STRUCTURAL BIOLOGY

ISSN 2059-7983

X-ray and EM structures of a natively glycosylated HIV-1 envelope trimer

Harry B. Gristick, Haoqing Wang and Pamela J. Bjorkman*

Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA.
*Correspondence e-mail: bjorkman@caltech.edu

Received 4 August 2017
Accepted 19 September 2017

Edited by J. L. Martin, Griffith University, Australia

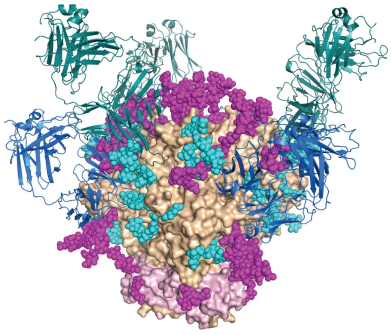
Keywords: HIV-1 envelope; N-linked glycans; X-ray crystallography; single-particle cryo-EM.

The structural and biochemical characterization of broadly neutralizing anti-HIV-1 antibodies (bNAbs) has been essential in guiding the design of potential vaccines to prevent infection by HIV-1. While these studies have revealed critical mechanisms by which bNAbs recognize and/or accommodate N-glycans on the trimeric envelope glycoprotein (Env), they have been limited to the visualization of high-mannose glycan forms only, since heterogeneity introduced from the presence of complex glycans makes it difficult to obtain high-resolution structures. 3.5 and 3.9 Å resolution crystal structures of the HIV-1 Env trimer with fully processed and native glycosylation were solved, revealing a glycan shield of high-mannose and complex-type N-glycans that were used to define the complete epitopes of two bNAbs. Here, the refinement of the N-glycans in the crystal structures is discussed and comparisons are made with glycan densities in glycosylated Env structures derived by single-particle cryo-electron microscopy.

1. Introduction

The trimeric HIV-1 envelope glycoprotein (Env), the only target of neutralizing antibodies, is among the most heavily glycosylated proteins ever characterized (Lasky *et al.*, 1986). It includes glycans constituting up to 50% of its mass attached to 30 ± 3 potential N-linked glycosylation sites (PNGSs) per gp120–gp41 protomer. PNGSs can be easily identified in protein sequences as a three-residue sequon: Asn-*X*-Ser/Thr. During transit through the endoplasmic reticulum, high-mannose forms of N-linked glycans are attached to the Asn; the high-mannose glycans are usually modified to complex-type N-glycans during subsequent trafficking though the Golgi apparatus (Fig. 1). Because of steric constraints that limit the activities of endoplasmic reticulum and Golgi carbohydrate-processing enzymes, the HIV-1 Env glycoprotein includes regions of under-processed N-glycans in oligomannose forms ($Man_{5–9}GlcNAc_2$), especially in the intrinsic mannose patch on gp120, which forms portions of the epitopes for many characterized HIV-1 broadly neutralizing antibodies (bNAbs; Doores, 2015). Although oligomannose glycans dominate parts of HIV-1 Env such as the $N332_{gp120}$ glycan-associated region on gp120, processed complex-type N-glycans predominate at N-linked glycosylation sites on gp41 and gp41-proximal regions of gp120 (Behrens *et al.*, 2016) and are thought to protect the host receptor (CD4) binding site (CD4bs) and the V3 loop of gp120 (Binley *et al.*, 2010).

Viral glycans are generally non-immunogenic because they are assembled by host cell machinery; thus, the carbohydrates decorating the surface of HIV Env constitute a 'glycan shield' that reduces access to underlying protein epitopes (Kwong *et*



al., 1998). Structural studies of bNAbs bound to HIV Env trimers revealed mechanisms by which bNAbs targeting various epitopes penetrate the glycan shield to either accommodate or include N-glycans in their epitopes (Julien *et al.*, 2013; Pancera *et al.*, 2014; Scharf *et al.*, 2015; Garces *et al.*, 2015; Lee *et al.*, 2015, 2016; Stewart-Jones *et al.*, 2016). However, because heterogeneous glycosylation generally prevents the formation of well ordered crystals, all HIV Env crystal structures had been solved using glycoproteins produced in exclusively high-mannose forms (Diskin *et al.*, 2011, 2013; Kwon *et al.*, 2015; Garces *et al.*, 2015; Julien *et al.*, 2013; Kong *et al.*, 2015; Pancera *et al.*, 2014; Scharf *et al.*, 2014, 2015; Stewart-Jones *et al.*, 2016; Zhou *et al.*, 2010, 2013, 2015; Kwong *et al.*, 1998). Therefore, little was known about the structure of the native HIV-1 Env glycan shield that includes both complex-type and oligomannose N-glycans, and the natively glycosylated epitopes of important HIV-1 bNAb classes, such as $Asn332_{gp120}$ glycan/V3 loop and CD4bs bNAbs, remained

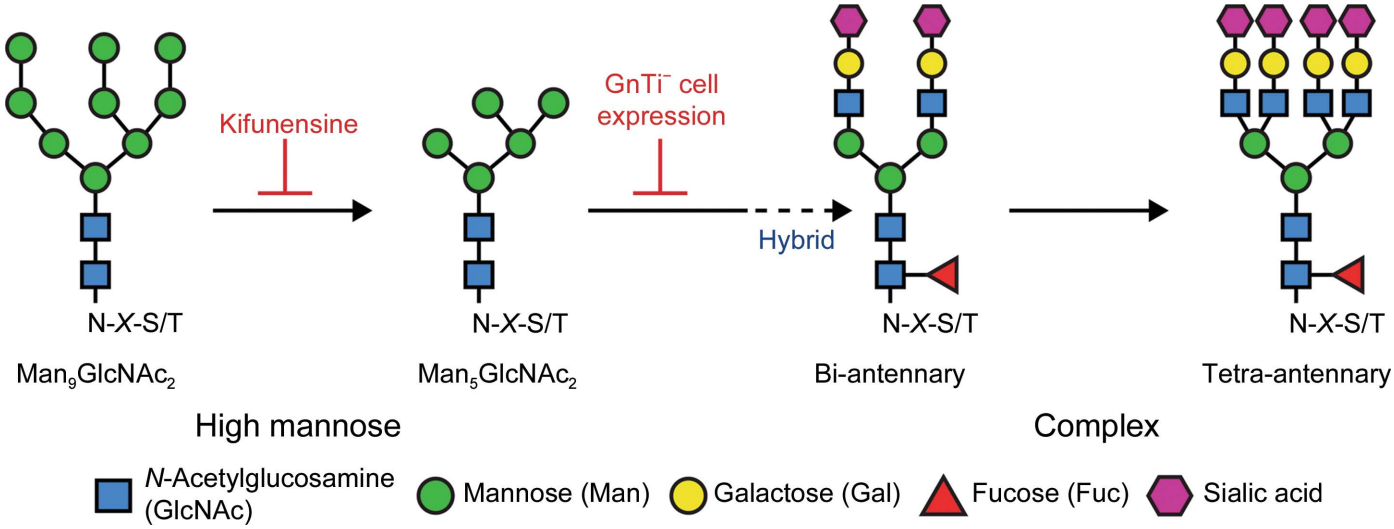


Figure 1
High-mannose N-glycans are added to Asn-*X*-Ser/Thr sequons in the endoplasmic reticulum (first two panels) and can be modified into complex-type glycans in the Golgi apparatus (third and fourth panels). Inhibitors such as kifunensine or growth in mutant mammalian cell lines (GnTI$^-$ expression) limits N-linked glycan addition to high-mannose forms. Complex-type glycans are distinguished from high-mannose glycans by the addition of a core fucose at the base and negatively charged sialic acids at the tips.

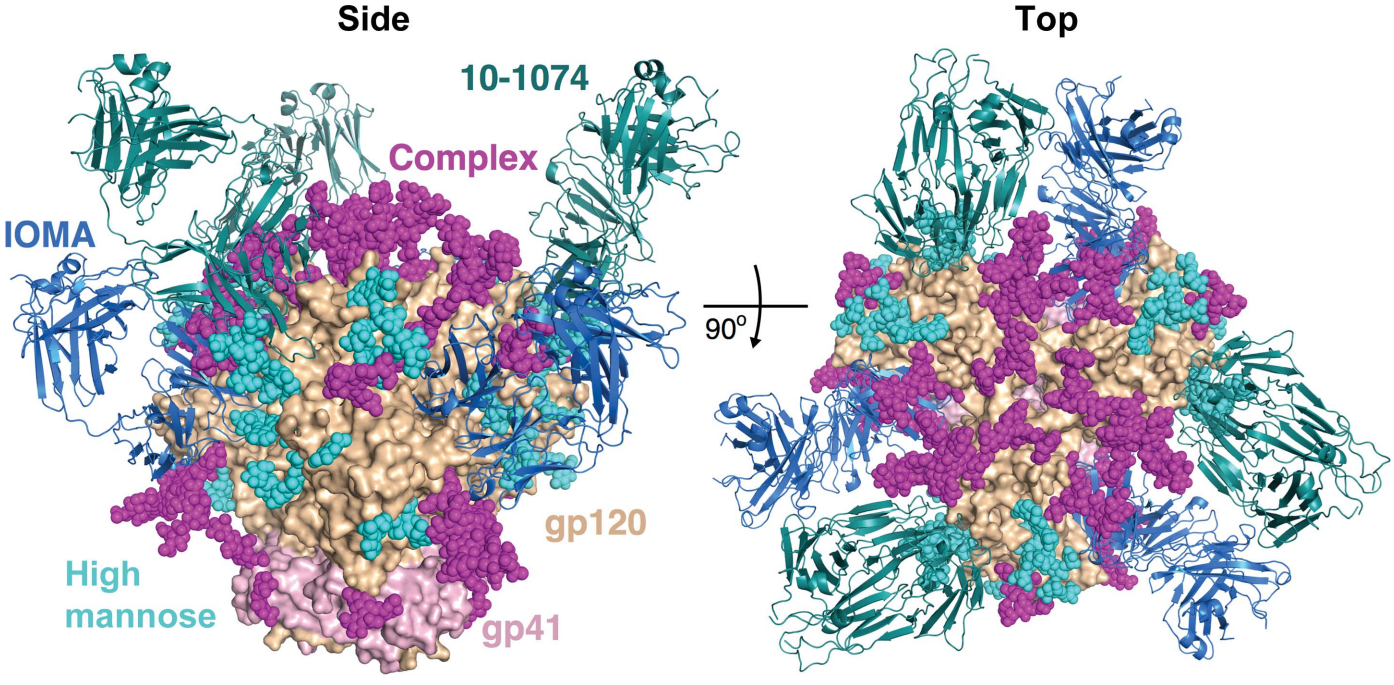


Figure 2
Side and top views of PDB entry 5t3x: 3.9 Å resolution crystal structure of natively glycosylated Env trimer (gp120, brown; gp41, pink) in complex with bNAb Fabs IOMA (blue) and 10-1074 (green). PNGSs modeled as high-mannose (cyan) or complex glycans (magenta) are shown as spheres.

incompletely characterized. We recently described crystal structures of a natively glycosylated Env trimer bound to two HIV-1 bNAbs (10-1074 and IOMA), which recognize the gp120 V3 loop and CD4bs, respectively (Gristick *et al.*, 2016). Analysis of the native glycan shield on HIV-1 Env allowed the first full description of the interplay between heterogeneous untrimmed high-mannose and complex-type N-glycans within the CD4bs, V3-loop and other epitopes on Env, revealing antibody-vulnerable glycan holes and roles of complex-type N-glycans on Env that are relevant to vaccine design. Here, we describe the crystallographic refinement of N-glycans in these structures in more detail and compare them with lower resolution cryo-EM structures of natively glycosylated HIV-1 Env.

2. Crystal structures of natively glycosylated Env trimer–Fab complexes

Apart from two exceptions (Scharf *et al.*, 2015; Stewart-Jones *et al.*, 2016), previous HIV Env trimer crystal structures used proteins produced in cells that attached only high-mannose-type N-glycans (Julien *et al.*, 2013; Pancera *et al.*, 2014; Scharf *et al.*, 2015; Garces *et al.*, 2015; Stewart-Jones *et al.*, 2016) that were then further enzymatically trimmed to reduce the glycans to single *N*-acetylglucosamines (GlcNAcs) at accessible PNGSs. Our crystals were obtained from a natively glycosylated Env trimer (BG505 SOSIP.664; Sanders *et al.*, 2013) that was prepared from human embryonic kidney cells (HEK 293 6E cells) that attached both complex-type and high-mannose N-glycans. The Env trimer was complexed with Fabs from the CD4bs bNAb IOMA and from 10-1074, a V3 loop/N332$_{gp120}$ glycan-directed bNAb (Mouquet *et al.*, 2012). We solved independent IOMA–10-1074–BG505 complex structures at resolutions of 3.5 Å (PDB entry 5t3z) and 3.9 Å (PDB entry 5t3x) using BG505 protein prepared from different size-exclusion chromatography fractions (Gristick *et al.*, 2016). These structures revealed an Env trimer bound to three 10-1074 and three IOMA Fabs (Fig. 2). 19 N-glycans (one GlcNAc up to complex-type tetra-antennary) were visible per gp120–gp41 protomer, forming glycan architectures extending ∼30 Å from the trimer surface.

3. Glycan interpretation and refinement

The overall glycan geometry was validated using programs including *PDB CArbohydrate REsidue check* (*pdb-care*; http://www.glycosciences.de/tools/pdb-care/), *CArbohydrate Ramachandran Plot* (*carp*; http://www.glycosciences.de/tools/carp/) and *Privateer* (Agirre *et al.*, 2015; Agirre, 2017). Glycans were built into the initial models for the 3.9 and 3.5 Å resolution IOMA–10-1074–BG505 crystal structures using $2F_o - F_c$ maps calculated with model phases and using composite-annealed OMIT maps calculated with phases from which the model was omitted (Adams *et al.*, 2010). Well characterized glycans known to exist in a homogeneous population (*e.g.* N332$_{gp120}$) were modeled using structural information from published BG505 trimer structures. Glycans at these positions were examined for correct geometry and then merged into the appropriate position in our models. Once imported, these glycans were then manually fitted into our density in *Coot* (Emsley *et al.*, 2010).

For glycans lacking previous structural information, we performed *de novo* modeling into the initial electron-density maps using the carbohydrate-building module in *Coot*. Owing to a lack of heavily glycosylated protein structures in the PDB, the majority of glycosylation sites were modeled in this way. Modeling was performed one glycosylation site at a time followed by refinement in *PHENIX* (Adams *et al.*, 2010) using torsion restraints to prevent the glycans from puckering and adopting an incorrect conformation. This was carried out until glycans were modeled into all of the visible density within the glycosylation sites.

Following multiple rounds of building and refinement, the glycan geometry was examined using *Privateer* within the *CCP*4 suite (Agirre *et al.*, 2015). As demonstrated, even with torsion restraints present during refinement in *PHENIX*, our fully glycosylated initial model contained numerous glycans

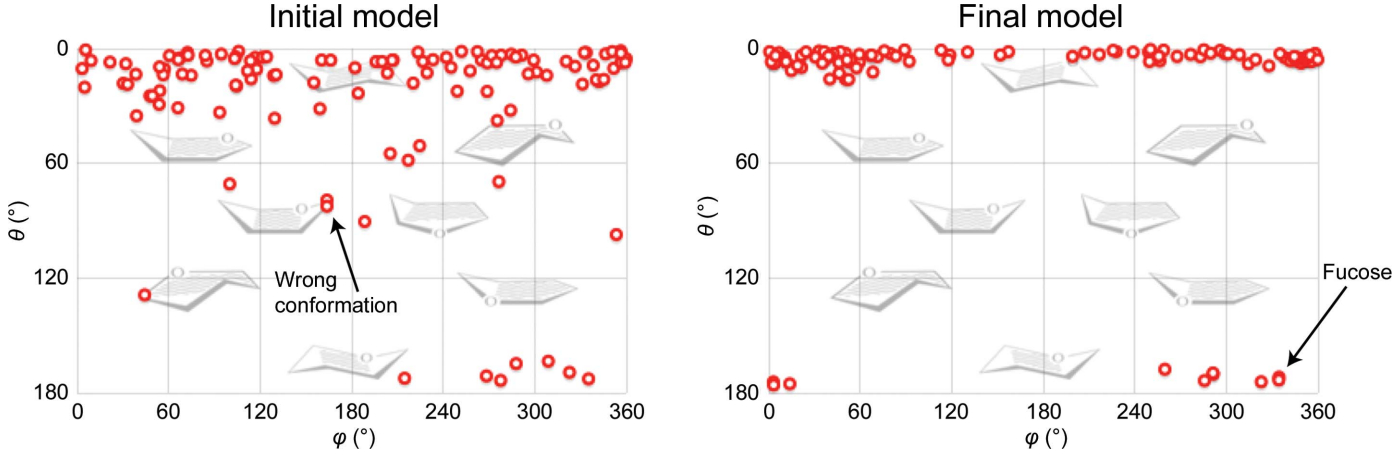


Figure 3
Cremer–Pople analysis of the initial model (left) and the final model (PDB entry 5t3x) after iterative building and refinement (right). Note that a larger fraction of the glycans in the initial model were in high-energy conformations with significant puckering, which shifted to the more common lower energy conformations following refinement.

with incorrect geometries and conformations (Fig. 3, left panel). Analysis with *Privateer* generated a new cif file containing updated restraints for each glycan type present in our structure. This cif file proved to be better at maintaining the correct glycan geometries and conformations when used in place of the default *PHENIX* restraints during subsequent refinements, which along with iterative manual modeling corrected the glycan geometries as seen in the final deposited model, PDB entry 5t3x (Fig. 3, right panel).

The initial unbiased maps, calculated using a model lacking glycans, displayed clear density for 19 glycans. However, the nature and extent of this density varied greatly between each PNGS. PNGSs containing heterogenous populations of glycans, such as N156$_{gp120}$ and N276$_{gp120}$, contained weaker density present in the initial electron-density maps that became more apparent after subsequent building and refinement (Fig. 4, first and second rows). In contrast, PNGSs containing homogenous populations of glycans, such as N332$_{gp120}$ and N386$_{gp120}$, had clear density that remained unchanged during refinement (Fig. 4, third and fourth rows). Therefore, less defined maps at PNGSs can be indicative of the presence of complex N-glycans since these are generally heterogeneous.

Glycans could sometimes be assigned at individual PNGSs as complex-type or high mannose by the presence of a core fucose ring that is found in complex-type, but not in high-mannose, N-glycans. The 3.9 Å resolution structure generally showed more density for individual BG505 N-glycans than the

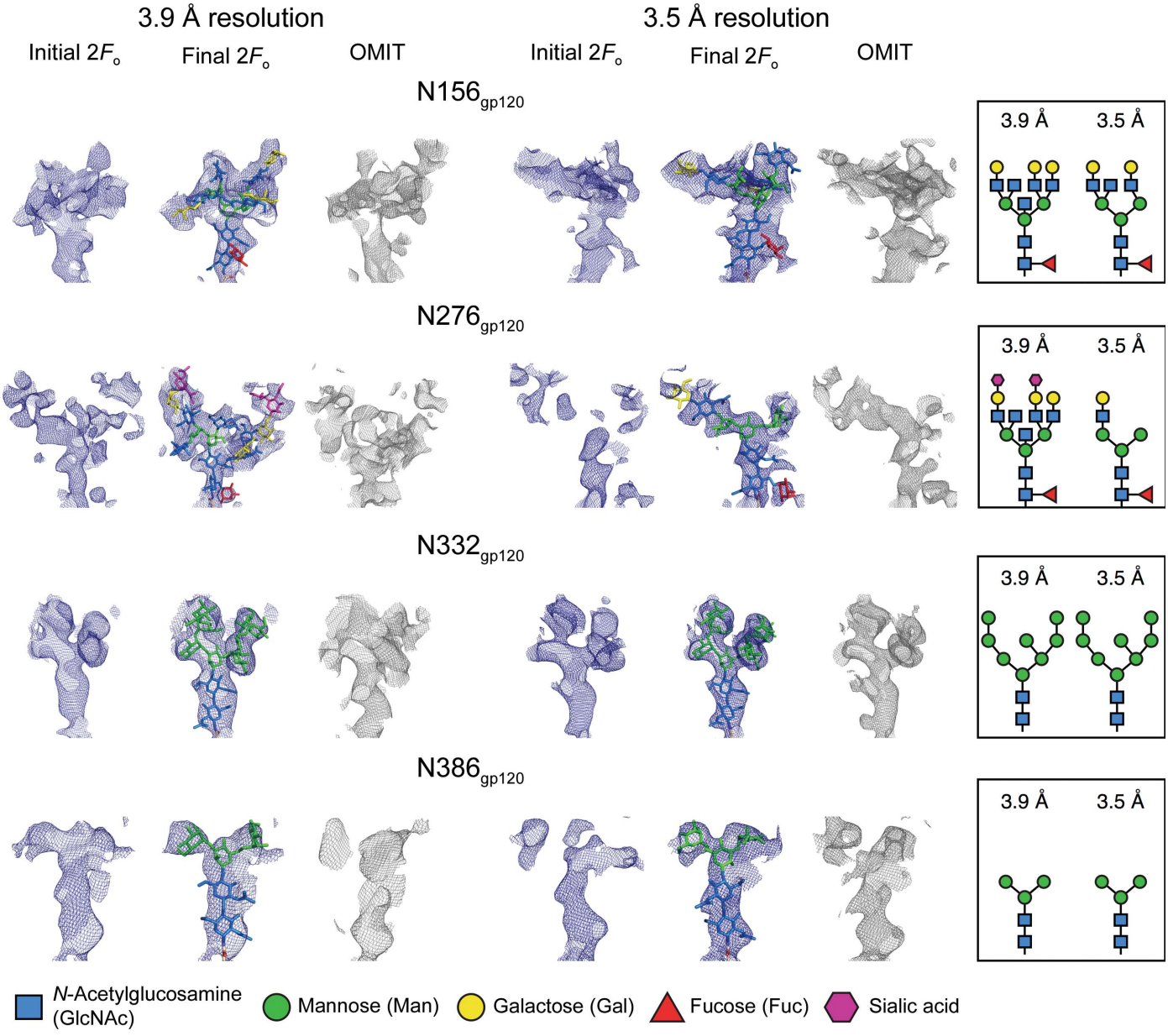


Figure 4
Comparison of the initial 2F_o (left), final 2F_o (center) and OMIT (right) maps for the 3.9 and 3.5 Å resolution structures at PNGSs N156$_{gp120}$ (first row), N276$_{gp120}$ (second row), N332$_{gp120}$ (third row) and N386$_{gp120}$ (fourth row). All maps are contoured at 0.8σ.

3.5 Å resolution structure, consistent with the presence of larger, more branching complex N-glycan architectures in the 3.9 Å resolution structure. Glycans were assigned as complex-type if there was density for a core fucose and/or based on mass-spectrometry assignments (Behrens *et al.*, 2016; Go *et al.*, 2011). A core fucose was sometimes visible in one structure but not in the other. Glycans at some individual PNGSs were therefore interpreted with different compositions in the two structures; this type of heterogeneity is consistent with multiple glycoforms at single PNGSs in preparations of the BG505 SOSIP.664 protein (Behrens *et al.*, 2016).

Although the composite-annealed OMIT maps displayed additional density near PNGSs compared with the maps calculated with model phases, some of the extra densities were not interpretable. Therefore, glycan residues were not built into these regions so as to avoid overfitting owing to a lack of supporting information (for example mass spectrometry). In contrast to the presence of these larger, branching glycans, some glycans were only partially ordered. Thus, a complex-type glycan could appear to be a small high-mannose glycan in our electron-density maps if the core fucose and residues beyond the core pentasaccharide were disordered, since this portion of an N-glycan is common to both high-mannose and complex-type N-glycans. In other cases, the interpretation of our electron-density maps was partially based on supporting experimental data: *e.g.* for the $N301_{gp120}$ glycan a core fucose was not ordered in our maps, but the $N301_{gp120}$ glycan was modeled as complex-type in our structures based on mass-spectrometric data (Go *et al.*, 2011). As expected given the large degree of glycan heterogeneity in HIV-1 Env (Doores, 2015), the glycan density was sometimes ambiguous. Consistent with the notion that uninterpretable glycan density results from sample heterogeneity rather than from refinement problems, we found unambiguous density at positions that should be homogenous, *e.g.* at $N332_{gp120}$, a high-mannose-only site (Behrens *et al.*, 2016), whereas sites predicted to be more heterogeneous, such as $N156_{gp120}$ (Behrens *et al.*, 2016; Go *et al.*, 2011), exhibited some uninterpretable heterogeneous electron density. Although the relatively low resolution of our crystal structures and heterogeneous glycosylation compounded inherent difficulties in making unambiguous glycan assignments, we built coordinates into very extensive densities (*e.g.* a complex glycan attached to

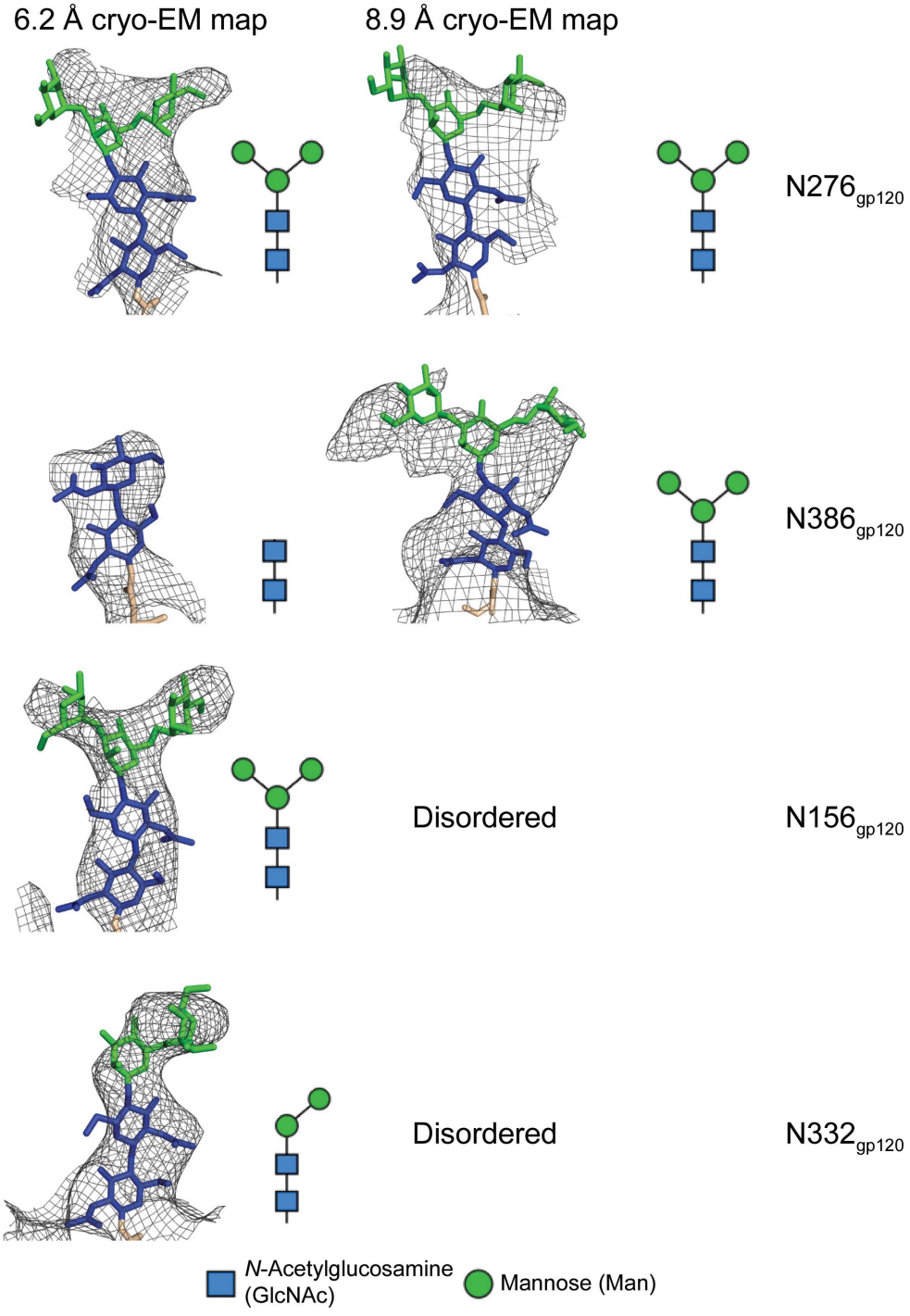


Figure 5
Comparison of the density at selected PNGSs from the 6.2 and 8.9 Å resolution cryo-EM maps.

$Asn276_{gp120}$), even if the exact structure of the glycan was uncertain, to demonstrate the extent of glycosylation at each PNGS. Additional confidence in electron-density interpretation was provided by comparing the independently refined 3.9 Å and 3.5 Å resolution IOMA–10-1074–BG505 structures. Although the refined glycan *B* factors in the final models are higher than in other published glycosylated Env structures, our structures are not only the most extensively glycosylated, but also represent the only complexes containing complex-type N-glycans. Thus, the higher *B* factors are most likely to be a reflection of the degree of glycosylation and the increased heterogeneity owing to the presence of complex-type glycans.

4. Glycans in cryo-EM maps

We recently solved single-particle cryo-EM structures of the Env trimer in order to visualize Env conformations that do not readily form well ordered crystals. One structure, reconstructed at 8.9 Å resolution, was of a complex between the Env trimer, the host receptor CD4 and Fabs from two different antibodies (Wang *et al.*, 2016). A second structure at 6.2 Å resolution was an asymmetric complex of the Env trimer bound to two, rather than three, apex-binding bNAb Fabs (Wang *et al.*, 2017). Density for ordered N-glycans at some PNGSs was visible in both maps, and coordinates for ordered N-linked glycans from our natively glycosylated Env trimer structure (Gristick *et al.*, 2016) were fitted separately as rigid bodies at PNGSs at which EM density was apparent, and geometric restraints for N-glycans were generated using *Privateer* (Agirre *et al.*, 2015). Fig. 5 compares N-linked glycan densities in the 8.9 and 6.2 Å resolution maps.

5. Conclusions

Glycoprotein crystallography has generally been limited to proteins with relatively low levels of N-linked glycans for decades. Indeed, the unwritten (but discussed at conferences) rule of thumb used to be that if your protein contained more than one PNGS per 100 amino acids, you would need to either remove selected PNGSs, use enzymatic deglycosylation, produce a high-mannose form of the protein or perform all three in order to obtain well ordered crystals. The HIV-1 Env trimer was regarded as a difficult target for crystallization, as depending on the viral strain it contains 30 ± 3 PNGSs per subunit. Thus, the first crystal structures of the HIV-1 Env trimer were derived from proteins produced in high-mannose-only forms that were usually further deglycosylated (Julien *et al.*, 2013; Pancera *et al.*, 2014; Scharf *et al.*, 2015; Garces *et al.*, 2015; Stewart-Jones *et al.*, 2016). However, we recently showed that it is possible to obtain crystal structures of fully and natively glycosylated HIV-1 Env trimers despite using trimers containing both complex-type and high-mannose glycans (Gristick *et al.*, 2016). In this case, crystallization was facilitated by adding antibody Fab fragments, which formed the crystal lattice (Gristick *et al.*, 2016), suggesting that other heavily glycosylated proteins might be crystallized without glycan trimming by adding nonglycosylated binding partners. The heavy and heterogeneous glycosylation of the Env trimer in our structures resulted in challenges for crystallographic refinement that are described in more detail here than in our original publication (Gristick *et al.*, 2016).

Unlike structure determination by X-ray crystallography, single-particle cryo-EM can be used for structures of natively glycosylated proteins, no matter how heavily glycosylated. Our cryo-EM structures of the natively glycosylated Env trimer (Wang *et al.*, 2016, 2017) were solved at low resolutions (6.2 and 8.9 Å), precluding the interpretation of glycan conformations. However, densities for N-glycans were apparent and we were able to use glycan coordinates from our higher resolution crystal structures to interpret the lower resolution density maps.

While these structures represent important steps towards furthering our understanding of how antibodies interact with complex-type glycans, advances in sample preparation, crystal handling and/or data collection are required to further increase the resolution (<3.0 Å) of structures containing complex-type glycans. A wealth of information about N-glycan geometry and interactions with other proteins awaits the solution of these technical problems.

Acknowledgements

We thank Christopher O. Barnes for careful proofreading and useful discussions while writing the manuscript. We also thank the beamline staff at Stanford Synchrotron Radiation Lightsource (SSRL) and Jens Kaiser and the Molecular Observatory at Caltech for assistance with data processing, Zhiheng Yu, Chuan Hong and Rick Huang (Janelia Farm) for assistance with cryo-EM data collection and motion correction, and Alasdair McDowall and Songye Chen for training in cryo-EM techniques and data processing.

Funding information

This research was supported by National Institutes of Health Grant 2 P50 GM082545-06 (to PJB) and National Institute Of Allergy and Infectious Diseases of the National Institutes of Health Grant HIVRAD P01 AI100148 (to PJB).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Agirre, J. (2017). *Acta Cryst.* D**73**, 171–186.
Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G. J., Wilson, K. S. & Cowtan, K. D. (2015). *Nature Struct. Mol. Biol.* **22**, 833–834.
Behrens, A. J. *et al.* (2016). *Cell. Rep.* **14**, 2695–2706.
Binley, J. M., Ban, Y.-E. A., Crooks, E. T., Eggink, D., Osawa, K., Schief, W. R. & Sanders, R. W. (2010). *J. Virol.* **84**, 5637–5655.
Diskin, R., Klein, F., Horwitz, J. A., Halper-Stromberg, A., Sather, D. N., Marcovecchio, P. M., Lee, T., West, A. P. Jr, Gao, H., Seaman, M. S., Stamatatos, L., Nussenzweig, M. C. & Bjorkman, P. J. (2013). *J. Exp. Med.* **210**, 1235–1249.
Diskin, R., Scheid, J. F., Marcovecchio, P. M., West, A. P. Jr, Klein, F., Gao, H., Gnanapragasam, P. N., Abadir, A., Seaman, M. S., Nussenzweig, M. C. & Bjorkman, P. J. (2011). *Science*, **334**, 1289–1293.
Doores, K. J. (2015). *FEBS J.* **282**, 4679–4691.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Garces, F., Lee, J. H., de Val, N., Torrents de la Pena, A., Kong, L., Puchades, C., Hua, Y., Stanfield, R. L., Burton, D. R., Moore, J. P., Sanders, R. W., Ward, A. B. & Wilson, I. A. (2015). *Immunity*, **43**, 1053–1063.
Go, E. P., Hewawasam, G., Liao, H.-X., Chen, H., Ping, L.-H., Anderson, J. A., Hua, D. C., Haynes, B. F. & Desaire, H. (2011). *J. Virol.* **85**, 8270–8284.
Gristick, H. B., von Boehmer, L., West, A. P. Jr, Schamber, M., Gazumyan, A., Golijanin, J., Seaman, M. S., Fätkenheuer, G., Klein, F., Nussenzweig, M. C. & Bjorkman, P. J. (2016). *Nature Struct. Mol. Biol.* **23**, 906–915.
Julien, J.-P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Ward, A. B. & Wilson, I. A. (2013). *Science*, **342**, 1477–1483.
Kong, L., Torrents de la Peña, A., Deller, M. C., Garces, F., Sliepen, K., Hua, Y., Stanfield, R. L., Sanders, R. W. & Wilson, I. A. (2015). *Acta Cryst.* D**71**, 2099–2108.
Kwon, Y. D. *et al.* (2015). *Nature Struct. Mol. Biol.* **22**, 522–531.
Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Nature (London)*, **393**, 648–659.
Lasky, L. A., Groopman, J. E., Fennie, C. W., Benz, P. M., Capon, D. J., Dowbenko, D. J., Nakamura, G. R., Nunes, W. M., Renz, M. E. & Berman, P. W. (1986). *Science*, **233**, 209–212.
Lee, J. H., de Val, N., Lyumkis, D. & Ward, A. B. (2015). *Structure*, **23**, 1943–1951.
Lee, J. H., Ozorowski, G. & Ward, A. B. (2016). *Science*, **351**, 1043–1048.
Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J. F., Halper-Stromberg, A., Gnanapragasam, P. N., Spencer, D. I., Seaman, M. S., Schuitemaker, H., Feizi, T., Nussenzweig, M. C. & Bjorkman, P. J. (2012). *Proc. Natl Acad. Sci. USA*, **109**, E3268–E3277.
Pancera, M. *et al.* (2014). *Nature (London)*, **514**, 455–461.
Sanders, R. W. *et al.* (2013). *PLoS Pathog.* **9**, e1003618.
Scharf, L., Scheid, J. F., Lee, J. H., West, A. P. Jr, Chen, C., Gao, H., Gnanapragasam, P. N., Mares, R., Seaman, M. S., Ward, A. B., Nussenzweig, M. C. & Bjorkman, P. J. (2014). *Cell. Rep.* **7**, 785–795.
Scharf, L., Wang, H., Gao, H., Chen, S., McDowall, A. W. & Bjorkman, P. J. (2015). *Cell*, **162**, 1379–1390.
Stewart-Jones, G. B. E. *et al.* (2016). *Cell*, **165**, 813–826.
Wang, H., Cohen, A. A., Galimidi, R. P., Gristick, H. B., Jensen, G. J. & Bjorkman, P. J. (2016). *Proc. Natl Acad. Sci. USA*, **113**, E7151–E7158.
Wang, H., Gristick, H. B., Scharf, L., West, A. P., Galimidi, R. P., Seaman, M. S., Freund, N. T., Nussenzweig, M. C. & Bjorkman, P. J. (2017). *Elife*, **6**, e27389.
Zhou, T. *et al.* (2010). *Science*, **329**, 811–817.
Zhou, T. *et al.* (2013). *Immunity*, **39**, 245–258.
Zhou, T. *et al.* (2015). *Cell*, **161**, 1280–1292.