











# Extended antibody-framework-to-antigen distance observed exclusively with broad HIV-1-neutralizing antibodies recognizing glycan-dense surfaces

Myungjin Lee <sup>1,4</sup>, Anita Changela<sup>1,4</sup>, Jason Gorman <sup>1,4</sup>, Reda Rawi <sup>1</sup>, Tatsiana Bylund<sup>1</sup>, Cara W. Chao<sup>1</sup>, Bob C. Lin<sup>1</sup>, Mark K. Louder <sup>1</sup>, Adam S. Olia<sup>1</sup>, Baoshan Zhang<sup>1</sup>, Nicole A. Doria-Rose<sup>1</sup>, Susan Zolla-Pazner <sup>2</sup>, Lawrence Shapiro <sup>1,3</sup>, Gwo-Yu Chuang<sup>1</sup>  & Peter D. Kwong <sup>1,3</sup> 

Antibody-Framework-to-Antigen Distance (AFAD) – the distance between the body of an antibody and a protein antigen – is an important parameter governing antibody recognition. Here, we quantify AFAD for ~2,000 non-redundant antibody-protein-antigen complexes in the Protein Data Bank. AFADs showed a gaussian distribution with mean of 16.3 Å and standard deviation ( $\sigma$ ) of 2.4 Å. Notably, antibody-antigen complexes with extended AFADs ( $>3\sigma$ ) were exclusively human immunodeficiency virus-type 1 (HIV-1)-neutralizing antibodies. High correlation ( $R^2 = 0.8110$ ) was observed between AFADs and glycan coverage, as assessed by molecular dynamics simulations of the HIV-1-envelope trimer. Especially long AFADs were observed for antibodies targeting the glycosylated trimer apex, and we tested the impact of introducing an apex-glycan hole (N16OK); the cryo-EM structure of the glycan hole-targeting HIV-1-neutralizing antibody 2909 in complex with an N16OK-envelope trimer revealed a substantially shorter AFAD. Overall, extended AFADs exclusively recognized densely glycosylated surfaces, with the introduction of a glycan hole enabling closer recognition.

<sup>1</sup>Vaccine Research Center, NIAID, National Institutes of Health, Bethesda, MD 20892, USA. <sup>2</sup>Department of Medicine and Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. <sup>3</sup>Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA. <sup>4</sup>These authors contributed equally: Myungjin Lee, Anita Changela, Jason Gorman. email: [gwo-yu.chuang@nih.gov](mailto:gwo-yu.chuang@nih.gov); [pdkwong@nih.gov](mailto:pdkwong@nih.gov)

Antibody recognition of protein surfaces forms the basis of the humoral immune response. Such recognition is a subset of more general protein–protein interactions, and substantial analysis has been carried out to understand the fundamental properties that characterize a protein–protein interface, including the size and shape of the interface, the complementarity between interacting surfaces, the propensity of residues to form interfaces, the hydrophobicity–polarity of interfaces, the amount of interfacial solvent, the segmentation and secondary structure of interfacing regions, and the conformational alterations that occur upon complex formation<sup>1,2</sup>.

With antibodies, protein recognition is encoded by recombination and developed through immune processes of selection and maturation<sup>3–5</sup>. These immune processes enable the development of high-affinity recognition against virtually all protein antigens, through elaborations of six complementarity determining regions (CDRs), three on the antibody-heavy chain and three on the antibody-light chain, which comprise the antibody-recognition surface<sup>6</sup>. Because the highly variable CDRs extend from  $\beta$ -sheets that comprise the conserved body of the antibody, it is possible to analyze properties of the conserved body of the antibody in the context of overall antibody–protein recognition.

Here we explore an important parameter governing antibody recognition, Antibody-Framework-to-Antigen Distance (AFAD)—the distance between the conserved body of the antibody and the recognized protein antigen—and quantify AFAD for 1879 non-redundant antibody–protein antigen complexes in the Protein Data Bank<sup>7</sup>. To provide insight into results, we performed molecular dynamics (MD) simulations on glycosylated human immunodeficiency virus-type 1 (HIV-1) envelope (Env) structures and determined the cryo-EM structure of an HIV-1 Env trimer in complex with monoclonal antibody 2909, which recognizes a glycan hole at the trimer apex. We found long AFADs to be exclusive to broadly neutralizing antibodies targeting HIV and their AFADs to correlate with density of glycan coverage. Overall, the results provide insight into the requirements for antibody recognition of highly glycosylated antigens.

## Results

**Analysis of AFAD for all antibody complexes in the PDB.** We defined AFAD as the closest distance between any heavy atom on the antigen and any of the heavy atoms on the four conserved cysteines on the variable domain (Fv)—residues C22 and C92 on the heavy chain, and C23 and C88 on the light chain—of the antibody (Fig. 1a, see “Methods” for details). Using this definition, we calculated AFAD for 1879 non-redundant antibody–antigen complexes from the Protein Data Bank<sup>7</sup>. We observed AFADs to form a Gaussian distribution, with a mean distance from cysteines of 16.3 Å and a standard deviation ( $\sigma$ ) of 2.4 Å (Fig. 1b).

Short AFADs ( $\sim 3\sigma$  below the mean) were observed for ligands that recognize the side of the antibody (Supplementary Fig. 1a left), with a peptide that penetrated into the CDR cleft (Supplementary Fig. 1a middle), or for ligands with large interactive surfaces that extended to the side of the antibody (Supplementary Fig. 1a right). Average AFADs were observed with both peptides and proteins that utilized CDRs for recognition and did not penetrate through or drape around CDRs (Supplementary Fig. 1b). Extended AFADs ( $\sim 3\sigma$  above the mean) were observed for antibodies binding highly glycosylated surfaces, especially those on the HIV-1 Env glycoprotein trimer (Supplementary Fig. 1c, Supplementary Table 1).

We compared the AFADs of 23 HIV-1 Env targeting broadly neutralizing antibodies, each representing a unique antibody class<sup>8</sup>, and observed 8 of the 23 antibodies to have AFADs longer

than the mean of the general distribution by  $3\sigma$  (7.3 Å) (Fig. 1b). Notably, none of the antibody–antigen complexes for Ebola virus glycoprotein, SARS-CoV-2 spike, influenza A virus hemagglutinin, Dengue/Zika envelope exhibited entries with such high levels of AFAD (Fig. 1b) even though antibodies against these pathogens (except for SARS-CoV-2) were found in the top 50 for extended AFADs relative to the mean (Supplementary Table 1). Only a subset of the top 50 was significantly different from the mean (AFAD  $> 3\sigma$ ), corresponding to the top 10 in Supplementary Table 1, which target densely glycosylated epitopes.

The 1879 PDB collection is composed of 823 viral pathogens and 1056 non-viral pathogens, with viral pathogens representing slightly less than half of the total PDB collection (Supplementary Fig. 2a). We evaluated the distribution of each viral antigen–antibody complex (e.g., HIV-1 non-broadly neutralizing, HIV-1 broadly neutralizing, and antibodies targeting influenza A virus, hepatitis C virus, Dengue/Zika virus, SARS-CoV-2, and Ebola virus) in Supplementary Fig. 2b and found HIV-1 antibody complexes (broadly neutralizing and non-broadly neutralizing combined) to dominate in the viral antigen dataset.

To examine a potential systematic bias toward observing longer AFADs with certain structure determination methods, we separated the entire PDB dataset by X-ray crystal structures and electron microscopy (EM) structures and analyzed the AFADs. The AFADs calculated from crystal structures (Supplementary Fig. 2c) and EM structures (Supplementary Fig. 2d) distribute in a similar manner and follow the total PDB collection distribution, suggesting there is no systematic bias toward observing longer AFADs with certain structure determination methods. This could not be factored when determining “deviation from the mean”. The deviation from the mean by what we observed is thoroughly from the shorter/longer distance of antibody–antigen recognition.

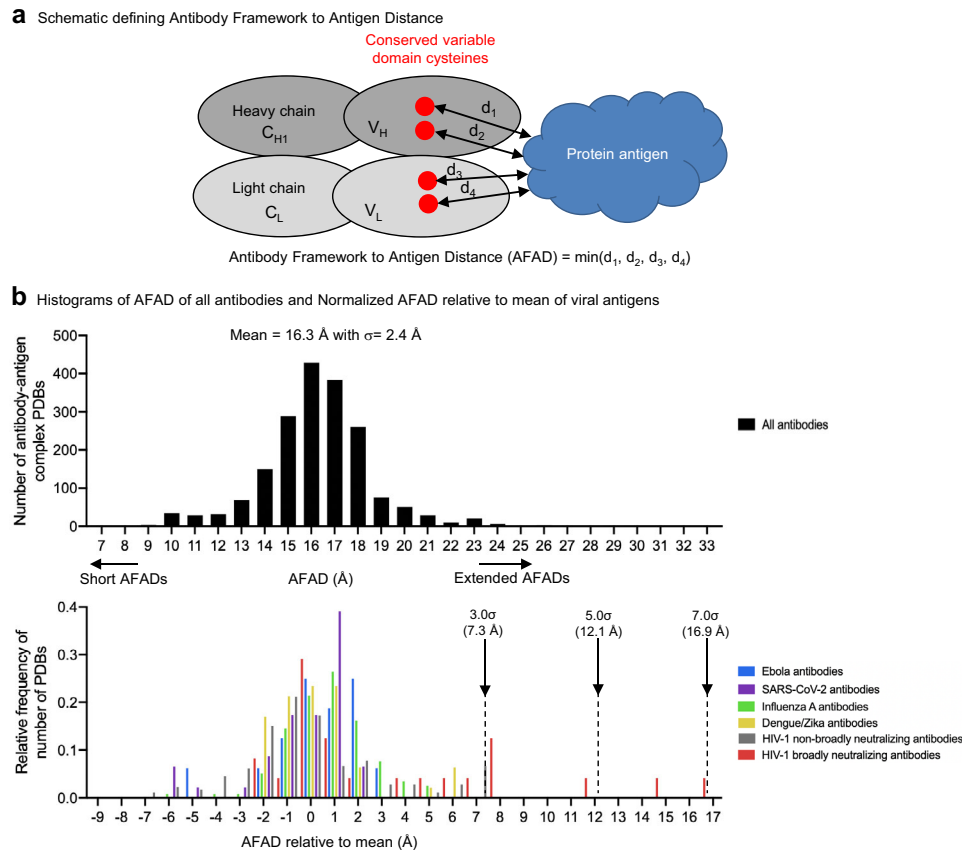
We calculated CDR H3/L3 lengths of all antibodies in the PDB dataset and show the distribution in Supplementary Fig. 2e and Supplementary Fig. 2f. The CDR H3 length distribution followed a Gaussian shape similar to AFAD distribution, whereas CDR L3 length showed a narrow distribution centered on 9 amino acids. We analyzed the correlation between AFAD and CDR H3 or CDR L3 length for 1879 non-redundant antibody–protein antigen complexes and found low correlations for CDR H3 length ( $R^2 = 0.1652$ ) (Supplementary Fig. 2g) and for CDR L3 length ( $R^2 = 0.0086$ ) (Supplementary Fig. 2h). However, the correlation between AFAD and CDR H3 length was higher ( $R^2 = 0.4593$ ) when analyzing antibodies with long CDR H3 ( $> 24$  amino acids) (Supplementary Fig. 2i).

### Only select categories of HIV-1 broadly neutralizing antibodies have extended AFADs relative to mean.

Examining the AFAD of HIV-1 antibodies based on their epitope categories (V1V2, glycan-V3, CD4-binding site, silent face, fusion peptide, subunit interface)<sup>8</sup>, we observed the V1V2 class broadly neutralizing antibodies to possess the longest AFAD relative to mean on average (12.9 Å), followed by silent face antibodies (7.0 Å), and glycan-V3 antibodies (5.1 Å) (Fig. 2). Specifically, CAP256-VRC26.25 possessed the longest AFAD relative to mean (17.3 Å), followed by PGT145 (15.2 Å), PG9 (11.6 Å), 2G12 (7.9 Å), VRC-PG05 (7.6 Å), VRC38.01 (7.6 Å), PGT128 (7.3 Å), and SF12 (6.4 Å).

### Extended AFADs are observed with antibodies recognizing densely glycosylated regions.

To examine the relationship between AFADs and glycan coverage on HIV-1 Env surface, we performed MD simulations of a BG505 wild-type (WT) Env trimer and quantified the number of glycans associated with each residue over the course of simulations (Fig. 3a). We observed that V1V2, glycan-



**Fig. 1 Analysis of Antibody-Framework-to-Antigen Distance (AFAD) for all antibody complexes in the PDB reveals outliers with extended recognition distances to be exclusively HIV-1 broadly neutralizing antibodies.** **a** Schematic of AFAD calculated from the shortest protein antigen atom-distance to each of the four conserved variable domain cysteines (red). Only protein atoms were used for calculations, excluding distances for cofactors, glycans, and other non-protein atoms. **b** Histogram of AFAD for  $n = 1879$  antibody complexes in the PDB (black). Normalized AFADs relative to mean are also shown for Ebola virus (blue), SARS-CoV-2 (purple), influenza A virus (green), Dengue/Zika virus (yellow), human immunodeficiency virus-1 (HIV-1) non-broadly neutralizing antibodies (gray), and broadly neutralizing antibodies targeting HIV-1 (red). AFAD relative to mean for  $n = 1879$  PDB complexes formed a normal distribution as mean values  $16.3 \pm 2.4$  Å SD. Select HIV-1 broad neutralizing antibody-recognition distances can be extended by more than 3 $\sigma$  longer. AFADs in the bottom plot are normalized relative to mean (16.3 Å). See also Supplementary Table 1.

V3, and silent face regions are the most densely glycosylated (colored in green to dark green), while the rest of the regions are covered with less glycans (in light green to white) (Fig. 3b).

We also found that antibodies with long AFADs recognized regions of the Env trimer with dense glycan shield coverage, whereas antibodies with short AFADs recognized regions of the Env trimer with sparse glycan coverage. A high correlation ( $R^2 = 0.8110$ ,  $p < 0.0001$ ) was observed between AFAD relative to mean and glycan coverage for 23 classes of HIV-1 broadly neutralizing antibodies (Fig. 3c). Densely glycosylated areas such as V1V2, silent face, and glycan-V3 regions were observed to have glycan coverage beyond  $1.39$  counts/Å<sup>2</sup> with equal or longer than  $1.45$  Å extended AFADs relative to mean, whereas the CD4-binding site, fusion peptide, and subunit interface had glycan coverages of lower than  $1.39$  counts/Å<sup>2</sup>, with shorter than  $1.45$  Å to negative AFADs relative to mean (except 35O22) (Supplementary Table 2).

We also examined the correlation between AFADs relative to mean and various antigen-antibody interaction properties (e.g., CDR H3 paratope surface area percentage, epitope total/protein surface area, CDR H3 length, epitope-glycan percentage, neutralization potency, epitope conservation, paratope surface area, somatic hypermutation, neutralization breadth) (Supplementary Fig. 3). While statistically significant, the correlation values were found to be less than that of glycan coverage versus AFAD relative to the mean. Notably, the paratope surface area

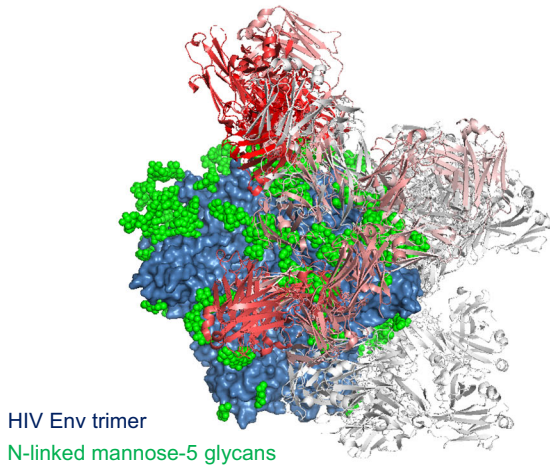
percentage of CDR H3 was found to have the second strongest correlation ( $R^2 = 0.6420$ ) after AFAD versus glycan coverage, suggesting that the paratope of antibodies with long AFADs are dominated by CDR H3 surface area. The CDR H3 contact with antigen is mostly related to long AFAD while other CDRs are interacting with protruding glycan residues.

In addition, we calculated the correlation of glycan coverage with the same set of properties, and similarly, none showed a correlation as high as the correlation between AFAD relative to mean and glycan coverage (Supplementary Fig. 4).

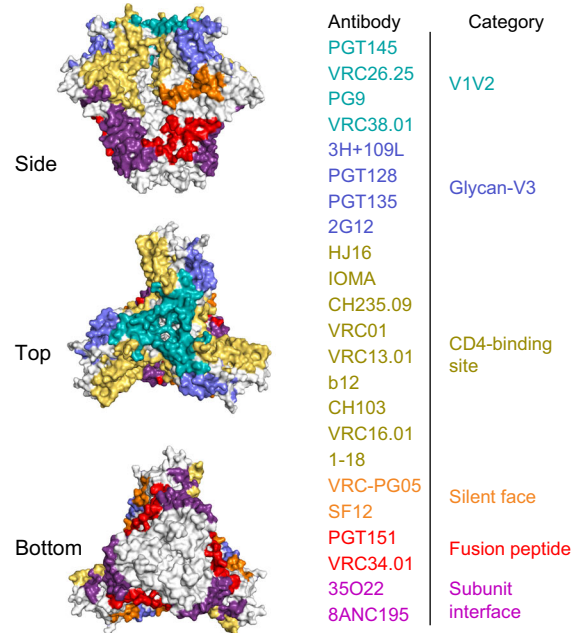
When analyzing antibodies with long AFADs or dense epitope-glycan coverage (above the average) only, the correlation between AFADs and glycan coverage or CDR H3 length was similar or stronger than that observed with the full set of HIV-1 broadly neutralizing antibodies ( $R^2 = 0.7444$ ,  $0.5545$ , respectively) (Supplementary Fig. 5a, b). Also, the correlation between AFAD and glycan coverage when epitope-glycan coverage is above average (Supplementary Fig. 5a) was found to be the same as glycan coverage versus CDR H3 (Supplementary Fig. 5c). Therefore, AFAD relates to glycan coverage ( $R^2 = 0.7444$ ) slightly stronger than the CDR H3 when CDR H3 is long ( $R^2 = 0.5545$ ) (Supplementary Fig. 5a and b).

**Structure of N160K glycan-hole recognizing antibody 2909 with Env trimer reveals similar recognition to PGT145, but**

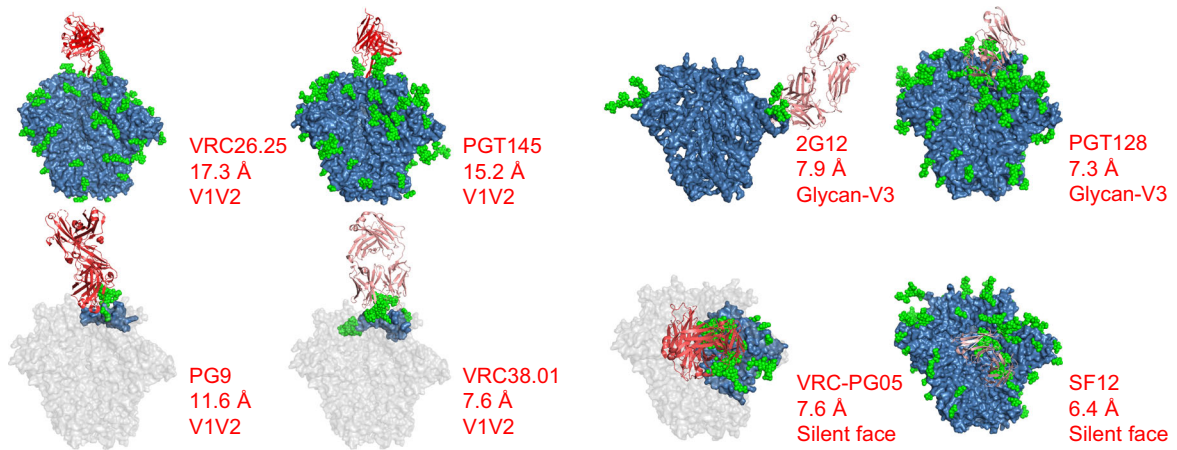
HIV-1 broadly neutralizing antibodies, colored according to their AFAD relative to mean



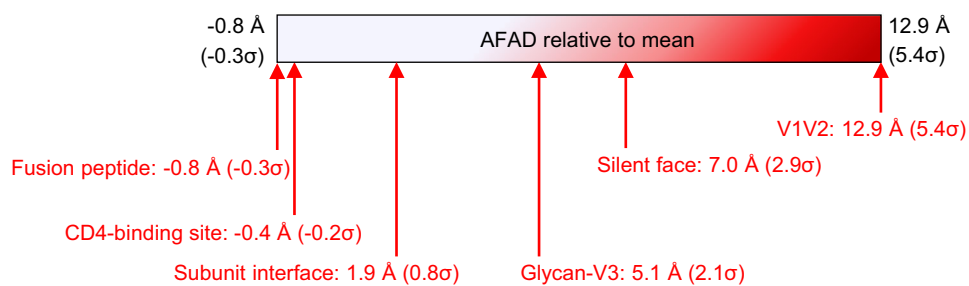
Protein surface colored by antibody class and its category



Specific HIV-1 broadly neutralizing antibodies with extended AFAD relative to mean

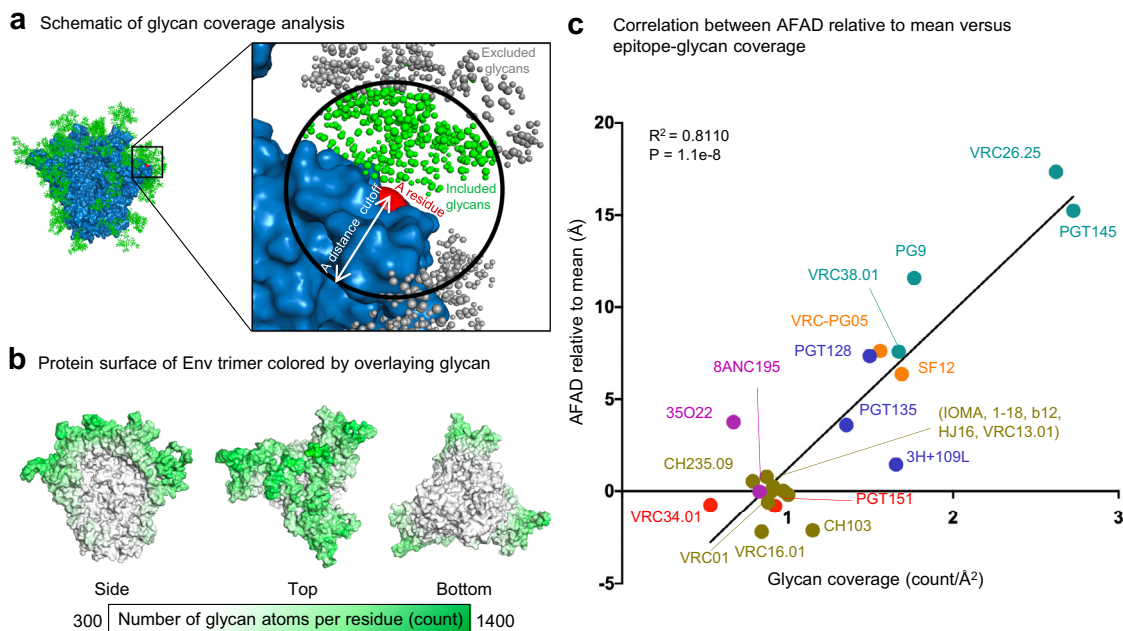


Categories of HIV-1 broadly neutralizing antibodies and their average AFAD relative to mean



**Fig. 2 Only select categories of HIV-1 broadly neutralizing antibodies have extended AFADs relative to mean.** HIV-1 broadly neutralizing antibodies, colored according to their AFAD relative to mean and the associated scale is shown at the bottom. Protein surface of Env trimer colored by epitope category (top right). See also Supplementary Table 2. Average AFAD relative to mean for classes of HIV-1 broadly neutralizing antibodies. The antigens are colored in blue if structures are available, otherwise in gray (middle).





**Fig. 3** AFAD correlated significantly with epitope-glycan coverage. **a** Schematic figure for defining associated glycan atoms per residue. Glycan atoms within a cutoff are counted as number of glycan atoms per residue (green) except those are blocked by protein region (gray). Glycans outside of the cutoff are excluded in the analysis (gray). The protein surface is colored in blue and the residue of interest on the surface is colored in red. **b** Protein surface of BG505 WT Env trimer colored by number of overlaying glycan atoms. Number of glycan atoms overlaying each exposed protein surface residue was calculated by the algorithm described in **(a)** and averaged over 500 ns molecular dynamics simulations of fully glycosylated HIV-1 BG505 Env trimer. **c** Correlation of AFAD relative to mean versus glycan coverage overlaying antibody epitopes of BG505 WT. A simple linear regression is calculated with  $n = 22$  antibodies. 2G12 is not presented in the plot due to absence of protein epitope residues. The antibody epitopes are colored as shown in Fig. 2 top right panel.

with  $\sim 13$  Å closer approach. We next investigated the effect of introducing a glycan hole on the HIV-1 Env trimer on AFAD. Human antibody 2909 is a strain-specific, HIV-1 neutralizing antibody targeting a quaternary V1V2 apex epitope containing a rare N160K substitution that removes a glycan<sup>9–11</sup>. To evaluate 2909 interactions with Env trimer containing an N160K glycan hole, we produced an N160K variant of the prefusion-closed stabilized CAP256.wk34.c80 SOSIP.RnS2 Env trimer<sup>12</sup>. Biolayer interferometry analysis confirmed 2909 binding to the CAP256-N160K glycan-knockout variant and not to CAP256 Env trimer containing the N160 glycan while the N160K mutation abolished recognition by PGT145 (Fig. 4a).

To provide a structural basis for 2909 recognition of the N160K glycan hole, we determined the 3.91 Å cryo-EM structure of the antigen-binding fragment (Fab) of 2909 in complex with the CAP256.wk34.c80 SOSIP.RnS2.N160K trimer bound to 3BNC117 Fab (Fig. 4b, Supplementary Fig. 6 and Supplementary Table 3). 3BNC117 was used to purify the trimer, to increase the mass and orientation distribution of particles for cryo-EM analysis, and to provide an analogous comparison to the PGT145-Env complex bound by 3BNC117. The CDR H3 of 2909 formed an extended hairpin, with sulfated tyrosine at its tip, highly similar to the previously determined ligand-free monoclonal antibody structure<sup>13–15</sup>. This extended hairpin inserted into the hole at the trimer apex to interact electrostatically with Lys121, in a manner similar to antibody PGT145<sup>16,17</sup>, which also possesses a CDR H3 which forms an extended hairpin, with sulfated tyrosine at its tip<sup>13,18</sup> (Fig. 4b). Notably, the distance of the antibody 2909 from the N160K Env trimer was 13.3 Å closer than the distance of antibody PGT145 from the WT trimer.

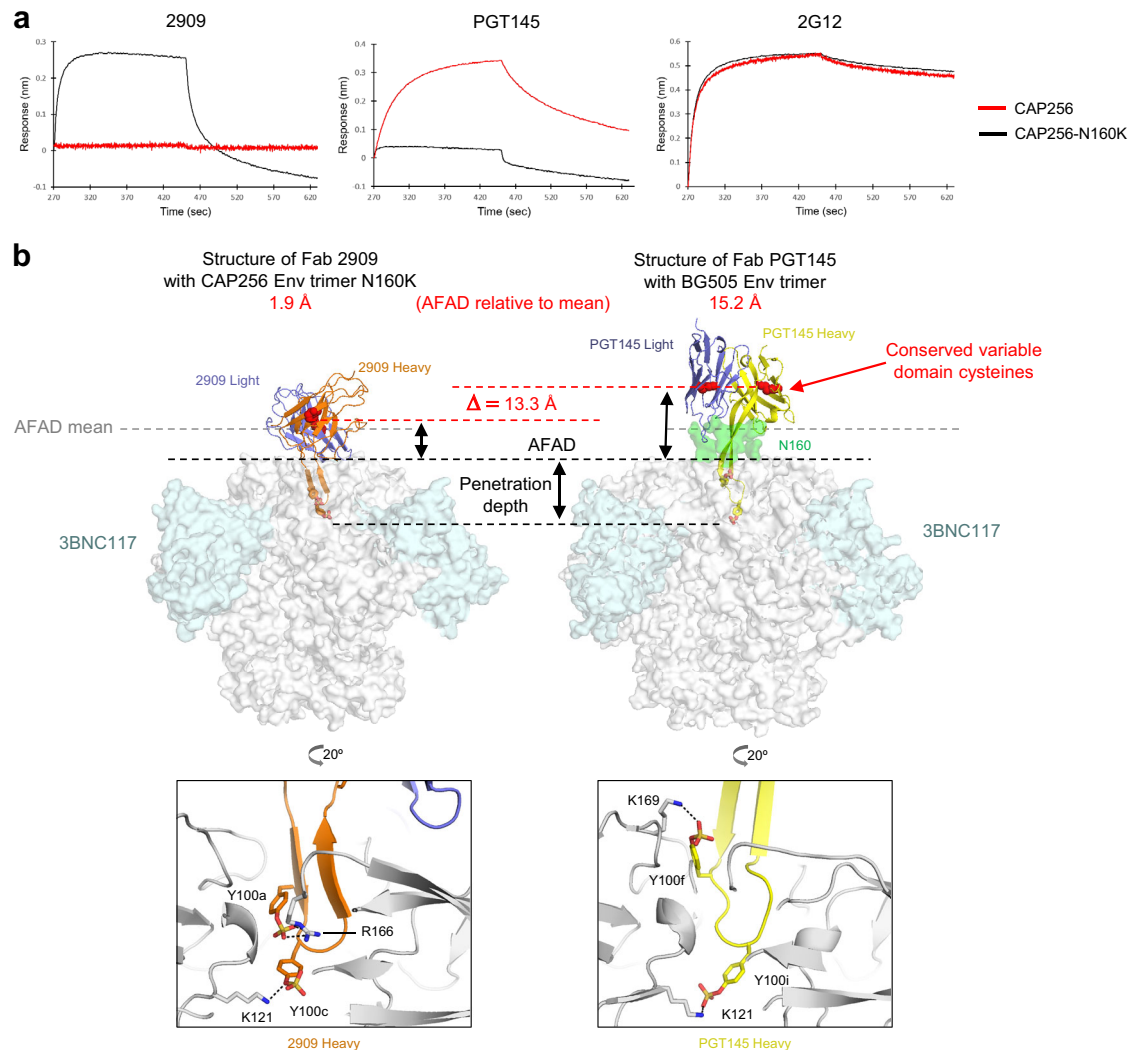
**Antibody 2909 targeting the glycan hole of the N160K Env trimer induces a shorter AFAD relative to the mean but a**

**larger protein epitope surface area compared to antibody PGT145.** The removal of glycans at position N160 in the V1V2 region generated a glycan hole that led to a significant difference in the number of glycan atoms per residue on the Env trimer between WT and N160K (Fig. 5a bottom). This further resulted in a substantial difference in epitope-glycan coverage between WT and N160K Env trimer for antibodies approaching this region (e.g., VRC26.25, PGT145, PG9, VRC38.01) (Fig. 5a top).

Interestingly, in spite of PGT145 and 2909 binding to the same V1V2 region, we observed far different glycan coverage from these two antibodies' epitopes (Fig. 5a). The main reason for the discrepancy between these two epitopes arises from the definition of glycan coverage—the number of glycan atoms per surface area of the epitope. In other words, the epitopes of two different antibodies targeting the same region with similar number of glycan atoms may have significantly different surface areas corresponding to their AFADs relative to the mean. Antibody 2909 has a much larger protein epitope surface than PGT145 (Fig. 5b and Supplementary Table 4), while there was only a subtle discrepancy in the average number of glycan atoms between WT and N160K epitopes of PGT145 (392.79) and 2909 (323.92) (Supplementary Table 4). Consequentially, this leads to the 2909 epitope having less than half the glycan coverage compared to the PGT145 epitope.

## Discussion

In this study, we systematically analyzed the AFADs of all antibody-antigen complexes available in the Protein Data Bank<sup>7</sup>, and showed that the longest AFADs were from antibodies targeting the HIV-1 Env. Specifically, these HIV-1 antibodies with long AFADs target the V1V2, glycan-V3, and silent face sites of HIV-1 Env, all of which are densely glycosylated. We found that only broadly neutralizing HIV-1 antibodies show the extended



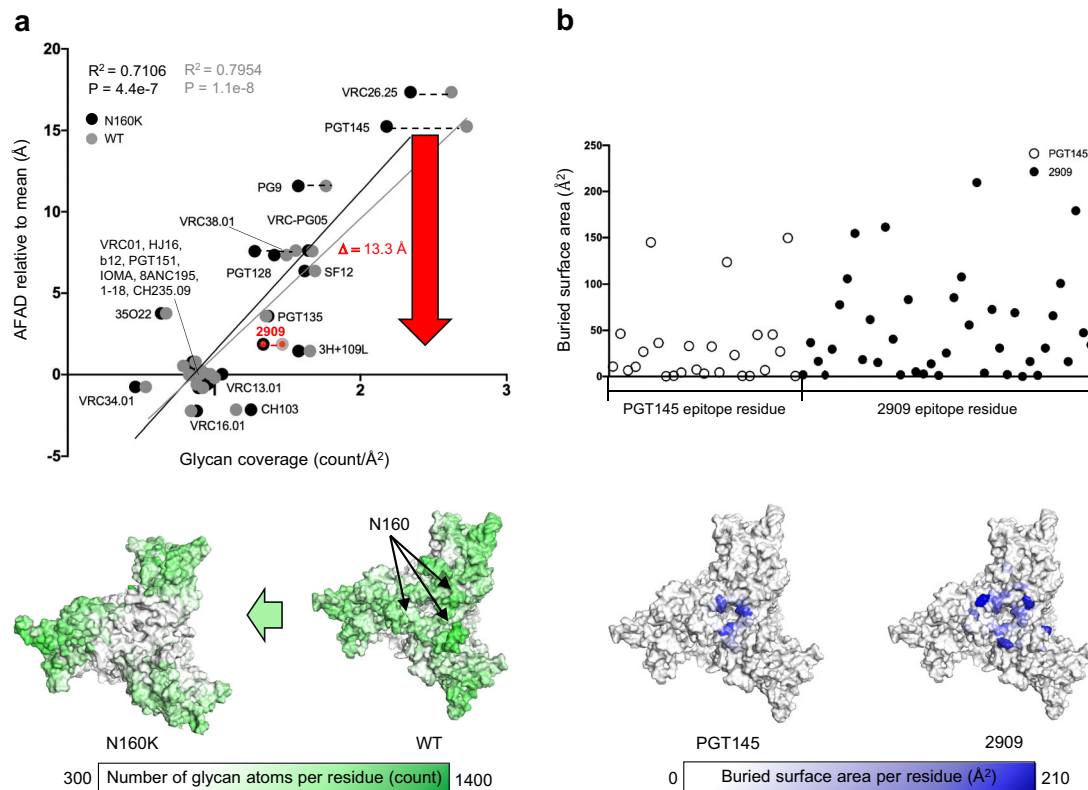
**Fig. 4 Structure of N160K glycan-hole recognizing antibody 2909 with Env trimer reveals similar recognition to PGT145, but with -13 Å closer approach.** **a** CAP256.wk34.c80 SOSIP.RnS2 Env trimer variants containing the native N160 glycan site (CAP256-WT) or the N160K mutation (CAP256-N160K) were assessed for binding to 2909, PGT145, and 2G12 antibodies using biolayer interferometry. **b** Cryo-EM structure of CAP256-N160K Env trimer (surface representation in gray) in complex with the Fabs from antibodies 2909 and 3BNC117; inset shows CDR H3 region of 2909 penetrating the apex of the trimer with two sulfated tyrosines, Tys100c shown interacting with K121; CDR H3 length is 21 in Kabat (left). Cryo-EM structure of BG505 Env trimer (surface representation in gray) in complex with the Fabs from antibodies PGT145 and 3BNC117 (PDB: 5V8L); inset shows CDR H3 region of PGT145 penetrating the apex of the trimer with two sulfated tyrosines, Tys100i shown interacting with K121. CDR H3 length is 31 in Kabat (right). See also Supplementary Fig. 6 and Supplementary Table 3.

AFADs relative to the mean, while none of the other antibodies targeting viral antigens, such as Ebola virus glycoprotein, influenza A virus hemagglutinin, Dengue/Zika envelope, or SARS-CoV-2 spike, had significantly extended AFADs relative to mean. HIV-1 Env has the most dense glycan coverage (Supplementary Fig. 7), especially in certain regions (V1V2, glycan-V3, and silent face sites), which explains why the exceptionally long AFADs were only observed for HIV-1 antibodies. The significant correlation between AFAD and glycan density demonstrated the dependency of these two properties.

We have used the distance between conserved Fv cysteines and antigens to define AFAD. A main caveat of this method is that an exceptionally short AFAD relative to the mean can result from an antigen interacting with the side of the antibody (Supplementary Fig. 1). Also, we built Man-9 models on each of the N-linked glycosylation sequons for MD simulations and this might not precisely capture the actual glycan coverage<sup>19</sup>. Finally, we have shown one example with 2909, an antibody targeting a strain-

specific glycan hole, which has a shorter AFAD relative to the mean at the densely glycosylated V1V2 site. The removal of three N160 glycans from the apex of the trimer mainly led to a considerably shorter AFAD relative to the mean. Shorter AFADs relative to the mean were also observed for antibodies targeting glycan holes in the V3 region, such as antibodies R56<sup>20</sup> (-4.38 Å) and R20<sup>20</sup> (3.51 Å), which were elicited from immunization of rabbits with gp120 of clade B strain JR-FL, supporting the observation that antibodies binding to glycan holes generally do not have extended AFADs relative to mean.

Of note, there was only a low correlation observed between epitope-glycan content and glycan coverage (Supplementary Fig. 4d), and the correlation of AFAD to epitope-glycan content (Supplementary Fig. 3d) was also much lower than that to glycan coverage (Fig. 3c). Glycan coverage quantifies the density of glycan shield that an antibody needs to overcome (e.g., through extended AFAD), but antibodies, such as fusion peptide-targeting VRC34.01 and PGT151, can certainly have short AFADs and



**Fig. 5 Antibody 2909 targeting to the glycan hole of N160K Env trimer induces shorter AFAD relative to mean but larger surface area compared to antibody PGT145.** **a** Plot of AFAD relative to mean versus epitope-glycan coverage overlaying antibody epitopes of BG505 WT (dotted in gray) and BG505 N160K (dotted in black). Simple linear regressions are calculated with  $n = 23$  antibodies on both WT and N160K. 2G12 is not presented in the plot due to the absence of protein epitope residues. 2909 is added to compare the difference between WT and N160K. The dash lines are plotted to the antibody epitopes that have the difference of glycan coverage higher than 0.2 (counts/Å<sup>2</sup>) between WT and N160K. The protein surface of BG505 WT (right) and N160K (left) Env trimer colored by number of overlaying glycan atoms (bottom). The red arrow indicates the difference of AFAD between PGT145 and 2909 which is 13.3 Å. **b** The buried surface area of each epitope residue for PGT145 and 2909 (top). The protein surface of epitope residues of PGT145 and 2909 are colored according to their buried surface area (bottom). See also Supplementary Table 4.

have substantial contact with glycans when binding, given that the glycan shield is not too dense (i.e., low glycan coverage).

Understanding the relationship between antibody properties and antigen recognition is critical for the design of B cell-based vaccines. This is particularly important for HIV-1 vaccine development as there are many obstacles, such as sequence diversity, conformational flexibility, and dense glycosylation of the HIV-1 Env trimer that have impeded the development of an effective HIV-1 vaccine<sup>21</sup>. Previously, we have analyzed several antibody properties, such as CDR H3 length, number of antibody protruding loops, level of somatic hypermutation, number of independent epitope segments and epitope-glycan components, and have found that epitope protein surface correlated with neutralization breadth and epitope glycan component inversely correlated with neutralization breadth<sup>8</sup>. Continuing research of antibody properties, quantification of AFAD at the molecular level in this work provides insight into antibody-guided vaccine design. The remarkable correlation between AFAD relative to the mean and glycan density suggests that it might be reasonable to focus on vaccine response toward glycan-depleted regions<sup>22–24</sup>, as antibodies targeting glycan-dense region often requires longer AFAD, which may be difficult to elicit. Alternatively, we note that the three antibodies with the largest AFADs targeting the most glycan-dense sites were some of the most potent broadly neutralizing antibodies thus far identified; despite their being difficult to elicit, the field may want to investigate vaccine regimes that can

coax the immune system into eliciting antibodies with long AFADs capable of effectively targeting glycan-dense regions.

**Methods**

**Antibody-antigen complexes.** More than 4500 PDBs of antibody-antigen complexes with high resolution (<3 Å) were obtained from SAbDab<sup>25</sup> on November 5, 2020. To remove redundancy of ~4500 PDBs, CD-HIT<sup>26,27</sup> was used to cluster the antibody-antigen complexes. A single input file that includes a full antibody set of heavy and light chain concatenated sequences was entered. CD-HIT clustered the entire set and yielded 1879 non-redundant antigen-antibody complexes using 95% sequence identity and 99% of alignment coverage for the longer sequence, while excluded gaps when calculating identity. The computational time spent ~1 s with one core of CPU, 2.4 GHz (Intel Xeon E5-2680v4). For HIV-1 antigen-antibody complexes specifically, we used the 20 complexes as described in Chuang et al.<sup>8</sup>, plus the addition of SF12 (PDB:6OKP)<sup>28</sup>, CAP256-VRC26.25 (PDB:6VTT)<sup>12</sup>, and 1-18 (PDB:6UDJ)<sup>29</sup> (Supplementary Table 5).

**Calculation of AFAD.** We measured the distances between any heavy atom on the antigen and any heavy atom on the four conserved cysteines—C22 and C92 on the heavy chain, and C23 and C88 on the light chain—of the antibody four conserved cysteines—C22 (D<sub>1</sub>) and C92 (D<sub>2</sub>) on the heavy chain, and C23 (D<sub>3</sub>) and C88 (D<sub>4</sub>) on the light chain—of the antibody.

$$D_i = \sqrt{(x_{ia} - x_{ib})^2 + (y_{ia} - y_{ib})^2 + (z_{ia} - z_{ib})^2}, \tag{1}$$

where  $(x_{ia}, y_{ia}, z_{ia})$  are the coordinates of the closest atom in antigen,  $(x_{ib}, y_{ib}, z_{ib})$  are the coordinates of the  $i$ -th conserved cysteine in variable domain of antibody. Then, AFAD was defined as the minimum value of four distances.

$$\text{AFAD} = \min(D_1, D_2, D_3, D_4), \tag{2}$$



**Glycan coverage analysis.** We used an in-house program, GLYCO, to quantify the glycan shield of Env trimer. The method counts the number of glycan atoms of glycosylated protein structures within a certain radial distance cutoff (e.g., 26 Å) from surface protein residues, while excluding ones that are blocked by protein region. The counted glycan atom numbers are averaged over 500 frames (one frame per one ns) of 500 nanosecond (ns) MD simulations. The cutoff was determined by analyzing the longest length of the glycan in Env trimer averaged over 500 ns of MD simulations. The surface residues were evaluated by their solvent accessible surface area (SASA) by FreeSASA<sup>30</sup> with 5 Å of probe radius. The residues with larger than 30 Å<sup>2</sup> of their surface area were selected as surface residues.

Epitope-glycan coverage was calculated by counting the number of glycan atoms of epitope residues of the glycosylated protein and dividing by surface area of the epitopes. The number of glycan atoms per epitope residue was analyzed with the same method and same parameters described above. The summation of the epitope residues' buried SASA was used as surface area of the epitope to normalize the average number of glycan atoms (see more details in "Epitope residue, epitope total/protein surface area, and paratope surface area determination" section).

**Molecular dynamics simulations.** Molecular dynamics (MD) simulations were carried out to estimate the conformational flexibility of HIV-1 Env protein under physiological conditions. Initial atomistic model of HIV-1 Env protein was generated using YASARA software's (<http://www.yasara.org/>) homology modeling plugin with default option. HIV-1 Env was modeled using clade A BG505 wild-type sequence as target and 5 trimeric Env structures as templates (PDBs: 5FYJ, 5FYK, 5FYL, 6CK9, 6MUF). Lassa virus GPC using one template (PDB ID: 5VK2), SARS-CoV-2 spike using one template (PDB ID: 6VXX), influenza virus HA using one template (PDB ID: 6CF7), RSV F using one template (PDB ID: 4MMS), and Zika virus E protein using one template (PDB ID: 5JHM). All potential N-linked glycosylation sites were modeled with Man-9 glycans (Figs. 3 and 5) and Man-5 glycans (Supplementary Fig. 7) using an in-house program Glycosylator<sup>31</sup>. Subsequently, the modeled trimeric Env structure was solvated in a 15 Å padding water box and neutralized by the addition of NaCl at a concentration of 150 mM. The final systems were composed of 555,778 (HIV-1 Env), 224,785 (Lassa virus GPC), 759,908 (SARS-CoV-2 spike), 265,711 (influenza virus HA), 258,677 (RSV F), and 150,576 (Zika virus E) atoms. The MD was performed using NAMD2.13 engine<sup>32</sup>, with CHARMM36 force field<sup>33,34</sup>. TIP3P water parameterization was utilized to describe the water molecules<sup>35</sup>. The periodic electrostatic interactions were computed using particle-mesh Ewald (PME) summation<sup>36</sup> with a grid spacing smaller than 1 Å. Constant temperature was imposed by using Langevin dynamics with a damping coefficient of 5.0 ps. The system was first minimized by 20,000 conjugate gradient steps and then equilibrated by using a linear temperature gradient, which heated up the system from 100 to 310 K in 2 ns plus additional 10 ns. The length of all bonds involving hydrogen atoms was constrained with the RATTLE algorithm<sup>37</sup>, thus allowing a time step of 2 fs. Unrestrained molecular dynamics were additionally performed up to 500 ns and used in the analysis of glycan conformations.

**Expression and purification of CAP256 Env trimer.** An N160K variant of the CAP256.wk34.c80 SOSIP.RnS2 trimer<sup>12</sup> was generated using site-directed mutagenesis (GeneImmune Biotechnology, Rockville, MD) and soluble Env trimers were produced by transient transfection in FreeStyle 293-F cells<sup>38</sup>. The FreeStyle 293-F cells were transfected with 600 µg of Env trimer plasmid along with 150 µg of human furin plasmid DNA per liter of cells using 293Fectin (Thermo Scientific) according to the manufacturer's directions. Transfected cells were incubated in a shaker at 120 rpm, 37 °C, and 9% CO<sub>2</sub>. The following day, 80 ml HyClone SFM4HEK293 medium and 80 ml FreeStyle 293 Expression Medium were added to each liter of cells. Culture supernatants were harvested by centrifugation 6 days following transfection and filtered. CAP256 Env trimers were purified using either 2G12 affinity chromatography, with low pH elution, followed by gel filtration on a Superdex 200 column in PBS<sup>12</sup> or by binding to a cleavable version of the CD4-binding site antibody, 3BNC117<sup>39</sup> to facilitate purification. In this case, 250 ml of cell supernatant, harvested from transiently transfected cells expressing 3BNC117 antibody with a HRV3C cleavage site in the heavy-chain hinge region, as described below, was added to the clarified supernatant prior to loading onto a protein A column. Following capture of the Env trimer-3BNC117 complex onto protein A resin, HRV3C protease was added to generate an Env trimer-3BNC117 Fab complex that was eluted from the column in PBS buffer. Env trimers and complexes were purified by gel filtration using a Superdex S-200 column equilibrated in PBS. Antigenicity and structural integrity of purified trimers were confirmed by biolayer interferometry binding and negative-stain electron microscopy analyses.

**Expression and purification of antibodies.** Variable regions for 2909<sup>9</sup>, 3BNC117<sup>39</sup>, PGT145<sup>16,17</sup>, and 2G12<sup>40</sup> were codon-optimized (Supplementary Table 6), synthesized, and subcloned (Gene Universal Inc, Newark, DE) into pVRC8400 (CMV/R) expression vectors containing human IgG constant regions. A heavy-chain plasmid containing an HRV3C cleavage site in the hinge region was used for Fab production. Corresponding heavy and light chain plasmids were co-transfected in Expi293F cells using Turbo293 transfection reagent<sup>41</sup>. Briefly,

0.75 ml of Turbo293 transfection reagent (Speed BioSystems) were added to 12.5 ml Opti-MEM medium (Life Technology) and incubated for 5 min at room temperature. Meanwhile, 125 µg of heavy chain and 125 µg of light chain plasmid DNA were added to 12.5 ml of Opti-MEM medium in another tube. The Opti-MEM medium containing Turbo293 was then added to plasmid DNAs, incubated for 15 min at room temperature, and added to 200 ml of Expi293 cells (Life Technology) at 2.5 million cells/ml. The transfected cells were cultured in shaker incubator at 120 rpm, 37 °C, 9% CO<sub>2</sub>. Culture supernatant was collected 5 days after transfection and loaded onto a protein columns. After washing the columns with PBS, IgG was eluted using a low pH buffer and immediately neutralized with 10% 1 M Tris-Cl, pH 8.0. To produce 2909 Fab, purified IgG was cleaved by HRV3C protease and the digest was passed over a protein A column to separate the Fab fragments. The Fab fragments were purified further by size-exclusion chromatography using a Superdex 200 column in a buffer containing 5 mM HEPES, pH 7.5, and 150 mM NaCl.

**Biolayer interferometry assay.** Binding of CAP256 Env trimers to HIV-1 antibodies was assessed using an Octet HTX instrument, as previously described<sup>38</sup>. Briefly, all assays were carried out with agitation set to 1000 rpm in PBS using solid black 96-well plates (Geiger Bio-One) at 30 °C. For the assessment, IgG antibodies were immobilized onto an AHC capture tip for 180 s at 50 µg/ml in PBS. Biosensor tips were then equilibrated for 300 s prior assessment of binding to the HIV-1 trimer molecules in solution (at 50 µg/ml). Association was allowed to proceed for 180 s followed by dissociation for 300 s. Data analysis was performed in ForteBio Data Analysis 11.0.

**Epitope residue, epitope total/protein surface area, and paratope surface area determination.** FreeSASA<sup>30</sup> was used to determine the epitope residues, epitope total surface area, epitope protein surface area, and paratope surface area. The epitope residues were defined by the residues having non-zero value of their SASA between antibody-antigen complex structure and antigen structure. Summation of the SASA of epitope residues is epitope total surface area and summation of only protein epitope residues is epitope protein surface area. The paratope surface area was calculated in the same manner as epitope surface area but between antibody-antigen complex structure and antibody structure.

**Epitope-glycan percentage.** The epitope-glycan percentage was determined by the dividing the SASA of glycan epitope residues by the SASA of the total epitope residues.

**Epitope conservation.** Residue conservation (percentage of the most prevalent residue type) was calculated based on the filtered web alignment of the year 2018 from the Los Alamos HIV sequence database (<http://www.hiv.lanl.gov/>). Alignment gap was treated as an additional residue type. Epitope conservation was determined based on average of the residue conservation among the epitope residues.

**CDR H3, CDR L3 length, and somatic hypermutation (SHM).** Antibody CDR H3 and L3 length were identified with IMGT numbering scheme by ANARCI<sup>42</sup>. The sequences between the second conserved C and WGXX (where X is any residue) were counted as CDR H3 length. CDR L3 length was defined by the sequences between the conserved C and FGXX (where X is any residue). The SHM was examined with the IGBLAST webserver (<https://www.ncbi.nlm.nih.gov/igblast/>).

**Virus neutralization.** Single-round-of-replication Env pseudoviruses were prepared, titers were determined, and the pseudoviruses were used to infect TZMbl target in an optimized and qualified automated 384-well format<sup>43</sup>. Briefly, antibodies were serially diluted, a constant amount of pseudovirus added, and plates incubated for 60 min; followed by addition of TZMbl cells which express luciferase upon viral infection. The plates were incubated for 48 h and then lysed, and luciferase activity measured. Percent neutralization was determined by the equation: (virus only) - (virus+antibody)/(virus only) multiplied by 100. Data are expressed as the antibody concentration required to achieve 50% neutralization (IC<sub>50</sub>) and calculated using a dose-response curve fit with a 5-parameter nonlinear function (Supplementary Data 1). We used a previously described panel<sup>8,44-46</sup> of 208 geographically and genetically diverse Env pseudoviruses representing the major subtypes and circulating recombinant forms.

**Cryo-EM data collection and processing.** The CAP256.wk34.c80 SOSIP.RnS2 trimer with the N160K mutation complexed with 3BNC117 was incubated with molar excess of 2909 Fab fragments. A volume of 2.1 µl of the complex at 2 mg/ml concentration was deposited on a C-flat 1.2/1.3 grid (protochip.com) then vitrified using an FEI Vitrobot Mark IV with a wait time of 30 s, blot time of 3 s, and blot force of 1. Data collection was performed on a Titan Krios electron microscope with Leginon<sup>47</sup> using a Gatan K2 Summit direct detection device. Exposures were collected in movie mode for a 10 s with the total dose of 66.88 e<sup>-</sup>/Å<sup>2</sup> fractionated over 50 raw frames. Images were pre-processed using Appion<sup>48,49</sup>; individual



frames alignment and dose-weighted were executed using MotionCor2<sup>50</sup>. CTFFind4<sup>51,52</sup> was used to estimate the CTF and DoG Picker<sup>48,49</sup> was used to pick particles. RELION<sup>53</sup> was used for particle extraction and the particle stack was transferred to cryoSPARC. cryoSPARC 2.15<sup>54</sup> was used for 2D classifications, ab initio 3D reconstruction, homogeneous refinement, and nonuniform 3D refinement. Initial 3D reconstruction and final refinements were performed using C1 symmetry, confirming three 3BNC117 Fab molecules and one 2909 Fab per trimer. The final resolution of the C1 nonuniform refinement was 3.91 Å.

Coordinates from PDB IDs 6VRW<sup>12</sup> and 3PIQ<sup>14</sup> were used for initial fit to the reconstructed map. This was followed by simulated annealing and real-space refinement in Phenix<sup>55</sup> with the sharpened map from cryoSPARC and with a density modified map from Phenix Resolve<sup>56</sup> and then iteratively processed with manual fitting of the coordinates in Coot<sup>57</sup>. Geometry and map fitting were evaluated throughout the process using Molprobity<sup>58</sup> and EMRinger<sup>59</sup>. PyMOL ([www.pymol.org](http://www.pymol.org)) and ChimeraX<sup>60</sup> were used to generate figures.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Cryo-EM maps and fitted coordinates have been deposited with EMDB entry ID EMD-23589 and PDB entry ID 7LY9, respectively. More than 4500 antibody dataset is available at SabDab. Source data are provided with this paper.

## Code availability

The in-house program GLYCO is available at <https://github.com/myungjinlee/GLYCO>.

Received: 25 February 2021; Accepted: 29 September 2021;

Published online: 09 November 2021

## References

- Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20 (1996).
- Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
- Schatz, D. G., Oettinger, M. A. & Schlissel, M. S. V(D)J recombination: molecular biology and regulation. *Annu. Rev. Immunol.* **10**, 359–383 (1992).
- Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
- Victoria, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
- Wu, T. T. & Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **132**, 211–250 (1970).
- Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- Chuang, G. Y. et al. Structural survey of broadly neutralizing antibodies targeting the HIV-1 Env trimer delineates epitope categories and characteristics of recognition. *Structure* **27**, 196–206 (2019).
- Gorny, M. K. et al. Identification of a new quaternary neutralizing epitope on human immunodeficiency virus type 1 virus particles. *J. Virol.* **79**, 5232–5237 (2005).
- Honnen, W. J. et al. Type-specific epitopes targeted by monoclonal antibodies with exceptionally potent neutralizing activities for selected strains of human immunodeficiency virus type 1 map to a common region of the V2 domain of gp120 and differ only at single positions from the clade B consensus sequence. *J. Virol.* **81**, 1424–1432 (2007).
- Wu, X. et al. Immunotypes of a quaternary site of HIV-1 vulnerability and their recognition by antibodies. *J. Virol.* **85**, 4578–4585 (2011).
- Gorman, J. et al. Structure of super-potent antibody CAP256-VRC26.25 in complex with HIV-1 envelope reveals a combined mode of trimer-apex recognition. *Cell Rep.* **31**, 107488 (2020).
- McLellan, J. S. et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336–343 (2011).
- Changela, A. et al. Crystal structure of human antibody 2909 reveals conserved features of quaternary structure-specific antibodies that potently neutralize HIV-1. *J. Virol.* **85**, 2524–2535 (2011).
- Spurrier, B. et al. Structural analysis of human and macaque mAbs 2909 and 2.5B: implications for the configuration of the quaternary neutralizing epitope of HIV-1 gp120. *Structure* **19**, 691–699 (2011).
- Liu, Q. et al. Quaternary contact in the initial interaction of CD4 with the HIV-1 envelope trimer. *Nat. Struct. Mol. Biol.* **24**, 370–378 (2017).
- Lee, J. H. et al. A broadly neutralizing antibody targets the dynamic HIV envelope trimer apex via a long, rigidified, and anionic beta-hairpin structure. *Immunity* **46**, 690–702 (2017).
- Huang, C. C. et al. Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proc. Natl Acad. Sci. USA* **101**, 2706–2711 (2004).
- Struwe, W. B. et al. Site-specific glycosylation of virion-derived HIV-1 Env is mimicked by a soluble trimeric immunogen. *Cell Rep.* **24**, 1958 (2018).
- Pan, R. M. et al. Rabbit anti-HIV-1 monoclonal antibodies raised by immunization can mimic the antigen-binding modes of antibodies derived from HIV-1-infected humans. *J. Virol.* **87**, 10221–10231 (2013).
- Schief, W. R., Ban, Y. E. A. & Stamatatos, L. Challenges for structure-based HIV vaccine design. *Curr. Opin. HIV AIDS* **4**, 431–440 (2009).
- Crooks, E. T. et al. Effects of partially dismantling the CD4 binding site glycan fence of HIV-1 envelope glycoprotein trimers on neutralizing antibody induction. *Virology* **505**, 193–209 (2017).
- McCoy, L. E. et al. Holes in the glycan shield of the native HIV envelope are a target of trimer-elicited neutralizing antibodies. *Cell Rep.* **16**, 2327–2338 (2016).
- Wagh, K., Hahn, B. H. & Korber, B. Hitting the sweet spot: exploiting HIV-1 glycan shield for induction of broadly neutralizing antibodies. *Curr. Opin. HIV AIDS* **15**, 267–274 (2020).
- Dunbar, J. et al. SABDab: the structural antibody database. *Nucleic Acids Res.* **42**, D1140–D1146 (2014).
- Li, W. Z. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
- Fu, L. M., Niu, B. F., Zhu, Z. W., Wu, S. T. & Li, W. Z. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
- Schoofs, T. et al. Broad and potent neutralizing antibodies recognize the silent face of the HIV envelope. *Immunity* **50**, 1513 (2019).
- Schommers, P. et al. Restriction of HIV-1 escape by a highly broad and potent neutralizing antibody. *Cell* **180**, 471–489 e422 (2020).
- Mitternacht, S. FreeSASA: an open source C library for solvent accessible surface area calculations. *F1000Research* **5**, 189 (2016).
- Lemmin, T. & Soto, C. Glycosylator: a Python framework for the rapid modeling of glycans. *BMC Bioinformatics* **20**, 513 (2019).
- Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
- Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
- Guvench, O. et al. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **7**, 3162–3180 (2011).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- Di Pierro, M., Elber, R. & Leimkuhler, B. A stochastic algorithm for the isobaric-isothermal ensemble with Ewald summations for all long range forces. *J. Chem. Theory Comput.* **11**, 5624–5637 (2015).
- Andersen, H. C. Rattle—a velocity version of the shake algorithm for molecular-dynamics calculations. *J. Comput. Phys.* **52**, 24–34 (1983).
- Kwon, Y. D. et al. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat. Struct. Mol. Biol.* **22**, 522–531 (2015).
- Scheid, J. F. et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
- Trkola, A. et al. Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70**, 1100–1108 (1996).
- Kwon, Y. D. et al. Surface-matrix screening identifies semi-specific interactions that improve potency of a near pan-reactive HIV-1-neutralizing antibody. *Cell Rep.* **22**, 1798–1809 (2018).
- Dunbar, J. et al. SABPred: a structure-based antibody prediction server. *Nucleic Acids Res.* **44**, W474–W478 (2016).
- Sarzotti-Kelsoe, M. et al. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J. Immunol. Methods* **409**, 131–146 (2014).
- Cale, E. M. et al. Virus-like particles identify an HIV V1V2 apex-binding neutralizing antibody that lacks a protruding loop. *Immunity* **46**, 777–791 e710 (2017).
- Georgiev, I. S. et al. Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. *Science* **340**, 751–756 (2013).
- Wu, X. et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).

47. Suloway, C. et al. Automated molecular microscopy: the new Legino system. *J. Struct. Biol.* **151**, 41–60 (2005).
48. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
49. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
50. Zheng, S. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods.* **14**, 331–332 (2017).
51. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
52. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
53. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
54. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
55. Adams, P. D. et al. Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11**, 53–55 (2004).
56. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
57. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D. Biol. Crystallogr* **60**, 2126–2132 (2004).
58. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
59. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
60. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

## Acknowledgements

We thank C.-H. Shen for calculating antibody SHM, J. Stuckey for assistance with figures, and members of the Structural Biology Section and Structural Bioinformatics Core, Vaccine Research Center, for discussions or comments on the manuscript. We thank J. Baalwa, D. Ellenberger, F. Gao, B. Hahn, K. Hong, J. Kim, F. McCutchan, D. Montefiori, L. Morris, E. Sanders-Buell, G. Shaw, R. Swanstrom, M. Thomson, S. Tovanabutra, C. Williamson, and L. Zhang for contributing the HIV-1 envelope plasmids used in the 208-strain panel. We thank K. McKee, C. Moore, S. O'Dell, G. Padilla, S.D. Schmidt, C. Whittaker, and A.B. McDermott for assistance with neutralization assessments on the 208-strain panel. This work utilized the computational resources of the NIH HPC Biowulf cluster (NIH HPC). This study used the Office of Cyber Infrastructure and Computational Biology (OCICB) High Performance Computing (HPC) cluster at the National Institute of Allergy and Infectious Diseases (NIAID), Bethesda, MD. Support for this work was provided by the Intramural Research Program of the Vaccine Research Center, National Institute of Allergy and Infectious Diseases. Some of this work was performed at the Simons Electron Microscopy Center and National Resource for

Automated Molecular Microscopy located at the New York Structural Biology Center, supported by grants from the Simons Foundation (SF349247), NYSTAR, and the NIH National Institute of General Medical Sciences (GM103310).

## Author contributions

M.L. analyzed AFADs and glycan coverage and wrote the paper. A.C. and J.G. characterized antibody 2909 and wrote the paper. R.R. performed MD simulations. T.B. contributed AFAD analysis. A.S.O. produced trimer and performed Octet. C.W.C. prepared 2909 Fab. B.Z. contributed antibody expression. S.Z.-P. provided antibody 2909. B.C.L., M.K.L., and N.A.D.-R. provided neutralization from 208-HIV isolate panel. L.S. conceptualized problems. G.-Y.C. defined AFAD, conceptualized informatics, and wrote the paper. P.D.K. analyzed the data and wrote the paper, with all authors providing revisions.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26579-z>.

**Correspondence** and requests for materials should be addressed to Gwo-Yu Chuang or Peter D. Kwong.

**Peer review information** *Nature Communications* thanks Shuqun Liu, Matthew Raybould and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021