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. Cotrell[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) and computational modeling to probe glycan shield structure and behavior at multiple levels. We found that dynamics lead to an extensive network of inter-glycan interactions and drive higher-order structuring within the glycan shield. This structure defines diffuse boundaries between buried and exposed protein surface and provides a mapping of potentially immunogenic sites on Env. Analysis of the same Env across a range of glycosylation states revealed that subtle changes in glycan occupancy, composition, and dynamics can impact glycan shield structure and epitope accessibility. We also performed site-specific mass-spectrometry analysis on the same samples and show how cryo-EM can complement such studies. Finally, we found that highly connected glycan sub-domains are resistant to enzymatic digestion and help stabilize the pre-fusion trimer state, suggesting functionality 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 thus 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 ~1/21 extracellular residues being glycosylated, accounting for ~1/2 the mass of the molecule[4-8]. This sugar coat, referred to as the “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]. Given the importance of the glycan shield to Env immunogenicity, 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 densely glycosylated proteins.

Structural analysis of proteins that are heavily glycosylated with N-linked glycans can be difficult because they are relatively large and flexible post-translational modifications that vary in length, connectivity, and chemical composition [11]. In general, they are composed of 10-20 sugar rings connected in a tree like structure to the corresponding asparagine (Asn/N) residue (Figure 1E). The inherent heterogeneity of surface exposed glycans hinders crystallization, therefore single-particle electron cryomicroscopy (cryo-EM) offers a more versatile solution for structural analysis. Thus, the first structures of the fully glycosylated soluble SOSIP, as well as transmembrane derived HIV-1 Env trimers were solved with cryo-EM[12,13]. In these reconstructions, most glycans were not resolved beyond the first or second sugar ring except when stabilized by antibodies, indicating the glycan shield is highly dynamic. This lack of immediately discernable structural features did not inspire closer analysis and structural studies remained focused on Env-antibody interactions and epitope mapping, while experimental investigations into the glycan shield shifted primarily towards chemical analysis via mass-spectrometry (MS) [11]. One crystallography study aimed specifically at addressing glycan shield structure was recently published, and the data shows evidence for highly stabilized glycans engaging in a wide range of glycan-glycan contacts [6]{27114034), which is at odds with the cryo-EM results. Such a disagreement between the two techniques highlights the difficulty in studying the glycan shield with a conventional structure-function approach.

Cryo-EM is not just a method for high-resolution structural analysis however, it is also a powerful tool for studying heterogeneous systems such as Env[14]. Datasets are composed of thousands to millions of images of individual molecules embedded in vitreous ice that ideally represent snapshots of that molecule in random states drawn from its equilibrium ensemble. Unfortunately, low signal to noise ratios in the raw projection images limit their interpretability and necessitates averaging to recover high-resolution features. This procedure boosts signal and flattens noise in regions of high spatial homogeneity but causes blurring around flexible domains. Despite this, heterogeneity can often be reduced to obtain higher resolution reconstructions or studied directly through classification and other more sophisticated algorithms[14-16]. In general, these algorithms work best when heterogeneity is discrete, and/or the variability exists between a few large structured domains. Surface exposed glycans on the other hand, are both small (compared to larger protein domains) and their fluctuations are continuous in nature. On the surface of Env there are on average ~95 glycans[6], each undergoing continuous thermal fluctuation. Such heterogeneity is not possible to resolve at the atomic scale with cryo-EM given the huge number of degrees of freedom. Despite this, cryo-EM maps still contain useful ensemble-average information about relative dynamics and other sources of heterogeneity in the form of variations in signal intensity and resolution, and such features can be analyzed quantitatively[17-24].

Given the current limitations of cryo-EM, computational modeling can provide alternatives to augment these studies[25]. Molecular dynamics (MD) simulations have been particularly useful as a complementary approach for characterizing glycoprotein structure and dynamics[26-29], including HIV Env[6,30-32]. Three sets of atomistic dynamics simulations of fully glycosylated Env trimers were recently reported and these predicted that conformational dynamics would lead to extensive shielding of antigenic surface and the formation of complicated networks among interacting glycans[6,31,32]. Although highly informative, MD simulations for systems as large as this require substantial computational resources to access biologically relevant spatial and temporal scales. Therefore, the development of more computationally efficient methods for modeling fully glycosylated Env is needed. In spite of advances in modeling the glycosylated proteins however, there is still no established methodology for validating the theoretical predictions with experimental data.

Here, we describe an integrated approach combining cryo-EM and computational modeling aimed at illuminating glycan shield structure and behavior at multiple resolutions, from global (low resolution) to local (high resolution) dynamics. First, we introduce a model system for cryo-EM analysis of the native glycan shield and examine it's features with progressive filtering and 3-D variance analysis. In parallel, we developed a high-throughput modeling pipeline for rapidly generating diverse ensembles of fully glycosylated Env models 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 synthetic cryo-EM datasets, which facilitated the development of new analysis tools and enabled the validation of theoretical models against cryo-EM experiments.

Using this integrated approach, we show that the glycan shield is highly dynamic but exhibits variation in regional dynamics between glycans due in part to crowding and other local geometric and energetic constraints. We found dynamics give rise to networks 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 provides a mapping of potentially antigenic sites on Env. Using Env expressed in three common cell lines, we show how differences in glycan composition and occupancy between them result in changes to glycan shield structure that affect the accessibility of epitopes on the surface. To better inform our results we present site-specific mass-spectrometry (MS) data for the same samples and show how cryo-EM can complement such studies. Finally, by exposing Env to slow endoglycosidase digestion and capturing reaction intermediates with cryo-EM in a time-resolved manner, we found that highly connected glycan sub-domains 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 mostly widely studied of the soluble, stabilized Envs[33,34]. Also, it is currently in the first ever human clinical trials of a trimeric

104 subunit-based HIV vaccine (https://clinicaltrials.gov/ct2/show/NCT03699241). In Figure 1 we present a cryo-EM
105 reconstruction of BG505 SOSIP.664 expressed and purified from HEK293F cells in complex with three copies of the
106 non-neutralizing fragment antigen binding domain (Fab) derived from the antibody RM20A3, which was included to
107 improve orientation bias (as illustrated by the 3-dimensioinal (3-D) angular distribution histogram - Figure 1F). This
108 complex will be referred to as BG505_293F from here on. In addition, RM20A3 binds to the base of the SOSIP at a
109 neo-epitope only accessible on the soluble trimer (Figure 1B) that is devoid of glycans (Figure 1D). This leaves the
110 native glycan shield unperturbed, making the SOSIP/base-Fab complex ideal for this study.
111
112 **Conventional cryo-EM analysis confirms the glycan shield is highly dynamic and that dynamics vary among glycans**
113 Cryo-EM maps contain structural information that varies across both resolution and intensity scales. Yet the typical
114 workflow involves isolating just a single point within this scale-space for interpretation and model building. Usually, a
115 map is filtered to the reported global resolution, then sharpened to enhance the highest-resolution features. Finally,
116 the data is thresholded, or contoured, at some intensity level so that no voxel of lesser value is displayed. The surface
117 constructed from this thresholded map is called an *isosurface*. This way of viewing and interpreting cryo-EM
118 reconstructions is well suited for the typical goals of structural biology, but when the reconstruction contains
119 variations in resolution and intensity, as is often the case for flexible and dynamic molecules and molecular complexes,
120 a single map processed in this manner is of limited value.
121 The limitations of this single-map approach are well illustrated in the context of fully glycosylated Env, as
122 shown in Figure 1. The global resolution of this C3-symmetric reconstruction determined by Fourier shell correlation
123 (FSC) was ~3.1Å (Supplemental Figure 1). The bulk of the Env protein structure is at or below the global resolution
124 (Figure 1C) and ~80% of the Env amino acid (aa) residues could be confidently built into this map, with missing residues
125 occurring around the flexible variable loops 1, 2, 4, and 5, the HR1 helix, and the N and C-termini. Most of the N-linked
126 glycans (colored green) on the other hand, are ill-defined beyond the core N-acetylglucosamine (NAG), and we can
127 only account for ~15% of the total glycan mass in this sharpened map (Figure 1D). Glycans located on the poorly
128 resolved flexible loops (N185e, N185h, N339, N398, N406, N411, N462) could not be identified at all, however MS
129 analysis confirms that all the potential N-linked glycosylation sites (PNGS) predicted from the sequence are indeed
130 glycosylated[35] (Supplemental Figure 12A). This indicates that the glycan shield is highly dynamic with respect to the
131 protein core but exhibits some relative differences in ordering between glycans. In a later section, we show how to
132 quantify these differences more precisely. At a low threshold (Figure 1C), noise appears surrounding Env's well
133 resolved protein core where the missing glycan mass should be, however it is not apparent if there exists any
134 meaningful structure within this noise.
135
136 **Scale-space and 3-D variance analysis reveal interconnectivity and higher-order structure within the glycan shield**
137 To look for potential structure within the glycan shield we first examined its properties across resolution scales by
138 progressively smoothing the map with a Gaussian kernel of increasing standard deviation (SD) and visualizing them at

their *noise threshold* (Figure 2A). We use a simple topological metric for determining the noise threshold based on *connected components*. A connected component is any connected and isolated group of 'on' voxels in a binary volume or isosurface. In the language of topology, the number of connected components defines the *0th Betti number* of the map at that threshold[36]. When the log of this number is plotted as a function of intensity for the series of filtered maps shown in Figure 2A we see the same general trend (Figure 2B and Supplemental Figure 7C). The initial rise-and-fall corresponds to high-intensity features emerging, growing, and merging, while the second sharp peak corresponds to low-intensity noise. We defined the noise threshold as the lowest intensity point in the local minimum between the signal and noise peaks (red circles Figure 2B).

Examination of this series of filtered maps reveals progressively more structural features in the glycan shield. At higher SD, there is a dramatic expansion of non-protein signal. In Figure 2C we plot the noise threshold as a function of SD as well as the total volume at the noise threshold. Both curves have a sigmoidal shape, with a rapid reduction in the noise threshold and concurrent increase in volume, followed by a plateau around ~1.5-2 SD. We interpret this plateau as indicating there is negligible gain in signal with further smoothing. Although volume dilation with decreasing threshold is normal for cryo-EM maps, the amount of dilation surrounding regions known to be glycosylated is greater than around regions that are not. This is well illustrated by plotting cross-sections through the six maps (Figure 2D).

Having identified the proper resolution scale for analyzing the glycan shield, we examined how its features change with intensity. The series of thresholded maps shown in Figure 2E illustrate how the topology of the glycan shield changes dramatically with intensity, becoming increasingly interconnected. Signal from individual glycans begins as isolated components then merges with neighboring glycans to form higher-order, multi-glycan structures. While the threshold needed to best capture the structure and connectivity of the glycan shield is not well defined, the higher order structure formed by neighboring glycans is clearly seen as the threshold is decreased.

As a complementary approach for visualization and characterization of the glycan shield, we performed 3-D variance analysis on the same dataset using the function *sx3dvariability* from the SPARX software package[37,38] (Supplemental Figure 2). We see high variability around the constant domains of the Fabs (clipped from view) as well as loops and the exterior of the protein surface at the sites of N-linked glycans. In particular, variability in the glycans is highest at the distal ends of the glycan stalks and expands outward as the threshold is reduced. Top and side views of the SPARX 3-D variability map are shown at three intensity thresholds and it exhibits features similar to the Gaussian filtered maps. In particular, non-uniform intensity across the surface (between glycans) and interconnectivity at low threshold. It is clear from the 3-D variability map how the higher-order structure within the glycan shield acts to occlude nearby non-glycosylated protein surfaces (red arrow - Supplemental Figure 2B).

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

Cryo-EM cannot capture atomistic details of individual molecules, and all-atom MD simulations of fully glycosylated Env requires extensive computational power and time. To overcome these limitations, we employed a high

174 throughput atomistic modeling (HT-AM) pipeline based around the ALLOSMOD package of the MODELLER software
175 suite. The glycan modeling functionality in the ALLSOMOD package was developed for generating ensembles of
176 heterogeneous glycoprotein models to simulate small-angle X-ray scattering (SAXS) profiles[39] and has since been
177 used to generate small ensembles glycosylated Env[40]. This methodology accounts for both energetic and spatial
178 constraints on glycan sampling by a combination of empirical energy minimization and simulated annealing (described
179 in detail by Guttman et al. [39]). Here, we have built on this by integrating a stereochemical check on the protein
180 backbone as an intermediate step for rigorous selection of the underlying scaffold and utilized it to generate much
181 larger ensembles of 1000 fully glycosylated Env structures (Figure 3). Ten such models are shown in Figure 3C, one
182 from each protein scaffold, while the full set of models at a single PNGS is shown in Figure 3D. We also repeated the
183 simulation with uniform mannose-5 (Man5) glycosylation for comparison.

184 To determine if our simulations converged, we first calculated the root mean-squared fluctuation (RMSF)
185 for each glycan (see Methods) across all 1000 models after aligning the protein scaffold (Figure 3E), then compared it
186 to the average of randomly selected equally sized subsets (Supplemental Figure 3D). We see that the mean RSMF
187 values between the subsets are nearly identical and the standard deviations from the means are very small, indicating
188 convergence. A similar trend can be seen for the per-glycan sampled volumes (Supplemental Figure 3E).
189

190 **HT-AM generated models captures spatial and energetic constraints on glycan flexibility**

191 Having verified that our simulations converged, we could then analyze the RMSF values more closely to assess how
192 spatial and energetic constraints affect the flexibility of individual glycans. It is clear from Figure 3E that our method
193 captures variability between glycans, and the additional flexibility imparted by the 10 starting protein scaffolds can be
194 appreciated by comparison to the glycan only RMSF (Figure 3F). For example, glycans located on the flexible loops
195 have a much higher RMSF than all the other glycans (185e, 185h as well as 406 and 411) whereas glycans at the N262,
196 N301, N332, N448, and N611 sites for example, have lower RMSF. This leads to a large difference in sampled volumes
197 between the most and least dynamic glycans (Supplemental Figure 5A). We also see an increase in average RMSF
198 within a single glycan as a function of glycan residue number starting from the first NAG (Supplemental Figure 3C),
199 which is in line with the cryo-EM results showing reduced resolution beyond this residue (Figure 1D).

200 We see a similar trend in the average RMSF values from the Man5 ensemble (Supplemental Figure 4A-B).
201 Interestingly however, by comparing the average RMSF values of the 5th mannose residue in both the Man5 and Man9
202 ensembles, we find a slight increase in RMSF and volume at most sites (Figure 3G and Supplemental Figure 4D). We
203 attribute this effect to restricted sampling in the glycan canopy from the more massive Man9 glycans. In support of
204 this, we see a positive correlation between glycan flexibility and a measure of local glycan density (see Methods),
205 however only when considering relatively large neighborhoods (Supplemental Figure 6A). A similar trend was
206 observed by Stewart-Jones et al. which they attribute to different "shells" of influencing glycans[6]. Such higher-order
207 dynamic effects could be possible in light of our structural observations.

208 However, crowding from neighboring glycans is not the only factor that can influence glycan flexibility, it can
209 also be influenced by the local protein structure. In our modeling pipeline, the protein backbone was kept
210 harmonically restrained close to the template structure to allow for extensive sampling of glycan conformations using
211 simulated annealing, without leading to unfolding of the underlying protein. Thus, we see that the Asn sidechains of
212 residues 88, 160, 197, 234 and 262 all have very low RMSF (Supplemental Figure 3F), possibly stemming to some
213 extent from limited torsional space available during modeling. The glycosylated Asn residues in gp41 have relatively
214 low RMSF as well (N611, N618, N625, and N637), being situated on stable helical bundles (Supplemental Figure 5B).
215 This ultimately results in a relative reduction of the glycan dynamics at some of these sites (Figure 3C). Correcting for
216 the contribution to fluctuations coming from the underlying protein, we observed that the RMSF between the
217 different glycans are comparable, ranging from 3 Å to 5 Å, with similar scale of SD (Figure 3F).
218
219 **Synthetic cryo-EM maps generated from the HT-AM ensembles reproduce global features of the glycan shield**
220 **from experimental cryo-EM data**
221 We showed the HT-AM pipeline converges computationally and captures variation in the flexibility of individual
222 glycans. However, to assess if it converges on physiologically relevant dynamics, the results need to be validated
223 against experiments that can capture the same ensemble-average observables. Given the single-molecule nature of
224 cryo-EM datasets, the heterogeneous ensemble generated by the HT-AM pipeline can be seen as representing the
225 individual particles that went into the 3-D reconstruction. With that in mind, we established a protocol for
226 transforming the HT-AM ensembles into synthetic cryo-EM dataset (Figure 4A). Each of the 1000 models was first
227 transformed into an *mrc* volume then projected (with or without noise) at 100 uniformly distributed angles. This
228 combined dataset of 100,000 projections was then refined and reconstructed in RELION[41] (Figure 4B). For
229 comparison, we replicated the protocol for the Man5 ensemble as well as a protein-only ensemble (referred to as
230 BG505_Man9, BG505_Man5, and BG505_PO respectively).
231 The simulated cryo-EM map reproduced some of the defining features of the experimental data. Like in the
232 experimental map, refinement was dominated by the stable protein core and only the first few sugar residues at each
233 site are defined at the global FSC resolution and high intensity thresholds (Figure 4B). We replicated the scale-space
234 analysis from Figure 2C on the simulated maps and observed a similar trend, with the plateau again appearing around
235 1.5-2 SD (Supplemental Figure 7A-B). In fact, the volume of the 1.5 SD map contoured at its noise threshold closely
236 approximates the total sampled volume measured directly from the atomic models (dashed line). When comparing
237 BG505_Man9 to the curves produced from the BG505_Man5 and BG505_PO reconstructions, the additional glycan
238 volume recovered by the filtering and thresholding process is apparent, indicating cryo-EM can detect global changes
239 in glycosylation between reconstructions. After filtering, we observed a similar threshold-dependent evolution
240 towards a more connected topology (Figure 4C), where at the lowest thresholds the majority of the protein surface is
241 occluded, and the glycan shield is completely interconnected. We even replicated the SPARX 3-D variability analysis
242 on the simulated Man9 dataset and observed very similar results (Supplemental Figure 8A). By comparing the SPARX

243 method to the true per-voxel 3-D variance calculated without projection and refinement, we confirmed there is
244 negligible difference between the two (Supplemental Figure 8B). These results suggest the HT-AM pipeline can
245 capture globally similar features to the experimental data, however we would like to assess the accuracy of the models
246 at a more local level.
247
248 **Measuring individual glycan dynamics from synthetic cryo-EM maps**
249 Cryo-EM maps represent the average structure of all the particles that went into their construction after alignment
250 to a common reference, and therefore if each molecule is chemically identical, the intensity at a particular voxel will
251 correlate with the probability of atoms occupying that location. Thus, measuring intensity around individual glycans
252 should allow us to assess dynamics and make direct comparison to the RMSF values calculated form the HT-AM
253 ensemble. To do this, we first built and relaxed idealized glycan stalks into the 1.5 SD Gaussian filtered BG505_Man9
254 simulated map (Figure 5A) at every PNGS we could confidently identify (23/28 possible glycans per monomer). We
255 could not identify the V2 and V4 loop glycans N185e, N185h, N406, and N411, which have the highest RMSF values,
256 and the glycan at N339, which has a lower RMSF but projects directly towards the heterogeneous V4 loop. The average
257 X, Y, and Z coordinate of each β-mannose (BMA) residue (third sugar ring) was used to analyze the local environment
258 around each glycan (Figure 5A). By comparing the normalized inverse RMSF and normalized mean intensity at the
259 location of each BMA residue for the Man9 dataset we observe a strong positive correlation between intensity and
260 inverse RMSF (Figure 5B), with a correlation coefficient of ~0.89 (p=8e-8) occurring at a probe radius of 1.725Å using
261 a 0.5 SD Gaussian filtered map (see Methods). Thus, local intensity around BMA residues accurately captures ground-
262 truth differences in relative flexibility between glycans and can be used to validate our results against the experimental
263 data.
264
265 **The HT-AM pipeline reproduces physiologically relevant trends in glycan dynamics measured by cryo-EM**
266 With a method in place for measuring glycan flexibility from cryo-EM maps, we could then make direct comparisons
267 to the experimental data. First, we built and relaxed glycan stalks into the 1.5 SD Gaussian filtered BG505_293F map
268 as described above using the refined model as a scaffold. We could identify clear signal at 21/28 PNGS, two less than
269 from the simulated map, suggesting the glycans at N398 and N462 are more dynamic than captured by the simulation.
270 However, the fact that the other V2 and V4 loop glycans could not be identified in both maps means the simulated
271 dynamics agree with our experimental data at least up to the detection limit of the method. Overall, we found the
272 HT-AM pipeline captures a similar trend in ordering with a correlation coefficient between the two of ~0.46 (p = 0.03)
273 (Figure 5D).
274 We observe deviations from the experimental results around a few glycans. In addition to the V4 and V5 loop
275 glycans at N398 and N462, we also see a large deviation at N137 on the V1 loop. The V1 and V5 loops are both dynamic
276 in the experimental data (as determined by reduced resolution) but were not modeled in the simulations, therefore
277 differences are to be expected at these sites. Outside of V loops, two major deviations occur the N262 and N301

278 glycans. In the BG505_293F map, the glycan at N262 is the most ordered due to its stabilizing contacts with the gp120
279 core, and these interactions may not be accurately captured by the simulation given the restricted protein dynamics.
280 The glycan at N301 on the other hand, is dynamic in the BG505_293F map but showed both low Asn RMSF and low
281 glycan RMSF during the simulation (Supplemental Figure 3F and Figure 3E). In gp41, a large deviation also occurred at
282 the N611 glycan, which is dynamic in the BG505_293F map. We attribute these differences to restrictive sampling at
283 the protein backbone level as previously discussed (Supplemental Figure 5). To complicate the comparison, the N618
284 and N625 sites are significantly under-occupied as revealed by MS (Supplemental Figure 10). As we show in the next
285 section, sub-occupancy will cause a reduction in local signal intensity due to averaging, which will corrupt
286 measurements of glycan dynamics and even affect the dynamics and processing of neighboring glycans.
287 Overall, the positive correlation between the experimental data and theoretical predictions shows that the
288 HT-AM pipeline can capture physiologically relevant glycan sampling when protein dynamics are small. In addition, it
289 establishes that cryo-EM can be useful for validating theoretical predictions of glycan dynamics from atomistic
290 simulations.
291
292 **Detecting site-specific changes in of glycan dynamics, occupancy and chemical composition from synthetic cryo-**
293 **EM maps.**
294 Using the methodology presented above we should be able to detect site-specific changes in dynamics and occupancy
295 between differentially glycosylated Env. To test this, we performed a comparative analysis between the BG505_Man9
296 and BG505_Man5 synthetic reconstructions. Given that Man9 and Man5 glycans are identical up to the 5th mannose
297 residue, the only changes in intensity around the BMA residue should arise from differences in dynamics alone. On
298 average, we see a ~17% reduction in intensity indicative of increased dynamics, which is in line the RMSF data (Figure
299 6A). We also accurately detect the largest increase and decrease in dynamics at the N262 and N234 sites respectively.
300 To verify we could detect changes in occupancy, we removed the glycan at the N625 site from half of the
301 models and re-refined the data (referred to as BG505_Man9HO for "half occupancy"). Not surprisingly, we see a near
302 50% reduction in mean intensity from the fully occupied reconstruction around this site (Figure 6A). The glycan at
303 N625 is one of the most dynamic, so the relative intensities do not change much, however, this shows how sub-
304 occupancy can affect measurements of flexibility, and thus should be taken into account when making comparisons
305 to theoretical estimates.
306 Another technique that should be sensitive to subtle changes between similar cryo-EM maps is difference
307 mapping, which involves subtracting one cryo-EM map from another. Indeed, the change in occupancy at the N625
308 site is apparent in the BG505_Man9 - BG505_Man9HO difference map (Figure 6B). At high intensity, the signal is
309 localized around the glycan stalk and extends to the protein surface. Even at low threshold there is still no other
310 detectable difference between the two independent reconstructions. In the BG505_Man9 - BG505_Man5 difference
311 map however (Figure 6C), the difference signal is strongest where the distal tips of the Man9 glycans would be, but
312 expands to include the entire additional sampled volume at low threshold. Shown below for comparison is the

313 BG505_Man9 - BG505_PO difference map, which isolates the full contribution of the glycan shield to the cryo-EM
314 reconstruction (Figure 6D). These results establish cryo-EM as a tool for measuring glycan dynamics as well as changes
315 in chemical composition and occupancy.
316
317 **Insights gained from analysis of synthetic cryo-EM data allow improved characterization of cell-type specific**
318 **differences in glycan shield structure, dynamics, and chemical composition**
319 We illustrated using simulated data that cryo-EM is capable of capturing subtle changes in glycan structure, dynamics,
320 and chemical composition, so next we sought to test this experimentally. To do so, we collected cryo-EM data on
321 complexes of the RM20A3 Fab and BG505 SOSIP.664 expressed in two additional common cell lines that produce
322 major and minor changes in glycosylation; HEK293S (GnT-) cells, and a stable CHO cell line (referred to as BG505_293S
323 and BG505_CHO respectively). To complement this analysis, we also present site-specific MS data on all three samples
324 (Supplemental Figure 12). The CHO cell line expressed BG505 SOSIP.664 sample was provided to use by KBI Biopharma
325 as part of the GMP test-run conducted prior to the currently on-going human clinical trials and therefore should be
326 identical to the final vaccine product[42].
327 Both the BG505_CHO and BG505_293S datasets refined to ~3Å-resolution (Supplemental Figure 1) and we
328 observe nearly identical Cα positions between the three structures (Supplemental Figure 1). In addition, we
329 performed the same scale-space analysis shown in Figure 2A on each map (Supplemental Figure 7D-F), as well as
330 SPARX 3-D variability analysis (Supplemental Figure 11), and observed very similar results. Changes in site-specific
331 mean intensity highlight primarily the glycans on gp41 as having the most cell-type specific differences (Figure 7A), in
332 line with the MS data.
333 The 293S sample lacks complex glycans (Supplemental Figure 12) and should therefore allow detection of
334 complex or hybrid type glycans via difference mapping. Indeed, we see strong difference signal around the gp41
335 glycans in both the BG505_293F - BG505_293S and BG505_CHO - BG505_293S difference maps (Figure 7C-D), which
336 are primarily complex and the most under-occupied as seen by MS. We also observe signal around the primarily
337 complex glycan at N88 (Figure 7G). At low thresholds, this difference signal forms a large belt around the base of the
338 trimer, illustrating the impact of glycoform distribution and occupancy on glycan shield structure. In addition, clear
339 difference signal appears around the V2, V4, and V5 loops, specifically near the glycans at N185e and h, N398, and
340 N462, all of which are shown to be complex by MS. The MS data also shows subtle difference in both occupancy and
341 percentage of complex glycans at the N398 site between 293F and CHO cell lines, which could explain why there is
342 strong difference signal in one difference map and not the other (Figure 7C-D). The only significant differences in
343 occupancy between the 293F and 293S samples occur at the N137, N133, and N611 sites. We detect subtle difference
344 signal around the N137 site but not N611.
345 The MS data also shows reduced occupancy of the CHO sample compared to 293F at multiple sites
346 (Supplemental Figure 12). Indeed, upon closer examination we see that the difference signal around the gp41 glycans
347 at N611, N618, and N625 in the BG505_293F – BG505_CHO difference map extends all the way to the protein surface

348 (Figure 7H), indicative of changes in occupancy. In this difference map as well, there is clear signal around the N137
349 site (confirmed by MS) and to a lesser extent at the tip of the N332 glycan stalk (Figure 7I). Given the proximity of
350 N137 and N332, it is plausible that sub-occupancy at one is driving changes in dynamics and/or glycan distribution at
351 the other. Additionally, as the CHO sample came from the pre-trial GMP test-run, these observations have clear clinical
352 implications.
353
354 **Development of a probabilistic glycan-glycan interaction network reveals highly connected glycan sub-domains**
355 So far in this study we have primarily focused on quantifying and validating dynamics at the individual glycan level,
356 but we were also interested in examining concerted behavior within of the glycan shield as a whole. Both the
357 experimental and simulated cryo-EM data show extensive interconnectivity among glycans and higher-order
358 structuring, so we sought to quantify these more precisely using the HT-AM ensembles. With each glycan in our
359 models sampling a particular volume in space, neighboring glycans can explore overlapping volumes. The fraction of
360 this overlap gives a measure of the interaction probability between glycans. Figure 8A shows a heat map of the
361 normalized glycan-glycan volume overlap. It can be seen that overall, there are three main regions of overlap, – the
362 V1/V2 apex, the gp41 base, and the densely occupied gp120 outer and inner domains that includes the high-mannose
363 patch (HMP). The inter-protomer overlaps are contributed mainly by the V1/V2 glycans. The nomenclature we use
364 here was inspired by the established nomenclature for gp120 structure[43], however our definitions of each domain
365 were adapted to better capture glycan shield structure (Supplemental Figure 15).
366 To confirm this apparent structural organization, we employed graph theory to capture the glycan shield
367 topological network by interpreting the volume overlap matrix as the adjacency matrix. For our network, each glycan
368 was defined as a node of the graph, and two nodes are connected by an edge if there is at least 5% overlap (Figure
369 8B). Only those glycans from the neighboring protomers are considered that have an inter-protomer edge. Additional
370 details on network construction are provided in the Methods section. Figure 8C shows the obtained network with
371 respect to the Man9 ensemble unfolded in 2-D for the ease of visualization. The nodes in the central region around
372 the V4 loop, which is the general location of the HMP, are very highly interconnected. While in the V1/V2 apex domain
373 is also reasonably well-connected. The connections at the base near gp41 are sparse, both locally and globally with
374 the rest of the network. The analysis was repeated for the Man5 ensemble and we observe reduced overall
375 connectivity as measured by the mean node degree that goes down from ~7 to ~5 and the maximum network
376 diameter that increases from 5 to 8 hops, consistent with their smaller size (Supplemental Figure 13). Glycan 355
377 remains as the only connecting node between the apex/HMP and base regions of the Man5 network.
378 With a network in place we can analyze the relative influence of each glycan on the whole system and
379 examine its long-range structure. To do this we calculated the relative eigenvector centrality of the nodes, which is
380 given by the sum of the centrality values of the nodes connected to it. Effectively, importance of each node is
381 determined by the total importance of all its direct neighbors. The normalized eigencentrality of the glycans are
382 projected on the network as a colormap in Figure 8C. We see that the high mannose patch glycans at N332, N339,

383 N363, N386, and N392, which are densely connected within the network, have strong eigencentrality measures,
384 presumably as a reflection of their dense packing. At the apex region at the V1/V2 loop, glycans 133, 160 and 185e/h
385 have relatively high eigencentrality, and have clear connections with the other glycans in the shield. On the other
386 hand, glycans 88, 234, 276, and those in gp41 have low eigencentrality, reflecting low interaction probabilities.
387 Incorporated intrinsically into the network is a set of stable sub-graphs that represent highly connected
388 glycan clusters, as is evident in the adjacency matrix (Figure 8A). To illustrate the hierarchy of these clusters we
389 progressively stripped the network using tighter overlap cutoffs (Supplemental Figure 14). As the network is
390 degraded, we first see the formation of two large sub-graphs; one composed of the V1/V2 apex and the gp120 outer
391 domains regions; and the second composed of the gp120 inner domain along with gp41 base glycans. With a more
392 stringent threshold, the sparsely connected glycans with low eigencentrality separate out and the two sub-graphs
393 split again to form four domains; a V1/V2 apex domain; a gp120 outer domain; a gp120 inner domain; and lastly, the
394 group of unconnected gp41 base glycans (Supplemental Figure 15).
395
396 **Highly connected sub-domains within the glycan shield are resistant to enzymatic digestion and critical to the**
397 **stability of the pre-fusion trimer**
398 Experimentally validating the proposed interaction networks and sub-domain structure is not as straightforward as
399 validating individual glycan dynamics. However, we hypothesized that highly connected glycans would be protected
400 from enzymatic digestion and conversely that sparsely connected glycans would be more susceptible. If confirmed, it
401 could provide indirect validation of our network models. To test this, we exposed the BG505_293S sample (already
402 complexed with RM20A3) to digestion by Endoglycosidase H (Endo H) and performed cryo-EM on samples after 2hrs
403 and 16 hrs of digestion. Endo H cleaves only high-mannose type glycans between the first and second residues, leaving
404 the core NAG attached (Figure 9A). The datasets, referred to as BG505_EndoH2 and BG505_EndoH16, reconstructed
405 to ~3.2Å and 3.5Å respectively, with similar overall quality, resulting in highly similar atomic models (Supplemental
406 Figure 1).
407 Using the methods presented above, we characterized the glycosylation state at each PNGS from the two
408 digestion intermediates. Indeed, we found that digestion occurs non-uniformly between glycans (Figure 9B). This was
409 confirmed by the difference maps (Figure 9C) and SPARX 3-D variability analysis (Supplemental Figure 11). SDS-PAGE
410 and SEC chromatography also confirmed a gradual reduction in molecular weight with Endo H treatment
411 (Supplemental Figure 16), while scale-space analysis of the cryo-EM maps confirms an overall reduction in glycan
412 shield volume (Supplemental Figure 7D-F).
413 If we assume a linear relationship between intensity and occupancy and use the MS data to determine the
414 initial occupancy at each site, we can calculate the percent occupancy after digestion (Figure 9B). After 2 hours we
415 see complete digestion of the gp41 glycans (N611-637) while some glycans, particularly those at the high-mannose
416 patch (N197, N295, N332, N363, N386, N392, and N448), remain mostly intact. Apparent in the 0-2hr difference map
417 only is signal around the V2, V4, and V5 loop glycans (Figure 9C). This indicates the dynamic V-loop glycans are highly

susceptible to digestion. We also found partial occupancy at a few sites, suggesting non-uniform digestion between the particles. For example, the apex glycans at N156 and N160 as well as the glycans at N133, N197, and N234 all showed partial signal reduction. After 16 hours we saw almost complete digestion of the glycan shield, with only reduced occupancy detected around the previously discussed cluster composed of the N363, N386, N137, and N197 glycans, as well as a cluster composed of the N295, N332, and N448 glycans. In addition, the highly protected glycan at N262 remained completely intact after 16 hrs.

By quantifying the degree of protection from Endo H (see Methods) and comparing it to the predicted eigencentralities from the proposed network model, we obtain a correlation coefficient of ~0.8 (p=1.14e-05), suggesting highly connected glycans are resistant to enzymatic digestion (Figure 9E). Also evident is the similarity between the persistent glycan clusters and the sub-graphs presented in Supplemental Figure 14 and 15. The stripping protocol used to define the sub-graphs can be seen as mimicking the gradual digestion by Endo H, and the stable subnetworks that persist closely match the glycan clusters remaining after digestion (Figure 9D). Thus, confirming our initial hypothesis, and providing indirect experimental validation of our proposed network models.

Surprisingly, 3-D classification of the Endo H digestion intermediate datasets (Supplemental Figures 17-19) revealed an increasing degree of protein unfolding and subunit dissociation that appeared to initiate from the V1-3 loops in the trimer apex. We identified 4 distinct classes; a stable trimer, unfolding V1-3 loops, dissociated gp120, and a monomer/dimer class (Figure 10A). A "junk" class was also detected, which could be more highly degraded sample or misclassified contaminants left over from picking. As the reaction progressed, the percentages of the unfolded trimer classes increased, while the percentage of stably folded trimers decreased (Figure 10B). Although it cannot be easily confirmed, the unfolded trimers within these datasets are likely to be more completely de-glycosylated than the particles that make up the stable trimeric classes. Thus, it would appear that the highly connected glycan sub-domains are not just resistant to Endo H digestion but are also critical to maintaining the stability of the pre-fusion trimer structure, suggesting glycan shield may function beyond immune evasion.

Discussion

The HIV-1 Envelope glycan shield is a remarkable example of post-translational modification. Composed on average of ~95 N-linked glycans per Env trimer, it nearly doubles the mass of the molecule, making Env one of the most densely glycosylated proteins known. Despite its central role in steering the development of neutralizing antibodies against Env, structural and physical models of the glycan shield are incomplete. Towards that goal, we presented an integrated experimental and theoretical approach aimed at illuminating glycan shield structure and behavior at multiple levels.

Prior to this study, there were conflicting reports of glycan shield structure from cryo-EM and X-ray crystallography measurements[6,12,44]. Our results support the previous conclusions from cryo-EM experiments that the native glycan shield is highly dynamic with respect to the protein core. We did not observe stabilized glycan-glycan interactions like those reported in crystal structures of the same clone[6]. In contrast, we only observed diffuse higher-

452 order structure at low resolutions, indicative of averaging over a large ensemble of heterogenous conformations. The
453 topology of this ensemble-average structure is highly connected; however, this does not prove glycans engage in
454 specific interactions, only that they sample overlapping volumes and can potentially interact. There were several
455 methodological differences between the crystallography and our cryo-EM studies that could contribute to the
456 apparently contradictory results. For example, the crystal structures included two bound bnAb per monomer to
457 facilitate crystal packing, both of which engage multiple glycans and disrupt the native dynamics and higher-order
458 structure. Desolvation of mobile waters embedded within the glycan shield during crystallization could also potentially
459 induce the stabilized glycan-glycan contacts observed. A favorable interpretation is that the crystal structures capture
460 physiologically relevant interactions and conformations, however artificially stabilized. This interpretation would fit
461 within the conclusions of this paper as well as the results of previous atomistic simulations [6,31,32], which show that
462 glycans can interact with one another, however more dynamically than the crystal structures suggest.

463 It is known that immune responses to Env preferentially target glycan-depleted surface area[9,10], and the
464 results presented here provide the first experimentally determined mapping of this surface. As the maps are
465 contoured from high to low intensity the glycan volume expands and the topology becomes more connected, while
466 the accessible surface area shrinks. Because intensity scales with the probability of a glycan occupying that region of
467 space, the "strength" of the shielding effect will too, and thus the boundaries delineating shielded from exposed
468 surface are diffuse. Assuming ergodic dynamics, the ensemble-average structure captured by cryo-EM should be
469 identical to the time-averaged dynamics of a single Env molecule at the temperature prior to vitrification. The
470 dynamics of highly flexible glycans on a single Env create a cloud/shield over the protein, similar to the continuous
471 appearance of a rapidly rotating fan. Thus, an approaching antibody with relatively torpid dynamics will effectively
472 "see" this ensemble-average structure. Eventually though, respective timescales will determine whether the antibody
473 can break through, similar to opportunistically inserting an object between the rotating fan blades. The cryo-EM maps
474 presented here do not perfectly capture the in-vivo structure of a single fluctuating Env because they are ensembles
475 averages of chemically heterogeneous molecules. Nevertheless, if one considers the net serum response to a vaccine
476 that itself contains the exact same heterogeneity as sample we analyzed here, then our cryo-EM structure still
477 provides an accurate mapping of the average surface exposure.

478 Expanding on our structural techniques, we demonstrated that cryo-EM is capable of detecting and
479 quantifying changes in glycan dynamics, as well as occupancy and chemical composition between differentially
480 glycosylated structures. The latter of which were exclusively provided by MS prior to this study. Thus, cryo-EM could
481 provide additional validation for these measurements while contributing valuable insight into the structural impact of
482 changes in glycosylation. For instance, semi-quantitative methods to assess the distribution of unoccupied sites have
483 only recently become available[45], and here we showed that cryo-EM is highly sensitive to occupancy. Our results
484 suggest there may be more substantial differences between the CHO and 293F samples than observed by MS, and
485 given the known impact of changes in glycosylation, in particular occupancy, on the immunogenicity of Env, this has
486 clear implications for the currently on-going human clinical trials that are based on the CHO purified Env trimer. In

487 addition, we found that reduced occupancy at the N137 site opens up a large glycan hole at V1 and appears to impact
488 the composition and/or dynamics at the neighboring N332 site in V3 (Figure 7I). Similar behavior was previously
489 observed via MS by comparing glycoform distributions before and after knocking out a particular glycan[46-48]. Such
490 higher-order effects on glycan processing and dynamics can be easily explained by our structural and computational
491 findings. Beyond the detection of changes in glycosylation, we found that glycan dynamics measured by local intensity
492 in the cryo-EM maps is strongly correlated with the extent of glycan processing measured by the percentage of high-
493 mannose type glycans at each site (Supplemental Figure 12B-C). Indicating the extent of processing at each site can
494 be predicted from cryo-EM maps alone. Processing was also correlated with susceptibility to Endo H digestion
495 (correlation coefficient = 0.6167, p = 0.0029). The relationship between these variables reflects their mutual
496 dependence upon local glycan density and protein structure, both of which can reduce dynamics and restrict access
497 of Endo H and glycan processing enzymes. Indeed, we also see a strong correlation between processing and the
498 measure of glycan crowding introduced earlier (Supplemental Figure 6D).

499 We showed the ALLOSMOD-based HT-AM pipeline is capable of reproducing key features of the
500 experimental data and obtaining physiologically relevant sampling at most glycan sites. Our work represents the first
501 case of using cryo-EM to validate atomistic models of glycoprotein ensembles, and in turn, their use in guiding cryo-
502 EM analysis. We found that glycan flexibility deviated from the experimental observations when the dynamics of the
503 underlying protein are greater than can be captured by the current pipeline. Thus, exploring solutions for enhanced
504 sampling of the protein backbone represents a clear route for improving accuracy of the pipeline. Another logical next
505 step is to use the experimental cryo-EM maps to steer the glycan modeling process, which has been successful at the
506 level of protein ensembles[21,49,50]. On the experimental side, our ability to accurately capture differences in glycan
507 dynamics from cryo-EM maps is limited by additional uncertainties that we did not quantify in this paper and could
508 therefore be contributing to the observed deviations. Improving upon the cryo-EM methodology will be equally as
509 important moving forward. Again, performing flexible fitting directly into the map could provide more accurate and
510 comprehensive assessments of variations in relative dynamics throughout the glycan shield in addition to improving
511 model accuracy. Finally, utilizing sequence engineering and different expression systems to achieve uniform
512 glycosylation, or conversely, modeling heterogeneous glycans at each site, could both allow for a more accurate
513 comparison between the simulations and cryo-EM maps.

514 The glycan interaction networks gleaned from our HT-AM ensembles are more challenging to validate
515 directly with cryo-EM. However, by capturing the gradual enzymatic digestion of the glycan shield in a time-resolved
516 manner, we found that the level of protection from Endo H digestion correlated strongly with the importance of each
517 glycan within the network as measured by eigencentrality. This observation makes sense considering high centrality
518 glycans are generally in densely packed regions within the glycan shield where Endo H would have a difficult time
519 accessing (Supplemental Figure 6E). Transient deflections of neighboring glycans away from a site can temporarily
520 expose it to digestion, but the probability of this occurring will be influenced by the local density and the dynamics of
521 the surrounding glycans. As the peripheral glycans were digested away it revealed a core set of highly connected

522 glycan sub-domains which showed close resemblance to the sub-graphs generated by defining a more stringent
523 overlap cutoff within the network. Although a direct comparison between our probabilistic network model and the
524 networks constructed from the MD simulations[31,32] would not be entirely accurate, we found they show a similar
525 overall structure. Taken together, these results lend strong support to the general accuracy of our network model,
526 which is an important validation considering the HT-AM pipeline does not robustly sample the protein backbones, nor
527 does it capture temporal dynamics.

528 The observation that enzymatic de-glycosylation leads to progressive destabilization of the Env trimer was
529 somewhat surprising. The most recent investigations into the effect of glycan knockouts and de-glycosylation on Env
530 stability and viral infectivity suggest the glycan shield has little to no effect on either[51,52], however other studies
531 have found the opposite[53-55]. There is also ample evidence in the literature of glycans influencing protein dynamics
532 and folding, even in the case of Env[27,56-67], but the physical mechanisms responsible for these effects are diverse
533 and not fully understood. One potential mechanism is stabilizing interactions between the core NAG and neighboring
534 side chains, which are observed throughout Env and are common in other glycoproteins. However, Endo H leaves the
535 core NAG attached, so the stability must arise by another mechanism. We hypothesize that the dense packing of
536 glycans serves to dampen the underlying protein dynamics. In line with this, we observed an increase in glycan RMSF
537 when using Man5 in place of the larger Man9, which we interpret as more crowding in the glycan canopy (above the
538 stalks) leading to reduced dynamics closer to the protein surface. Furthermore, the correlation we observed between
539 Endo H protection and network eigencentrality suggests glycan-glycan interactions might also play a role. Even though
540 a single interaction may be weak, the combined effect of many such interactions in a densely glycosylated region
541 could become significant. The Endo H experiments provide an explicit mapping of the importance of each glycan
542 toward maintaining the integrity of the glycan shield and the overall stability of Env. This has clear implications for
543 rational vaccine design, where adding and removing glycans could have undesired effects on protein stability and the
544 structure of glycan networks. In addition, we found that instability originates from the trimer apex, around the V1-3
545 loops. This region of the structure is known to be metastable and to change conformation upon CD4 and co-receptor
546 binding[43,68]. Therefore, it is plausible the glycans in this region are contributing to maintaining stability of the pre-
547 fusion and pre-CD4 bound state. Finally, glycan induced stabilization of Env may enable hyper-mutation of the
548 underlying protein surface, thereby endowing Env with the ability to escape immune pressure with otherwise
549 deleterious mutations.

550 Looking beyond HIV-1 Env, the integrated cryo-EM and molecular modeling approach can be easily extended
551 to glycan shields from other viral fusion proteins such as Influenza, Ebola, Lassa, and Coronaviruses, and it represents
552 a potentially powerful tool for studying the structure and biophysical properties of glycoproteins in general.

553

554 ## Materials and Methods

555

556 **BG505 SOSIP.664v3 expression and purification in HEK293F, 293S, and CHO cells**

557 BG505 SOSIP.664v3 was expressed and purified from HEK293F and HEK293S suspension cell cultures. Briefly, cell
558 cultures were expanded then transiently co-transfected with Env and Furin plasmid DNA. Expression was carried out
559 for 6 days followed by supernatant harvesting and protein purification via PGT121 antibody and size-exclusion
560 chromatograph. BG505 SOSIP.664 GMP was expressed and purified from a stable CHO cell-line as described in Dey,
561 et al., 2017 [42] and was provided to us by KBI Biopharma.
562
563 **RM20A3 IgG isolation and Fab purification**
564 The monoclonal antibody RM20A3 was isolated from a BG505 SOSIP.664 immunized rhesus macaque (rh2011) at
565 week 53, which was 1 week after the 6th immunization[69], using single cell FACS sorting and antibody cloning as
566 previously described[70]. A full description of the antibody will be presented in elsewhere. The RM20A3 Fab was
567 expressed and purified from HEK293F cells as described previously[71].
568
569 **Site-specific mass spectrometry**
570 The BG505 SOSIP.664 samples expressed in HEK293F and CHO cell lines were prepared for MS analysis as previously
571 described[35], and the HEK293S sample was prepared with slight modifications on that protocol. In brief, since glycans
572 from HEK293S cells are all oligomannose (Man5-Man9) BG505 SOSIP.664 samples peptide mixtures were
573 deglycosylated only with endoglycosidase PNGase F in 100 mM ammonium bicarbonate prepared with O18-water.
574 The samples were analyzed on an Q Exactive HF-X mass spectrometer (Thermo). Samples were injected directly onto
575 a 25 cm, 100 µm ID column packed with BEH 1.7 µm C18 resin (Waters). Samples were separated at a flow rate of 300
576 nL/min on a nLC 1200 (Thermo). Solutions A and B were 0.1% formic acid in 5% and 80% acetonitrile, respectively. A
577 gradient of 1–25% B over 160 min, an increase to 40% B over 40 min, an increase to 90% B over another 10 min and
578 held at 90% B for 30 min was used for a 240 min total run time. Column was re-equilibrated with solution A prior to
579 the injection of sample. Peptides were eluted directly from the tip of the column and nanosprayed directly into the
580 mass spectrometer by application of 2.8 kV voltage at the back of the column. The HFX was operated in a data
581 dependent mode. Full MS1 scans were collected in the Orbitrap at 120k resolution. The ten most abundant ions per
582 scan were selected for HCD MS/MS at 25NCE. Dynamic exclusion was enabled with exclusion duration of 10 s and
583 singly charged ions were excluded.
584 The MS data were processed as previously[35]. The data were searched against the proteome database and
585 quantified using peak area in Integrated Proteomics Pipeline-IP2. For samples produced in HEK293F and CHO cells,
586 glycosites (N-X-T/S) with N + 203 were identified as sites with high mannose glycans removed by the initial Endo H
587 treatment (high mannose), the glycosites with N + 3 were identified as sites whose glycans were complex type glycans
588 removed by PNGase F, and glycosites with N+0 were identified as sites that had no glycan prior to endoglycosidase
589 treatments. Since samples produced in HEK293S are only high mannose, and were treated only with PNGase F, sites
590 with N+3 were identified as sites with high mannose.
591

592 **Endoglycosidase H digestion of BG505 SOSIP.664 for cryo-EM experiments**

593 1.4 mg of purified BG505 SOSIP.664 from HEK293S cells in complex with RM20A3 was mixed with 80,000 units of
594 endoglycosidase H (Endo H; New England Biolabs) in non-denaturing reaction buffer to a final volume of 1ml and
595 incubated at 37°C for 2 hrs and 16 hrs (two separate reactions from the same master mix). To quench the reaction
596 and purify the sample for cryo-EM experiments, it was run over an SEC column and fractions were pulled and
597 concentrated. When the sample was used for MS or SDS PAGE, the reaction was heat quenched at 100°C for 10min.

598

599 Cryo-EM sample preparation for BG505 SOSIP.664 in complex with RM20A3

600 Purified BG505 SOSIP.664v3 from either HEK293F, HEK293S, or CHO cells was incubated overnight at 4°C in a 1:6
601 molar ratio of SOSIP trimer to purified RM20A3 Fab. The complex was purified from the remaining free Fab by SEC,
602 then concentrated to ~6mg/ml in 1X TBS buffer. Purified complex was then mixed with n-dodecyl β-D-maltoside
603 (DDM; Anatrace) to a final concentration of 0.06 mM immediately prior to vitrification. 3ul of this solution was then
604 applied to plasma cleaned Cflat 1.2x1.3 holey carbon grids and plunge frozen at 4°C and 100% chamber humidity with
605 an FEI Vitrobot MarkIV (Thermo-Fisher) after blotting for 7 s. Grids were transferred in liquid nitrogen to an autoloader
606 and into one of the two microscopes for imaging. The Endo H digestion samples were prepared in the same way after
607 purification from the reaction mixture.

608

609 **Cryo-EM data collection**

610 Imaging was performed on either a FEI Titan Krios or Talos Arctica (Thermo-Fisher), operating at 300KeV and 200KeV
611 respectively. The microscopes were taken through a standard coma-free alignment protocol before each imaging
612 session. Movie micrographs were collected on a Gatan K2 Summit direct electron detector (Gatan) operating in
613 counting mode. Imaging was adjusted to achieve the same total dose for both microscopes of ~50 e-/Å^2. Movies were
614 collected at a frame rate of 1/250 ms. The final calibrated pixel sizes were 1.03 Å for the Krios and 1.15 Å for the
615 Arctica. During data collection, frames were aligned and dose weighted in real-time with MotionCor2 [72] and CTF fits
616 were performed using Gctf[73] to monitor image quality.

617

618 **Cryo-EM data processing**

619 All non-custom cryo-EM data processing, which includes particle picking, 2-D and 3-D classification, refinement, per-
620 particle CTF refinement, and postprocessing were performed using a combination RELION-2/3[41,74] and CryoSparc
621 v1[75]. Micrograph sorting and carbon masking were conducted with EMHP[76] when necessary. The following
622 general processing protocol was used for all datasets. After global CTF estimation, all micrographs with resolution
623 estimates greater than 6Å were excluded from the dataset. Particle picking was then performed using the Gaussian
624 disk template in RELION-2. Extracted particles were then transferred to CryoSparc 1 servers for initial 2-D and 3-D
625 processing. The initial processing usually involved 2 rounds of 2-D classification and subset selection followed by 1
626 round of ab-initio classification into 4 classes. Clean classes were then pooled and refined using the class average as

627 a template. All refinements were run with C3 symmetry unless stated otherwise. At this stage, refinement meta-data
628 was downloaded from CryoSparc and re-formatted for processing in RELION. Following refinement in RELION, per-
629 particle CTF and beam-tilt refinement were performed, followed by another round of refinement, and 1 or more
630 rounds of 3-D classification using a tight mask and local angular searches only. For the Endo H treated samples,
631 subtractive classification in RELION-3 was also performed at this stage to focus classification to the heterogenous
632 apex. Final sorted particles were then refined using C3 symmetry and sharpened with a data-derived B-factor in
633 RELION using the mask from refinement. Map segmentations were performed with Segger[77] in UCSF Chimera[78].
634 Local resolution plots were calculated in RELION using the method presented in Kucukelbir et. al., 2014 [17].
635

636 **Model building and refinement**

637 Model building and refinement was initially performed on the sharpened 3.1Å-resolution BG505_293F map using the
638 following protocol. First, models of the complete gp120, gp41, and the RM20A3 heavy and light chains, were
639 generated with SWISS-MODEL[79] using PDBID:5ACO as the Env template. These models were then docked into the
640 sharpened map and saved as a single PDB file using UCSF Chimera[78]. This model was then refined with C3 symmetry
641 in Rosetta[80], asking for 300 models. Each model was scored with MolProbity [81] and EMRinger [82], and the one
642 with the best overall score was selected. N-linked glycans were then modeled into the map unit using Coot[83] and
643 the glycosylated model was re-refined in Rosetta using the newly introduced glycan force fields [84]. After refinement,
644 coordinates outside of clear density were deleted and manual adjustments to the structure were made with Coot if
645 necessary followed by refinement in Rosetta. This final BG505_293F model then served as the input for a single round
646 of Rosetta refinement into the BG505_CHO, BG505_293S, BG505_EndoH2, and BG505_EndoH16 maps. Ca RMSDs
647 were calculated in PyMol[85] {The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.}.
648

649 **Scale-space analysis**

650 Sharpened cryo-EM maps were loaded into MatLab (2018b) {MATLAB and Image Processing Toolbox Release 2018b,
651 The MathWorks, Inc., Natick, Massachusetts, United States.} using the *ReadMRC.m* script obtained from the following
652 GitHub: https://github.com/Sarofi/PythonImages/blob/master/Matlab/FredSigworth/EMIODist/ReadMRC.m. All
653 subsequent analysis was performed in MatLab and the MatLab ImageProcessing Toolbox. To determine the noise
654 threshold, the map was binarized at a series of intensities, ranging from the maximum to minimum intensity (at 100-
655 200 evenly spaced intervals), and the number of connected components of each binary volume was recorded. This
656 function has a characteristic double maximum, where the high intensity peak corresponding to signal and the low
657 intensity peak to noise. The noise threshold was then defined as the global minimum between these two peaks. If the
658 function plateaus at this global minimum, then the lowest intensity point in the plateau defined the noise threshold.
659 The total volume at the noise threshold was calculated as the sum of the volumes of all the connected components.
660 Noise threshold and total volume were then calculated as a function of Gaussian filter SD to create the plot shown in
661 Figure 2B and Supplemental Figure 7. The process was performed before and after resampling and intensity

662 equalization, also with the Fabs masked out to better capture the relevant signal (Supplemental Figure 7). The mask
663 used to delete the Fab here and for difference mapping has been deposited to the EMDB with the BG505_293F map.
664
665 **SPARX 3-D variability analysis**
666 3-D variance analysis was carried out using the function *sx3dvariability* from the SPARX software package[37].
667 Refinement *data.star* files from the final refinement in RELION (before per-particle CTF refinement) for each dataset
668 were converted into SPARX format with the function *sxrelion2sparx*. 3-D variability maps were then calculated with
669 the function *sx3dvariability* using C3 symmetry. Prior to calculating 3-D variability, projections were internally down
670 sampled 2x and lowpass filtered to twice the binned Nyquist frequency to reduce noise in the resulting variability
671 map. The method is referred to as 3-D "variability" because the resulting map is not a true variance since positivity is
672 enforced when reconstructing the 3-D map from the 2-D variance fields. Using the simulated cryo-EM data from the
673 HT-AM Man9 ensemble we calculated the true per-voxel variance map and compared the results to the equivalent
674 SPARX 3-D variability map. We found a negligible difference between the two.
675
676 **Cryo-EM difference mapping**
677 First, maps were re-sampled onto the same grid and equalized with the RELION-3 function *adjust_power.* Because
678 data was collected on two microscopes operating at different magnifications, all volumes were re-sampled and
679 adjusted to the larger pixel size of 1.15Å. Difference maps were calculated by subtracting one map from another in
680 the order mentioned in the text followed by smoothing with a Gaussian kernel for visualization. These operations
681 were performed either in UCSF Chimera[78] or in MatLab.
682
683 **Site-specific local intensity analysis from cryo-EM maps**
684 To measure map intensity around each glycan, first the sharpened map was filtered with a 1.5 SD Gaussian kernel and
685 saved and opened in Coot, along with the refined atomic model for the BG505_293F dataset. At each PNGS where a
686 glycan could be confidently identified, a full idealized N-linked glycan stalk was added if not already present and
687 adjusted to fit the density with the *refine tree* function[83]. The volumes were then resampled and equalized as
688 described above and loaded into MatLab. The average X,Y, and Y coordinates of each BMA residue were then used to
689 localize each glycan within the volume, and the average voxel intensity of all positive intensity voxels within a sphere
690 around each BMA residue was calculated for a range of probe diameters along with the Pearson correlation between
691 the normalized inverse RMSF and normalized mean intensities (Supplemental Figure 9B). The cumulative absolute
692 deviation from the mean across all glycans was calculated for each individual glycan residue, which suggested the
693 BMA residue closely approximates the full-glycan RMSF (Supplemental Figure 9A). The correlation analysis was
694 repeated using each of the 11 individual glycan residues in place of the average and we saw negligible differences
695 (Supplemental Figure 9C). This same model was used to analyze the BG505_CHO data, while minor adjustments to
696 the N611 glycan were made to for the BG505_293S map. This same procedure was used to analyze all the simulated

697 cryo-EM maps. To quantify the percent occupancy at each site after Endo H digestion we assumed a linear relationship
698 between signal intensity and occupancy, with any intensity <= 0 being considered fully digested. Initial occupancies
699 were determined by MS. To calculate the Endo H protection score we added the percent occupancy at each site from
700 the EndoH2 and EndoH16 maps and normalized the results between 0 and 1. This can be seen as the integral of
701 occupancy with respect to reaction time. All local intensity analysis was repeated before and after intensity
702 equalization in RELION for comparison (Supplemental Figure 10).
703
704 **Glycan crowding analysis**
705 We defined a simple crowding score as the number of glycans within a spherical radius around a glycan. To localize
706 each glycan, we calculated the average location of the BMA oxygen atom across all 1000 models in the Man9
707 ensemble. Then, the number of glycans within a radius around each point was recorded as a function of radius (0Å-
708 135Å with 135Å being the maximum pairwise distance) along with the Pearson correlation between the normalized
709 crowding score and either the normalized inverse RMSF, normalize local map intensity, or normalized percent high-
710 mannose.
711
712 **ALLOSMOD-based high-throughput atomistic modeling (HT-AM) of fully glycosylated Env**
713 The computational atomistic ensemble modeling method that we employed is as follows. First, we generated a robust
714 ensemble of atomistic models of BG505 SOSIP.664 by implementing the ALLOSMOD[39,86] package of
715 MODELLER[87,88] in a streamlined pipeline. The BG505 SOSIP.664 protein scaffold was homology modeled and the
716 missing residues in the hypervariable and dynamic V2 and V4 loops (residues 186-189 and residues 400-410) were
717 modeled *ab initio*. All the disulfide bonds were maintained as additional restraints. 100 models were generated, and
718 the best 10 were selected (Figure 3A) based on MODELLER optimization scores and stereochemistry scores as
719 determined by PROCHECK[89]. For each of the 10 selected protein structures, glycans were initially added at the
720 known glycosylation sites based on ideal geometries as dictated by CHARMM36 [90,91] force field internal
721 coordinates. For simplicity, we chose to apply uniform glycosylation of mannose-9 because the MS data suggests this
722 is the most common glycan type on BG505 SOSIP.664[47]. This was followed by a 1Å random deviation added to
723 the overall atomic coordinates. This template-free glycan modeling method optimizes an energy function given by a
724 combination of spatial restraints and CHARMM36 glycan forcefield terms, to enforce proper stereochemistry. The
725 generated structures were further relaxed (Figure 3B) with several steps of conjugate gradient minimization followed
726 by simulated annealing as described by Guttman et. al. [39]. The glycans and the loop regions were kept flexible during
727 the refinement steps. 100 fully glycosylated structures were modeled from each of the 10 selected protein models,
728 resulting in the final ensemble containing 1000 different poses.
729
730 **RMSF and sampling volume calculations from HT-AM ensembles**

731 For both the native and all-man9 glycosylated models, the root mean square fluctuations (RMSF) of each glycan *n* was
732 calculated as an average over all its heavy atoms, by the following equation:
733

734 $$RMSF_n = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{M}\sum_{m=1}^{M}|\vec{r}_{mnk} - \langle\vec{r}_{nk}\rangle|^2}$$

735
736
737 where $\vec{r}_{mnk}$ is the atomic position of heavy atom *k* of glycan *n* in snapshot *m*, $\langle\vec{r}_{nk}\rangle = (\frac{1}{M})\sum_{m=1}^{M}\vec{r}_{mnk}$ is the
738 average atomic position of heavy atom *k* in glycan *n*. K is the total number of heavy atoms in the glycan (127 for Man9).
739 The ensemble for each model contains 1000 snapshots, making *M* = 1000 snapshots for each of the two models. The
740 standard deviations (s.d.) were obtained by dividing the 1000 snapshots into 4 sets of M=250, and calculating the
741 square root of the variations from the four sets of RMSF values. The single-model average RMSF was similarly
742 calculated, where M reduces to 100. The value obtained for each of the 10 models were averaged to get the final per-
743 model RMSF value. For calculating the sampled volumes, volumetric surface maps were created for each individual
744 glycan using the volmap plugin of VMD[92] (http://www.ks.uiuc.edu/Research/vmd/), including all 1000 models, with
745 1Å grid size. The enclosed volume of these generated surface models was calculated using UCSF CHIMERA [78]
746 (http://www.rbvi.ucsf.edu/chimera).
747
748 **Graph theoretic analysis of ALLOSMOD ensembles**
749 The inter-glycan overlap is calculated as the total fraction of heavy atoms from the two glycans that come within 5Å
750 of each other. A single mannose-9 glycan has 127 heavy atoms. Since our ensemble is composed of 1000 possible
751 structures, there are effectively 127,000 heavy atoms per ensemble of mannose-9 at one position. The fraction of the
752 total number of heavy atoms from two neighboring ensembles that come within contact distance defines the overlap
753 fraction. An overlap greater than or equal to 50% of heavy atoms from two neighboring mannose-9 glycans is assigned
754 as 1. This overlap matrix is used to define the adjacency matrix for our network analysis. Each glycan functions as a
755 node of the graph, and two nodes are connected by an edge if there is at least 5% overlap as per our overlap definition
756 given above. The edge length is inversely proportional to the overlap value, i.e., the larger the overlap, the closer two
757 nodes (glycans) are in the graph. Only those glycans from the neighboring protomers are considered, that have an
758 inter-protomer edge. All graph theory and network-based analysis were performed using Python {Python Software
759 Foundation, https://www.python.org/} and Matlab_R2018a packages.
760
761 **Simulated cryo-EM data creation and processing from ALLOSMOD ensembles**
762 First, each model generated from the HT-AM pipeline was converted into *mrc* volume file with the same box and pixel
763 sizes as the experimental data (in this case we used the lower magnification from the Talos Arctica of 1.15Å/pixel)

using the EMAN2[93] function e2*pdb2mrc*, with the max resolution set as twice the voxel size. Next, each volume was projected at 100 uniformly distributed angles, either with or without white noise, using the RELION function *relion_project*. All 1000 *particles.star* files were then combined (100K total projections) and refinement was initiated using the BG505_293F map with Fabs removed as an initial model. In all cases the resolution converged around the Nyquist frequency (2.3Å), and the final maps were sharpened using a data-derived B-factor and soft mask in RELION. The absolute intensity scales of the simulated maps are different from the experimental maps however we observed minimal change with intensity equalization. SPARX 3-D variability maps were calculated in the same manner as the experimental data however without including CTF information. Because SPARX re-extracts particles, the center of each projection was given in place of the micrograph coordinates for each particle. True per-voxel average and variance maps were calculated by taking the mean and variance of each voxel across all 1000 *mrc* volumes used for synthetic data projections.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The electron potential maps have been deposited in the Electron Microscopy Data Bank under accession codes XXXXX. Atomic models have been deposited in the Protein Data Bank under accession codes XXXX.

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 KBI Biopharma, Antu Dey, and John P. Moore for supplying GMP material and MS analysis of GMP material. 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

797 The authors declare no competing interests.

798

799 **Funding**

800 This work was supported by the National Institute of Allergy and Infectious Diseases grants: UM1 AI100663 (to A.B.W.,
801 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
802 by the Bill and Melinda Gates Foundation through the Collaboration for AIDS Vaccine Discovery (CAVD)
803 grant OPP1115782 (A.B.W.). This work was funded by grants from the NIH (Center for HIV/AIDS Vaccine Immunology
804 and immunogen Discovery, UM1 AI100645; Consortia for HIV/AIDS Vaccine Development, UM1 AI144371). S.C. was
805 also partially supported by the Center for Nonlinear Studies (CNLS) at Los Alamos National Laboratory (LANL).

806

807 **Author contributions**

808 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.;
809 Protein expression and purification, Z.T.B., C.A.C, and J.L.T; Antibody isolation, M.J.vG; Mass-spectrometry
810 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,
811 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.,
812 S.G. and A.B.W.; Supervision, S.G. and A.B.W.

813

814 **Author ORCIDs**

815 Zachary T. Berndsen: https://orcid.org/0000-0003-4531-9907
816 Srirupa Chakraborty: https://orcid.org/0000-0002-7155-2764
817 Xiaoning Wang: https://orcid.org/ 0000-0003-0153-324X
818 Christopher A. Cottrell: https://orcid.org/0000-0002-3364-3083
819 Jonathan L. Torres: https://orcid.org/0000-0003-0137-8497
820 Jolene K. Diedrich: https://orcid.org/ 0000-0001-6489-4558
821 Cesar A. López: https://orcid.org/0000-0003-4684-3364
822 John R. Yates: https://orcid.org/ 0000-0001-5267-1672
823 Marit J. van-Gils: https://orcid.org/ 0000-0003-3422-8161
824 James C. Paulson: https://orcid.org/ 0000-0003-4589-5322
825 S. Gnanakaran: https://orcid.org/0000-0002-9368-3044
826 Andrew B. Ward: https://orcid.org/0000-0001-7153-3769

827

828 ## References

829

830 1. Fenouillet E, Gluckman JC, Jones IM. Functions of HIV envelope glycans. Trends in Biochemical Sciences. 1994;19:
831 65–70. doi:10.1016/0968-0004(94)90034-5

832 2. Grundner C, Pancera M, Kang J-M, Koch M, Sodroski J, Wyatt R. Factors limiting the immunogenicity of HIV-1
833 gp120 envelope glycoproteins. Virology. 2004;330: 233–248. doi:10.1016/j.virol.2004.08.037
834 3. Dacheux L, Moreau A, Ataman-Onal Y, Biron F, Verrier B, Barin F. Evolutionary dynamics of the glycan shield of
835 the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12
836 epitope. J Virol. 2004;78: 12625–12637. doi:10.1128/JVI.78.22.12625-12637.2004
837 4. Lasky LA, Groopman JE, Fennie CW, Benz PM, Capon DJ, Dowbenko DJ, et al. Neutralization of the AIDS retrovirus
838 by antibodies to a recombinant envelope glycoprotein. Science. American Association for the Advancement of
839 Science; 1986;233: 209–212. doi:10.1126/science.3014647
840 5. Doores KJ, Bonomelli C, Harvey DJ, Vasiljevic S, Dwek RA, Burton DR, et al. Envelope glycans of immunodeficiency
841 virions are almost entirely oligomannose antigens. Proc Natl Acad Sci USA. National Academy of Sciences;
842 2010;107: 13800–13805. doi:10.1073/pnas.1006498107
843 6. Stewart-Jones GBE, Soto C, Lemmin T, Chuang G-Y, Druz A, Kong R, et al. Trimeric HIV-1-Env Structures Define
844 Glycan Shields from Clades A, B, and G. Cell. Elsevier Inc; 2016;165: 813–826. doi:10.1016/j.cell.2016.04.010
845 7. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications
846 of contemporary HIV-1 variation. Br Med Bull. 2001;58: 19–42. doi:10.1093/bmb/58.1.19
847 8. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds
848 and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus
849 envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. Journal of Biological Chemistry.
850 1990;265: 10373–10382.
851 9. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, et al. Antibody neutralization and escape by HIV-1. Nature.
852 Nature Publishing Group; 2003;422: 307–312. doi:10.1038/nature01470
853 10. Zhou T, Doria-Rose NA, Cheng C, Stewart-Jones GBE, Chuang G-Y, Chambers M, et al. Quantification of the Impact
854 of the HIV-1-Glycan Shield on Antibody Elicitation. Cell Rep. 2017;19: 719–732. doi:10.1016/j.celrep.2017.04.013
855 11. Ward AB, Wilson IA. The HIV-1 envelope glycoprotein structure: nailing down a moving target. Immunol Rev.
856 2017;275: 21–32. doi:10.1111/imr.12507
857 12. Lyumkis D, Julien J-P, Potter CS, Burton DR, Wilson IA, Ward AB. Cryo-EM structure of a fully glycosylated soluble
858 cleaved HIV-1 envelope trimer. Science. 2013;342: 1484–1490. doi:10.1126/science.1245627
859 13. Lee JH, Ozorowski G, Ward AB. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer.
860 Science. American Association for the Advancement of Science; 2016;351: 1043–1048.
861 doi:10.1126/science.aad2450
862 14. Jonić S. Cryo-electron Microscopy Analysis of Structurally Heterogeneous Macromolecular Complexes. Comput
863 Struct Biotechnol J. 2016;14: 385–390. doi:10.1016/j.csbj.2016.10.002
864 15. Murata K, Wolf M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules.
865 Biochim Biophys Acta Gen Subj. 2018;1862: 324–334. doi:10.1016/j.bbagen.2017.07.020

866 16. Bai X-C, McMullan G, Scheres SHW. How cryo-EM is revolutionizing structural biology. Trends in Biochemical
867 Sciences. 2015;40: 49–57. doi:10.1016/j.tibs.2014.10.005
868 17. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. Nat Methods.
869 Nature Publishing Group; 2014;11: 63–65. doi:10.1038/nmeth.2727
870 18. Pintilie G, Chen D-H, Haase-Pettingell CA, King JA, Chiu W. Resolution and Probabilistic Models of Components in
871 CryoEM Maps of Mature P22 Bacteriophage. Biophysical Journal. 2016;110: 827–839.
872 doi:10.1016/j.bpj.2015.11.3522
873 19. Herzik MA, Fraser JS, Lander GC. A Multi-model Approach to Assessing Local and Global Cryo-EM Map Quality.
874 Structure. 2019;27: 344–358.e3. doi:10.1016/j.str.2018.10.003
875 20. Jakobi AJ, Wilmanns M, Sachse C. Model-based local density sharpening of cryo-EM maps. Elife. eLife Sciences
876 Publications Limited; 2017;6: 213. doi:10.7554/eLife.27131
877 21. Bonomi M, Pellarin R, Vendruscolo M. Simultaneous Determination of Protein Structure and Dynamics Using
878 Cryo-Electron Microscopy. Biophysical Journal. 2018;114: 1604–1613. doi:10.1016/j.bpj.2018.02.028
879 22. Hadden JA, Perilla JR, Schlicksup CJ, Venkatakrishnan B, Zlotnick A, Schulten K. All-atom molecular dynamics of
880 the HBV capsid reveals insights into biological function and cryo-EM resolution limits. Elife. eLife Sciences
881 Publications Limited; 2018;7: 165101. doi:10.7554/eLife.32478
882 23. Ramírez-Aportela E, Vilas JL, Glukhova A, Melero R, Conesa P, Martínez M, et al. Automatic local resolution-based
883 sharpening of cryo-EM maps. Cowen L, editor. Bioinformatics. 2019;66: 213. doi:10.1093/bioinformatics/btz671
884 24. Beckers M, Jakobi AJ, Sachse C. Thresholding of cryo-EM density maps by false discovery rate control. IUCrJ.
885 International Union of Crystallography; 2019;6: 18–33. doi:10.1107/S2052252518014434
886 25. López CA, Swift MF, Xu X-P, Hanein D, Volkmann N, Gnanakaran S. Biophysical Characterization of a Nanodisc with
887 and without BAX: An Integrative Study Using Molecular Dynamics Simulations and Cryo-EM. Structure. 2019;27:
888 988–999.e4. doi:10.1016/j.str.2019.03.013
889 26. Kumar S, Cieplak P. Role of N-glycosylation in activation of proMMP-9. A molecular dynamics simulations study.
890 Yang C-M, editor. PLoS ONE. Public Library of Science; 2018;13: e0191157. doi:10.1371/journal.pone.0191157
891 27. Lee HS, Qi Y, Im W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis
892 and molecular dynamics simulation study. Sci Rep. Nature Publishing Group; 2015;5: 8926.
893 doi:10.1038/srep08926
894 28. Dong C, Lee J, Kim S, Lai W, Webb EB, Oztekin A, et al. Long-ranged Protein-glycan Interactions Stabilize von
895 Willebrand Factor A2 Domain from Mechanical Unfolding. Sci Rep. Nature Publishing Group; 2018;8: 16017–11.
896 doi:10.1038/s41598-018-34374-y
897 29. Hang I, Lin C-W, Grant OC, Fleurkens S, Villiger TK, Soos M, et al. Analysis of site-specific N-glycan remodeling in
898 the endoplasmic reticulum and the Golgi. Glycobiology. 2015;25: 1335–1349. doi:10.1093/glycob/cwv058

899 30. Ferreira R-C, Grant OC, Moyo T, Dorfman JR, Woods RJ, Travers SA, et al. Structural Rearrangements Maintain the
900 Glycan Shield of an HIV-1 Envelope Trimer After the Loss of a Glycan. Sci Rep. Nature Publishing Group; 2018;8:
901 15031–15. doi:10.1038/s41598-018-33390-2
902 31. Yang M, Huang J, Simon R, Wang L-X, MacKerell AD. Conformational Heterogeneity of the HIV Envelope Glycan
903 Shield. Sci Rep. Nature Publishing Group; 2017;7: 4435–15. doi:10.1038/s41598-017-04532-9
904 32. Lemmin T, Soto C, Stuckey J, Kwong PD. Microsecond Dynamics and Network Analysis of the HIV-1 SOSIP Env
905 Trimer Reveal Collective Behavior and Conserved Microdomains of the Glycan Shield. Structure. 2017;25: 1631–
906 1639.e2. doi:10.1016/j.str.2017.07.018
907 33. Sanders RW, Derking R, Cupo A, Julien J-P, Yasmeen A, de Val N, et al. A next-generation cleaved, soluble HIV-1
908 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing
909 antibodies. Trkola A, editor. PLoS Pathog. Public Library of Science; 2013;9: e1003618.
910 doi:10.1371/journal.ppat.1003618
911 34. Sanders RW, Moore JP. Native-like Env trimers as a platform for HIV-1 vaccine design. Immunol Rev. 2017;275:
912 161–182. doi:10.1111/imr.12481
913 35. Cao L, Pauthner M, Andrabi R, Rantalainen K, Berndsen Z, Diedrich JK, et al. Differential processing of HIV envelope
914 glycans on the virus and soluble recombinant trimer. Nat Commun. 2nd ed. Nature Publishing Group; 2018;9:
915 3693. doi:10.1038/s41467-018-06121-4
916 36. Robins V. Algebraic Topology. 2013.
917 37. Hohn M, Tang G, Goodyear G, Baldwin PR, Huang Z, Penczek PA, et al. SPARX, a new environment for Cryo-EM
918 image processing. J Struct Biol. 2007;157: 47–55. doi:10.1016/j.jsb.2006.07.003
919 38. Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. Cell. 2015;161:
920 438–449. doi:10.1016/j.cell.2015.03.050
921 39. Guttman M, Weinkam P, Sali A, Lee KK. All-atom ensemble modeling to analyze small-angle x-ray scattering of
922 glycosylated proteins. Structure. 2013;21: 321–331. doi:10.1016/j.str.2013.02.004
923 40. Liang Y, Guttman M, Williams JA, Verkerke H, Alvarado D, Hu S-L, et al. Changes in Structure and Antigenicity of
924 HIV-1 Env Trimers Resulting from Removal of a Conserved CD4 Binding Site-Proximal Glycan. Silvestri G, editor. J
925 Virol. 2016;90: 9224–9236. doi:10.1128/JVI.01116-16
926 41. Kimanius D, Forsberg BO, Scheres SH, Lindahl E. Accelerated cryo-EM structure determination with parallelisation
927 using GPUs in RELION-2. Elife. eLife Sciences Publications Limited; 2016;5: 19. doi:10.7554/eLife.18722
928 42. Dey AK, Cupo A, Ozorowski G, Sharma VK, Behrens A-J, Go EP, et al. cGMP production and analysis of BG505
929 SOSIP.664, an extensively glycosylated, trimeric HIV-1 envelope glycoprotein vaccine candidate. Biotechnol
930 Bioeng. 2017. doi:10.1002/bit.26498
931 43. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope
932 glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. Nature Publishing
933 Group; 1998;393: 648–659. doi:10.1038/31405

934 44. Lee JH, de Val N, Lyumkis D, Ward AB. Model Building and Refinement of a Natively Glycosylated HIV-1 Env Protein
935 by High-Resolution Cryoelectron Microscopy. Structure. 2015;23: 1943–1951. doi:10.1016/j.str.2015.07.020
936 45. Cao L, Diedrich JK, Kulp DW, Pauthner M, He L, Park S-KR, et al. Global site-specific N-glycosylation analysis of HIV
937 envelope glycoprotein. Nat Commun. Nature Publishing Group; 2017;8: 14954. doi:10.1038/ncomms14954
938 46. Behrens A-J, Kumar A, Medina-Ramírez M, Cupo A, Marshall K, Cruz Portillo VM, et al. Integrity of Glycosylation
939 Processing of a Glycan-Depleted Trimeric HIV-1 Immunogen Targeting Key B-Cell Lineages. J Proteome Res.
940 American Chemical Society; 2018;: acs.jproteome.7b00639. doi:10.1021/acs.jproteome.7b00639
941 47. Behrens A-J, Vasiljevic S, Pritchard LK, Harvey DJ, Andev RS, Krumm SA, et al. Composition and Antigenic Effects
942 of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. Cell Rep. 2016;14: 2695–2706.
943 doi:10.1016/j.celrep.2016.02.058
944 48. Pritchard LK, Spencer DIR, Royle L, Bonomelli C, Seabright GE, Behrens A-J, et al. Glycan clustering stabilizes the
945 mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. Nat Commun. Nature
946 Publishing Group; 2015;6: 7479. doi:10.1038/ncomms8479
947 49. Bonomi M, Hanot S, Greenberg CH, Sali A, Nilges M, Vendruscolo M, et al. Bayesian Weighing of Electron Cryo-
948 Microscopy Data for Integrative Structural Modeling. Structure. 2019;27: 175–188.e6.
949 doi:10.1016/j.str.2018.09.011
950 50. Bonomi M, Vendruscolo M. Determination of protein structural ensembles using cryo-electron microscopy. Curr
951 Opin Struct Biol. 2019;56: 37–45. doi:10.1016/j.sbi.2018.10.006
952 51. Rathore U, Saha P, Kesavardhana S, Kumar AA, Datta R, Devanarayanan S, et al. Glycosylation of the core of the
953 HIV-1 envelope subunit protein gp120 is not required for native trimer formation or viral infectivity. J Biol Chem.
954 American Society for Biochemistry and Molecular Biology; 2017;292: 10197–10219.
955 doi:10.1074/jbc.M117.788919
956 52. Depetris RS, Julien J-P, Khayat R, Lee JH, Pejchal R, Katpally U, et al. Partial enzymatic deglycosylation preserves
957 the structure of cleaved recombinant HIV-1 envelope glycoprotein trimers. J Biol Chem. American Society for
958 Biochemistry and Molecular Biology; 2012;287: 24239–24254. doi:10.1074/jbc.M112.371898
959 53. Wang W, Nie J, Prochnow C, Truong C, Jia Z, Wang S, et al. A systematic study of the N-glycosylation sites of HIV-
960 1 envelope protein on infectivity and antibody-mediated neutralization. Retrovirology. BioMed Central; 2013;10:
961 14. doi:10.1186/1742-4690-10-14
962 54. Lee WR, Syu WJ, Du B, Matsuda M, Tan S, Wolf A, et al. Nonrandom distribution of gp120 N-linked glycosylation
963 sites important for infectivity of human immunodeficiency virus type 1. Proceedings of the National Academy of
964 Sciences. 1992;89: 2213–2217. doi:10.1073/pnas.89.6.2213
965 55. Ohgimoto S, Shioda T, Mori K, Nakayama EE, Hu H, Nagai Y. Location-specific, unequal contribution of the N
966 glycans in simian immunodeficiency virus gp120 to viral infectivity and removal of multiple glycans without
967 disturbing infectivity. J Virol. American Society for Microbiology (ASM); 1998;72: 8365–8370.

968 56. Helenius A, Aebi M. Intracellular functions of N-linked glycans. Science. 2001;291: 2364–2369.
969 doi:10.1126/science.291.5512.2364
970 57. Solá RJ, Rodríguez-Martínez JA, Griebenow K. Modulation of protein biophysical properties by chemical
971 glycosylation: biochemical insights and biomedical implications. Cell Mol Life Sci. Birkhäuser-Verlag; 2007;64:
972 2133–2152. doi:10.1007/s00018-007-6551-y
973 58. Shental-Bechor D, Levy Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization.
974 Proc Natl Acad Sci USA. National Academy of Sciences; 2008;105: 8256–8261. doi:10.1073/pnas.0801340105
975 59. Imperiali B, O'Connor SE. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. Curr Opin
976 Chem Biol. 1999;3: 643–649. doi:10.1016/s1367-5931(99)00021-6
977 60. Tian J, López CA, Derdeyn CA, Jones MS, Pinter A, Korber B, et al. Effect of Glycosylation on an Immunodominant
978 Region in the V1V2 Variable Domain of the HIV-1 Envelope gp120 Protein. Rein A, editor. PLoS Comput Biol.
979 2016;12: e1005094. doi:10.1371/journal.pcbi.1005094
980 61. Wood NT, Fadda E, Davis R, Grant OC, Martin JC, Woods RJ, et al. The influence of N-linked glycans on the
981 molecular dynamics of the HIV-1 gp120 V3 loop. Wu Y, editor. PLoS ONE. 2013;8: e80301.
982 doi:10.1371/journal.pone.0080301
983 62. Gerken TA, Butenhof KJ, Shogren R. Effects of glycosylation on the conformation and dynamics of O-linked
984 glycoproteins: carbon-13 NMR studies of ovine submaxillary mucin. Biochemistry. American Chemical Society;
985 1989;28: 5536–5543. doi:10.1021/bi00439a030
986 63. Mer G, Hietter H, Lefèvre JF. Stabilization of proteins by glycosylation examined by NMR analysis of a fucosylated
987 proteinase inhibitor. Nat Struct Biol. Nature Publishing Group; 1996;3: 45–53. doi:10.1038/nsb0196-45
988 64. Joao HC, Dwek RA. Effects of glycosylation on protein structure and dynamics in ribonuclease B and some of its
989 individual glycoforms. Eur J Biochem. John Wiley & Sons, Ltd (10.1111); 1993;218: 239–244. doi:10.1111/j.1432-
990 1033.1993.tb18370.x
991 65. Lu D, Yang C, Liu Z. How hydrophobicity and the glycosylation site of glycans affect protein folding and stability: a
992 molecular dynamics simulation. J Phys Chem B. American Chemical Society; 2012;116: 390–400.
993 doi:10.1021/jp203926r
994 66. Shental-Bechor D, Levy Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation
995 code. Curr Opin Struct Biol. 2009;19: 524–533. doi:10.1016/j.sbi.2009.07.002
996 67. Hanson SR, Culyba EK, Hsu T-L, Wong C-H, Kelly JW, Powers ET. The core trisaccharide of an N-linked glycoprotein
997 intrinsically accelerates folding and enhances stability. Proc Natl Acad Sci USA. National Academy of Sciences;
998 2009;106: 3131–3136. doi:10.1073/pnas.0810318105
999 68. Ozorowski G, Pallesen J, de Val N, Lyumkis D, Cottrell CA, Torres JL, et al. Open and closed structures reveal
1000 allostery and pliability in the HIV-1 envelope spike. Nature. Nature Research; 2017;15: 690–363.
1001 doi:10.1038/nature23010

1002 69. Sanders RW, van Gils MJ, Derking R, Sok D, Ketas TJ, Burger JA, et al. HIV-1 VACCINES. HIV-1 neutralizing antibodies
1003 induced by native-like envelope trimers. Science. American Association for the Advancement of Science;
1004 2015;349: aac4223–aac4223. doi:10.1126/science.aac4223
1005 70. Sok D, van Gils MJ, Pauthner M, Julien J-P, Saye-Francisco KL, Hsueh J, et al. Recombinant HIV envelope trimer
1006 selects for quaternary-dependent antibodies targeting the trimer apex. Proc Natl Acad Sci USA. National Acad
1007 Sciences; 2014;111: 17624–17629. doi:10.1073/pnas.1415789111
1008 71. Steichen JM, Lin Y-C, Havenar-Daughton C, Pecetta S, Ozorowski G, Willis JR, et al. A generalized HIV vaccine
1009 design strategy for priming of broadly neutralizing antibody responses. Science. American Association for the
1010 Advancement of Science; 2019;: eaax4380. doi:10.1126/science.aax4380
1011 72. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-
1012 induced motion for improved cryo-electron microscopy. Nat Methods. Nature Publishing Group; 2017;14: 331–
1013 332. doi:10.1038/nmeth.4193
1014 73. Zhang K. Gctf: Real-time CTF determination and correction. J Struct Biol. 2016;193: 1–12.
1015 doi:10.1016/j.jsb.2015.11.003
1016 74. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, et al. New tools for automated high-resolution
1017 cryo-EM structure determination in RELION-3. Elife. eLife Sciences Publications Limited; 2018;7: 163.
1018 doi:10.7554/eLife.42166
1019 75. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure
1020 determination. Nat Methods. 2017;14: 290–296. doi:10.1038/nmeth.4169
1021 76. Berndsen Z, Bowman C, Jang H, Ward AB. EMHP: An accurate automated hole masking algorithm for single-
1022 particle cryo-EM image processing. Bioinformatics. 2017;33: 3824–3826. doi:10.1093/bioinformatics/btx500
1023 77. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC. Quantitative analysis of cryo-EM density map segmentation
1024 by watershed and scale-space filtering, and fitting of structures by alignment to regions. J Struct Biol. 2010;170:
1025 427–438. doi:10.1016/j.jsb.2010.03.007
1026 78. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization
1027 system for exploratory research and analysis. J Comput Chem. 2004;25: 1605–1612. doi:10.1002/jcc.20084
1028 79. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling
1029 of protein structures and complexes. Nucleic Acids Res. 2018;46: W296–W303. doi:10.1093/nar/gky427
1030 80. Wang RY-R, Song Y, Barad BA, Cheng Y, Fraser JS, DiMaio F. Automated structure refinement of macromolecular
1031 assemblies from cryo-EM maps using Rosetta. Elife. eLife Sciences Publications Limited; 2016;5: 352.
1032 doi:10.7554/eLife.17219
1033 81. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure
1034 validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. International Union of
1035 Crystallography; 2010;66: 12–21. doi:10.1107/S0907444909042073

1036 82. Barad BA, Echols N, Wang RY-R, Cheng Y, DiMaio F, Adams PD, et al. EMRinger: side chain-directed model and
1037 map validation for 3D cryo-electron microscopy. Nat Methods. Nature Research; 2015;12: 943–946.
1038 doi:10.1038/nmeth.3541
1039 83. Emsley P, Crispin M. Structural analysis of glycoproteins: building N-linked glycans with Coot. Acta Crystallogr D
1040 Struct Biol. International Union of Crystallography; 2018;74: 256–263. doi:10.1107/S2059798318005119
1041 84. Frenz B, Rämisch S, Borst AJ, Walls AC, Adolf-Bryfogle J, Schief WR, et al. Automatically Fixing Errors in
1042 Glycoprotein Structures with Rosetta. Structure. 2019;27: 134–139.e3. doi:10.1016/j.str.2018.09.006
1043 85. Schrodinger,, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
1044 86. Weinkam P, Pons J, Sali A. Structure-based model of allostery predicts coupling between distant sites. Proc Natl
1045 Acad Sci USA. National Academy of Sciences; 2012;109: 4875–4880. doi:10.1073/pnas.1116274109
1046 87. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein
1047 structure modeling using Modeller. Bateman A, Pearson WR, Stein LD, Stormo GD, Yates Iii JR, editors. Curr Protoc
1048 Bioinformatics. John Wiley & Sons, Ltd; 2006;Chapter 5: Unit–5.6. doi:10.1002/0471250953.bi0506s15
1049 88. Sali A. Comparative protein modeling by satisfaction of spatial restraints. Mol Med Today. 1995;1: 270–277.
1050 doi:10.1016/s1357-4310(95)91170-7
1051 89. Laskowski RA, MacArthur MW, Moss DS, Thornton JM, IUCr. PROCHECK: a program to check the stereochemical
1052 quality of protein structures. J Appl Crystallogr. International Union of Crystallography; 1993;26: 283–291.
1053 doi:10.1107/S0021889892009944
1054 90. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein
1055 force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ(1) and χ(2) dihedral angles. J Chem
1056 Theory Comput. American Chemical Society; 2012;8: 3257–3273. doi:10.1021/ct300400x
1057 91. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR
1058 data. J Comput Chem. John Wiley & Sons, Ltd; 2013;34: 2135–2145. doi:10.1002/jcc.23354
1059 92. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14: 33–8– 27–8.
1060 doi:10.1016/0263-7855(96)00018-5
1061 93. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, et al. EMAN2: an extensible image processing suite for
1062 electron microscopy. J Struct Biol. 2007;157: 38–46. doi:10.1016/j.jsb.2006.05.009

Figures and Figure Legends

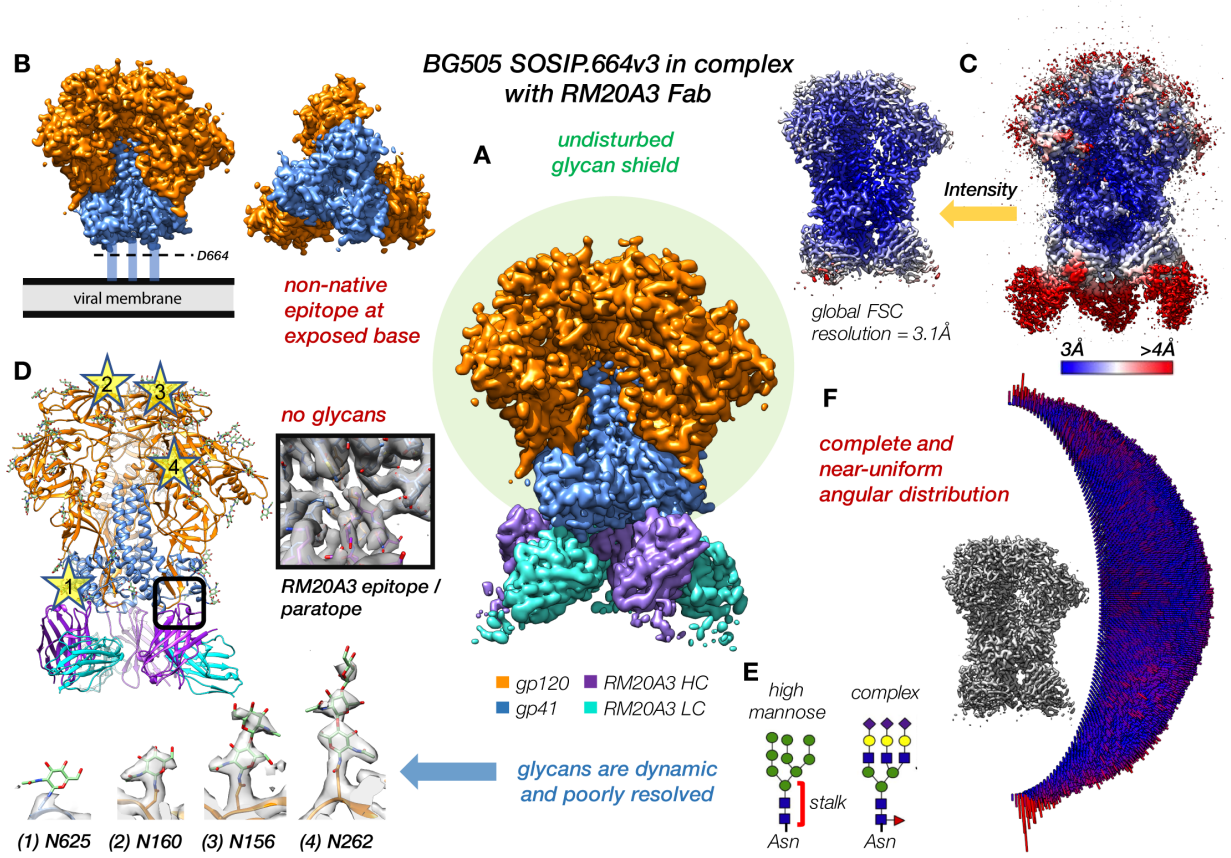

Figure 1 | Soluble SOSIP in complex with a base-specific Fab as a model system for cryo-EM analysis of the native HIV-1 Env glycan shield. (A) Gaussian filtered and segmented cryo-EM map of BG505 SOSIP.664 purified from HEK293F cells in complex with three copies RM20A3 Fab. **(B)** Segmented map with Fabs removed showing neo-epitope formed at the exposed base of the soluble trimer. **(C)** Sharpened 3.1Å-resolution map visualized at a high and low intensity threshold and colored by local resolution. **(D)** Refined atomic model with fly-out of RM20A3 epitope/paratope and stars highlighting four different N-linked glycans displayed below. **(E)** Cartoon models of the two main types of N-linked glycans on BG505 SOSIP.664. **(F)** 3-D angular distribution histogram for BG505_293F.

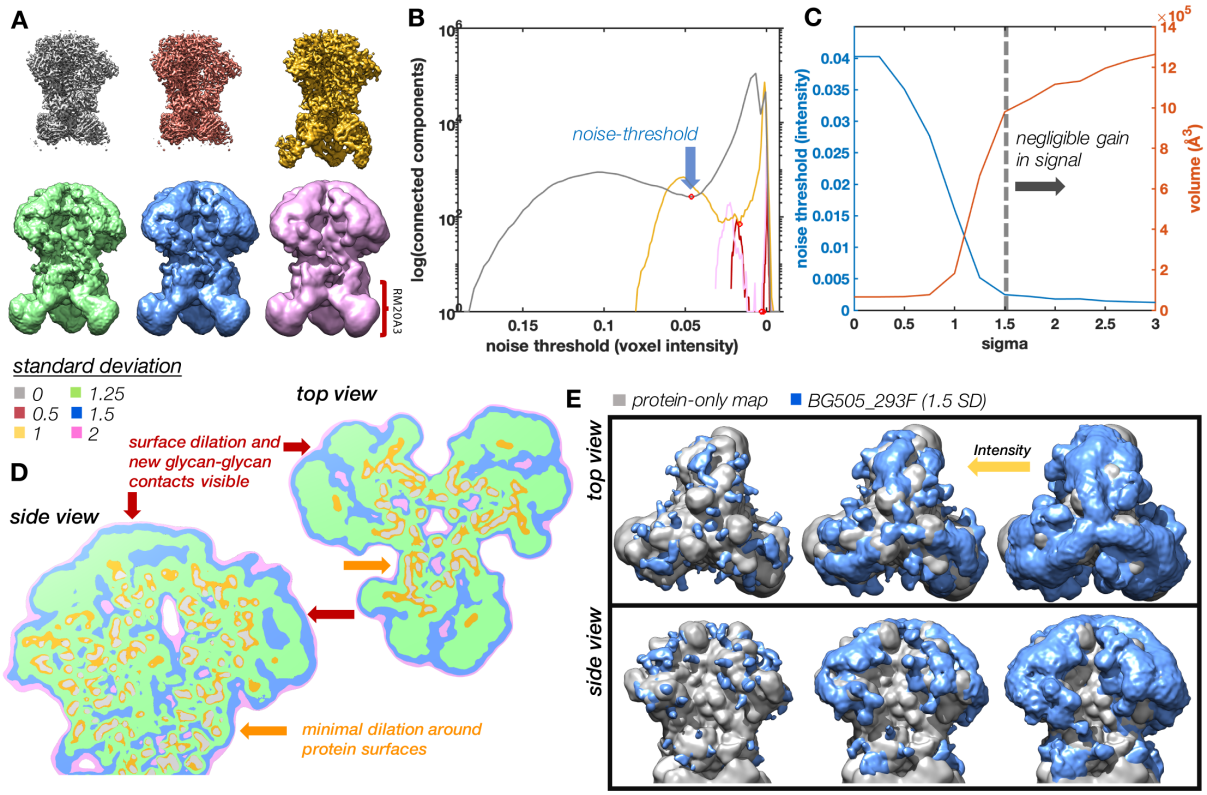

Figure 2 | Scale-space analysis of BG505_293F reveals low resolution structure in the glycan shield. (A) Scale-space representation of the cryo-EM map using a Gaussian kernel (0-2 SD). **(B)** Logarithm of the number of connected components of the map as a function of intensity threshold for a series of Gaussian filtered maps with red circles indicating the noise-thresholds. **(C)** Noise threshold and volume at the noise threshold as a function of SD. **(D)** Slices through the side and top views of the 6 maps displayed in (A) highlighting the volume expansion around glycosylated surfaces (red arrow) but not protein surfaces (yellow arrow). **(E)** Top and side views of the 1.5 SD Gaussian filtered BG505_293F map (blue) at three intensity thresholds with a simulated protein-only map underneath (gray).

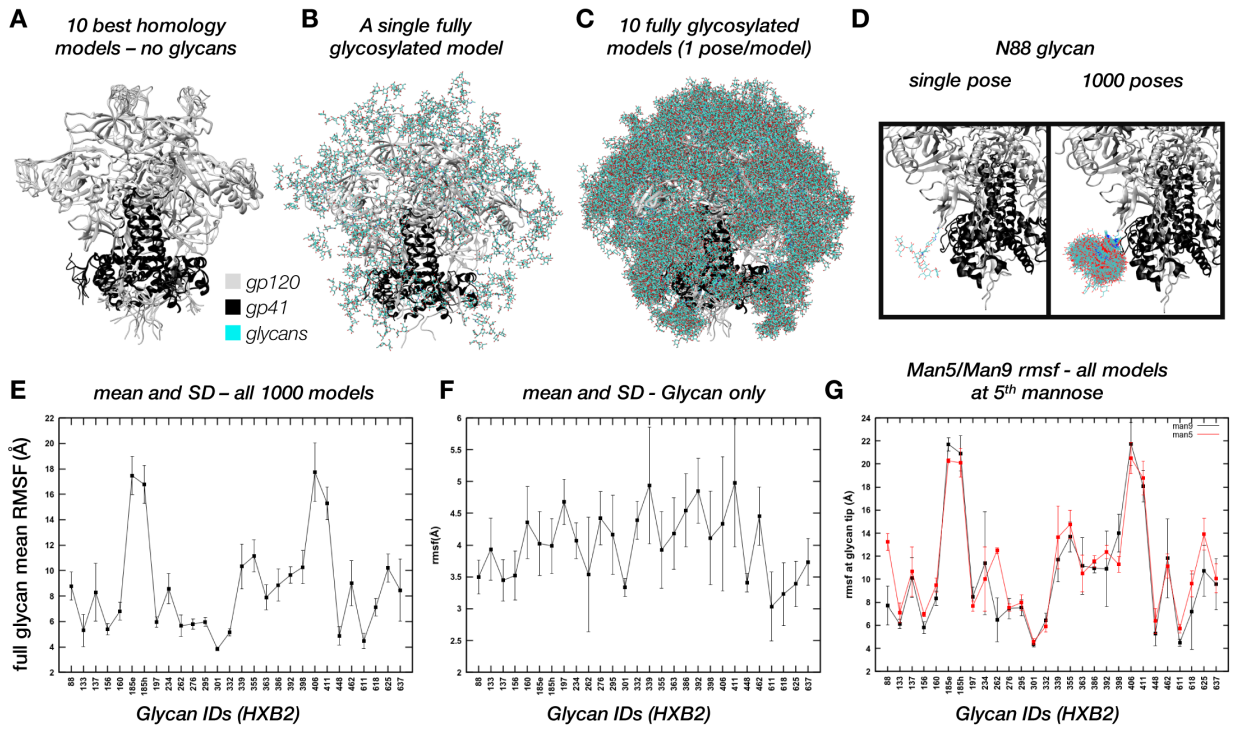

Figure 3 | ALLOSMOD-based high-throughput atomistic modeling pipeline for fast and robust sampling of fully glycosylated Env. (A) Ribbon diagram of the 10 best scoring homology models used as protein scaffolds, showing flexible modeling of the V2 and V4 loops (L). One of the ten models with a single relaxed Man9 glycan at each PNGS (M). One fully glycosylated and relaxed model for each of the ten protein scaffolds (R). **(B)** Close up of the N88 PNGS with 1 and all 1000 glycan poses. **(C)** Full glycan mean RMSF for all 1000 poses (100 glycan poses/scaffold). **(D)** Glycan only RMSF for all 1000 poses. **(E)** Full glycan mean RMSF for all 1000 poses up to the 5th mannose residue for both the Man9 and Man5 ensembles showing increased dynamics in Man5 ensemble.

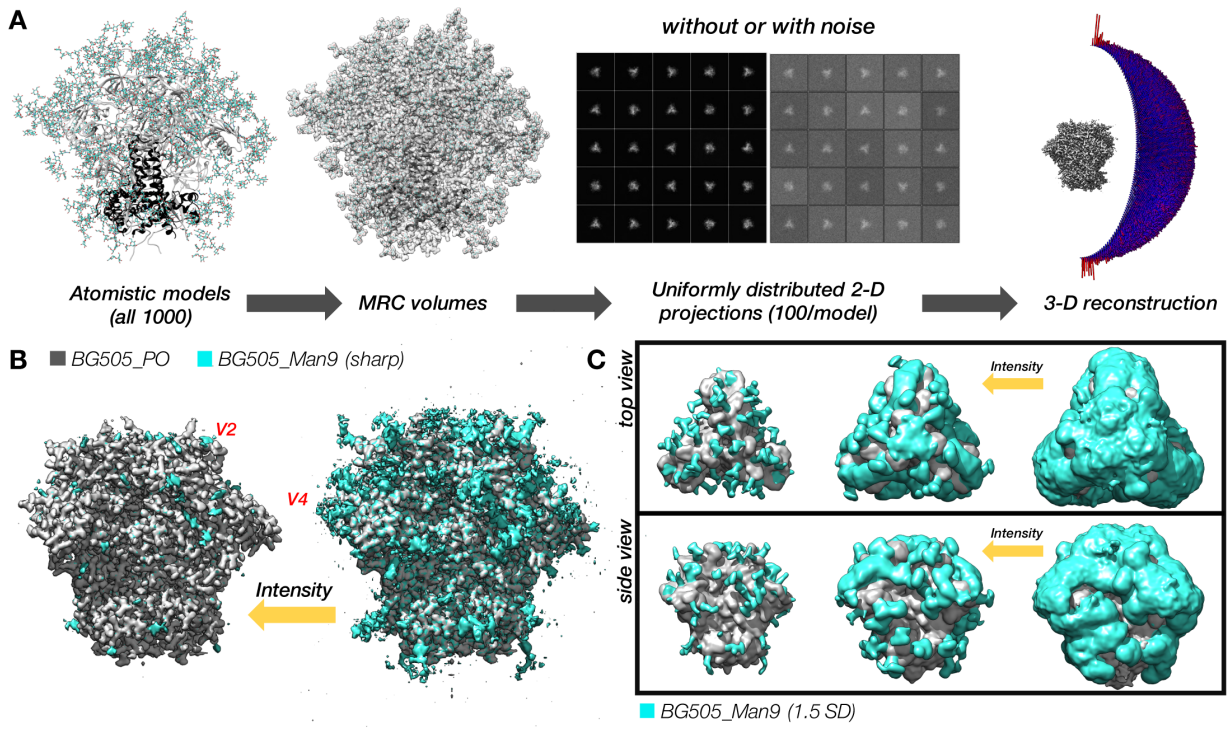

Figure 4| Synthetic cryo-EM maps from ALLOSMOD ensembles reproduce global features of the experimental data. (A) Schematic of the synthetic data creation pipeline. First, each of the 1000 models is converted into an MRC volume file then projected at 100 random and uniformly distributed angles either with or without added noise (100 SD white noise on the right). A total of 100,000 projections are then refined in RELION. **(B)** 3-D reconstruction from synthetic Man9 dataset at two different intensity thresholds (cyan) and a protein-only map displayed for contrast (grey). *Ab initio* modeled V2 and V4 loops are not visible at these resolutions. **(C)** 1.5 SD Gaussian filtered BG505_Man9 and BG505_PO reconstructions viewed from the side and top at three intensity thresholds.

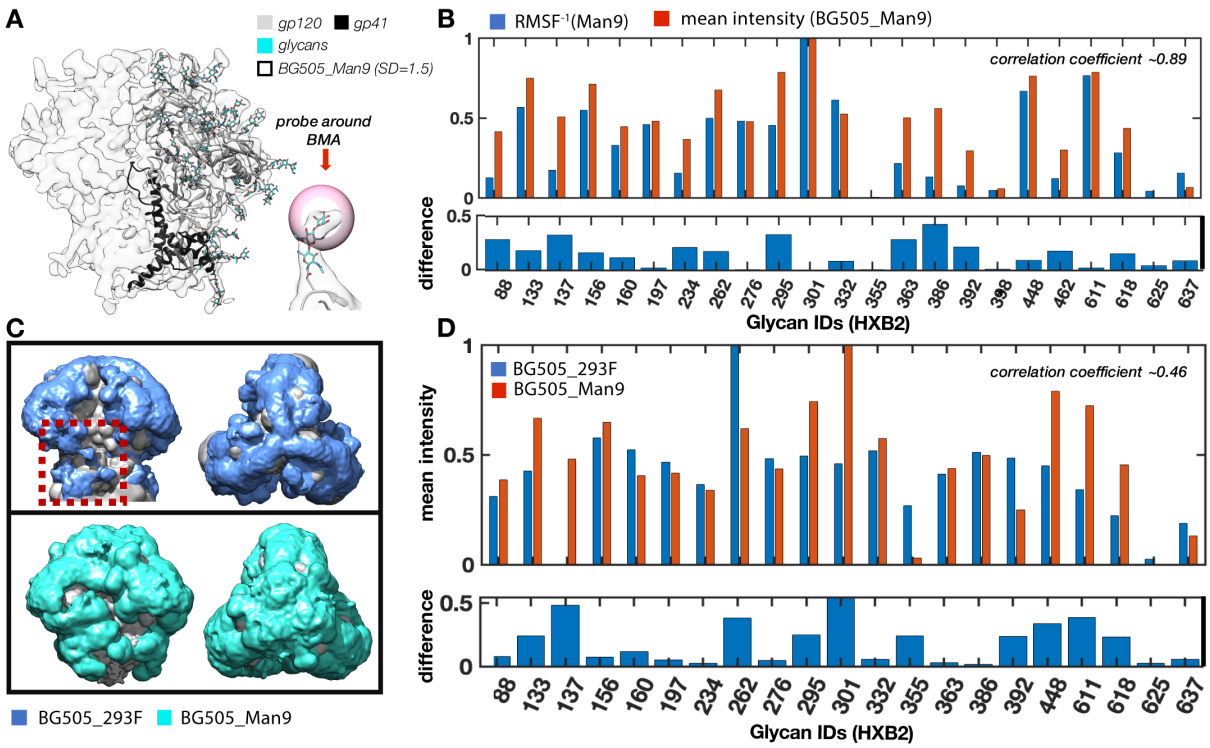

Figure 5 | Measuring individual glycan dynamics from cryo-EM maps reveals close agreement between theory and experiment. (A) 1.5 SD Gaussian filtered BG505_Man9 synthetic cryo-EM map (transparent) with the glycan stalk model used in the local intensity analysis displayed within. Also shown is a close up of a glycan stalk with a sphere around the BMA residue representing the local intensity probe. **(B)** Bar plot showing relationship between normalized mean intensity around BMA residues from the 0.5 SD Gaussian filtered BG505_Man9 reconstruction using a probe radius of a 1.725Å and the inverse normalized full glycan mean RMSF from the Man9 ensemble. Pearson correlation coefficient ~0.89 (p=8e-8). **(C)** 1.5 SD BG505_293F and BG505_Man9 maps visualized from the top and side and contoured to their noise thresholds. Red box highlighting effects of sub-occupancy in the gp41 region. **(D)** Bar plot showing relationship between the normalized site-specific mean local intensity at from the BG505_293F and BG505_Man9 maps along with the absolute difference at each site. Pearson correlation coefficient ~0.46 (p = 0.03).

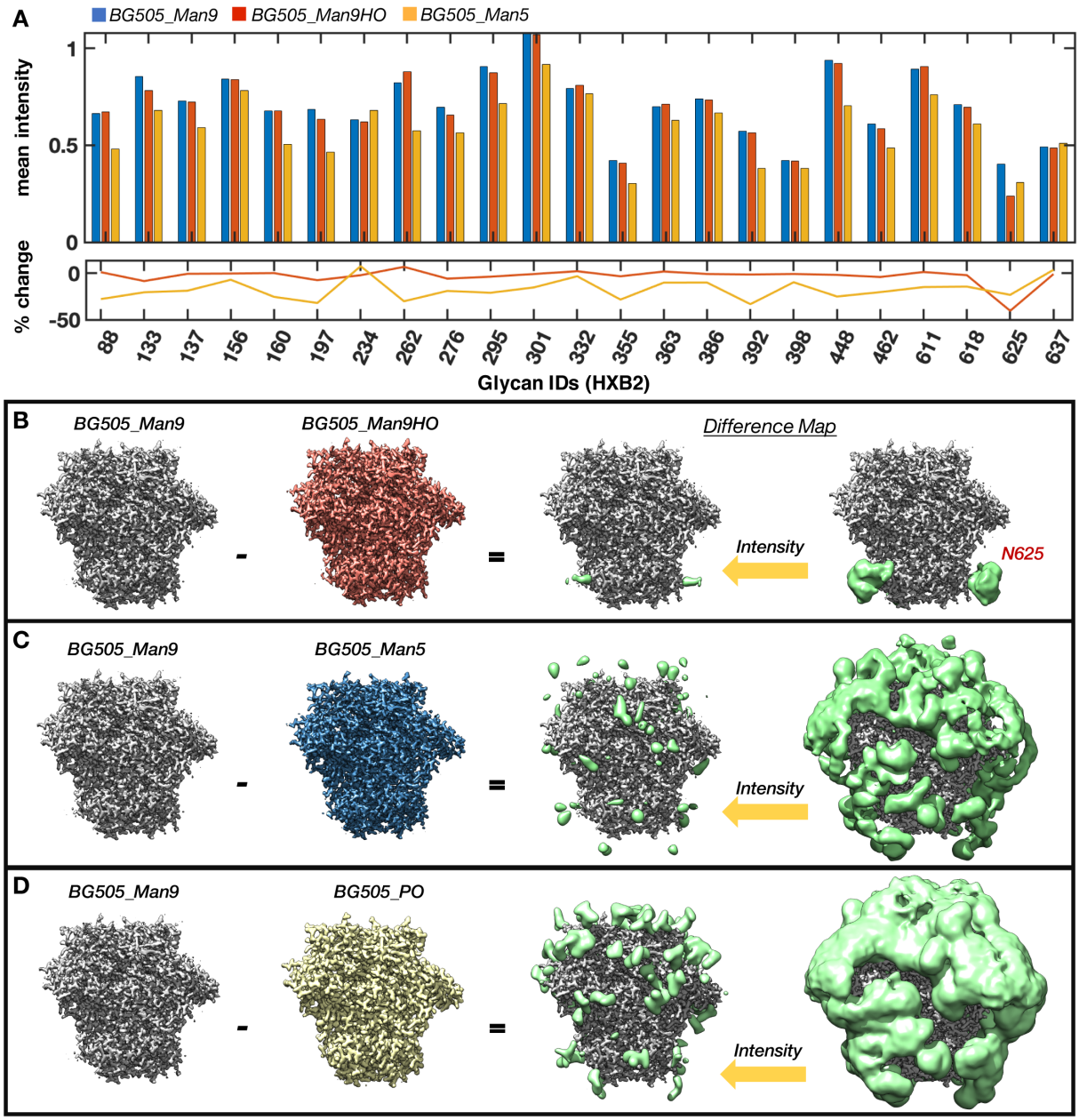

Figure 6 | Detecting changes in glycan dynamics, occupancy, and chemical composition from synthetic cryo-EM maps. (A) Un-normalized mean local intensity at each PNGS for the BG505_Man9, BG505_Man5, and BG505_ManHO maps along with the percent change from the BG505_Man9 values. **(B)** BG505_Man9 - BG505_Man9HO, **(C)** BG505_Man9 - BG505_Man5, and **(D)** BG505_Man9 - BG505_PO Gaussian filtered (SD=2) difference maps at two intensity thresholds (R - green) along with high-resolution sharpened maps (L).

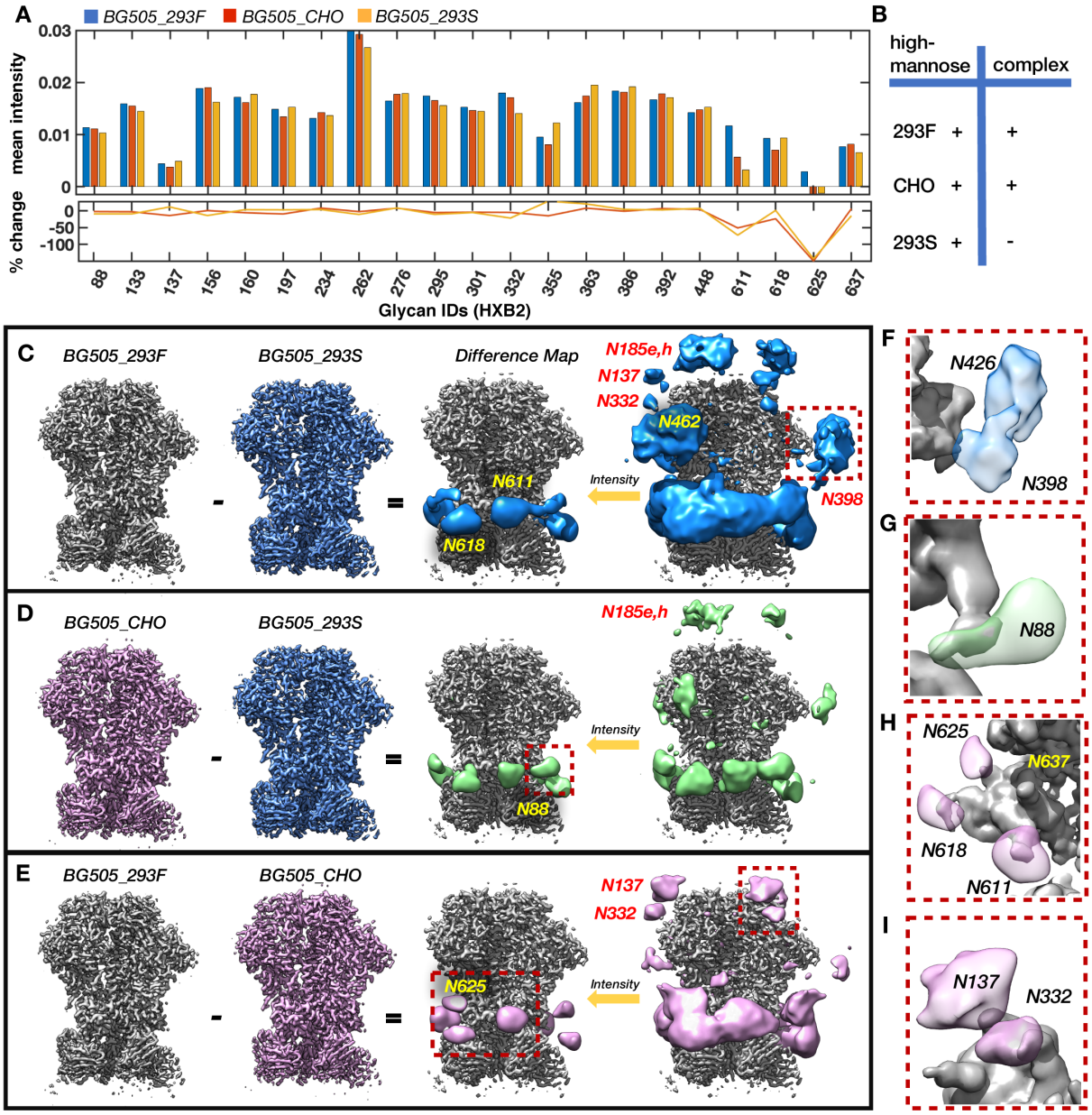

Figure 7 | Detection of cell-type specific changes in glycan shield composition and dynamics with cryo-EM. (A) Unnormalized mean local intensity at each PNGS for the BG505_293F, BG505_293S, and BG505_CHO cryo-EM maps along with percent change from the BG505_293F values. **(B)** Truth table illustrating glycosylation content of each sample. **(C)** BG505_293F – BG505_293S, **(D)** BG505_CHO – BG505_293S, and **(E)** BG505_293F - BG505_CHO difference maps at two intensity thresholds. All difference maps were multiplied by a soft mask around the RM20A3 Fabs prior to Gaussian filtering. **(F-I)** Magnifications of the red boxed areas of the difference maps.

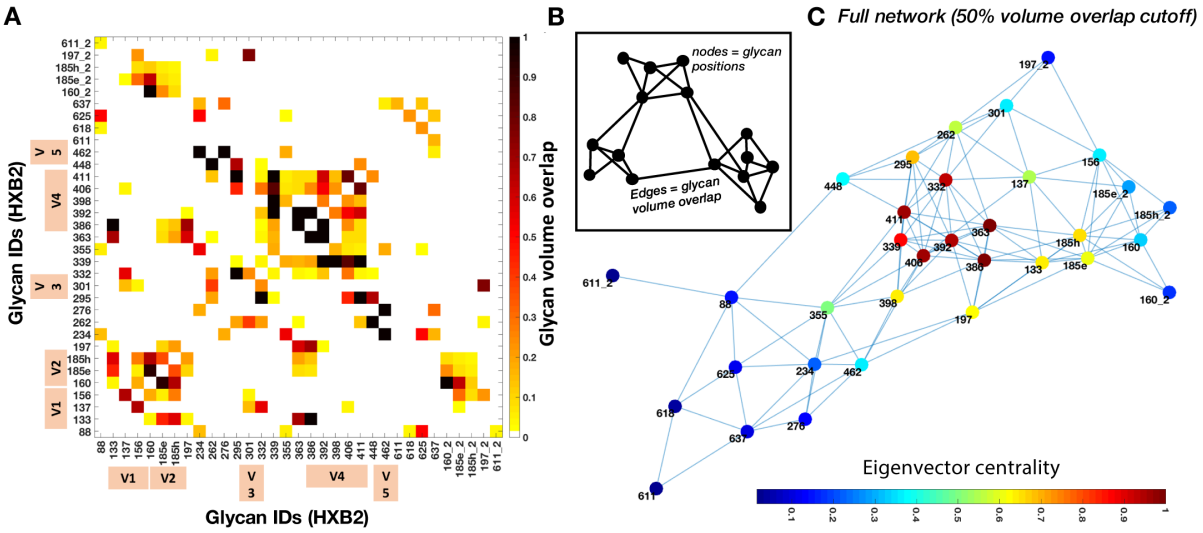

Figure 8 | Probabilistic glycan-glycan interaction network for the Man9 ensemble. (A) Glycan-glycan volume overlap matrix. Inter-protomer overlap is given by suffix "_2". The overlap fraction is normalized with at least 50% overlap being designated as 1. The HXB2 numberings of the glycosylated asparagines are given in the X and Y axes. The glycans in the variable loop regions are indicated by tan color bars. **(B)** Cartoon model of the glycan-glycan interaction network generated by interpreting the overlap matrix as an adjacency matrix (inset). The edge length is drawn inversely proportional to the overlap value, i.e., the larger the overlap, the closer two nodes are in the graph. **(C)** Glycan-glycan interaction network calculated from the matrix in (A). Edges represent overlap between two glycans represented by nodes. All nodes are colored by eigenvector centrality. The centrality values are normalized with the highest value assigned as 1. Higher the centrality values mean greater its influence over the network.

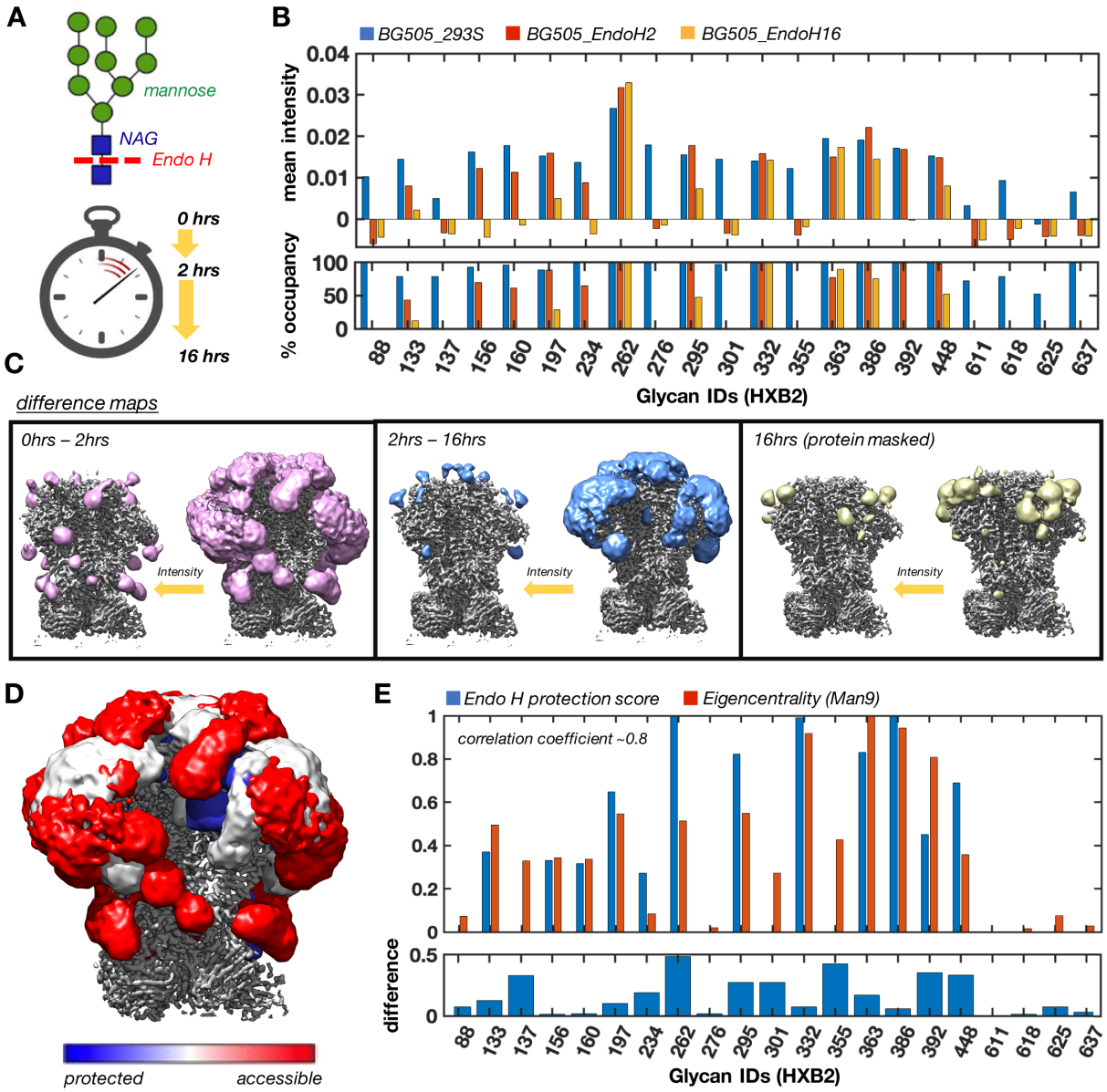

Figure 9 | Highly connected glycan clusters are resistant to enzymatic digestion. (A) Cartoon schematic of Endo H digestion of high-mannose glycans and illustration of the three reaction times captured for cryo-EM analysis. **(B)** Unnormalized mean local intensity at each PNGS for the BG505_293S, BG505_EndoH2, and BG505_EndoH16 cryo-EM reconstructions along with the percent occupancy at each site after 2 and 16 hrs of digestion assuming 100% initial occupancy. **(C)** BG505_293S – BG505_EndoH2 and BG505_EndoH2 – BG505_EndoH16 3 SD Gaussian filtered difference maps at two intensity thresholds. Also shown is the BG505_EndoH16 cryo-EM after multiplication by a soft protein-only mask. **(D)** Overlay of three maps from (C) colored according to susceptibility to Endo H digestion with red being the most and blue being the least. **(E)** Bar plot showing the normalized Endo H protection score (see Methods) and normalized eigencentrality from the Man9 ensemble. Pearson correlation coefficient between the two ~0.8 (p=1.14e-05).

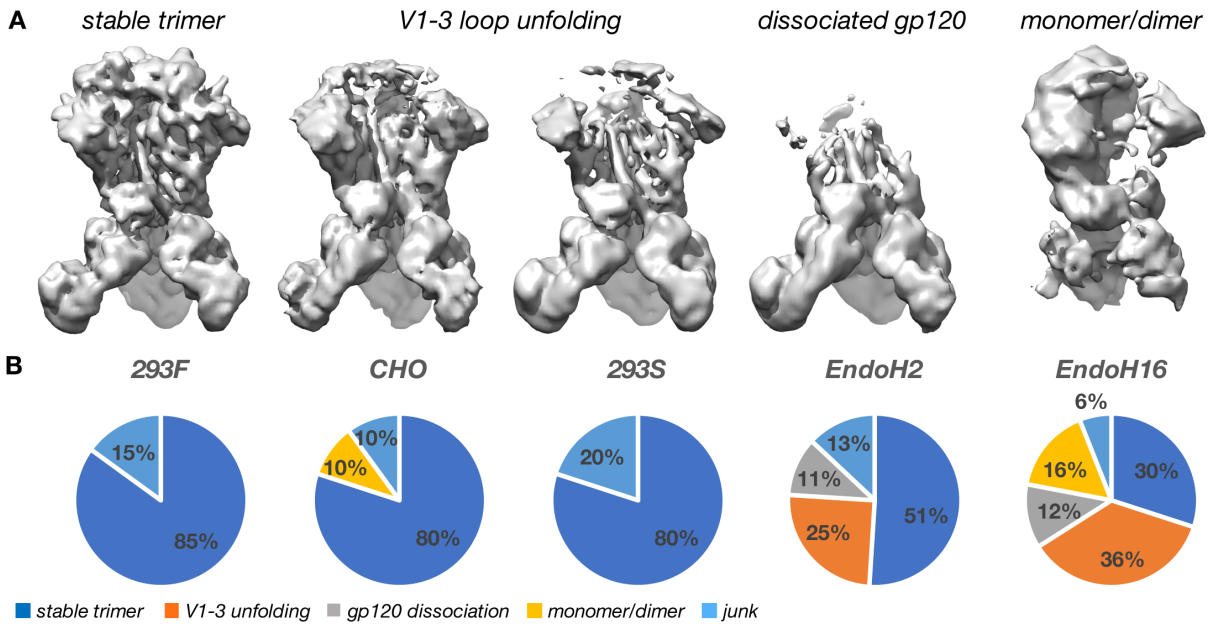

Figure 10 | 3-D classification of cryo-EM data reveal enzymatic digestion of the glycan shield leads to progressive destabilization of the pre-fusion trimer. (A) Representative cryo-EM maps of the four distinct states captured by 3-D classification. **(B)** Pie charts for each of the 5 experimental BG505 SOSIP.664 datasets showing the relative percentages of each class shown in (A) and a junk class.