupancy between the CHO sample compared to the 293F
307 sample at multiple sites (*SI Appendix*, Fig. S2*A*). Indeed, upon closer examination we found that the difference signal
308 around the gp41 glycans at N611, N618, and N625 extends all the way to the protein surface (Fig. 7*G*), indicative of
309 changes in occupancy. In this difference map, there is also clear signal around the N137 site, and to a lesser extent at
310 the tip of the N332 glycan stem (Fig. 7*H*). Given the proximity of N137 and N332, it is plausible that sub-occupancy at
311 one is driving changes in dynamics and/or glycan distribution at the other. Similar higher-order effects have also been

observed via MS when comparing glycoform distributions on Env before and after knocking out specific glycans (26, 32, 33).

A probabilistic glycan-glycan interaction network reveals highly connected glycan clusters

Both the experimental and simulated cryo-EM maps showed extensive interconnectivity among glycans, so we sought to quantify this more precisely. With each glycan in our models sampling a particular region of space, neighboring glycans can explore overlapping volumes, and the fraction of this overlap gives a measure of their interaction probability. Fig. 8*A* shows a heat map of the normalized glycan-glycan volume overlap matrix. By interpreting this as an adjacency matrix we can model the glycan shield as a network. Here, each glycan was defined as a node, and two nodes were connected by an edge if there was at least 5% overlap between them (Fig. 8*B*). The only glycans from the neighboring protomers that were considered are those with an inter-protomer edge. Fig. 8*C* shows the network obtained for the Man9 ensemble unfolded in 2-D. It can be seen that overall there are three main regions of overlap, – the V1/V2 apex, the gp41 base, and the densely occupied gp120 outer and inner domains that includes the high-mannose patch (HMP). The inter-protomer overlaps are contributed mainly by the V1/V2 glycans. The nomenclature was inspired by the established nomenclature for gp120 domains (34), however, our definitions were adapted to better capture glycan shield structure (*SI Appendix*, Fig. S10*A*). The analysis was repeated for the Man5 ensemble and we observed a reduced overall connectivity as measured by the mean node degree (from ~7 to ~5) and the maximum network diameter (from 5 to 8 hops), consistent with their smaller size (*SI Appendix*, Fig. S*11*).

With a network in place we could then analyze the relative influence of each glycan on the whole system and examine its long-range structure. To do this we calculated the relative eigenvector centrality of the nodes, which is a measure of importance in the network, and the results are projected on the network as a colormap in Fig. 8*C*. We see that the gp120 outer domain glycans at N332, N339, N363, N386, and N392, which are densely connected within the network, have strong eigencentrality measures. In the apex region, the V1/V2 loop glycans at N133, N160 and N185e/h also have relatively high eigencentrality due to their increased flexibility, while the glycans at N88, N234, N276, and those in gp41 have low eigencentrality, reflecting low interaction probabilities. Incorporated intrinsically into the network is a set of stable sub-graphs that represent highly connected glycan clusters. To illustrate this structural hierarchy, we progressively stripped the network using tighter overlap cutoffs (*SI Appendix*, Fig. S10*B*). As the network is degraded, we see the formation of two large sub-graphs; one composed of the V1/V2 apex and the gp120 outer domains; and the second composed of the gp120 inner domain along with gp41 base glycans. With an even stricter cutoff, the sparsely connected glycans with low eigencentrality separate out and the two sub-graphs split again to form four sub-graphs; a V1/V2 apex domain; a gp120 outer domain; a gp120 inner domain; and lastly, the group of unconnected gp41 base glycans.

Highly connected glycan clusters are resistant to enzymatic digestion and help stabilize of the pre-fusion trimer

346 To determine if the glycan shield affects Env structure and dynamics we exposed BG505_293S (already in complex
347 with RM20A3 Fab) to digestion by endoglycosidase H (Endo H), which cleaves high-mannose glycans between the first
348 and second sugar rings (Fig. 9*A*), and collected cryo-EM data on the deglycosylated sample. In addition, we
349 hypothesized that highly connected glycans would be protected from enzymatic digestion, and vice versa. Therefore,
350 if we could measure the order in which glycans were cleaved by Endo H it could provide indirect validation of our
351 network models. To test this, we collected cryo-EM data on a partially deglycosylated sample as well. The reaction
352 times and Endo H concentrations for achieving partial (2hrs) and complete (16hrs) digestion were determined prior
353 to performing cryo-EM. The datasets, referred to as BG505_EndoH2 and BG505_EndoH16, reconstructed to ~3.2Å
354 and ~3.5Å-resolution respectively, with similar overall quality, resulting in nearly identical atomic models to the other
355 three datasets (*SI Appendix*, Fig. S1*A-B*). The progressive digestion of the glycan shield was confirmed by SDS-PAGE
356 and SEC chromatography (*SI Appendix*, Fig. S12), while scale-space analysis of the cryo-EM maps revealed a clear
357 progressive reduction of glycan signal (*SI Appendix*, Fig. S3*G-I*).

358 Indeed, analysis of local map intensity around each glycan revealed that digestion occurred non-uniformly
359 (Fig. 9*B-C*). This was also confirmed by changes in the 3-D variability maps (*SI Appendix*, Fig. S9*D-E*). By assuming a
360 linear relationship between intensity and occupancy and using the MS data to set the initial occupancy, we calculated
361 the occupancy at each site (Fig. 9*C*). After 2 hours the gp41 glycans (N611-637) were completely digested, while some
362 glycans, particularly those in the densely packed gp120 outer domain, remained mostly intact. Although we could not
363 identify unique signal for the individual V2, V4, and V5 loop glycans, there was clear signal around these sites in the
364 0-2hr difference map only (Fig. 9*B*), indicating the dynamic V-loop glycans were completely digested within 2hrs. We
365 also found partial signal reduction at a few sites, indicative of non-uniform digestion. For example, the apex cluster
366 composed of the N156 and N160 glycans as well as the glycans at N133, N197, and N234. After 16 hours the glycan
367 shield was almost completely digested, however, we still detected some residual glycan signal at the previously
368 discussed cluster composed of the N363, N386, and N197 glycans, as well as a cluster composed of the N295, N332,
369 and N448 glycans. In addition, the highly protected glycan at N262 remained completely intact. By quantifying the
370 degree of protection from Endo H (see Methods) and comparing it to the predicted network eigencentralities for the
371 Man9 ensemble (Fig. 9*E*), we obtained a correlation coefficient of ~0.8 (p=1.14e-05), suggesting highly connected
372 glycans are resistant to enzymatic digestion. Also evident is the similarity between the persistent glycan clusters and
373 the sub-graphs generated by progressively stripping the glycan overlap network (*SI Appendix*, Fig. S10*B*), which could
374 be seen as mimicking the gradual digestion by Endo H.

375 Interestingly, 3-D classification of the Endo H treated datasets revealed an increasing degree of protein
376 unfolding and subunit dissociation initiating at the V1-3 loops in the trimer apex. Classification consistently converged
377 to four 4 classes that appeared to capture found heterogeneous intermediates in the unfolding process (Fig. 10*A* and
378 *SI Appendix*, Fig. S13-17). We interpreted these as representing (1) stable trimers, (2) trimers with unfolded or partially
379 unfolded V1-V3 loops, (3) trimers with fully unfolded/dissociated gp120s, and (4) monomers/dimers. As the reaction
380 progressed, the percentages of unfolded trimers increased, while the percentage of stably folded trimers decreased

381 (Fig. 10*B*). Although it cannot be easily confirmed, the unfolded trimers in each dataset were likely more completely
382 deglycosylated than the particles that make up the stable trimeric classes. We also observed a small percentage of
383 dissociated gp120s in the CHO sample, perhaps as a result of the higher degree of sub-occupancy. When viewed in
384 context of the results presented above, we concluded that the highly connected glycans and glycan clusters that were
385 resistant to digestion are also critical to maintaining structural stability of the pre-fusion Env trimer, suggesting the
386 glycan shield may have additional functions beyond its role in immune evasion.

387

388 **Discussion**

389 Prior to this study, there were conflicting reports of glycan shield structure from cryo-EM and X-ray
390 crystallography measurements (6, 14, 35). Our results clearly show that the native glycan shield is highly dynamic
391 with respect to the protein core. We did not observe any stabilized glycan-glycan interactions like those reported in
392 crystal structures (6), suggesting methodological differences may be responsible for the apparently contradictory
393 results. For example, the crystal structures included two bound bnAbs per monomer to facilitate crystal packing, both
394 of which engage multiple glycans and disrupt the native dynamics and higher-order structure. Desolvation of mobile
395 waters embedded within the glycan shield during crystallization could also potentially induce the stabilized glycan-
396 glycan contacts observed. A favorable interpretation that also fits with the conclusions of this paper is that the crystal
397 structures captured artificially stabilized yet physiologically relevant conformations and are thus suggestive of the
398 types transient, of non-covalent interactions possible between glycans.

399 Our work represents the first case of using cryo-EM to validate glycoprotein ensembles, and in turn, their
400 use in guiding cryo-EM analysis. We found that the HT-AM pipeline captured physiologically relevant glycan
401 sampling at global and local scales, however, deviated from the experimental observations when the flexibility of the
402 underlying protein was greater than could be captured by the current pipeline. Thus, exploring methods to enhance
403 sampling of the protein backbone, particularly at Env variable loops, represents a clear route for improving the
404 accuracy of the pipeline (36). Another logical next step is to use the experimental cryo-EM maps to steer the
405 ensemble modeling process, which has been successful for non-glycosylated proteins (37-39), however, extending
406 such techniques to capture the extreme levels of heterogeneity present in the glycan shield will be challenging.
407 Experimentally, our ability to capture differences in glycan dynamics from cryo-EM maps is limited when using local
408 map intensity around BMA residues alone, and such limitations could also be contributing to the observed
409 discrepancies. Again, using the cryo-EM maps to steer the modeling process could provide more accurate and
410 comprehensive assessments of variations in relative dynamics between glycan in addition to improving modeling
411 accuracy. Finally, utilizing sequence engineering and/or different expression systems to achieve a more uniformly
412 glycosylated Env sample, or conversely, incorporating experimentally relevant levels of heterogeneity into the
413 modeling pipeline will allow for a more accurate comparison between theory and experiment.

414 Using graph theory, we mapped out the complex network of potential interactions within the glycan shield.
415 A similar network-based approach was introduced previously to capture concerted behavior within the glycan shield
416 during all-atom MD simulations (22). Although the HT-AM pipeline does not capture temporal dynamics, we found
417 that the probabilistic networks reported here closely resembled the ones derived from the MD simulations, and the
418 two lead to similar predictions. For instance, the clustering of glycans into larger microdomains (*SI Appendix*, Fig. S10*A-*
419 *B*). Moreover, the overlap probability between glycans gives a measure of the structural shielding rather than the
420 energetic interactions between glycans. At a qualitative level, the hierarchy of structural features we observed in the
421 Gaussian filtered and 3-D variability maps (Fig. 2*C-D*) could be seen as the ensemble-average manifestation of this
422 clustering behavior. Given the preference of bnAbs for targeting the interfaces between higher-order glycan clusters
423 (22), experimentally mapping such long-range structure could be important. At a more quantitative level, we found
424 that the average degree of protection from Endo H digestion correlated strongly with network eigencentrality for
425 those glycans we could identify in the cryo-EM maps, and that the order of digestion closely matched the order in
426 which glycans were removed from the network when applying stricter overlap thresholds. It should be noted,
427 however, that because of the somewhat simplified approach we have taken to modeling the most dynamic segments
428 of the variable loops, in particular in the V4 loop segment containing the N398, N406, and N411 glycans, the
429 eigencentrality of these glycans may be artificially high, leading to minor discrepancies between digestion
430 susceptibility and network centrality. Nonetheless, these correlations could be seen as providing indirect validation of
431 the network models, where the high centrality glycans are either located in densely packed regions where Endo H
432 would have a difficult time accessing (*SI Appendix*, Fig. S6*D-E*) or are protected by strong interaction probability with
433 their neighbors.
434 We demonstrated that cryo-EM is capable of detecting and quantifying changes in glycan dynamics, as well
435 as occupancy and chemical composition between differentially glycosylated Envs, the latter of which were exclusively
436 provided by MS prior to this study. Thus, cryo-EM provides validation for these measurements while contributing
437 insight into the structural impact of changes in glycosylation. For instance, semi-quantitative methods to assess the
438 distribution of unoccupied sites have only recently become available (13), and here we showed that cryo-EM is highly
439 sensitive to occupancy. Importantly, we found evidence for more extensive changes in glycosylation between the
440 293F and CHO expressed samples than detected by MS, which has clear relevance to the currently on-going human
441 clinical trials based on the CHO expressed Env. In addition to these results, we found that the extent of processing
442 measured by the percentage of high-mannose type glycans at each site was strongly correlated with ordering in the
443 cryo-EM map (*SI Appendix*, Fig. S2*C-D*), suggesting it may be possible to predict the degree of processing at each site
444 from cryo-EM maps alone. The relationship between these variables likely reflects their mutual dependence upon
445 glycan crowding (*SI Appendix*, Fig. S6*B-C*), as well as the properties surrounding protein structure, both of which can
446 reduce dynamics and restrict the access of Endo H and glycan processing enzymes in the cell.
447 The observation that enzymatic deglycosylation leads to progressive destabilization of the Env trimer was
448 somewhat surprising because recent investigations into the effect of glycan knockouts and deglycosylation on Env

449 stability and viral infectivity have come to different conclusions (40-44). However, there is ample biophysical
450 evidence that glycans influence protein stability, dynamics, and folding, even in the case of Env (17, 45-56),
451 although it is not obvious which mechanisms are important here. One possible mechanism is stabilizing interactions
452 between the core NAG and neighboring side chains, which are observed throughout Env and are common in other
453 glycoproteins. However, Endo H leaves the core NAG attached, so the stabilizing effect must arise by other
454 mechanisms. It has been proposed that glycans stabilize proteins primarily by destabilizing the unfolded state and by
455 increasing the height of the unfolding energy barrier (47). The same study also found that glycosylation reduced the
456 entropy of the folded state, possibly by imposing pressure on the protein and dampening dynamics, similar to the
457 effects of molecular crowding or confinement, although the effect was canceled out by an equivalent change in
458 enthalpy. Along these lines, the slightly lower resolution of the BG505_EndoH16 map compared the fully glycosylated
459 maps could be indicative of increased dynamics. In addition, our observation that the Man9 glycans were less dynamic
460 than the smaller Man5 glycans in our simulations could be attributed to more crowding in the canopy, which in turn
461 could further stabilize the underlying protein. Also, the stabilizing effects were found to depend strongly on the
462 number of glycans as well as their positions on the protein (47), which is in line with the results of our Endo H
463 digestion experiment. Therefore, given the high density of glycans on the surface of Env and the importance of their
464 tertiary arrangement into high density clusters, it is plausible these stabilizing effects could be amplified. Furthermore,
465 glycan-glycan interactions such as those observed in the crystal structures solved by Stewart-Jones et. al. (6) could
466 also play a role in stabilizing folded state. Although our results clearly show that such interactions must be more
467 dynamic than the crystal structures suggest, the combined effect of many such short-lived interactions could become
468 significant.

469 The time-resolved Endo H digestion experiment provide an explicit mapping of the importance of each glycan
470 toward maintaining the integrity of the glycan shield and the overall stability of Env. In addition to the glycan at N262,
471 the two highly connected clusters in the gp120 outer domain centered around the N386 and N295 glycans were the
472 most important for trimer stability, and to a lesser extent the apex cluster composed primarily of the N156 and N160
473 glycans. In support of this, previous studies have shown that knocking out the glycans at N386 or N262 individually
474 have dramatic effects on trimer stability (33), whereas deletion of up to 5 glycans at once peripheral to the CD4bs,
475 which are not part of these cluster, are well tolerated (10, 57, 58). In addition, the glycans at N386 and N295 in
476 particular were found to strongly control the processing of the other glycans near them, reinforcing their importance
477 in maintaining the integrity of large-scale glycan shield structure (33). These results have implications for rational
478 vaccine design where removing certain glycans could have undesired effects on protein stability and suggest that
479 there are limits to the number glycan deletions that can be tolerated, with particular emphasis on maintaining the
480 integrity of the aforementioned clusters. In addition, we found that instability originated at the trimer apex, around
481 the V1-3 loops. This region of the structure is known to be metastable and to change conformation upon CD4 and co-
482 receptor binding(34, 59). Therefore, it is plausible the glycans and glycan clusters in this region are contributing to

483 maintaining Env in a metastable state poised for receptor binding. Given the high conservation of the glycan clusters
484 identified here, our results are likely universal to all Env sequences. Lastly, it is possible that glycan induced
485 stabilization of Env may enable hyper-mutation of the underlying protein surface, thereby endowing Env with the
486 ability to escape immune pressure with otherwise deleterious mutations.

487 It is known that immune responses to Env preferentially target glycan-depleted surface area (9, 10), and
488 the results presented here provide the first experimentally determined mapping of this surface. Under reasonable
489 physical assumptions, the ensemble-averaged structure captured by cryo-EM should be identical to the time-averaged
490 structure of a single thermally fluctuating Env molecule at the temperature prior to vitrification. The dynamics of the
491 glycans create a cloud/shield over the protein, thus an approaching antibody with relatively torpid dynamics will
492 effectively "see" this blurry ensemble-average structure, which is consistent with the relatively slow on-rates of most
493 HIV bnAbs. We quantified the shielding effect with a simple "rolling sphere" method for BG505_CHO and visualized
494 the results as a colormap projected onto the surface of Env (Fig. 11). White surfaces are strongly masked by nearby
495 glycans and red surfaces are weakly masked, highlighting sites of potential vulnerability to neutralizing antibodies (see
496 the *SI Appendix* for details). Although the cryo-EM maps presented here do not perfectly capture the in-vivo structure
497 of a single fluctuating Env because they are ensemble averages of chemically heterogeneous molecules, if one
498 considers the net serum response to a vaccine that itself contains the exact same heterogeneity as the sample we
499 analyzed here (60, 61), they still provide an accurate representation of the average surface exposure of the entire
500 ensemble.

501 Looking beyond the results of this paper, our integrated approach can be easily extended to other Env clones
502 and to heavily glycosylated spike proteins from other viruses such as Influenza, Ebola, Lassa, and Coronaviruses, and
503 represents a potentially powerful approach for studying the structure and dynamics of glycoproteins in general.
504

505 Materials and Methods

506 BG505 SOSIP.664v3 was expressed and purified from HEK293F and HEK293S suspension cell culture in house while
507 the CHO cell-line derived BG505 SOSIP.664 was provided to us by IAVI and expressed and purified as described in
508 Dey, et al., 2017 (31). The monoclonal antibody RM20A3 was isolated from a BG505 SOSIP.664 immunized rhesus
509 macaque(62) and the Fab was expressed and purified from HEK293F cells as described previously(63). The BG505
510 SOSIP.664 samples expressed in HEK293F and CHO cell lines were prepared for MS analysis as described
511 previously(25), and the HEK293S sample was prepared with slight modifications on that protocol. All samples were
512 prepared for cryo-EM analysis via the same protocol after being complexed with RM20A3 Fab. Imaging was
513 performed on either an FEI Titan Krios or Talos Arctica (Thermo-Fisher) microscope. All non-custom cryo-EM data
514 processing steps were performed using a combination RELION-2/3(64, 65) and CryoSparc v1(66). Model building
515 and refinement was carried out with UCSF Chimera (67), Coot (68, 69), and Rosetta(70). All custom analysis of
516 cryo-EM maps were performed in MatLab (2018b) {MATLAB and Image Processing Toolbox Release 2018b, The

MathWorks, Inc., Natick, Massachusetts, United States.} 3-D variability analysis was performed in SPARX(27). The HT-AM was carried out by implementing the ALLOSMOD(29, 71) package of MODELLER(72, 73) in a streamlined pipeline. All graph theory and network-based analysis were performed using Python {Python Software Foundation, https://www.python.org/} and Matlab_R2018a packages. Simulated cryo-EM data generation and analysis was performed in RELION. All cryo-EM map and model visualization was performed with UCSF Chimera. A comprehensive description of all the methods can be found in the *SI Appendix*.

Acknowledgements

Z.T.B would like to acknowledge Bill Anderson and Hannah Turner for assistance with microscopy, Charles Bowman and JC Ducom for computational support, Pawel Penczek, Francisco Asturias, Jack Johnson, and Gabe Lander for helpful discussions on 3-D variance analysis and cryo-EM data processing, Jesper Pallesen and JH Lee for early support with data processing, Jeffrey Copps for help with protein expression and purification, Lauren Holden for critical reading of the manuscript, and finally IAVI, Antu Dey, and John P. Moore for supplying demo run material and MS analysis. S.C and S.G. acknowledge Timothy Travers for proposing ALLOSMOD for this work, Nicolas Hengartner for helpful discussions on graph theory and Bette Korber and Kshitij Wagh for insights on glycosylations in HIV. S.C. and S.G. acknowledge the LANL High Performance Computing Division for providing computational facilities. We thank Robin Sung Kyu Park and Titus Jung for help with analysis of MS data for site specific glycosylation.

Additional Information

Competing Interests

The authors declare no competing interests.

Funding

This work was supported by the National Institute of Allergy and Infectious Diseases grants: UM1 AI100663 (to A.B.W., J.C.P.), UM1 AI144462 (to A.B.W., J.C.P.), AI113867 (to J.C.P.), and P41GM103533 (to J.R.Y.). This work was supported by the Bill and Melinda Gates Foundation through the Collaboration for AIDS Vaccine Discovery (CAVD) grants OPP1100639 and OPP1115782 (A.B.W.). This work was supported by the IAVI Neutralizing Antibody Center through the Collaboration for AIDS Vaccine Discovery grant OPP1196345 funded by the Bill and Melinda Gates Foundation (A.B.W.). This work was funded by grants from the NIH (Center for HIV/AIDS Vaccine Immunology and immunogen Discovery, UM1 AI100645; Consortia for HIV/AIDS Vaccine Development, UM1 AI144371). S.C. was also partially supported by the Center for Nonlinear Studies (CNLS) at Los Alamos National Laboratory (LANL).

550 **Author contributions**

551 Conceptualization, Z.T.B, S.C., C.A.L, S.G. and A.B.W.; Methodology, Z.T.B, S.C., C.A.L, X.W., J.C.P., S.G. and A.B.W.;
552 Protein expression and purification, Z.T.B., C.A.C, and J.L.T; Antibody isolation, M.J.vG; Mass-spectrometry
553 experiments and data analysis, X.W, J.K.D, J.R.Y., and J.C.P.; Cryo-EM experiments, Z.T.B.; Modeling and Simulations,
554 S.C.; Formal Analysis, Z.T.B. and S.C.; Visualization, Z.T.B. and S.C.; Manuscript Preparation, Z.T.B, S.C., C.A.L, J.C.P.,
555 S.G. and A.B.W.; Supervision, S.G. and A.B.W.

556

557 **Data Availability**

558 All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary
559 Information. Additional data related to this paper may be requested from the authors. The electron potential maps
560 for BG505_293F, BG505_CHO, BG505_293S, BG505_EndoH2, and BG505_EndoH16 have been deposited in the
561 Electron Microscopy Data Bank under accession codes 22108, 22109, 22110, 22111, and 22112 respectively, along
562 with all corresponding half-maps, masks, and SPARX 3-D variability maps. Atomic models have been deposited in the
563 Protein Data Bank under accession codes 6X9R, 6X9S, 6X9T, 6X9U, 6X9V.

564

565 ## References

566 1. Fenouillet E, Gluckman JC, Jones IM (1994) Functions of HIV envelope glycans. *Trends in Biochemical*
567 *Sciences* 19(2):65–70.
568 2. Grundner C, et al. (2004) Factors limiting the immunogenicity of HIV-1 gp120 envelope glycoproteins.
569 *Virology* 330(1):233–248.
570 3. Dacheux L, et al. (2004) Evolutionary dynamics of the glycan shield of the human immunodeficiency virus
571 envelope during natural infection and implications for exposure of the 2G12 epitope. *J Virol* 78(22):12625–
572 12637.
573 4. Lasky LA, et al. (1986) Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope
574 glycoprotein. *Science* 233(4760):209–212.
575 5. Doores KJ, et al. (2010) Envelope glycans of immunodeficiency virions are almost entirely oligomannose
576 antigens. *Proc Natl Acad Sci USA* 107(31):13800–13805.
577 6. Stewart-Jones GBE, et al. (2016) Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and
578 G. *Cell* 165(4):813–826.
579 7. Korber B, et al. (2001) Evolutionary and immunological implications of contemporary HIV-1 variation. *Br*
580 *Med Bull* 58(1):19–42.
581 8. Leonard CK, et al. (1990) Assignment of intrachain disulfide bonds and characterization of potential
582 glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein
583 (gp120) expressed in Chinese hamster ovary cells. *Journal of Biological Chemistry* 265(18):10373–10382.

584 9. Wei X, et al. (2003) Antibody neutralization and escape by HIV-1. *Nature* 422(6929):307–312.
585 10. Zhou T, et al. (2017) Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell*
586 *Rep* 19(4):719–732.
587 11. Ward AB, Wilson IA (2017) The HIV-1 envelope glycoprotein structure: nailing down a moving target.
588 *Immunol Rev* 275(1):21–32.
589 12. Schwarz F, Aebi M (2011) Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct*
590 *Biol* 21(5):576–582.
591 13. Cao L, et al. (2017) Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nat Commun*
592 8:14954.
593 14. Lyumkis D, et al. (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer.
594 *Science* 342(6165):1484–1490.
595 15. Lee JH, Ozorowski G, Ward AB (2016) Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1
596 envelope trimer. *Science* 351(6277):1043–1048.
597 16. Kumar S, Cieplak P (2018) Role of N-glycosylation in activation of proMMP-9. A molecular dynamics
598 simulations study. *PLoS ONE* 13(1):e0191157.
599 17. Lee HS, Qi Y, Im W (2015) Effects of N-glycosylation on protein conformation and dynamics: Protein Data
600 Bank analysis and molecular dynamics simulation study. *Sci Rep* 5(1):8926.
601 18. Dong C, et al. (2018) Long-ranged Protein-glycan Interactions Stabilize von Willebrand Factor A2 Domain
602 from Mechanical Unfolding. *Sci Rep* 8(1):16017–11.
603 19. Hang I, et al. (2015) Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the
604 Golgi. *Glycobiology* 25(12):1335–1349.
605 20. Ferreira R-C, et al. (2018) Structural Rearrangements Maintain the Glycan Shield of an HIV-1 Envelope
606 Trimer After the Loss of a Glycan. *Sci Rep* 8(1):15031–15.
607 21. Yang M, Huang J, Simon R, Wang L-X, MacKerell AD (2017) Conformational Heterogeneity of the HIV
608 Envelope Glycan Shield. *Sci Rep* 7(1):4435–15.
609 22. Lemmin T, Soto C, Stuckey J, Kwong PD (2017) Microsecond Dynamics and Network Analysis of the HIV-1
610 SOSIP Env Trimer Reveal Collective Behavior and Conserved Microdomains of the Glycan Shield. *Structure*
611 25(10):1631–1639.e2.
612 23. Sanders RW, et al. (2013) A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140,
613 expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog*
614 9(9):e1003618.
615 24. Sanders RW, Moore JP (2017) Native-like Env trimers as a platform for HIV-1 vaccine design. *Immunol Rev*
616 275(1):161–182.

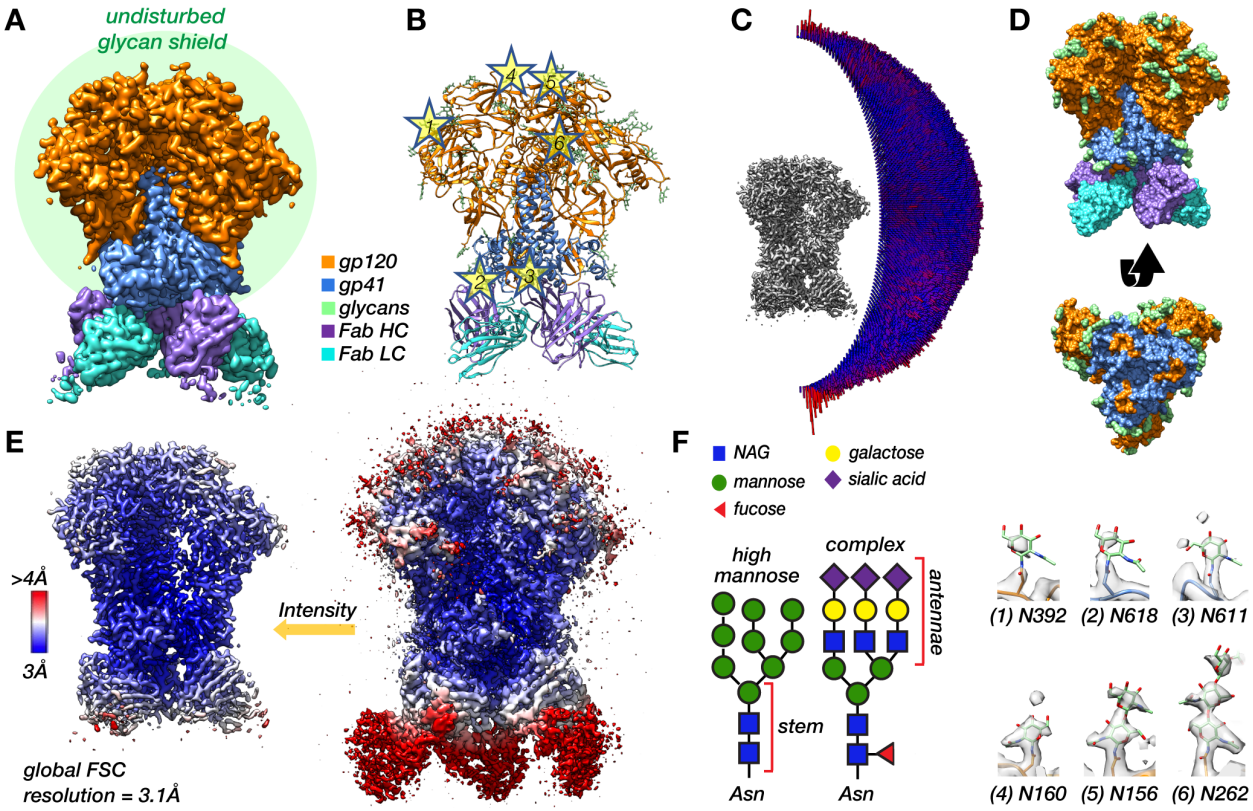
617 25. Cao L, et al. (2018) Differential processing of HIV envelope glycans on the virus and soluble recombinant
618 trimer. *Nat Commun* 9(1):3693.
619 26. Behrens A-J, et al. (2016) Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1
620 Envelope Glycoprotein. *Cell Rep* 14(11):2695–2706.
621 27. Hohn M, et al. (2007) SPARX, a new environment for Cryo-EM image processing. *J Struct Biol* 157(1):47–
622 55.
623 28. Cheng Y, Grigorieff N, Penczek PA, Walz T (2015) A primer to single-particle cryo-electron microscopy. *Cell*
624 161(3):438–449.
625 29. Guttman M, Weinkam P, Sali A, Lee KK (2013) All-atom ensemble modeling to analyze small-angle x-ray
626 scattering of glycosylated proteins. *Structure* 21(3):321–331.
627 30. Liang Y, et al. (2016) Changes in Structure and Antigenicity of HIV-1 Env Trimers Resulting from Removal of
628 a Conserved CD4 Binding Site-Proximal Glycan. *J Virol* 90(20):9224–9236.
629 31. Dey AK, et al. (2017) cGMP production and analysis of BG505 SOSIP.664, an extensively glycosylated,
630 trimeric HIV-1 envelope glycoprotein vaccine candidate. *Biotechnol Bioeng*. doi:10.1002/bit.26498.
631 32. Behrens A-J, et al. (2018) Integrity of Glycosylation Processing of a Glycan-Depleted Trimeric HIV-1
632 Immunogen Targeting Key B-Cell Lineages. *J Proteome Res*:acs.jproteome.7b00639.
633 33. Pritchard LK, et al. (2015) Glycan clustering stabilizes the mannose patch of HIV-1 and preserves
634 vulnerability to broadly neutralizing antibodies. *Nat Commun* 6:7479.
635 34. Kwong PD, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor
636 and a neutralizing human antibody. *Nature* 393(6686):648–659.
637 35. Lee JH, de Val N, Lyumkis D, Ward AB (2015) Model Building and Refinement of a Natively Glycosylated
638 HIV-1 Env Protein by High-Resolution Cryoelectron Microscopy. *Structure* 23(10):1943–1951.
639 36. Chakraborty S, et al. (2020) Quantification of the Resilience and Vulnerability of HIV-1 Native Glycan Shield
640 at Atomistic Detail. *bioRxiv*:846071.
641 37. Bonomi M, Pellarin R, Vendruscolo M (2018) Simultaneous Determination of Protein Structure and
642 Dynamics Using Cryo-Electron Microscopy. *Biophysical Journal* 114(7):1604–1613.
643 38. Bonomi M, et al. (2019) Bayesian Weighing of Electron Cryo-Microscopy Data for Integrative Structural
644 Modeling. *Structure* 27(1):175–188.e6.
645 39. Bonomi M, Vendruscolo M (2019) Determination of protein structural ensembles using cryo-electron
646 microscopy. *Curr Opin Struct Biol* 56:37–45.
647 40. Rathore U, et al. (2017) Glycosylation of the core of the HIV-1 envelope subunit protein gp120 is not
648 required for native trimer formation or viral infectivity. *J Biol Chem* 292(24):10197–10219.
649 41. Depetris RS, et al. (2012) Partial enzymatic deglycosylation preserves the structure of cleaved recombinant
650 HIV-1 envelope glycoprotein trimers. *J Biol Chem* 287(29):24239–24254.

651 42. Wang W, et al. (2013) A systematic study of the N-glycosylation sites of HIV-1 envelope protein on
652 infectivity and antibody-mediated neutralization. *Retrovirology* 10(1):14.
653 43. Lee J-Y, et al. (2016) Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature*
654 533(7604):561–564.
655 44. Ohgimoto S, et al. (1998) Location-specific, unequal contribution of the N glycans in simian
656 immunodeficiency virus gp120 to viral infectivity and removal of multiple glycans without disturbing
657 infectivity. *J Virol* 72(10):8365–8370.
658 45. Helenius A, Aebi M (2001) Intracellular functions of N-linked glycans. *Science* 291(5512):2364–2369.
659 46. Solá RJ, Rodríguez-Martínez JA, Griebenow K (2007) Modulation of protein biophysical properties by
660 chemical glycosylation: biochemical insights and biomedical implications. *Cell Mol Life Sci* 64(16):2133–
661 2152.
662 47. Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: a close look at thermodynamic
663 stabilization. *Proc Natl Acad Sci USA* 105(24):8256–8261.
664 48. Imperiali B, O'Connor SE (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure.
665 *Curr Opin Chem Biol* 3(6):643–649.
666 49. Tian J, et al. (2016) Effect of Glycosylation on an Immunodominant Region in the V1V2 Variable Domain of
667 the HIV-1 Envelope gp120 Protein. *PLoS Comput Biol* 12(10):e1005094.
668 50. Wood NT, et al. (2013) The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3
669 loop. *PLoS ONE* 8(11):e80301.
670 51. Gerken TA, Butenhof KJ, Shogren R (1989) Effects of glycosylation on the conformation and dynamics of O-
671 linked glycoproteins: carbon-13 NMR studies of ovine submaxillary mucin. *Biochemistry* 28(13):5536–
672 5543.
673 52. Mer G, Hietter H, Lefèvre JF (1996) Stabilization of proteins by glycosylation examined by NMR analysis of
674 a fucosylated proteinase inhibitor. *Nat Struct Biol* 3(1):45–53.
675 53. Joao HC, Dwek RA (1993) Effects of glycosylation on protein structure and dynamics in ribonuclease B and
676 some of its individual glycoforms. *Eur J Biochem* 218(1):239–244.
677 54. Lu D, Yang C, Liu Z (2012) How hydrophobicity and the glycosylation site of glycans affect protein folding
678 and stability: a molecular dynamics simulation. *J Phys Chem B* 116(1):390–400.
679 55. Shental-Bechor D, Levy Y (2009) Folding of glycoproteins: toward understanding the biophysics of the
680 glycosylation code. *Curr Opin Struct Biol* 19(5):524–533.
681 56. Hanson SR, et al. (2009) The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding
682 and enhances stability. *Proc Natl Acad Sci USA* 106(9):3131–3136.
683 57. Dubrovskaya V, et al. (2019) Vaccination with Glycan-Modified HIV NFL Envelope Trimer-Liposomes Elicits
684 Broadly Neutralizing Antibodies to Multiple Sites of Vulnerability. *Immunity* 51(5):915–929.e7.

685 58. Dubrovskaya V, et al. (2017) Targeted N-glycan deletion at the receptor-binding site retains HIV Env NFL
686 trimer integrity and accelerates the elicited antibody response. *PLoS Pathog* 13(9):e1006614.
687 59. Ozorowski G, et al. (2017) Open and closed structures reveal allostery and pliability in the HIV-1 envelope
688 spike. *Nature* 15(7663):690–363.
689 60. Bianchi M, et al. (2018) Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal
690 Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization. *Immunity* 49(2):288–300.e8.
691 61. Nogal B, et al. (2020) Mapping Polyclonal Antibody Responses in Non-human Primates Vaccinated with HIV
692 Env Trimer Subunit Vaccines. *Cell Rep* 30(11):3755–3765.e7.
693 62. Cottrell CA, et al. (2020) Mapping the immunogenic landscape of near-native HIV-1 envelope trimers in
694 non-human primates. *bioRxiv*:2020.02.05.936096.
695 63. Steichen JM, et al. (2019) A generalized HIV vaccine design strategy for priming of broadly neutralizing
696 antibody responses. *Science* 366(6470):eaax4380.
697 64. Kimanius D, Forsberg BO, Scheres SH, Lindahl E (2016) Accelerated cryo-EM structure determination with
698 parallelisation using GPUs in RELION-2. *Elife* 5:19.
699 65. Zivanov J, et al. (2018) New tools for automated high-resolution cryo-EM structure determination in
700 RELION-3. *Elife* 7:163.
701 66. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-
702 EM structure determination. *Nat Methods* 14(3):290–296.
703 67. Pettersen EF, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J*
704 *Comput Chem* 25(13):1605–1612.
705 68. Emsley P, Crispin M (2018) Structural analysis of glycoproteins: building N-linked glycans with Coot. *Acta*
706 *Crystallogr D Struct Biol* 74(Pt 4):256–263.
707 69. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol*
708 *Crystallogr* 60(Pt 12 Pt 1):2126–2132.
709 70. Wang RY-R, et al. (2016) Automated structure refinement of macromolecular assemblies from cryo-EM
710 maps using Rosetta. *Elife* 5:352.
711 71. Weinkam P, Pons J, Sali A (2012) Structure-based model of allostery predicts coupling between distant
712 sites. *Proc Natl Acad Sci USA* 109(13):4875–4880.
713 72. Eswar N, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics*
714 Chapter 5(1):Unit–5.6.
715 73. Sali A (1995) Comparative protein modeling by satisfaction of spatial restraints. *Mol Med Today* 1(6):270–
716 277.
717 74. Varki A, et al. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*
718 25(12):1323–1324.

719

720 **Figures**



721

722 **Fig. 1 | Soluble SOSIP in complex with a base-specific Fab as a model system for cryo-EM analysis of the native**
723 **HIV-1 Env glycan shield. (A)** Segmented cryo-EM map of BG505 SOSIP.664 in complex with three copies of the
724 RM20A3 Fab. The green circle is meant to represents the glycan shield. **(B)** Refined atomic model with stars
725 highlighting several N-linked glycans. **(C)** 3-D angular distribution histogram. **(D)** Surface representation of the refined
726 atomic model viewed from the side and from the bottom with the Fab chains removed. **(E)** Sharpened 3.1Å-resolution
727 cryo-EM map at high and low threshold colored by local resolution. **(F)** SNFG depiction (74) of a representative high-
728 mannose and complex type glycan (L). Cryo-EM map density for 6 representative glycans from the refined atomic
729 model (numbers corresponding to the stars in panel B).

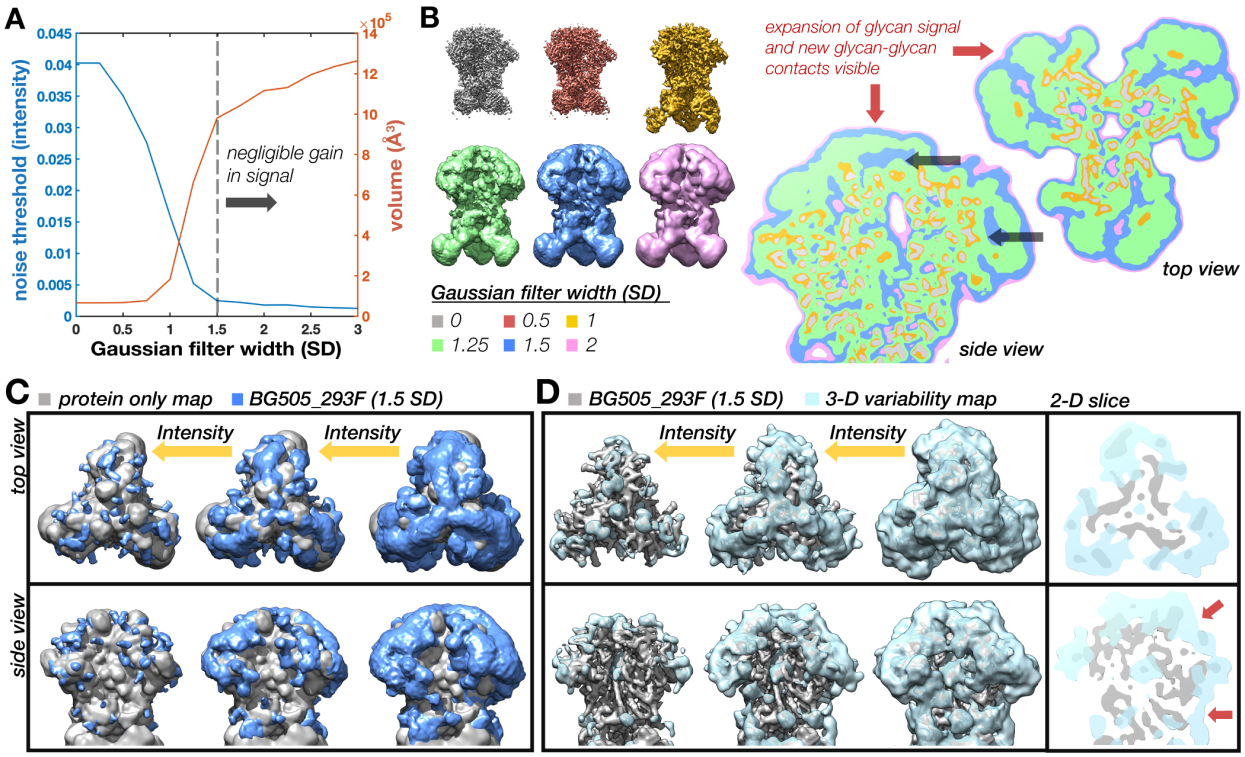


Fig. 2 | Scale-space and 3-D variability analysis reveal low resolution structure in the glycan shield. (A) Noise threshold and volume at the noise threshold as a function of Gaussian filter width measured in standard deviation (SD) showing the emergence of glycan signal at low resolutions and plateau around 1.5 SD. **(B)** BG505_293F cryo-EM map at 6 representative scales along with 2-D slices through the maps overlaid on top of one another showing the emergence of glycan signal and new glycan-glycan contacts visible at low resolutions and thresholds (red arrows). Black arrows are pointing to protein surfaces occluded by glycans. **(C)** Gaussian filtered (1.5 SD) BG505_293F map (blue) visualized at three intensity thresholds along with a protein-only map (gray). **(D)** SPARX 3-D variability map (light blue) visualized at three intensity thresholds along with the BG505_293F map (gray) and 2-D slices through the top and side of the map with red arrows highlighting glycan-glycan contacts.

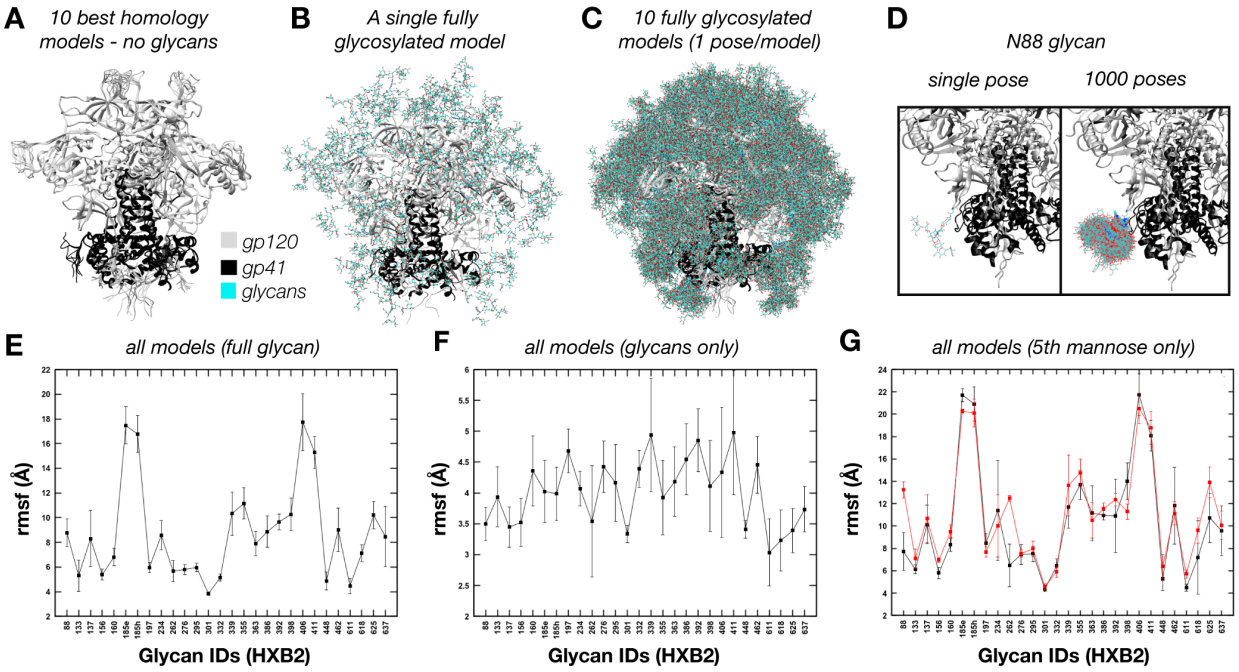


Fig. 3 | ALLOSMOD-based HT-AM pipeline for fast and robust sampling of fully glycosylated Env. (A) 10 homology models used as protein scaffolds. **(B)** One of the ten models with a single relaxed Man9 glycan at each site. **(C)** One fully glycosylated and relaxed model for each of the ten protein scaffolds. **(D)** Close up of the N88 glycan showing a single glycan pose (L) and all 1000 poses (R). **(E)** Full glycan average rmsf for all 1000 poses. **(F)** Glycan only rmsf for all 1000 poses. **(E)** Full glycan average rmsf for all 1000 poses at the 5th mannose residue for both the Man9 and Man5 ensembles.

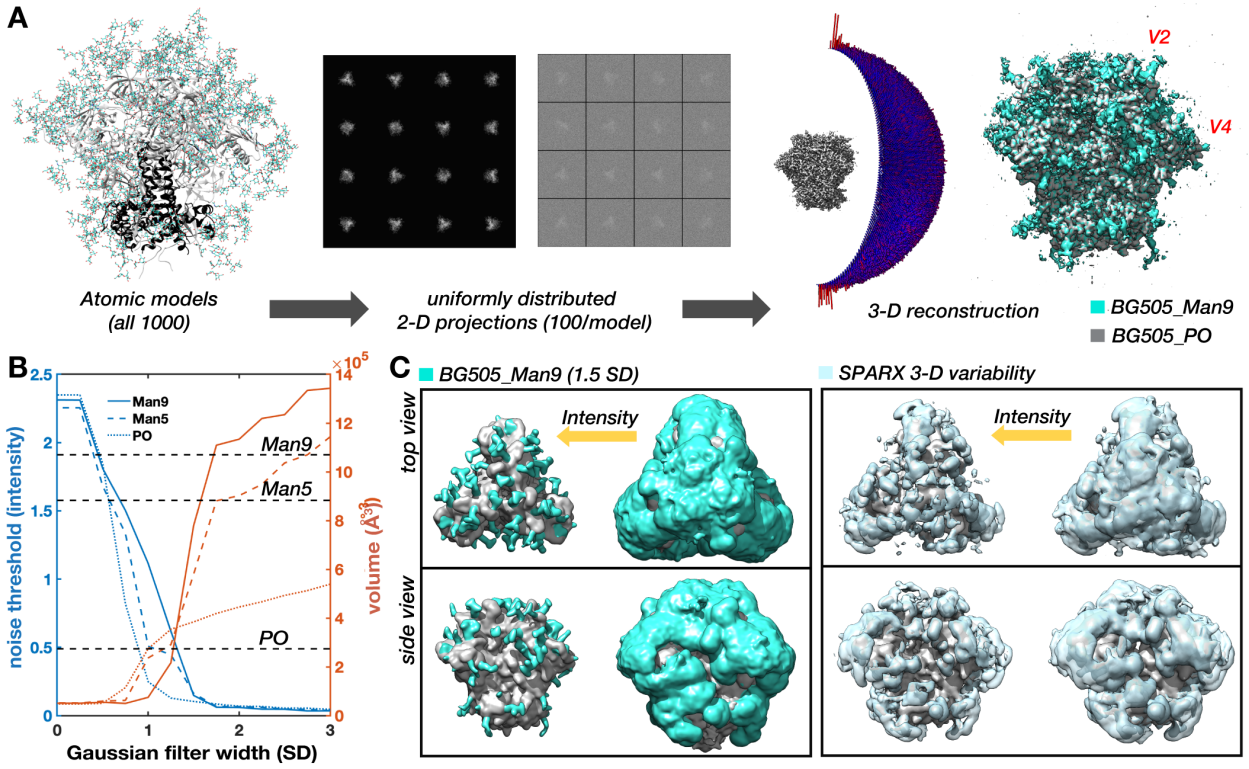


Fig. 4| Simulated cryo-EM maps from the HT-AM ensembles reproduce defining features of the experimental data. (A) Schematic of the simulated data creation pipeline. Each model is projected at 100 uniformly distributed angles (representative 2-D projections with and without noise are shown), then the combined dataset of 100K projections is refined using RELION (the resulting 3-D reconstruction and angular distribution histogram are shown). **(B)** Noise threshold and volume at the noise threshold as a function of Gaussian filter width (SD) for the BG505_Man9, BG505_Man5, and BG505_PO simulated cryo-EM maps with dashed lines indicating the volume sampled by each ensemble. **(C)** 1.5 SD Gaussian filtered BG505_Man9 (teal) and BG505_PO (gray) maps along with the SPARX 3-D variability map (light blue) viewed from the side and top at high and low ntensity thresholds.

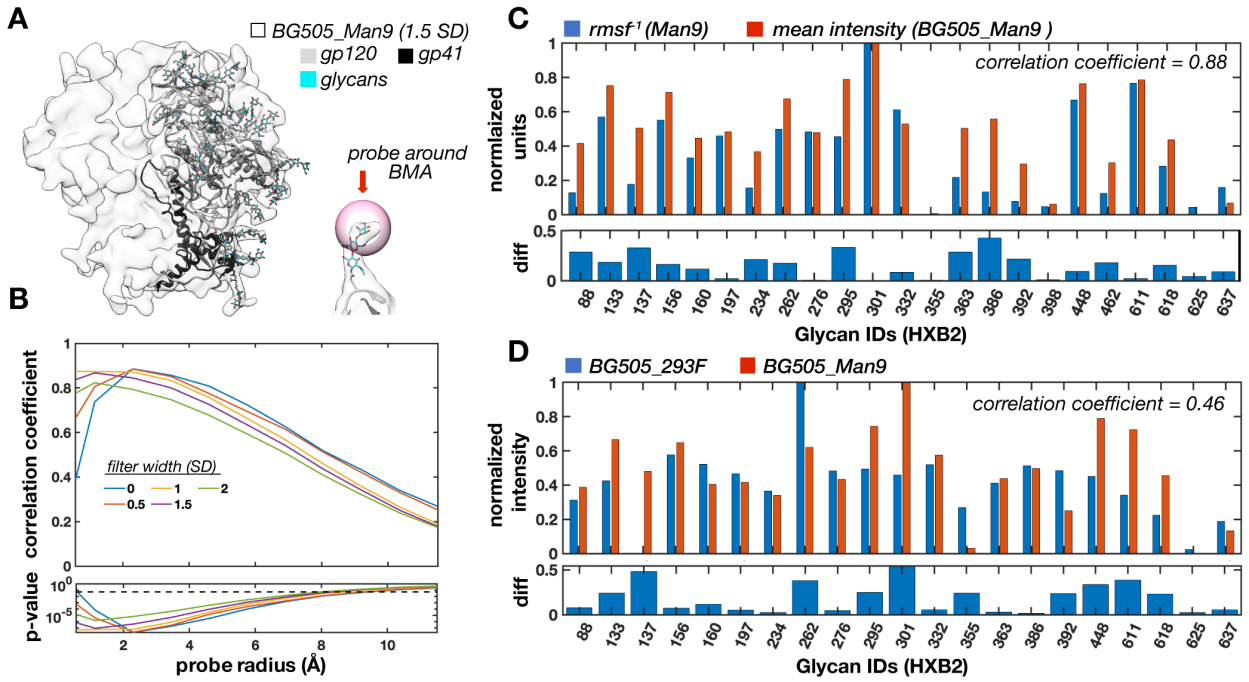


Fig. 5 | Measuring glycan dynamics from cryo-EM maps reveals close agreement between simulation and experiment. (A) 1.5 SD Gaussian filtered BG505_Man9 simulated cryo-EM map (transparent) and atomic model used for local intensity analysis. Also shown is a single glycan stem with a spherical probe around the BMA residue. **(B)** Pearson correlation coefficient and p-value between mean intensity and inverse rmsf as a function of probe radius for 5 different Gaussian filtered maps. The dashed line designates the confidence threshold (p-value = 0.05). **(C)** Normalized mean intensity around BMA residues for BG505_Man9 (probe radius = 2.3Å) and the inverse normalized full glycan average rmsf. Pearson correlation coefficient ~0.88 (p-value = 2.1e-8). **(D)** Normalized mean intensity around BMA residues for BG505_293F and BG505_Man9 along with the absolute difference at each site. Pearson correlation coefficient ~0.46 (p = 0.03).

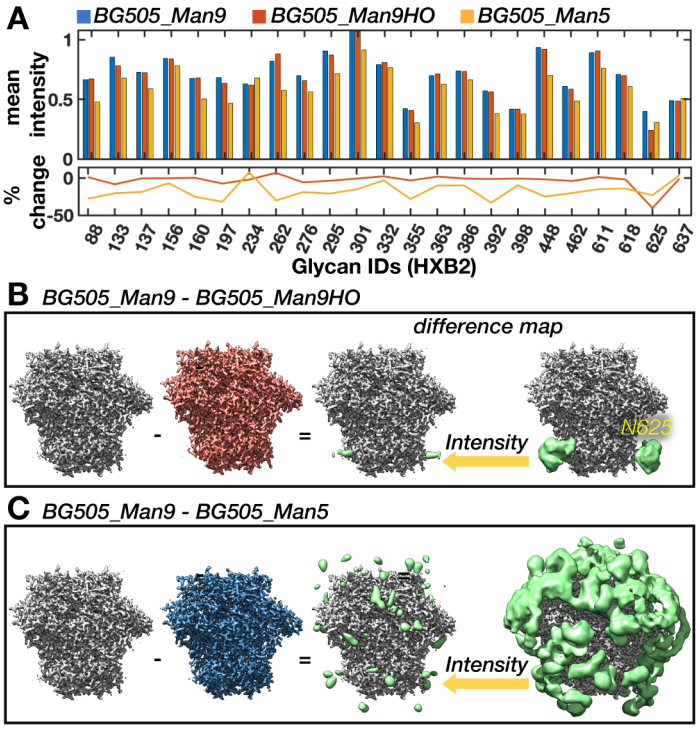


Fig. 6 | Detecting changes in glycan dynamics, occupancy, and chemical composition from simulated cryo-EM maps. (A) Mean intensity around each glycan BMA residue for the BG505_Man9, BG505_Man5, and BG505_Man9HO maps along with the percent change from BG505_Man9. **(B)** BG505_Man9 - BG505_Man9HO and **(C)** BG505_Man9 - BG505_Man5 difference maps (Gaussian filtered) at two intensity thresholds (right - green) along with high-resolution sharpened maps (left).

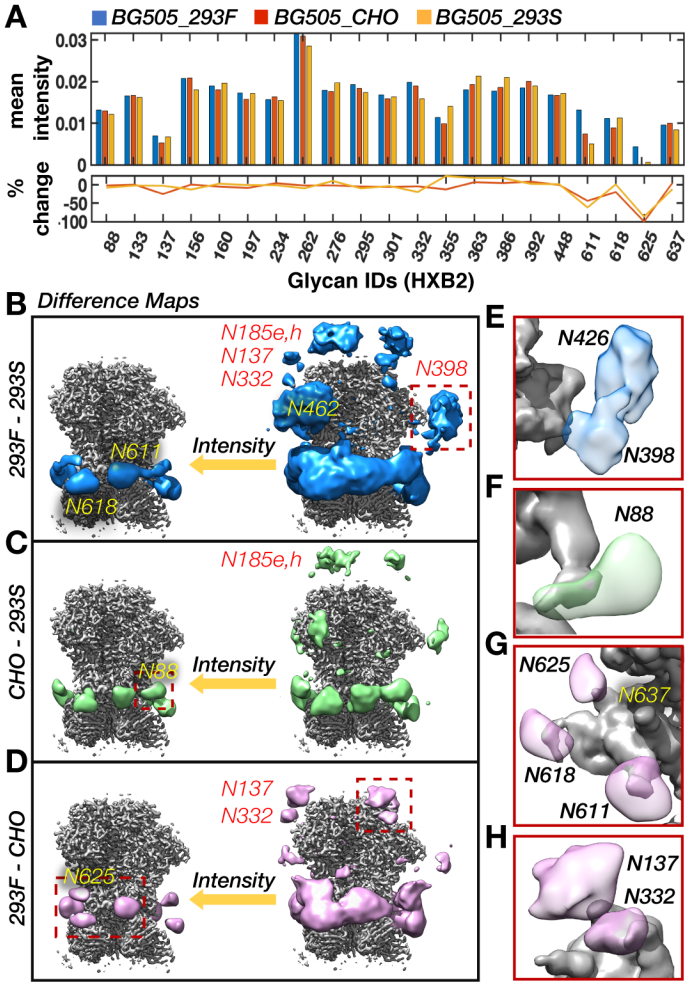


Fig. 7 | Detection of cell-type specific changes in glycan shield composition and dynamics with cryo-EM. (A) Mean intensity at each glycan BMA residue for the BG505_293F, BG505_293S, and BG505_CHO cryo-EM maps along with percent change from BG505_293F. **(B)** BG505_293F – BG505_293S, **(C)** BG505_CHO – BG505_293S, and **(C)** BG505_293F - BG505_CHO difference maps at two intensity thresholds. All difference maps were multiplied by a soft mask around the RM20A3 Fabs then smoothed with a 2 SD Gaussian filter. **(E-H)** Fly-outs of the regions outlined by dashed red boxes.

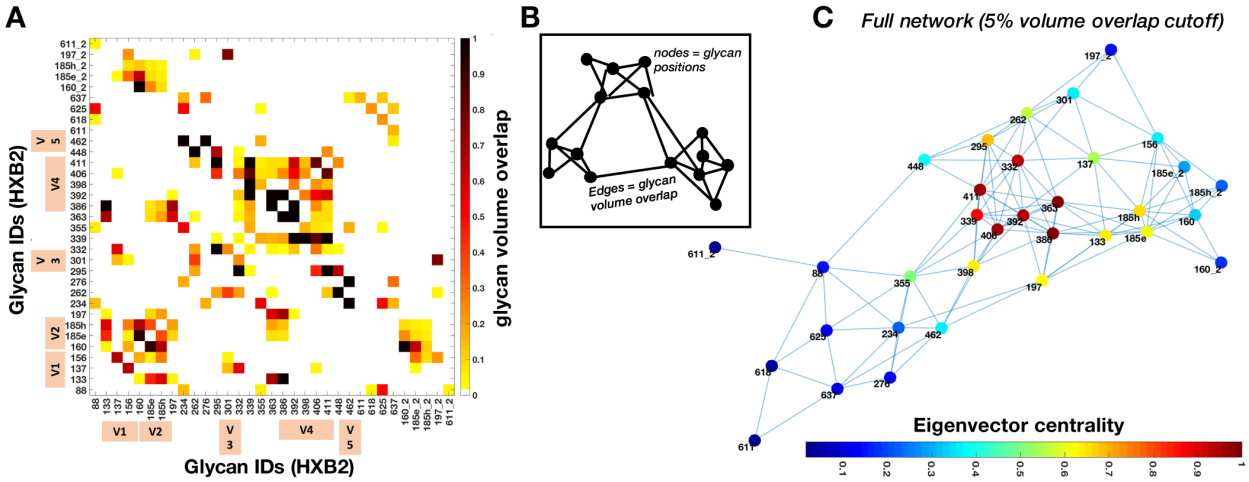


Fig. 8 | Probabilistic glycan-glycan interaction network for the Man9 ensemble. (A) Glycan-glycan volume overlap matrix. Inter-protomer overlap is given by suffix "_2" and the variable loop glycans are indicated by tan color bars. The overlap fraction is normalized with at least 50% overlap being designated as 1. **(B)** Cartoon model of the glycan-glycan interaction network generated by interpreting the overlap matrix as an adjacency matrix. The edge length is drawn inversely proportional to the overlap value. Edges represent overlap between two glycans represented by nodes. **(C)** Glycan-glycan interaction network calculated from the matrix in (A) with nodes colored by normalized eigenvector centrality.

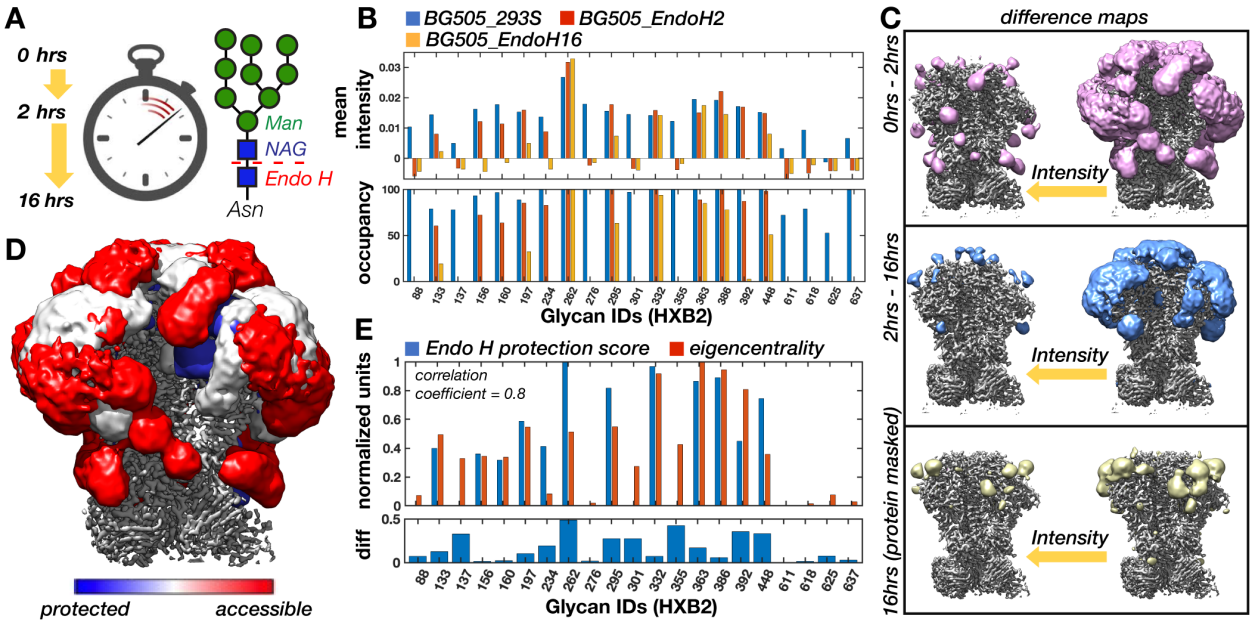


Fig. 9 | Time-resolved cryo-EM reveals that highly connected glycan clusters are resistant to enzymatic digestion. (A) Cartoon schematic of the Endo H digestion of high-mannose glycans and an illustration of the three reaction times captured by time-resolved cryo-EM. **(B)** Mean intensity around glycan BMA residues for the BG505_293S, BG505_EndoH2, and BG505_EndoH16 cryo-EM maps along with the percent occupancy at each site. Occupancy was calculated assuming a linear relationship between occupancy and intensity with the initial occupancy determined by MS. **(C)** BG505_293S – BG505_EndoH2 and BG505_EndoH2 – BG505_EndoH16 difference maps at two intensity thresholds. Difference maps were first multiplied by a soft mask around the RM20A3 Fabs and smoothed with a 2 SD Gaussian filter. Also shown is the residual glycan signal remaining after 16hrs of digestion isolated by masking out the protein density. **(D)** Overlay of the three maps from panel B colored according to susceptibility to Endo H digestion. **(E)** Normalized Endo H protection score measured as the cumulative occupancy at each site in the two digestion intermediates and the normalized eigencentrality from the Man9 volume overlap network. Pearson correlation coefficient of ~0.8 (p=1.14e-05).

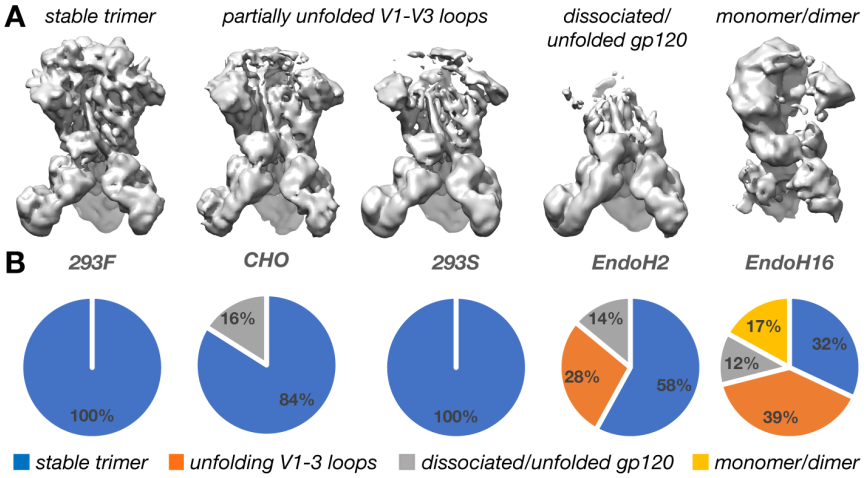


Fig. 10 | Digestion of the glycan shield leads to progressive destabilization of the Env timer. (A) Representative cryo-EM maps of the unfolding intermediates captured by 3-D classification. **(B)** Pie charts showing the relative percentages of each class present in the five cryo-EM datasets presented here.

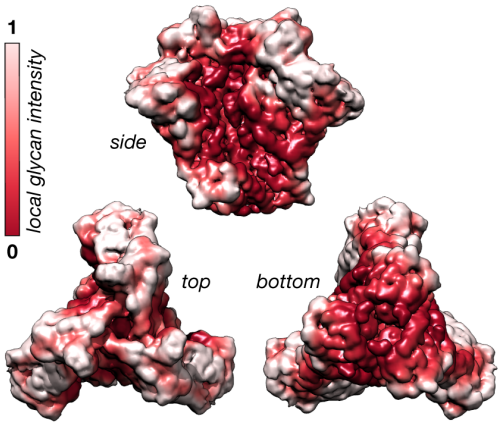


Fig. 11 | Quantifying the glycan shielding effect from cryo-EM maps. The normalized glycan shielding effect for BG505_CHO defined as the total glycan signal within a ~7Å radius spherical probe centered at each voxel. White surfaces are strongly shield by glycans.