

# Contemporary HIV-1 consensus Env with AI-assisted redesigned hypervariable loops promote antibody binding

Received: 26 June 2023

Accepted: 19 April 2024

Published online: 09 May 2024

 Check for updates

Hongjun Bai<sup>1,2</sup>, Eric Lewitus<sup>1,2</sup>, Yifan Li<sup>1,2</sup>, Paul V. Thomas<sup>2,3</sup>, Michelle Zemil<sup>1,2</sup>, Mélanie Merbah<sup>1,2</sup>, Caroline E. Peterson<sup>2,3</sup>, Thujitha Thuraiamy<sup>1,2</sup>, Phyllis A. Rees<sup>2,3</sup>, Agnes Hajduczki<sup>2,3</sup>, Vincent Dussupt<sup>1,2</sup>, Bonnie Slike<sup>1,2</sup>, Letzibeth Mendez-Rivera<sup>1,2</sup>, Annika Schmid<sup>1,2</sup>, Erin Kavusak<sup>1,2</sup>, Mekhala Rao<sup>1,2</sup>, Gabriel Smith<sup>1,2</sup>, Jessica Frey<sup>1,2</sup>, Alicea Sims<sup>1,2</sup>, Lindsay Wieczorek<sup>1,2</sup>, Victoria Polonis<sup>1</sup>, Shelly J. Krebs<sup>1</sup>, Julie A. Ake<sup>1</sup>, Sandhya Vasani<sup>1,2</sup>, Diane L. Bolton<sup>1,2</sup>, M. Gordon Joyce<sup>2,3</sup>, Samantha Townsley<sup>1,2</sup> & Morgane Rolland<sup>1,2</sup> ✉

An effective HIV-1 vaccine must elicit broadly neutralizing antibodies (bnAbs) against highly diverse Envelope glycoproteins (Env). Since Env with the longest hypervariable (HV) loops is more resistant to the cognate bnAbs than Env with shorter HV loops, we redesigned hypervariable loops for updated Env consensus sequences of subtypes B and C and CRF01\_AE. Using modeling with AlphaFold2, we reduced the length of V1, V2, and V5 HV loops while maintaining the integrity of the Env structure and glycan shield, and modified the V4 HV loop. Spacers are designed to limit strain-specific targeting. All updated Env are infectious as pseudoviruses. Preliminary structural characterization suggests that the modified HV loops have a limited impact on Env's conformation. Binding assays show improved binding to modified subtype B and CRF01\_AE Env but not to subtype C Env. Neutralization assays show increases in sensitivity to bnAbs, although not always consistently across clades. Strikingly, the HV loop modification renders the resistant CRF01\_AE Env sensitive to 10-1074 despite the absence of a glycan at N332.

HIV-1 remains a global health priority. Vaccines have historically been proven to be the most powerful tool to protect against viral pathogens. Despite many setbacks in HIV-1 vaccine efficacy trials, the search for a safe and effective HIV-1 vaccine should continue. Several unique HIV-1 features have hampered the development of an HIV-1 vaccine. The protein that allows HIV-1 entry, Env, is metastable, heavily glycosylated, and highly sequence-diverse. To counteract HIV-1 diversity, broadly neutralizing antibodies (bnAbs) must be elicited by a vaccine to cross-react with a variety of Env<sup>1,2</sup>. The vaccine candidates tested in the nine

HIV-1 vaccine efficacy trials conducted to date did not yield bnAbs. Structural biology studies have revealed the atomic details of the epitopes recognized by numerous HIV-1 bnAbs (e.g. as reviewed<sup>3-5</sup>), providing a roadmap for designing antigens that could integrate these bnAb epitopes. Known HIV-1 bnAbs cover the Env surface<sup>6</sup> and are grouped in categories based on the location of their epitopes, including Env variable loop 2 (V2)-apex antibodies, glycan supersite antibodies, and CD4 binding site (CD4bs) antibodies<sup>3-5</sup>. Those bnAb epitopes are under functional constraints. For example, the V2-apex

<sup>1</sup>U.S. Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA. <sup>2</sup>Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD 20817, USA. <sup>3</sup>Emerging Infectious Disease Branch, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA. ✉e-mail: [mrolland@hivresearch.org](mailto:mrolland@hivresearch.org)

epitope shields the V3 loop and stabilizes the prefusion Env trimer<sup>7,8</sup> and the CD4bs epitope is where the host receptor binds to<sup>9</sup>.

HIV-1 Env presents five variable loops and four of them (V1, V2, V4, and V5) contain a hypervariable segment, called the hypervariable (HV) loop. These HV loops can be adjacent to epitope sites recognized by bnAbs. The V1 HV loop (VIHV) is near the glycan supersite epitope (target of 10-1074<sup>10</sup>, PGT135<sup>11,12</sup>, PGT128<sup>11,13</sup> and 2G12<sup>14</sup>), the V2 HV loop (V2HV) is near the V2-apex epitope (target of PGT145<sup>15-17</sup>, VRC26.25<sup>18</sup>, PG9<sup>19</sup> and PGDM1400<sup>15</sup>), and the V5 HV loop (V5HV) is near the CD4bs epitope (target of VRC01<sup>20</sup> and 3BNC117<sup>21</sup>). These HV loops can occlude access to the adjacent bnAb epitopes and can be targeted by antibodies to elicit strain-specific responses rather than the more functional bnAb specificities they are shielding. Antibodies targeting these three epitopes accounted for more than half of the neutralizing antibodies in two natural infection cohorts: they corresponded to 57% of dominant specificities of the top 42 neutralizers in the IAVI Protocol C<sup>22</sup> and to 65% (33/51) of delineated epitope-specificities in the RV217 cohort<sup>23</sup>. Thus, modifying hypervariable loops to optimize access to adjacent bnAb epitopes on Env could improve HIV-1 Env vaccine candidates.

Previous studies analyzed HIV-1 antigens in which either V1, V2 or both loops were removed<sup>24-26</sup>. These variable loop deletions refocused immune responses, yet they also revealed that the removal of entire loops can lead to protein misfolding. Other studies used natural sequences with short HV loops, especially shorter VIHV, as candidate antigens. Zolla-Pazner and collaborators tested VIV2 from strain ZM53, which had a short V1, on scaffold proteins as vaccine antigens<sup>27,28</sup>. We selected a Circulating Recombinant Form (CRF) 01\_AE strain with short VIHV and V2HV to be loaded on a self-assembling nanoparticle, VIV2-SHB-SAPN<sup>29</sup>. However, a natural sequence does not necessarily have ideal bnAb epitopes as well as ideal HV loops for eliciting bnAbs.

In this work, we systematically redesigned HV loops to optimize access to bnAb epitopes and overcome limitations associated with natural HIV-1 sequences. Our goals were to reduce the shielding of bnAb epitopes caused by HV loops and to limit the targeting of HV loops by strain-specific antibodies. Publicly available circulating subtype B and C and CRF01\_AE sequences were analyzed to infer naturally occurring loop lengths and AlphaFold2<sup>30</sup> was used to predict protein structure to maintain structural integrity, loop anchoring, and the glycan shield. Env HV loops were redesigned for updated consensus sequences of subtypes B and C and CRF01\_AE<sup>31</sup> that we recently created using publicly available<sup>32</sup> independent sequences sampled only since 2010. We used consensus sequences because they are a better representation of HIV-1 diversity than any natural sequence and are thus optimal as vaccine candidates<sup>33-36</sup>. We hypothesize that subtype consensus sequences with redesigned HV loops reduce the shielding of bnAb epitopes caused by HV loops limit the targeting of HV loops by strain-specific antibodies and show improved antibody binding for Env with redesigned HV loops.

## Results

### Short HV loops were rare for subtypes B, C, and CRF01\_AE

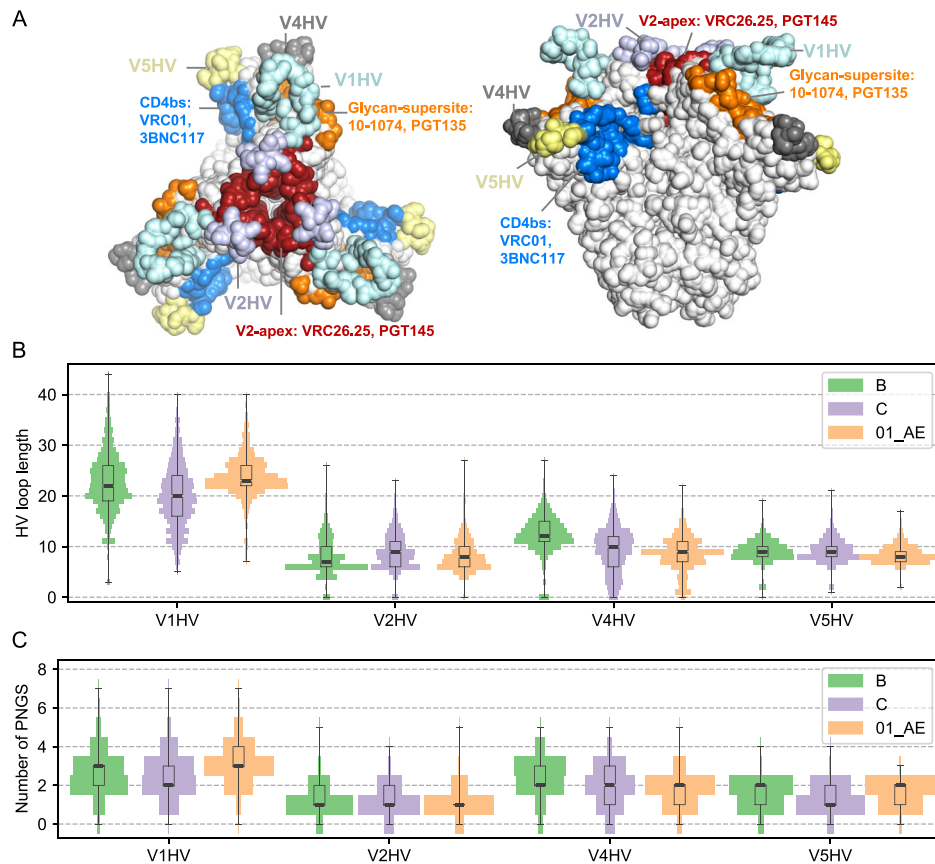
We curated publicly available circulating HIV-1 sequences to obtain a dataset of 4847 independent Env sequences sampled from people living with HIV-1 (PLWH) between 1979 and 2020<sup>31</sup>. We focused on subtype B ( $n = 2495$ ), subtype C ( $n = 1503$ ) and CRF01\_AE ( $n = 849$ ) (Table S1) sequences since these subtypes/CRF are most frequently sequenced, thereby enabling robust analyses. We analyzed the distribution of HV loop lengths and the number of potential N-linked glycosylation sites (PNGS) in four HV loops (Fig. 1, Tables S1 and S2). We focused on the HV loops, which correspond to a portion of the variable loops: VIHV, sites 132-152 within sites 131-157 (V1); V2HV, sites 185-190 within sites 157-196 (V2); V4HV, sites 395-412 within sites 385-418 (V4); V5HV, sites 460-467 within sites 460-470 (V5). V3 was omitted because it does not contain an HV segment. The location of the HV

loops, as well as adjacent bnAb epitopes, are shown on the structure in Fig. 1A.

VIHV loops showed a median length of 20 to 23 amino acids (AA), which was about twice as long as V2HV (7-9 AA), V4HV (9-12 AA), and V5HV (8-9 AA) loops (Fig. 1B, Table S1). The length distributions across subtypes/CRF were similar for V2HV and V5HV, while there were differences for VIHV and V4HV. CRF01\_AE showed fewer short VIHV loops than subtype B or subtype C: there were 8% of CRF01\_AE Env with VIHV  $\leq 20$  AA compared to 26% and 49% for B and C Env, respectively. Subtype B had fewer short V4HV loops than subtype C and CRF01\_AE: there were 13% of subtype B Env with V4HV  $\leq 10$  AA compared to 49% and 65% for C and CRF01\_AE Env, respectively. Similar patterns were observed for the number of PNGS on HV loops across subtypes/CRF (Fig. 1C, Table S2). Sequences with V1, V4, and V5 HV loops that were shorter than or equal to the 10th percentile were very rare representing 0.73%, 0.53%, and 0.32% of the dataset for subtypes B, C, and CRF01\_AE, respectively. We also determined the loop lengths of sequences that have been used as inserts in previous vaccine efficacy trials or that are under consideration for future trials. We found that these insert sequences usually had one or more long HV loops (Table 1)<sup>37-46</sup>. For example, the subtype C TVI sequence which was included in the vaccine tested in the HVTN702<sup>45</sup> vaccine efficacy trial had a VIHV loop that was 33 AA long while the median VIHV for subtype C is 20 AA. The insert Mosaic2.Env<sup>40</sup> used in the Mosaico and Imbokodo vaccine efficacy trials had a V2HV that was twice as long (16 AA) as the median for subtypes B and C (7 and 9 AA, respectively). We identified some exceptions with short HV loops: the subtype A1 92RW020<sup>41</sup> insert that was part of the trivalent vaccine tested in the HVTN505 trial and the subtype C CH505<sup>44</sup> insert that has been undergoing early clinical testing.

### Long HV loops associated with decreased neutralization sensitivity to bnAbs

The neutralization sensitivity to bnAbs was compared for HIV-1 Env with the shortest ( $\leq 5$  percentile) and longest ( $\geq 95$  percentile) HV loops. All neutralization and corresponding sequence data were obtained from CATNAP<sup>47</sup>. Six bnAbs belonging to three representative epitope groups were compared: V2-apex, PGT145 and VRC26.25; glycan supersite, 10-1074 and PGT135; and CD4bs, VRC01 and 3BNC117 (Fig. 2). To evaluate the effect of loop length on bnAb sensitivity, we compared short and long HV loops for the HV loop proximal to each bnAb (V2HV to V2-apex antibodies, VIHV to glycan supersite antibodies, and V5HV to CD4bs antibodies) and for the three other HV loops. The virus sequences with the longest adjacent HV loops had significantly higher half maximal inhibitory concentration (IC50) than the sequences with the shortest HV loops ( $p$ -value  $\leq 0.015$ , one-tailed Mann-Whitney U test) for the six bnAbs (Fig. 2), suggesting that longer HV loops obstruct bnAb epitope availability. When HV loops distant from the bnAb epitope (V5HV to V2-apex and glycan supersite antibodies and V2HV to CD4bs antibodies) were compared, there was no significant difference in sensitivity for short or long HV loops for five of the six bnAbs tested ( $p$ -value  $\geq 0.207$ ). Interestingly, there might be some longer-range interactions between HV loops. Sequences with long/short VIHV had a similar impact on IC50 values for V2-apex antibodies as V2HV, but there is no association between the length of VIHV and V2HV (Figure. S1), suggesting that interactions between VIHV and V2HV may also modulate the accessibility of V2-apex epitopes. For VRC01, sequences with shorter HV loops for V1, V2, and V5 were significantly associated with increased sensitivity when compared to sequences with longer HV loops (Fig. 2E). The strongest impact was seen for VIHV loops and the neutralization sensitivity to 10-1074. HIV-1 Envs with the shortest VIHV loops had a median IC50 greater than or equal to 5000 times lower than the strains with the longest VIHV loops (0.02  $\mu\text{g/ml}$  vs.  $\geq 100 \mu\text{g/ml}$ , one-tailed Mann-Whitney U test  $p$ -value = 0.00003). For V5HV loops, which are distant



**Fig. 1 | Characteristics of hypervariable (HV) loops.** **A** Location of HV loops and related HIV-1 bnAb epitopes on Env prefusion structure. **B** Distribution of lengths for V1, V2, V4, and V5 HV loops and **C** distribution of the number of potential N-linked glycosylation sites (PNGS) for subtypes B ( $n = 2495$ ) and C ( $n = 1503$ ) and CRF01\_AE ( $n = 849$ ) Env sequences. The width corresponds to the proportion of

sequences in the subtype/CRF sample of sequences and is normalized for each HV loop. The box plots depict the median and 25th/75th percentiles of the distribution, with the minima and maxima shown as whiskers. Source data for panels **B** and **C** are provided in the Source Data files.

from the 10-1074 epitope, there was no significant difference in sensitivity when short and long loops were compared (one-tailed Mann-Whitney U test  $p$ -value = 0.838). Importantly, these data masked some subtype-specific differences, as differences between short and long HV loops were not seen for subtype B (with  $p$ -values  $\geq 0.139$ ) except for 10-1074 ( $p = 3.6e-07$ ) (Figure. S2). Differences for subtype C were generally statistically significant, except for 10-1074 ( $p = 0.129$ ) and PGT135 ( $p = 0.128$ ). We repeated the analysis using 10–25% as the threshold to compare the extremes of the distributions (Figure. S3). Differences between short/long adjacent HV loops remained statistically significant for most bnAbs ( $p < 0.045$ , Figure. S3C-F) or borderline significant for VRC26.25 when the cutoff was set as 25% ( $p = 0.079$ , Figure. S3B), with PGT145 as an exception ( $p \geq 0.101$  with 10–25% as the threshold, Figure. S3A). Overall, these results indicated that a long HV loop can diminish the virus's sensitivity to a bnAb.

### Strategy to redesign HV loops

We redesigned the HV loops on Env to improve accessibility to typical bnAb epitopes. We used Env consensus sequences which were derived only from sequences that were sampled after 2010, thus reflecting currently circulating viruses<sup>31</sup>. Our goals were to: (1) make the loop as short as possible while maintaining structural integrity; (2) preserve the loop anchor sites, which are the relatively conserved start and end of the HV loops that form favorable interactions with the non-HV part of Env; and (3) maintain the glycan shield by ensuring there was at least

one PNGS on each redesigned loop. Using the V1HV loop of subtype B as an example, we analyzed the alignment of circulating sequences (Fig. 3A) and the AlphaFold2 predicted structure of the corresponding consensus sequence (Fig. 3B). We identified the relatively conserved TDL motif (sites 132-134) as N-anchor sites and the relatively conserved MEKG motif (sites 149-152) as C-anchor sites, with all sites in between considered spacer sites. In the TDL motif, D133 may form favorable charge-charge interactions with K151 in the C-anchor sites or with K155; L134 may form hydrophobic interactions with I154, L175, I323, and I326. In the MEKG motif, M149 may form hydrophobic interactions with I326 and I154; E150 may form a favorable charge-charge interaction with R419. These anchor sites were retained since they did not have direct contact with glycan supersite antibodies. To shorten the spacer, we evaluated different lengths iteratively using AlphaFold2 predicted structures by comparing the predicted local Difference Distance Test (IDDT) scores, which measure the confidence given by AlphaFold2 for each output structure. A specific spacer was selected when it maintained or increased the predicted IDDT for the anchor and surrounding sites of the HV loop, suggesting that the anchor stability will be maintained. For V1HV, the spacer was shortened from 15 AA to 5 AAs. This resulted in the V1HV loop length changing from 22 AA (50<sup>th</sup> percentile in subtype B sequences) to 12 AA (4<sup>th</sup> percentile in subtype B sequences) (Fig. 3B, C). The V1HV, V2HV, and V5HV loops of subtypes B, C, and CRF01\_AE were redesigned with the same strategy and are shown in Table 2 and Figure. S4–S11.

**Table 1 | Comparison of redesigned HV loops to those found in vaccine candidates**

Sequence	Subtype/CRF	Efficacy trials	V1HV		V2HV		V4HV		V5HV	
			N <sub>AA</sub> <sup>a</sup>	N <sub>PNGS</sub>	N <sub>AA</sub> <sup>a</sup>	N <sub>PNGS</sub>	N <sub>AA</sub> <sup>a</sup>	N <sub>PNGS</sub>	N <sub>AA</sub> <sup>a</sup>	N <sub>PNGS</sub>
A244	CRF01_AE	VAX003, RV144	28 (90%)	5	8 (62%)	1	9 (65%)	1	9 (75%)	2
MN	B	VAX003, VAX004, RV144	27 (81%)	3	6 (45%)	1	12 (52%)	2	10 (78%)	1
GENE8	B	VAX004	20 (36%)	2	6 (45%)	1	15 (83%)	3	7 (12%)	1
92TH023	CRF01_AE	RV144	23 (52%)	3	8 (62%)	1	9 (65%)	2	7 (27%)	2
92RW020	A1	HVTN505	9 (2%)	2	8 (61%)	1	12 (67%)	3	7 (16%)	1
HXB2	B	HVTN505	21 (43%)	2	6 (45%)	1	15 (83%)	2	8 (36%)	1
97ZA012	C	HVTN505	15 (23%)	2	11 (76%)	2	2 (9%)	0	8 (44%)	0
96ZM651	C	HVTN702	26 (86%)	4	14 (93%)	2	6 (27%)	1	12 (92%)	2
TV1	C	HVTN702	33 (98%)	4	6 (25%)	0	13 (87%)	2	9 (60%)	2
1086	C	HVTN702	12 (12%)	1	9 (62%)	1	12 (79%)	3	10 (76%)	2
Mosaic1.Env	–	HVTN705, HVTN706	25 (74%)	3	8 (61%)	2	16 (93%)	3	8 (43%)	1
Mosaic2.Env	–	HVTN705, HVTN706	21 (45%)	3	16 (97%)	2	9 (33%)	2	11 (89%)	1
CH505	C	–	12 (12%)	2	6 (25%)	1	15 (94%)	3	6 (6%)	1
WITO	B	–	19 (26%)	3	8 (68%)	1	11 (37%)	2	8 (36%)	2
DU422	C	–	21 (62%)	3	10 (70%)	1	9 (49%)	2	7 (18%)	1
BG505	A1	–	13 (8%)	2	14 (94%)	2	14 (83%)	3	8 (43%)	1
JR-FL	B	–	19 (26%)	3	6 (45%)	1	11 (37%)	2	8 (36%)	1
B_2010s <sup>b</sup>	B	–	12 (4%)	1	4 (11%)	1	11 (37%)	3	7 (12%)	2
C_2010s <sup>b</sup>	C	–	13 (16%)	1	5 (8%)	1	11 (69%)	3	7 (18%)	1
O1_AE_2010s.1 <sup>b</sup>	CRF01_AE	–	9 (1%)	1	5 (12%)	1	11 (84%)	3	7 (27%)	2
O1_AE_2010s.2 <sup>b</sup>	CRF01_AE	–	19 (8%)	1	5 (12%)	1	11 (84%)	3	7 (27%)	2

<sup>a</sup>HV loop length with the percentile for this loop length in corresponding subtypes shown in parenthesis, except for Mosaic1.Env, Mosaic2.Env, and subtype A1 sequences for which a pooled set of subtype B, C, and CRF01\_AE sequences was used to calculate the percentile for the specific loop length.

<sup>b</sup>Consensus sequences with redesigned HV loops. O1\_AE\_2010s.2 includes the 'IGNITD' segment in the V1HV and is 10 AA longer in V1HV than O1\_AE\_2010s.1.

Contrary to the strategy for V1HV, V2HV, and V5HV, for V4HV loops we did not seek to obtain the shortest V4HV loops in order to accommodate its specific structural context. Sites adjacent to V4HV (non-HV sites within 10 Å of the HV loops, relative surface area > 0.30) were diverse, and no major bnAbs are known to target this segment of Env. For subtype C, the median Shannon entropy of V4HV adjacent sites was 2.4 bits compared to 0.8, 0.7 and 1.1 bits, respectively, for the sites adjacent to V1HV, V2HV and V5HV (Fig. 4A). This high Shannon entropy for V4HV loops indicated that there were 4.5 ( $2^{2.4/1.1}$ ) to 10.8 ( $2^{2.4/0.7}$ ) times more AA possibilities on average for surface sites adjacent to V4HV compared to the other HV loops. In addition, the distance encompassing spacer sites was larger for V4HV than for V1HV, V2HV, and V5HV (20.8 Å for V4HV versus 10.8, 7.9, and 9.4 Å for V1HV, V2HV, and V5HV, respectively, based on the AlphaFold2 model of the subtype B consensus). This suggests that a short V4HV spacer may not be compatible with the structure. Hence, we chose a medium length for the V4HV loop; this loop length could also partially shield the highly diverse segment adjacent to the V4HV. For the subtype B consensus, we identified W395 as the N-terminal anchor site, as it is well conserved and forms hydrophobic interaction with nearby non-HV sites. We also assigned ND (sites 411–412) as the C-terminal anchor sites, as N is the most frequent option and D/E412 might form a charge-charge interaction with R335. The spacer was replaced with an 11-AA linker (Fig. 4B–D). The redesigned linker contained two PNGS, which was the median number for V4HV of subtypes B and C, and CRF01\_AE sequences (Fig. 1C). The V4HV loops of subtype C and CRF01\_AE were redesigned with the same strategy (Figure. S12, S13). One notable feature of the V4HV redesign for CRF01\_AE is that C395 (which was predicted to form a disulfide bond in our sequence) was replaced with the second most frequent residue at that site, W, and the spacer was extended to better shield the surrounding sites of V4HV (Figure. S13). Across all redesigned HV loops, the spacer consisted mainly of glycine

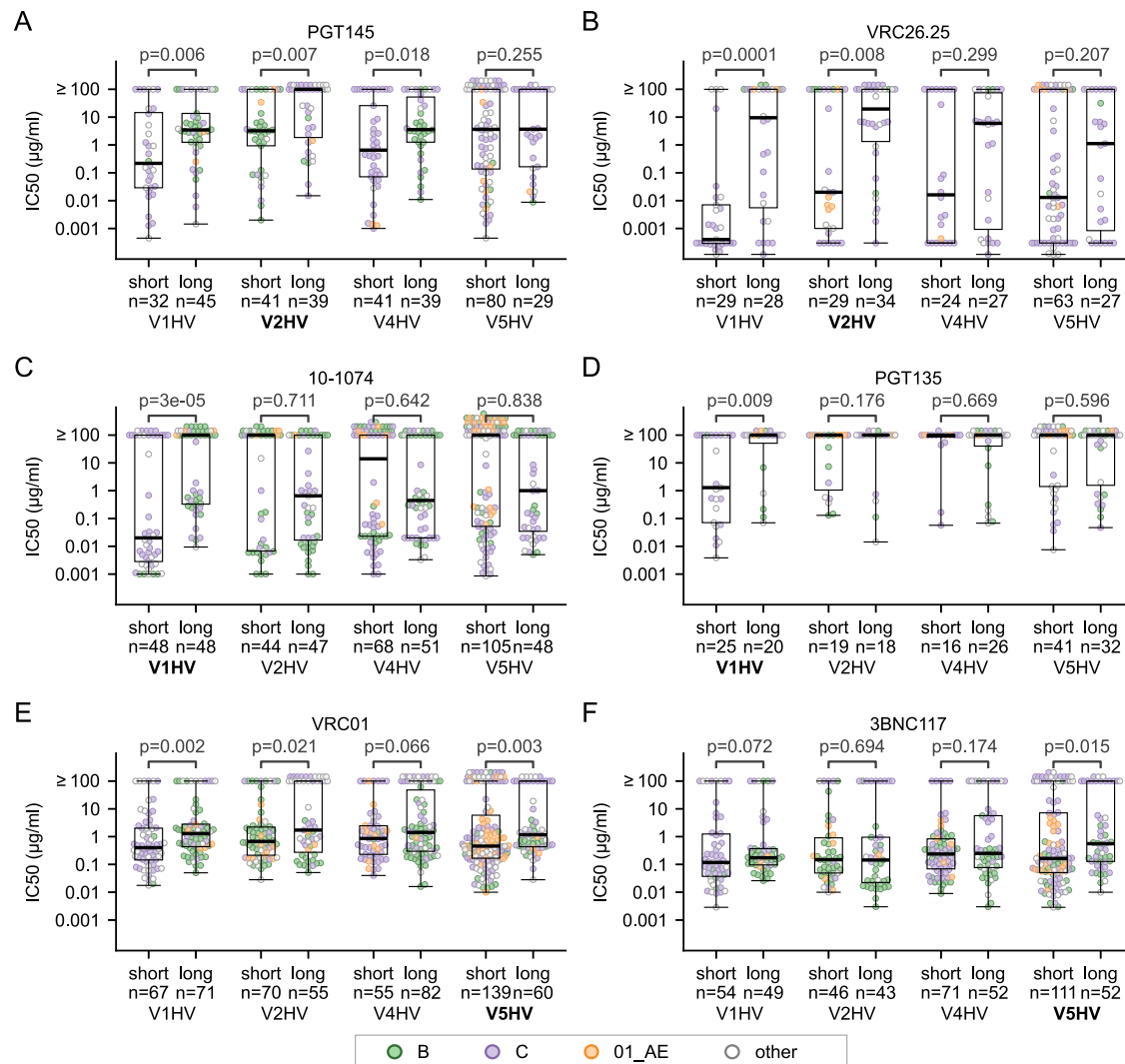
and serine which lack the side chains that are preferentially targeted by antibodies and may also increase the flexibility of the HV loops. The redesigned HV loops tended to have higher predicted IDDT scores from AlphaFold2 than the unmodified ones, except for the V4HV of the subtype C consensus (Figure. S14). All redesigned HV loops are listed in Table 2.

#### Updated Env consensus sequences are functional Env proteins

We produced pseudoviruses that expressed Env proteins for subtypes B, C, and CRF01\_AE with either modified or unmodified loops. We showed that the seven Env tested were infectious. When normalized by the Gag-p24 quantities, viral titers were lower for the pseudoviruses with redesigned Env loops compared to those with unmodified loops for subtypes B and C, while it was the opposite for CRF01\_AE (Fig. 5A).

Next, we produced our consensus sequences as gp140 glycoproteins. Negative-stain electron-microscopy (EM) showed that gp140s for subtypes B, C, and CRF01\_AE were mostly open trimers and monomers, with only monomers being averageable from negative stain EM (Figure. S15).

Finally, we stabilized Env in the prefusion conformation by combining previously reported modifications<sup>48,49</sup> and three extra mutations. As stabilized trimers, produced by transient transfection and purified by GNA-lectin and size-exclusion chromatography, consensus B and C Env were just under 50% trimer by 2D averaging of negative-stain EM data (42 and 45%, respectively); the HV loop redesign marginally increased the population of trimer observed to 59 and 55%, respectively (Fig. 5B). To determine whether the HV loop redesign affected the opening of the trimer, we used 3D classification and refinement without any applied symmetry. For the subtype B consensus, 76% of trimers classified with the reconstruction identifiable as the closed prefusion state, while 24% appeared open, but still in a prefusion state (Fig. 5C). Results were similar for the HV loop redesign



**Fig. 2 | Comparison of the bnAb sensitivity for HIV-1 Env with short ( $\leq 5$  percentile) or long ( $\geq 95$  percentile) hypervariable (HV) loops.** For each panel, comparisons of IC<sub>50</sub> values for short and long HV loops are shown for representative antibodies: PGT145 (**A**) and VRC26.25 (**B**), two V2 apex-targeting antibodies whose epitopes are close to V2HV loops; 10-1074 (**C**) and the PGT135 (**D**), two glycan supersite antibodies whose epitopes are close to V1HV loops; VRC01 (**E**) and 3BNC117 (**F**), two CD4 binding site antibodies with epitopes close to the V5HV

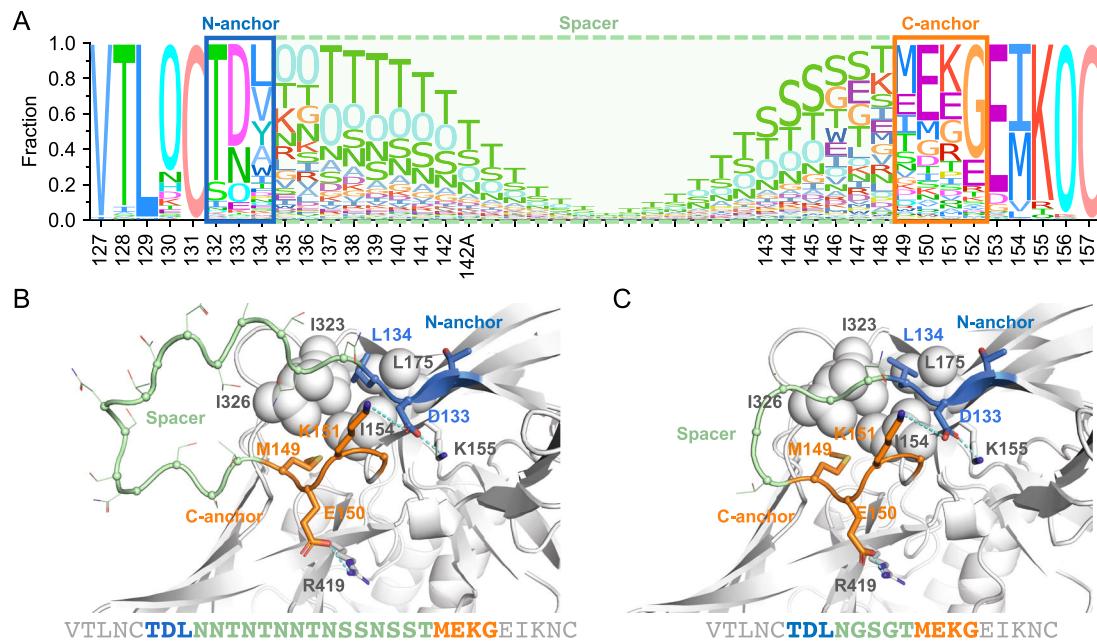
loops. The HV loop nearest to the epitope of each antibody is shown in bold font. The box plots depicted the median and 25th/75th percentiles of the distribution, with the minima and maxima as whiskers. The *p*-values from non-paired, non-parametric one-tailed Mann-Whitney U tests are provided. IC<sub>50</sub> values equal to or greater than 100  $\mu\text{g/ml}$  are grouped (upper limit of antibody neutralization measurements in the CATNAP database). Source data for all panels are provided in the Source Data files.

with 68% in the closed and 32% in the open prefusion state. In contrast, the subtype C consensus trimers were only at 31% in the closed prefusion state with 69% appearing open, with minimal changes with the HV loop redesign (29% in the closed and 71% in the open prefusion state). We cannot rule out that the conformational landscape of stabilized trimers is affected by the protein production and purification process as well as the adsorption to the carbon grid, which may explain the substantial fraction of monomer observed by negative-stain EM. Overall, these results suggest that the HV loop redesign marginally increased the trimer yield and had limited impact on the conformation of Env.

### Redesigned HV loops improved accessibility to bnAb epitopes in silico

To measure in silico if the redesign improved exposure of bnAb epitopes, we calculated a depth index that corresponds to the closest distance between an epitope site and the antibody probe. The antibody probe is a sphere with a radius corresponding to an antibody's Fv ( $\approx 30$  Å) that rolls around the Env structure (Fig. 6A). A short depth

indicates that the site is more easily accessible to the antibody. We calculated the depth corresponding to the epitopes of two glycan supersite bnAbs, 10-1074 and PGT135, and two CD4bs bnAbs, VRC01 and 3BNC117, before and after the HV loop redesign. We remodeled each loop 500 times via the kinematic closure (KIC) loop modeling protocol<sup>50</sup> from PyRosetta<sup>51</sup>. Among the 500 models, the 100 models with the lowest Rosetta energy were selected for the epitope depth calculation. The epitope depth decreased for the four bnAb epitopes of the Env consensus of subtypes B, C, and CRF01 AE (Fig. 6B–D). The median depth decreased by 0.8 to 1.6 Å for glycan supersite antibody epitope, while the median depth decreased by 0.2 to 0.8 Å for CD4bs antibody epitope. Besides the depth, the clashes between HV loops and corresponding antibodies were also calculated to show the impact of the HV loop redesign on the antibody accessibility of bnAb epitopes. The redesigned Env alleviated the clashes between 10-1074 and V1HV, but with limited effect on the clashes between CD4bs antibodies and V5HV (Fig. 6E, F). The greater depth and clash reduction for glycan-supersite epitope sites near V1HV was likely due to the fact that V1HV loops were longer than other HV loops and the HV loop



**Fig. 3 | Redesigned V1 hypervariable (HV) loop of the subtype B Env consensus sequence.** **A** Sequence logo of 2495 subtype B sequences around the V1HV loop, with the N-/C-anchors and spacer indicated by boxes. The height of each amino acid letter represents its frequency at the position. The letter ‘O’ indicates a potential N-linked glycosylation site. The consensus B Env with unmodified V1HV loop (**B**)

and redesigned V1HV loop (**C**) modeled by AlphaFold2 are shown with the N-anchor, C-anchor, and spacer sites colored blue, orange, and light green, respectively. Non-HV residues that interact with anchor sites are shown as spheres (hydrophobic interactions) or sticks (charge-charge interactions). Source data for panel **A** are provided in the Source Data file.

**Table 2 | Sequences of redesigned HV loops**

Variable loop	Subtype/CRF	Original sequence	Redesigned
V1HV	B	TDLNNTNTNNTNSSNSSTMEKG	TDLNGSGTMEKG
V1HV	C	TNVNNTNTNTNNTNNTNIKE	TNVNGSGNDTMKE
V1HV	CRF01_AE	TNANLNTNTNTNNTNNTNIGNITD	TNVNGSGGD (1) TNANLTGNGSGGSIGNITD (2)
V2HV	B	DNNNNTS	DNTS
V2HV	C	DNNNSSE	NNGSE
V2HV	CRF01_AE	DNNNSSE	GNSSE
V4HV	B	WNNNSTNTNNTNTNDT	WNGNGSGNGSGSNDT
V4HV	C	YNGNGNNTNNSNST	YNGNGSGNGSGSNST
V4HV	CRF01_AE	CIGNETTEGCNGT	WNGNGSGNGSGSNGT
V5HV	B	NNNTTET	NNGTET
V5HV	C	NNNNTTET	NNGTET
V5HV	CRF01_AE	NNNTNET	NNGTNET

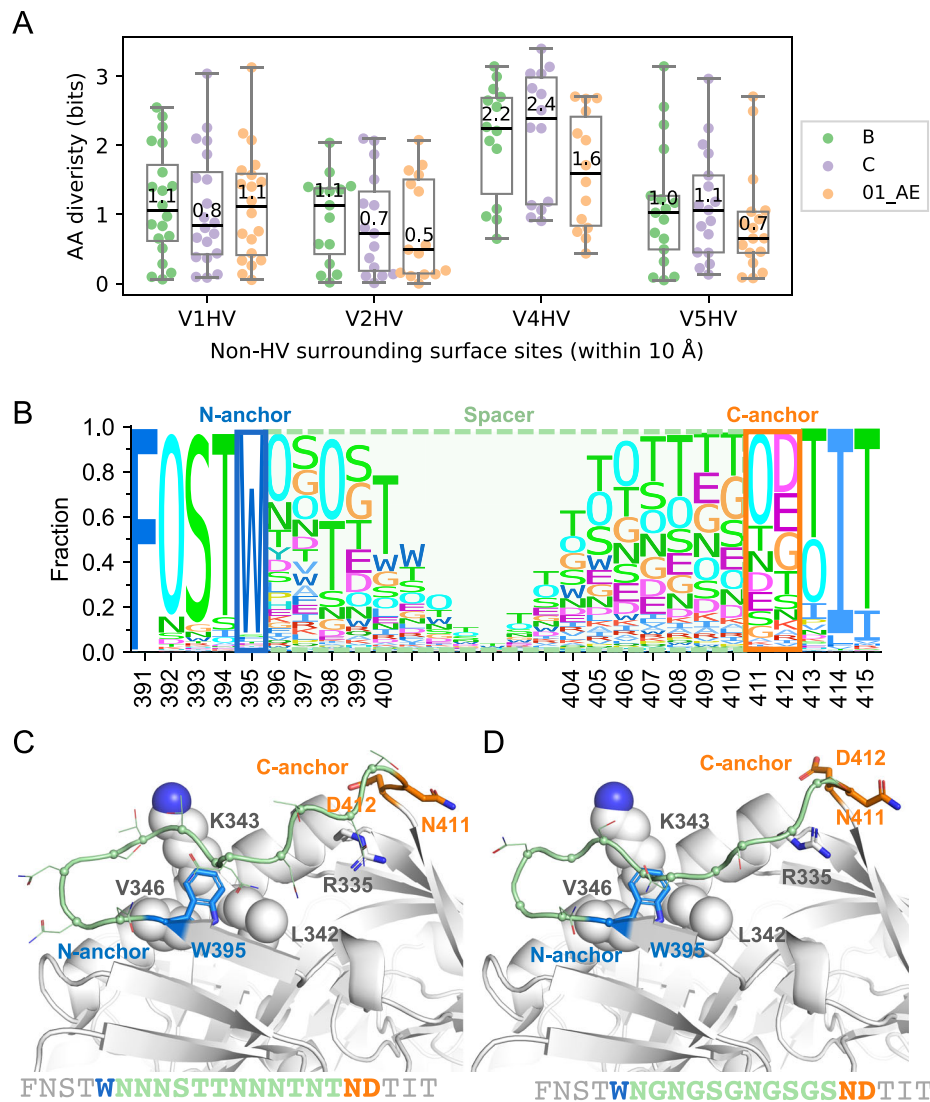
redesign removed more spacer residues than for other HV loops (Fig. 1A, Table 2). The apex antibodies, PGT145 and VRC26.25, were not tested because their epitopes are quaternary and require modeling of the entire Env trimer, which is currently beyond our computational power.

We further examined the modeled structure of CRF01\_AE V1HV because CRF01\_AE Env has longer V1HV than those of subtypes B and C (Fig. 1B) and shows high resistance to glycan supersite antibodies<sup>52</sup>. The consensus sequence had a ‘IGNITD’ motif (147-152) as its V1HV C-anchor sites (Figure. S5AB). AlphaFold2 predicted that I147 occupied the hydrophobic pocket formed by R327, K328, Y330 and P417, which corresponds to the site targeted by F<sup>100G</sup> on the CDR-H3 of 10-1074 (Fig. 6E). After the HV loops redesign, the 24 AA-long V1HV loop (58.2 percentile in CRF01\_AE) was replaced with a 9 AA-long loop (0.6 percentile in CRF01\_AE, 01\_AE.1) (Figure. S5C), hence, the median depth of 10-1074 epitope sites was reduced from 6.3 to 4.8 Å. The redesigned Env reduced the average number of clashes with 10-1074 from 120.8 to

9.2 atoms (Fig. 6E). A similar decrease in the number of clashes was observed for another glycan supersite bnAb, PGT135, but the impact was limited for CD4bs antibodies (VRC01 and 3BNC117) (Fig. 6F). Because the ‘IGNITD’ motif is highly conserved in CRF01\_AE viruses, another version of V1HV (01\_AE.2) was designed to preserve the ‘IGNITD’ motif, which may be important for CRF01\_AE specific bnAbs (Figure. S5D). This variant CRF01\_AE.2 design also improved antibody accessibility for glycan super site antibodies albeit to a lesser extent (Fig. 6F and S16).

### Improved antibody binding for Env-gp140s with redesigned HV loops

To test if the HV loop redesign improves antibody binding to Env, we analyzed binding to the Env-gp140 glycoproteins. Both versions of the consensus subtype B and C, plus the unmodified CRF01\_AE consensus and the redesigned V1HV with the conserved ‘IGNITD’ motif (01\_AE.2) were tested for binding to plasma pools and monoclonal antibodies



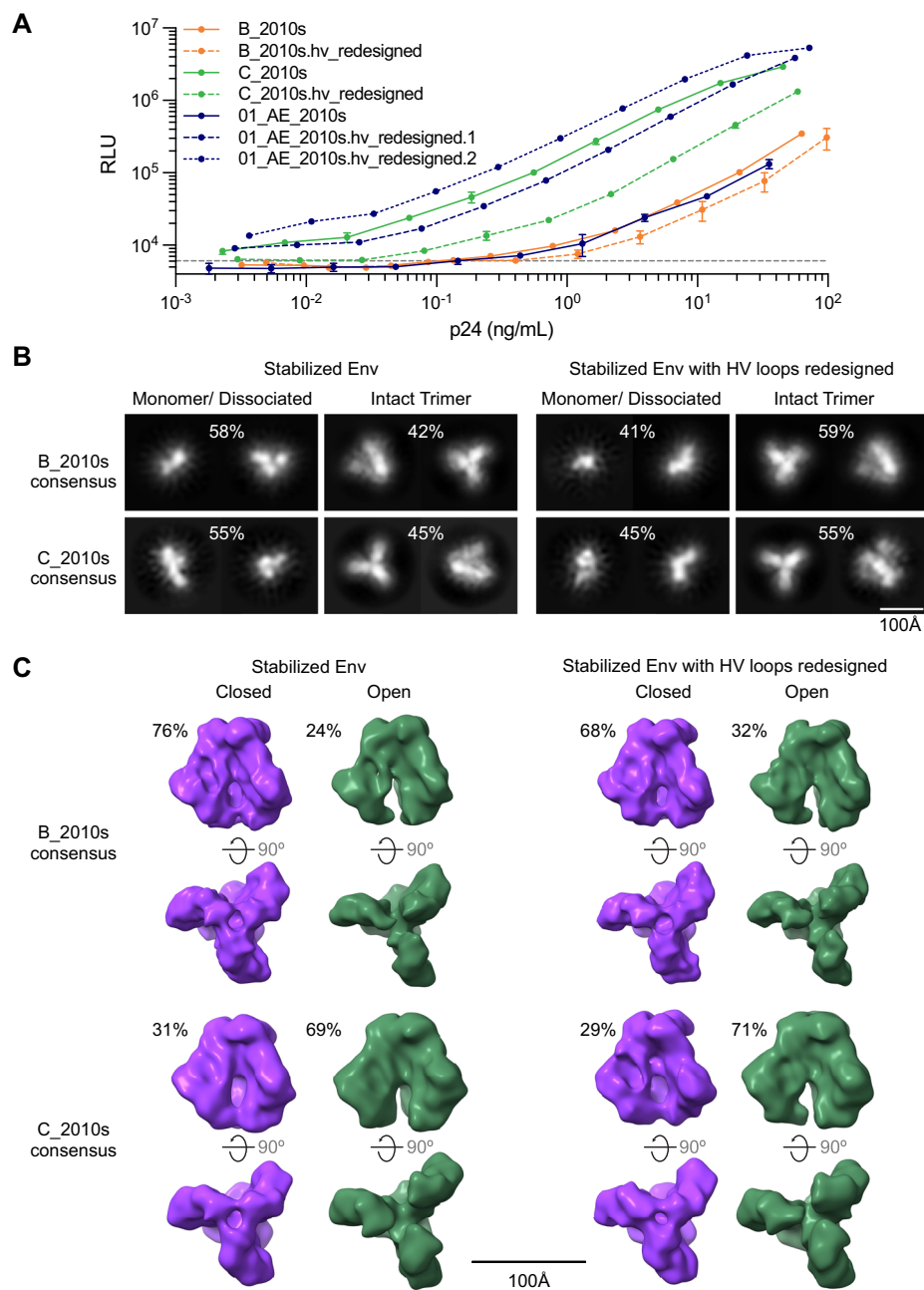
**Fig. 4 | Redesign of V4 hypervariable (HV) loop of the subtype B Env consensus sequence.** **A** The Shannon entropy of non-HV surrounding surface sites is shown for V1HV ( $n=20$ ), V2HV ( $n=15$ ), V4HV ( $n=14$ ) and V5HV ( $n=17$ ) of each subtype/CRF, with median values reported. The box plots depict the median and 25th/75th percentiles of the distribution, with the minima and maxima shown as whiskers. **(B)** Sequence logo of 2495 subtype B viruses around the V4HV loop, with the N-/C-anchor and spacer labeled. The height of each amino acid letter represents its

frequency at the position. The letter 'O' represents potential N-linked glycosylation sites. The consensus B Env with unmodified V4HV (**C**) and redesigned V4HV loop (**D**) modeled by AlphaFold2 are shown with the N-anchor, C-anchor, and spacer sites colored blue, orange, and light green, respectively. Non-HV residues that interact with anchor sites are shown as spheres (hydrophobic interactions) or sticks (charge-charge interactions). Source data for panels (**A**) and (**B**) are provided in the Source Data files.

with a bead-based multiplex immunoassay<sup>53,54</sup>). The pooled plasma samples represented 13 cohorts of PLWH spanning diverse geographic areas and subtype/CRF distributions. Plasma antibodies across multiple clade-specific pools recognized the consensus B Env with redesigned HV loops at a slightly higher magnitude of binding compared to the Env with unmodified HV loops (Fig. 7A). Similarly, monoclonal antibodies recognized the subtype B Env-gp140s with redesigned loops at a higher magnitude of binding, especially for antibodies targeting the CD4bs, V2 apex and glycan supersite epitopes (Fig. 7B). Summary measures for binding of pooled plasma or monoclonal antibodies demonstrated minor but significantly higher recognition of the subtype B Env-gp140s with redesigned HV loops when compared to the Env with unmodified loops ( $p \leq 0.008$ ) (Fig. 7C, D). A similar improvement of antibody binding was seen for the CRF01\_AE consensus with redesigned HV loops ( $p = 0.008$ ) (Fig. 7C, D, Figure. S17). The subtype C consensus proteins with or without redesigned loops showed no difference in antibody binding (Fig. 7C, D, Figure. S17), consistent with consensus C Env exhibiting high binding to plasma

relative to the other subtype/CRF before the loops were redesigned. Binding to monoclonal antibodies that target epitopes that are not presented on gp140 such as MPER antibodies and PGDM1400 (which targets the quaternary V2-apex epitope) was low as expected and was excluded from the statistical test. We compared the change in the predicted epitope depth to the change in antibody binding signal and found no significant correlation between them (Spearman's  $\rho = 0.52$ ,  $p$ -value = 0.15) (Figure. S18). In addition, we transfected 293 T cells with Env-gp160 plasmids corresponding to unmodified and redesigned HV loops and evaluated bnAb binding. For the two subtype B consensus sequences, we showed comparable Env cell surface expression, and staining with bnAbs such as 3BNC117 and PGDM1400 indicated that Env were expressed as closed trimers. Moreover, seven bnAbs tested showed increased binding to the Env with redesigned HV loops, suggesting improved epitope display (Figure. S19).

Finally, we evaluated how the Env consensus proteins (with or without modified HV loops) were neutralized by bnAbs. We used a pseudovirus neutralization assay in TZM-bl cells and tested ten bnAbs



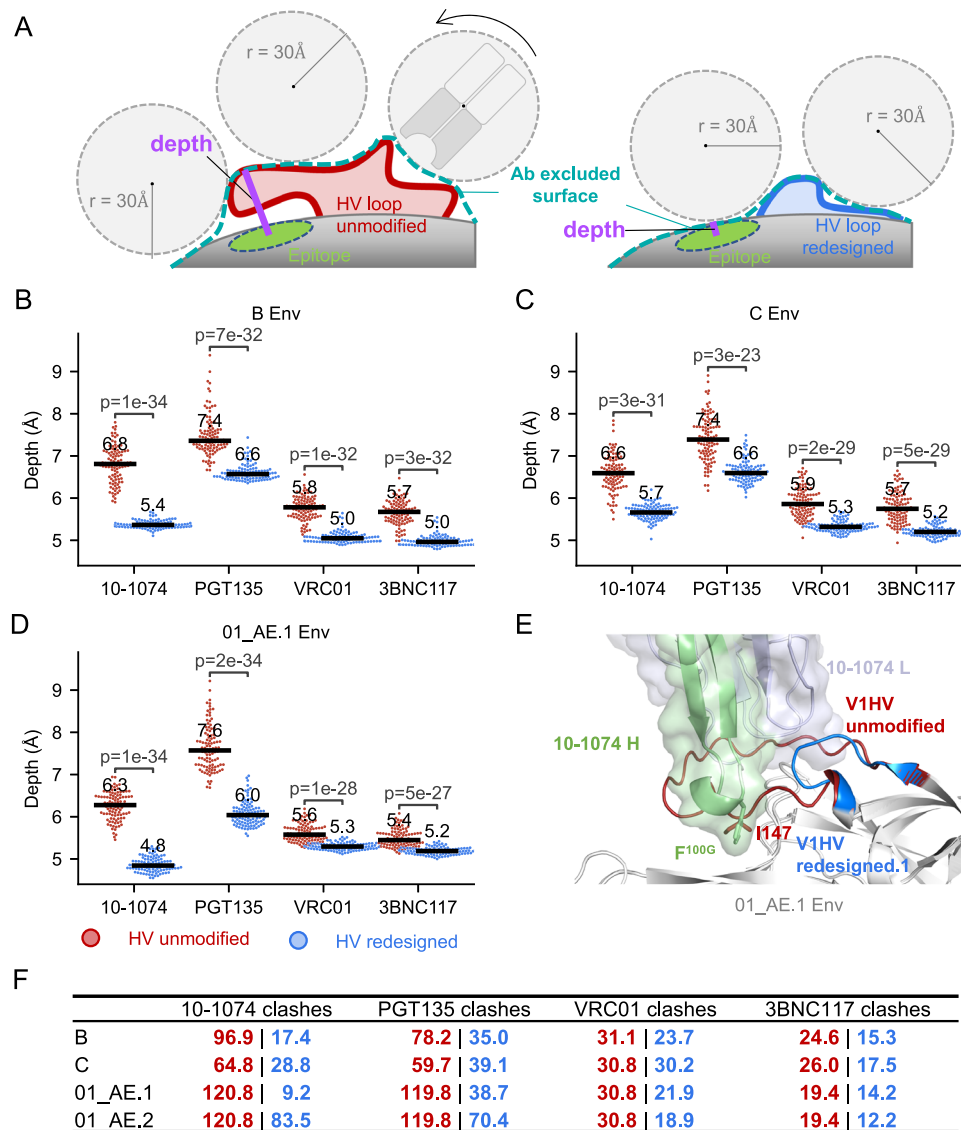
**Fig. 5 | Updated Env consensus proteins for subtypes B, C, and CRF01\_AE are functional.** **A** Viral titers for Env consensus proteins with or without redesigned hypervariable (HV) loops are presented as mean  $\pm$  SEM of three replicates. **B** Negative-stain electron microscopy for stabilized trimers with and without redesigned loops for consensus subtypes B and C. Representative 2D class averages of putative monomeric or disassembled Env and intact trimers for stabilized trimers with and without redesigned HV loops corresponding to the consensus for subtypes B and C. Percentages above the images indicate the relative proportion of

particles between monomer and trimer classes. **C** 3D reconstructions of particles classified as trimers in 2D with no symmetry applied (C1). Classes resembling the closed, prefusion conformation of HIV-1 Env are shown in purple while classes resembling the open, prefusion conformation are shown in green. Percentages above each map indicate the relative proportion between closed and open prefusion conformations. A side view for each reconstruction with the trimer apex on the bottom (top) and the trimer apex facing the viewer (bottom) is shown for each reconstruction. Source data of **A** are provided in the Source Data file.

targeting the CD4 binding site (VRC01 and 3BNC117), the V2-apex (PG9 and PGT145), the glycan supersite (10-1074 and PGT128), the MPER (10E8 and 4E10) and the carbohydrate-specific mannose-dependent (2G12) epitopes as well as the soluble CD4 (sCD4) (Fig. 7E, G for IC<sub>50</sub> values and S20 for IC<sub>80</sub> values). When comparing the consensus Env with the redesigned HV loop Env to those with the native HV loops, the neutralization sensitivity did not increase or decrease consistently for all the bnAbs (Fig. 7E, G and S20). When pooling data across the 10 bnAbs, there was no significant increase in neutralization sensitivity for subtype B and CRF01-AE with redesigned HV loops (one-tailed exact

Wilcoxon-Pratt signed-rank test,  $p > 0.201$ ) (Fig. 7F). In contrast, the subtype C Env with redesigned HV loops was more sensitive to the ten bnAbs than the subtype C Env with unmodified loops ( $p = 0.047$  for IC<sub>50</sub> values (Fig. 7F) and  $p = 0.016$  for IC<sub>80</sub> values (Figure. S20B)). One bnAb, PG9, showed a notable increase in sensitivity for the antigens with redesigned HV loops with a similar pattern across subtypes/CRF (Fig. 7G). Moreover, CRF01\_AE consensus proteins with the redesigned HV loop version 1 became sensitive to the glycan supersite bnAb 10-1074 (IC<sub>50</sub> = 0.544  $\mu$ g/ml) and to the sCD4 (IC<sub>50</sub> = 0.617  $\mu$ g/ml) while the original CRF01\_AE consensus was resistant (IC<sub>50/80</sub> > 50  $\mu$ g/ml)





**Fig. 6 | Impact of redesigned hypervariable (HV) loops on the antibody accessibility of epitope sites.** **A** Schematic showing how the minimum distance between an epitope site and the sphere probe (radius 30 Å) representing an antibody, defined as depth, is used to detect antibody accessibility differences of Env surface sites before (red) and after (blue) the HV loop redesign. **B–D** Comparison of the depth for four representative antibodies before and after the HV loop redesign. Median depth is labeled. The  $p$ -value from non-paired, non-parametric one-tailed

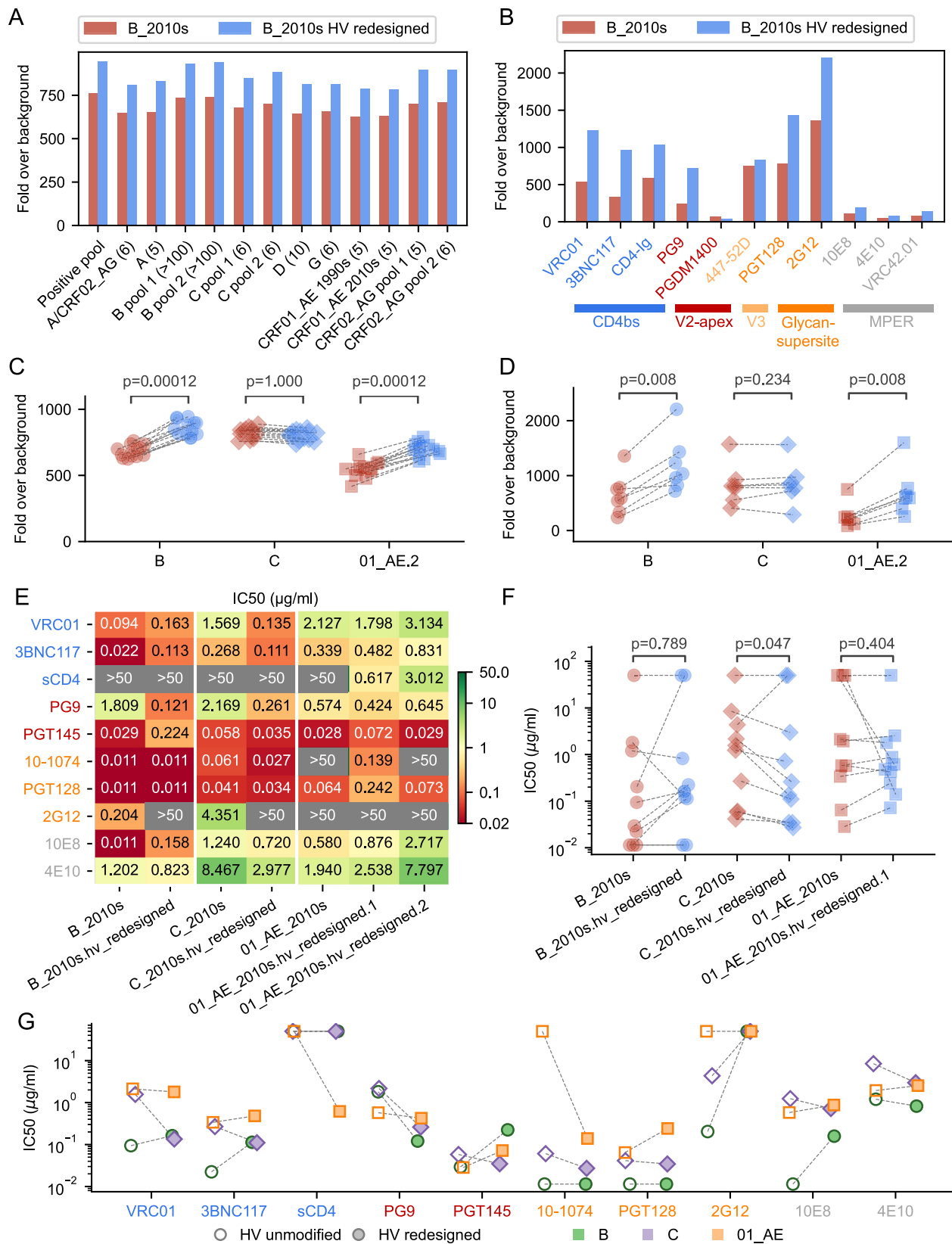
Mann-Whitney U tests is indicated ( $n = 100$  lowest energy conformations selected from 500 modeled conformations). **E** Consensus CRF01\_AE Env with unmodified and redesigned HV loops (AlphaFold2 prediction) are superimposed with Env in the 10-1074-Env complex. **F** Average decrease in clashes between antibody atoms and Env for unmodified (red) and redesigned (blue) cognate HV loops (based on 100 conformations). Source data for panels **B–D** are provided in the Source Data file.

(Fig. 7E, G, S20). Interestingly, the CRF01\_AE consensus with redesigned HV loop version 1 was sensitive to 10-1074 while the Env with version 2 of the redesign was resistant, in agreement with the antibody access predictions described above. The redesign version 1 is shorter in V1HV indicating that resistance to 10-1074 in CRF01\_AE is associated with the V1 loop length and is not necessarily mediated by the lack of glycosylation at site 332 as neither redesigned consensus Env had a glycan at site 332.

## Discussion

Here we redesigned the hypervariable loops V1HV, V2HV, V4HV, and V5HV using a structural modeling strategy to optimize access to critical bnAb epitopes. The redesign was done for subtypes B and C and CRF01\_AE. These subtypes/CRF were selected because large datasets of sequences were available (for the next most frequent subtype, A1, only 288 sequences were available). In addition, HIV-1

subtype B is the most thoroughly characterized subtype, subtype C is responsible for the majority of cases among PLWH, and CRF01\_AE is the virus that was circulating in the only HIV-1 vaccine trial where some efficacy was recorded<sup>39</sup>. We redesigned HV loops in Env consensus sequences that we had recently created; these consensus sequences reflected currently circulating viruses since they were derived only from sequences that were sampled after 2010<sup>31</sup>. The LANL database recently updated HIV-1 consensus sequences using all sequences sampled since the beginning of the epidemic<sup>55</sup>, hence, these 2021 consensus sequences tend to be biased towards past sequences which are more numerous than sequences sampled in the past decade. We chose to restrict our consensus sequences to sequences sampled since 2010 to ensure an accurate representation of recently circulating viruses. We note that HIV-1 proteins that are available at the HIV Reagent Program ([hivreagentprogram.org](http://hivreagentprogram.org)) and are commonly used as reagents in experimental assays were often



produced two decades ago. Consensus sequences were chosen because they represent the theoretically most fit virus, and, as such, they are more likely than a natural Env sequence to present favorable features for structural integrity as well as ideal bnAb epitopes<sup>33–36</sup>. We showed that our updated contemporary consensus sequences with

or without the redesigned HV loops were functional as demonstrated by infectivity assays with pseudoviruses and EM analyses of the corresponding stabilized trimers. In addition, the consensus sequences with redesigned HV loops promoted antibody binding when compared to the unmodified updated consensus.

**Fig. 7 | Improved antibody binding to subtype B and CRF01\_AE consensus Env with redesigned hypervariable (HV) loops.** Consensus Env sequences were expressed as gp140 glycoproteins. Comparison of the binding of **A** plasma pools from 13 cohorts of PLWH and **B** monoclonal antibodies to consensus B Env with unmodified (in red) and redesigned (in blue) HV loops. The number of plasma samples in each pool is shown in parenthesis in panel **A**. Binding of plasma samples ( $n = 13$ ) **C** and monoclonal antibodies ( $n = 7$ , with PGDM1400 and MPER antibodies excluded) (**D**) to the subtype B, C and CRF01\_AE consensus Env with unmodified HV loops (red) and redesigned HV loops (blue) is compared using paired, non-parametric one-tailed Wilcoxon signed-rank tests. MPER antibodies and PGDM1400 were not included in the statistical tests as their epitopes are not presented on the gp140 glycoproteins. **E** Neutralization sensitivity of the consensus

Env with or without redesigned HV loops for subtypes B, C, and CRF01\_AE measured for ten bnAbs that targeted the CD4 binding site (VRC01 and 3BNC117), V2-apex (PG9 and PGT145), glycan supersite (10-1074 and PGT128), MPER (10E8 and 4E10) and mannose-dependent (2G12) epitopes as well as the sCD4. The neutralization sensitivity is reported as IC50 values ( $\mu\text{g/ml}$ ) with the most sensitive viruses colored in red and the most resistant in gray. Neutralization sensitivity (IC50 values) are shown for the unmodified and redesigned HV loops and displayed separately for each Env clade ( $n = 10$ ) (**F**) or each bnAb (**G**). The increase in sensitivity with the redesigned HV loops is tested for each clade with one-tailed exact Wilcoxon-Pratt signed-rank tests. Only the O1\_AE\_2010s.hv\_redesigned.1 is shown in panels **F** and **G** as it has a shorter VIHV than O1\_AE\_2010s.hv\_redesigned.2. Source data for all panels are provided in the Source Data files.

Our analysis of a large set of sequences showed that Env with long HV loops were more resistant to bnAbs targeting cognate epitopes than Env with shorter loops. This is consistent with previous findings showing that the combined length of VIHV and V2HV was correlated with virus sensitivity to glycan-supersite or V3-glycan bnAbs, and that the length of V5 was related to virus sensitivity to CD4bs antibodies<sup>52</sup>. Interestingly, our structural modeling uncovered a novel mechanism for CRF01\_AE's remarkable resistance to glycan-supersite antibodies such as 10-1074. It was previously proposed that 10-1074 resistance was mediated by a shift of the glycan at site 332 to site 334<sup>52</sup>. Here, we showed that our redesigned Env CRF01\_AE consensus with the shortest VIHV (without the 'IGNITD' motif) was sensitive to 10-1074 despite having a glycan shifted from site 332 to 334 (Fig. 7E, F). The other (longer) redesigned CRF01\_AE VIHV with the 'IGNITD' motif (also with no PNGS at site 332) was resistant to 10-1074, indicating that the long VIHV was associated with 10-1074 resistance. In our model of CRF01\_AE VIHV N-anchor sites, the I147 in the 'IGNITD' motif was predicted to occupy the same location as the CDRH3 of 10-1074 (Fig. 6E and S5), a pocket composed of sites 328, 330, 415 and 417. In contrast, the pocket was not occupied by the modeled consensus B or C Env with unmodified VIHV. Furthermore, PGT128, a V3-glycan antibody whose epitope does not occupy this pocket<sup>13</sup> and is thus less occluded by the CRF01\_AE VIHV, retains the capacity to neutralize half of the CRF01\_AE viruses<sup>52</sup>. Our neutralization results confirmed that all the Env consensus sequences were sensitive to PGT128, while only the CRF01\_AE with the shortest VIHV was sensitive to 10-1074 (the two CRF01\_AE Envs that retained the 'IGNITD' motif remained resistant to 10-1074) (Fig. 7E, F). It is possible that other viruses with long VIHV use a similar mechanism to obstruct the binding of glycan-supersite bnAbs and divert antibody responses to these loops, highlighting that reducing the length of HV loops could play a key role in minimizing interference with bnAbs targeting adjacent epitopes. Importantly, the consensus sequences we used had HV loops of medium length due to how they are designed. Hence, redesigning the HV loop would likely have a more pronounced benefit in the context of a natural HIV-1 sequence.

Besides reducing the HV loop length, we focused on reducing the immunogenicity of the HV loop by using shorter spacers composed of fewer immunogenic amino acids. The redesigned spacers were composed mainly of glycine and serine residues except for the PNGS site. This is unlike natural HV loops, which consist of various AAs. Our goal was to minimize strain-specific responses. Since glycine and serine lack antibody-recognizable side chains, the redesigned loops are less likely to act as decoys attracting strain-specific antibody responses. A recent study found that long VIV2s in CRF01\_AE Env were associated with the more rapid development of autologous neutralization in participants from the RV217 cohort<sup>56</sup>; these results highlight the potential role of long HV loops in the development of strain-specific responses – these strain-specific responses could be considered undesirable as autologous neutralization was not associated with the development of neutralization breadth in this cohort.

Importantly, we noted that HV loop length, particularly for V1, has not been an important consideration for vaccine inserts used in vaccine efficacy trials or under consideration for future trials (Table 1). Most of the sequences used in previous vaccines had HV loops that were longer than the median HV loop length in the population, suggesting that candidate inserts for future vaccine trials could be optimized to minimize the HV loop lengths to improve immunogenicity profiles. We want to emphasize that the HV loop redesign can and should be combined with other strategies. First, while bnAb epitopes near HV-loops were more accessible after our redesign, their accessibility may still be outcompeted within the immune system, with preferential recognition at strain-specific glycan holes or of the base of the trimer, which is common when a soluble, stabilized prefusion Env is administered<sup>57–61</sup>. Thus, fixing strain-specific glycan holes, and shielding the Env trimer base (e.g. by mounting it on a liposome<sup>62</sup> or co-delivering it with Gag in an mRNA vaccine<sup>66</sup>) could also help presenting bnAb epitopes. Second, the epitope near the redesigned HV loops may also need to be optimized, though what constitutes an optimal epitope for the elicitation of bnAbs is debatable. We previously showed that bnAb epitopes were actually as diverse as non-bnAb epitopes among HIV-1 sequences; however, what distinguished the broadest bnAbs from the other nAbs targeting similar epitopes was their ability to focus on key epitope sites that were more conserved<sup>6</sup>. As a result, if rare mutations are found in key epitope sites, it would make sense to revert the mutated residues to the consensus AA at that site. Third, many unshielded epitopes are exposed on the Env monomer and post-fusion Env. Thus, for the HV loop modification to be effective, a stabilized prefusion Env trimer is required as the improved bnAb accessibility may be insignificant in the monomer/post-fusion context. This can be achieved with the repair and stabilization procedure which systematically fixes rare mutations and stabilizes HIV-1 Env in the prefusion conformation<sup>48,49</sup>.

There are some limitations to our computational approach to assess antibody accessibility. First, the glycan shield is absent while we calculate the depth of the epitope and the clashes between Env and bnAbs. That might lead to an overestimation of the antibody accessibility both before and after the HV-loop redesign, as the occlusion from glycans is omitted. Second, we modeled one subunit in the prefusion Env trimer by AlphaFold2 and modeled loop conformations based on them. The impact from neighboring subunits is absent in our evaluation. Third, when calculating the depth, a sphere with a radius of 30 Å was chosen as the antibody probe to simplify calculations, which is not necessarily a good approximation for Fv or Fab domain. As a result, our computational assessment of antibody binding should only be considered qualitatively correct in terms of capturing the antibody accessibility difference. Moreover, the neutralization sensitivity assays that compared how the different consensus proteins were neutralized by ten bnAbs showed that the Env with redesigned HV loops did not show increased sensitivity to all bnAbs across all subtypes. There were nonetheless striking effects regarding the increased sensitivity to PG9 across clades or the change from resistance to sensitivity to 10-1074 for

CRF01\_AE Env. These results illustrate that Env sensitivity to neutralization is not an all-or-none phenomenon and warrants further studies. Although a significant increase in sensitivity was not evidenced across all bnAbs and all clades, it remains possible that our redesigned Env when used as immunogens could promote the immunofocusing of bnAb responses to conserved epitopes due to the elimination of potentially interfering residues. While the EM characterization did not reveal any aberration for our redesigned consensus Env with four HV loops shortened simultaneously, it will be important to test the effect of each HV loop redesign independently and to generate higher-resolution empirical structural information.

In this work, we redesign new Env consensus sequences that reflect currently circulating sequences for subtypes B and C and CRF01\_AE with modified HV loops and use in vitro assays to test infectivity, neutralization sensitivity, and structural conformation to show that these computationally-derived sequences resulted in functional Env. We demonstrate that long HV loops interfered with bnAbs targeting adjacent epitopes and that our modifications alleviate the interference of HV loops on bnAb epitopes, especially for the glycan-supersite bnAbs. Since half of all neutralizing antibodies target epitopes around the four optimized HV loops<sup>22,23</sup>, we propose that this design strategy could yield more effective antigens than those currently used as reagents or vaccine candidates. This optimization strategy can be integrated with other vaccine design techniques, such as the 'repair and stabilize' of Env in the prefusion conformation<sup>48,49</sup>, shielding the base of Env<sup>46</sup>, or the presentation of an optimum combination of Envs<sup>63</sup>, to create the next-generation of HIV-1 vaccines for eliciting broadly protective antibodies.

## Methods

### Dataset

Env AA sequences of subtypes B, C, and CRF01\_AE were retrieved from the Los Alamos National Laboratory HIV-1 sequence database on 2022-11-07. Sequences were retained if they had a complete open reading frame, information on the year and location of collection, and were not hypermutated or problematic (as defined by LANL). Non-independent sequences were removed so that the final dataset included one sequence per person. The IC50 of PGT145, VRC26.25, 10-1074, PGT135, VRC01, and 3BNC117 against viruses that had the corresponding Env sequences available were downloaded from the CATNAP database as of 2022-10-04<sup>47</sup>. In total, there were 549, 423, 782, 310, 1032, and 818 IC50s and associated Env sequences for PGT145, VRC26.25, 10-1074, PGT135, VRC01 and 3BNC117, respectively.

### HV loop analysis

The HXB2 genome annotation from LANL HIV Database (<https://www.hiv.lanl.gov/content/sequence/HIV/MAP/annotation.html>) was followed to define the start and end of HV loops, which were 132-152, 185-190, 396-410 and 460-467, for VIH, V2HV, V4HV and V5HV, respectively. As an exception, V4HV was extended to include 395 and 411-412 anchor sites in the HV loop redesign. Structurally, the distance between the start and end of spacers was calculated as the distance between CA atoms before the N-terminal and after the C-terminal of the spacer loop. The HXB2 sequence was added and aligned to define the positions. MAFFT<sup>64</sup> was used to align sequences. For the logo plot of HVs, HV loops were split from the middle: the first half was left-aligned to the start position while the second half was right-aligned to the end position.

### Structure prediction and analysis

We used local ColabFold<sup>65</sup> which integrated AlphaFold-Multimer<sup>66</sup> with other tools and databases<sup>67-72</sup> to model unmodified and redesigned Env structures. During the prediction, one of the three subunits in the prefusion Env trimer is predicted by feeding one gp120 (31-511) and one gp41 (512-665) as input. The top model from the five predictions

was adopted for later analysis. Loop modeling was done with the KIC loop modeling protocol<sup>50</sup> via PyRosetta<sup>51</sup>. The four HV loops were modeled sequentially in random orders in individual runs.

The non-HV surrounding surface sites of HV loops were defined as Env sites that were not part of the HV loop which have relatively accessible surface area (RSA) greater or equal to 0.30 and have any atoms within 10 Å of any HV loop atoms in the predicted unmodified consensus B Env structure. The accessible surface area (ASA)<sup>73</sup> was approximated by the Shrake-Rupley method<sup>74</sup>. In brief, evenly distributed points ( $n = 256$ ) were placed on spheres that were centered at atoms in the protein. The radius of the spheres was the sum of the atoms' and the solvent probe's radius (1.5 Å). The accessible surface was defined after filtering out points on a sphere that were within neighboring spheres. Then, the RSA of a site was calculated as the ratio of ASA over the estimated maximum ASA<sup>75</sup>. The same approach was also used to estimate depth, only the probe radius was set to 30 Å, and the depth equals the minimal distance between a protein atom and the accessible surface minus the probe radius. Env structure 5FYJ<sup>76</sup> was used for the ASA estimation. It was also used in the illustration of the HV loop and nearby bnAb epitopes in Fig. 1.

The clashes between antibody and Env were determined as the number of antibody atoms within 2.5 Å of Env after superimposing the modeled Env on the Env in the solved Env-antibody complex. Bio.PDB<sup>77</sup> in Biopython<sup>78</sup> was used to superimpose structures. The Env-antibody complexes used, listed as PDB code, were 5V8L<sup>17</sup>, 6VTT<sup>18</sup>, 6UDJ<sup>79</sup>, 4JM2<sup>22</sup>, 6V8X<sup>80</sup> and 7PC2<sup>81</sup> for PGT145, VRC26.25, 10-1074, PGT135, VRC01, and 3BNC117, respectively. PyMol<sup>82</sup> was used for the structure visualization and figure generation. Hydrogen atoms, when present, were removed from the structure, in all analyses.

### Protein production and antibody binding

Uncleaved gp140 proteins were designed with R to S mutations in the gp120/gp41 furin cleavage site and incorporated the native leader peptide and full MPER sequence, followed by a short GGG linker sequence and a C-terminal AviTag. Sequences, codon-optimized for expression in human cells, were synthesized (Genscript) and cloned into a custom pcDNA3.4 (ThermoFisher) expression vector (plasmids are available with a Material Transfer Agreement). Proteins were expressed by transient transfections in Expi293F cells (ThermoFisher) according to the manufacturer's instructions. Gp140 proteins were purified from clarified cell culture supernatants 4 days post-transfection using Galanthus nivalis lectin (GNL) affinity chromatography followed by a Q-sepharose (GE Healthcare) polishing step to remove host cell protein contaminants. All gp140s were purified to 90% purity or higher, as assessed by SDS-PAGE and Coomassie staining in reducing and non-reducing conditions. Antibodies VRC01<sup>20</sup>, 3BNC117<sup>83</sup>, CD4-Ig<sup>84</sup>, PG9<sup>19</sup>, PGDM1400<sup>15</sup>, PGT128<sup>13</sup>, 2G12<sup>85</sup>, 10E8<sup>86</sup>, 4E10<sup>14,87</sup> were produced in-house from public sequences as IgG1 by transfection in Expi293F cells and purified by protein A affinity chromatography; antibodies 447-52D<sup>88</sup> and VRC42.01<sup>89</sup> were obtained from Polymun (Cat. No. AB014) and from Nicole Doria-Rose (VRC, NIH), respectively. For antibody binding and characterization, gp140 proteins were coupled to uniquely coded carboxylated magnetic microspheres (Luminex Corp., Austin TX) per manufacturer's protocol and as previously described and pooled into an antigen cocktail. Antibody binding and characterization were performed as previously described<sup>53,90,91</sup> with minor modifications. The antigen cocktail was incubated with an array of heat-inactivated human plasma pool samples from various clade-specific cohorts of PLWH, or with monoclonal antibodies targeting various regions of HIV-1 Env. Samples were incubated for 2 hours at room temperature, washed then incubated for 1 hour with phycoerythrin labeled anti-human IgG Fc (Southern Biotech, Birmingham AL) detection antibody. After a final wash to remove unbound detection reagent, microspheres were resuspended in 40 µL sheath fluid (Luminex Corp) and data was collected on a Bio-Plex<sup>®</sup>3D

Suspension Array system (Bio-Rad, Hercules CA) running xPONENT® v.4.2 (Luminex Corp). A multi-clade control plasma pool comprised of samples from HIV-1 Group O, subtypes B, C, and CRF01\_AE, CRF02\_AG, and CRF03\_AB (SeraCare, Milford MA) was used as a positive control ("Positive Pool") and a pool of plasmas from people without HIV-1 (PWOH) served as a negative control. The figures correspond to monoclonal antibodies at a concentration of 10 µg/mL and pooled plasma at 1:100 dilution. Responses are reported as fold-over PWOH for plasma samples and fold-over the ZIKA virus-specific monoclonal antibody MZ4<sup>92</sup>.

### HIV-1 Env expression and antibody binding

293 T cells were transfected with 1 µg of plasmid DNA encoding an HIV-1 env gene (PolyJet™ In Vitro DNA Transfection Reagent (SigmaGen Laboratories, Cat. No. SL100688) in 6-well plates and harvested at 48 h. Cells were split into separate staining reactions for monoclonal antibodies 3BNC117 (Cat. No. ARP-12474), 10-1074 (Cat. No. ARP-12477), 10E8 (Cat. No. ARP-12294), 447-52D (Cat. No. ARP-4030, all from BEI resources), b12 (Polymun, Cat. No. AB011), PGDM1400<sup>15</sup> and PGT151<sup>93</sup> (both kindly provided by Dr. Devon Sok and International AIDS Vaccine Initiative). 2G12 was produced in Expi293F cells from publicly available heavy and light chain sequences<sup>94</sup> and purified using Protein A affinity chromatography. Cells were washed and stained with anti-human IgG-PE at 1:250 dilution (SouthernBiotech, Cat. No. 2040-09) for 15 min and fixed in 0.5% paraformaldehyde. Fluorescence was measured on a FACSymphony A5 SE Cell Analyzer and analyzed in FlowJo v.9.9.6.

### Recombinant HIV-1 Env stabilized trimer production and purification

Stabilized, soluble trimers in a prefusion state were engineered by introducing previously reported<sup>48,49</sup> and additional modifications with ten stabilization mutations (M302, V304, L320, N535, P556, E588, V589, F651, I655, and V658) and three SOSIP mutations (C501, C605 and P559). Sites 508-511 were replaced with six arginines for more efficient furin cleavage. All Envs were terminated at site 664. Trimers were expressed in Expi 293 cells (Thermo Fisher Scientific) by transient transfection with plasmid DNA complexed with PEI. 800 µg of the HIV-1 Env expressing plasmid and 200 µg of furin expressing plasmid were used for each 1 L of cells. Cell-free culture supernatant was 0.2 µm filtered and purified by GNA Lectin (VectorLabs) before size exclusion chromatography on a Superdex 200 16/600 column (GE Healthcare Cat. No. 29323952) in PBS. Any shoulders to the roughly monodisperse Env trimer peak were removed, and protein was pooled and concentrated before adding glycerol to a final concentration of 5%. Trimeric HIV-1 Env protein was sterile-filtered, snap-frozen, and stored at -80 °C before further use.

### Negative-stain Electron Microscopy

Purified protein was thawed and diluted to 10 µg/ml with 20 mM HEPES, pH 7.5; 150 mM NaCl immediately prior to application on glow-discharged carbon-coated grids (Electron Microscopy Sciences #CF300H-Cu-50) and stained with 2% phosphotungstic acid, pH 7.4. Samples were imaged on a JEOL CryoARM 200 using SerialEM 4.1.0beta with a nominal magnification of 60kX and a pixel size of 0.873 Å/pixel on a Ametek Gatan K3 in counting mode. A 100-frame movie with a total dose of 225 electrons/Å<sup>2</sup> was motion corrected and CTF was estimated using patch CTF in CryoSPARC v4<sup>95</sup>. All further processing was done in Cryosparc v4. Particle picking was conducted with a circular Gaussian blob of 100-150 Å for all samples. Reference-free 2D classification was used to separate false picks from putative particles. A second round of 2D with a maximum alignment resolution of 15 Å and a maximum resolution of 10 Å on putative particles was used to separate monomer and trimer classes and estimate the relative percentages of each species. For the gp140 samples, no averageable trimer was observed.

For stabilized trimer samples, A C1 ab initio model was created for particles that classified as trimers in 2D for each of the protein samples imaged. This was used as an initial model for a C1 3D homogenous refinement with an initial low pass filter of 60 Å, GS-FSC split of 40 Å, and a maximum alignment resolution of 20 Å. The resulting consensus structure was 3D classified without alignment with 10 classes and a target resolution of 15 Å. The particles from the most open trimer class and most closed trimer class were selected and refined with 3D homogenous refinement using the volume from their classification with the same parameters as noted above in both C1 and C3. C1 and C3 closed and open models were then used as initial models for a two-class heterogeneous refinement with an initial resolution of 30 Å and a maximum resolution of 15 Å. The relative percentages and maps shown were estimated using this C1 heterogeneous refinement, but C3 refinements converged on the same approximate proportions.

This procedure was effective for all samples except for the subtype C consensus stabilized trimer without HV loop redesign. In this case, the classification without alignment did not produce a particle subset that was closed enough to separate closed from open trimer using heterogeneous refinement. In this case, an initial model of closed trimer from the subtype C stabilized trimer with HV loop redesign sample and an open trimer without HV loop redesign was used for the final C1 heterogeneous refinement. Data was visualized and figures were generated using ChimeraX<sup>96</sup> and EMAN2<sup>97</sup>.

### HIV-1 pseudovirus (PSV) production, titration and neutralization

HIV-1 PSVs were produced by co-transfection of HEK293T/17 cells (ATCC, Manassas, VA, USA) with a pSG3ΔEnv DNA plasmid encoding the HIV-1 backbone (24 µg DNA) and a plasmid encoding the HIV-1 env gene (8 µg DNA). Plasmids were combined with X-tremeGENE 9 DNA transfection reagent (Sigma-Aldrich, St. Louis, MO, USA) at a 1:3 ratio in serum-free DMEM, per manufacturer recommendations. The transfection mix was added to a T75 flask containing plated HEK293T/17 cells in a culture medium. The flask was incubated for 6 hours at 37 °C. Then the media was replaced. PSVs were harvested 72 hours after transfection, filtered through a 0.45 µm PVDF membrane (MilliporeSigma, Burlington, MA, USA), and stored in 3 ml aliquots at -80 °C.

PSV stocks were titrated on TZM-bl cells (NIH AIDS Research and Reference Reagent Program, Cat. no. 8129) to determine an acceptable dilution for neutralization assays. Virus stocks were serially diluted then added to 96-well plates (Thermo Fisher Scientific, Waltham, MA, USA) in a final 25 µL/well volume. An additional 25 µL of supplemented DMEM was added to each well. TZM-bl cells were added to plates in 50 µL of supplemented DMEM with 40 µg/mL of DEAE-Dextran (Sigma, St. Louis, MO, USA). Virus and cells were incubated together for 48 hours at 37 °C. Infectivity was measured by luminescence after adding 100 µL of the Bright-Glo Luciferase Assay System (Promega, Madison, WI, USA) to each well; luminescence was quantified using the EnVision luminometer (Perkin Elmer, Waltham, MA, USA).

Neutralizing antibody titers were determined using TZM-bl cells in a manual 96-well assay. BnAbs were obtained from the HIV reagent Program (VRC01 (Cat. No. ARP-12033, Lot 180231); 3BNC117 (Cat. No. ARP-12474, Lot 140211); sCD4 (Cat. No. ARP-4615, Lot 180011); PGT145 (Cat. No. ARP-12703, Lot 190527); 10-1074 (Cat. No. ARP-12477, Lot 190442); PGT128 (Cat. No. ARP-13352, Lot 190522); 2G12 (Cat. No. ARP-1476, Lot 180138); 10e8 (Cat. No. ARP-12294, Lot 170395) or from PolyMun (PG9 (Cat. No. AB015, Lot PG9\_021014A); 4e10 (Cat. No. AB004, Lot T531002-A)). BnAbs were prepared at a starting concentration of 25 µg/mL in supplemented DMEM. PSVs were diluted to a concentration to yield 100,000 relative light units (RLUs) and at least 3 times above the cell-only controls, as determined by the PSV titration assay. Neutralization reagents were serially diluted. Titered bNABs were transferred to 96-well culture plates and incubated with an equal volume of PSV for 1 hour at 37 °C. A non-specific virus control (MuLV)

was used to ensure neutralization by the bNAbs was HIV-specific. TZMbl cells ( $1 \times 10^4$  cell/well) mixed with 40  $\mu\text{g}/\text{mL}$  DEAE-dextran (Sigma, St. Louis, MO, USA) were added to each well and incubated for an additional 48 hours. RLUs were detected with the EnVision luminometer (Perkin Elmer, Waltham, MA, USA) using the Bright-Glo Luciferase Assay System (Promega Corporation, Madison, WI, USA). Percent neutralization (percentage reduction of relative light units in the presence of bNAb) was calculated for each bNAb concentration. Neutralization dose-response curves were fitted by nonlinear regression using the LabKey Server, and the final titer is reported as the concentration of bNAb (IC<sub>50</sub>/IC<sub>80</sub>) necessary to achieve 50% or 80% neutralization.

### Quantification and statistical analysis

Statistical analyses were performed using SciPy<sup>98</sup> and coin<sup>99</sup>. The non-paired, non-parametric Mann-Whitney U tests were used to compare the IC<sub>50</sub> of short/long nearby or distant HV loops in Fig. 2, S2 and S3. The paired, non-parametric Wilcoxon signed-rank tests were used to compare the depth difference before and after HV loop redesign (Fig. 6) and the binding responses before and after the HV loop redesign with pooled plasma samples or bNAbs (Fig. 7). The one-tailed exact Wilcoxon-Pratt signed-rank test is carried out using the coin package<sup>99</sup>. Differences corresponding to p-values of less than 0.05 were considered as significant.

### Data availability

The source data for all figures and tables in the manuscript and in the Supplementary Information are provided with this paper. Source data are provided with this paper.

### Code availability

The code used for the analysis is available at <https://www.hivresearch.org/node/951>.

### References

- Mascola, J. R. & Montefiori, D. C. The role of antibodies in HIV vaccines. *Annu. Rev. Immunol.* **28**, 413–444 (2010).
- Saunders, K. O., Rudicell, R. S. & Nabel, G. J. The design and evaluation of HIV-1 vaccines. *AIDS Lond. Engl.* **26**, 1293–1302 (2012).
- Wibmer, C. K., Moore, P. L. & Morris, L. HIV broadly neutralizing antibody targets. *Curr. Opin. HIV AIDS* **10**, 135–143 (2015).
- Kwong, P. D. & Mascola, J. R. HIV-1 vaccines based on antibody identification, B cell ontogeny, and epitope structure. *Immunity* **48**, 855–871 (2018).
- Sok, D. & Burton, D. R. Recent progress in broadly neutralizing antibodies to HIV. *Nat. Immunol.* **19**, 1179–1188 (2018).
- Bai, H., Li, Y., Michael, N. L., Robb, M. L. & Rolland, M. The breadth of HIV-1 neutralizing antibodies depends on the conservation of key sites in their epitopes. *PLoS Comput. Biol.* **15**, e1007056 (2019).
- O’Connell, R. J., Kim, J. H. & Excler, J.-L. The HIV-1 gp120 V1V2 loop: structure, function and importance for vaccine development. *Expert Rev. Vaccines* **13**, 1489–1500 (2014).
- Moore, P. L., Gorman, J., Doria-Rose, N. A. & Morris, L. Ontogeny-based immunogens for the induction of V2-directed HIV broadly neutralizing antibodies. *Immunol. Rev.* **275**, 217–229 (2017).
- Kwong, P. D. et al. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
- Gristick, H. B. et al. Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. *Nat. Struct. Mol. Biol.* **23**, 906–915 (2016).
- Walker, L. M. et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).
- Kong, L. et al. Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat. Struct. Mol. Biol.* **20**, 796–803 (2013).
- Pejchal, R. et al. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334**, 1097–1103 (2011).
- Buchacher, A. et al. Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Res. Hum. Retroviruses* **10**, 359–369 (1994).
- Sok, D. et al. Recombinant HIV envelope trimer selects for quaternary-dependent antibodies targeting the trimer apex. *Proc. Natl. Acad. Sci. USA.* **111**, 17624–17629 (2014).
- 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).
- 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).
- Walker, L. M. et al. Broad and potent neutralizing antibodies from an African donor reveal a New HIV-1 vaccine target. *Science* **326**, 285–289 (2009).
- Zhou, T. et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).
- Klein, F. et al. Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* **153**, 126–138 (2013).
- Landais, E. et al. Broadly neutralizing antibody responses in a large longitudinal sub-Saharan HIV primary infection cohort. *PLoS Pathog.* **12**, e1005369 (2016).
- Townsley, S. M. et al. B cell engagement with HIV-1 founder virus envelope predicts development of broadly neutralizing antibodies. *Cell Host Microbe* **29**, 564–578.e9 (2021).
- Srivastava, I. K., VanDorsten, K., Vojtech, L., Barnett, S. W. & Stamatatos, L. Changes in the immunogenic properties of soluble gp140 human immunodeficiency virus envelope constructs upon partial deletion of the second hypervariable region. *J. Virol.* **77**, 2310–2320 (2003).
- Silva de Castro, I. et al. Anti-V2 antibodies virus vulnerability revealed by envelope V1 deletion in HIV vaccine candidates. *iScience* **24**, 102047 (2021).
- Bontjer, I. et al. Comparative immunogenicity of evolved V1V2-deleted HIV-1 envelope Glycoprotein trimers. *PLoS One* **8**, e67484 (2013).
- Zolla-Pazner, S. et al. Rationally designed vaccines targeting the V2 region of HIV-1 gp120 induce a focused, cross-clade-reactive, biologically functional antibody response. *J. Virol.* **90**, 10993–11006 (2016).
- Jiang, X. et al. Rationally designed immunogens targeting HIV-1 gp120 V1V2 induce distinct conformation-specific antibody responses in rabbits. *J. Virol.* **90**, 11007–11019 (2016).
- Karch, C. P. et al. Design and characterization of a self-assembling protein nanoparticle displaying HIV-1 Env V1V2 loop in a native-like trimeric conformation as vaccine antigen. *Nanomed. Nanotechnol. Biol. Med.* **16**, 206–216 (2019).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Lewitus, E., Li, Y., Bai, H., Pham, P. & Rolland, M. HIV-1 Gag, Pol, and Env diversified with limited adaptation since the 1980s. *mBio* e0174923 <https://doi.org/10.1128/mbio.01749-23> (2024).
- Kuiken, C., Korber, B. & Shafer, R. W. HIV sequence databases. *AIDS Rev.* **5**, 52–61 (2003).
- Korber, B. et al. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **58**, 19–42 (2001).
- Gaschen, B. et al. Diversity considerations in HIV-1 vaccine selection. *Science* **296**, 2354–2360 (2002).

35. Rolland, M. et al. Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins. *J. Virol.* **81**, 8507–8514 (2007).
36. Rolland, M. HIV-1 phylogenetics and vaccines. *Curr. Opin. HIV AIDS* **14**, 227–232 (2019).
37. Gilbert, P. B. et al. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. *J. Infect. Dis.* **191**, 666–677 (2005).
38. Pitisuttithum, P. et al. Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J. Infect. Dis.* **194**, 1661–1671 (2006).
39. Rerks-Ngarm, S. et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **361**, 2209–2220 (2009).
40. Barouch, D. H. et al. Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nat. Med.* **16**, 319–323 (2010).
41. Hammer, S. M. et al. Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *N. Engl. J. Med.* **369**, 2083–2092 (2013).
42. Bekker, L.-G. et al. Subtype C ALVAC-HIV and bivalent subtype C gp120/MF59 HIV-1 vaccine in low-risk, HIV-uninfected, South African adults: a phase 1/2 trial. *Lancet HIV* **5**, e366–e378 (2018).
43. Barouch, D. H. et al. Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). *Lancet Lond. Engl.* **392**, 232–243 (2018).
44. LaBranche, C. C. et al. Neutralization-guided design of HIV-1 envelope trimers with high affinity for the unmutated common ancestor of CH235 lineage CD4bs broadly neutralizing antibodies. *PLoS Pathog* **15**, e1008026 (2019).
45. Gray, G. E. et al. Vaccine efficacy of ALVAC-HIV and bivalent subtype C gp120-MF59 in adults. *N. Engl. J. Med.* **384**, 1089–1100 (2021).
46. Zhang, P. et al. A multiclade env-gag VLP mRNA vaccine elicits tier-2 HIV-1-neutralizing antibodies and reduces the risk of heterologous SHIV infection in macaques. *Nat. Med.* **27**, 2234–2245 (2021).
47. Yoon, H. et al. CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res.* **43**, W213–W219 (2015).
48. Rutten, L. et al. A universal approach to optimize the folding and stability of prefusion-closed HIV-1 envelope trimers. *Cell Rep.* **23**, 584–595 (2018).
49. Rawi, R. et al. Automated design by structure-based stabilization and consensus repair to achieve prefusion-closed envelope trimers in a wide variety of HIV strains. *Cell Rep.* **33**, 108432 (2020).
50. Stein, A. & Kortemme, T. Improvements to robotics-inspired conformational sampling in rosetta. *PLoS One* **8**, e63090 (2013).
51. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinforma. Oxf. Engl.* **26**, 689–691 (2010).
52. Bricault, C. A. et al. HIV-1 neutralizing antibody signatures and application to epitope-targeted vaccine design. *Cell Host Microbe* **25**, 59–72.e8 (2019).
53. Mdululi, T. et al. RV144 HIV-1 vaccination impacts post-infection antibody responses. *PLoS Pathog.* **16**, e1009101 (2020).
54. Merbah, M. et al. A high-throughput multiplex assay to characterize flavivirus-specific immunoglobulins. *J. Immunol. Methods* **487**, 112874 (2020).
55. Linchangco, G. V., Foley, B. & Leitner, T. Updated HIV-1 consensus sequences change but stay within similar distance from worldwide samples. *Front. Microbiol.* **12**, 828765 (2021).
56. Kuriakose Gift, S. et al. Evolution of antibody responses in HIV-1 CRF01\_AE acute infection: founder envelope V1V2 impacts the timing and magnitude of autologous neutralizing antibodies. *J. Virol.* e0163522 <https://doi.org/10.1128/jvi.01635-22> (2023).
57. Klasse, P. J. et al. Sequential and simultaneous immunization of rabbits with HIV-1 envelope Glycoprotein SOSIP.664 Trimers from Clades A, B, and C. *PLoS Pathog.* **12**, e1005864 (2016).
58. Nogal, B. et al. Mapping polyclonal antibody responses in non-human primates vaccinated with HIV Env Trimer subunit vaccines. *Cell Rep.* **30**, 3755–3765.e7 (2020).
59. Reiss, E. I. M. M. et al. Fine-mapping the immunodominant antibody epitopes on consensus sequence-based HIV-1 envelope trimer vaccine candidates. *NPJ Vaccines* **7**, 152 (2022).
60. Hu, J. K. et al. Murine antibody responses to cleaved soluble HIV-1 envelope trimers are highly restricted in specificity. *J. Virol.* **89**, 10383–10398 (2015).
61. 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).
62. Shao, S. et al. Functionalization of cobalt porphyrin-phospholipid bilayers with his-tagged ligands and antigens. *Nat. Chem.* **7**, 438–446 (2015).
63. Lewitus, E., Hoang, J., Li, Y., Bai, H. & Rolland, M. Optimal sequence-based design for multi-antigen HIV-1 vaccines using minimally distant antigens. *PLoS Comput. Biol.* **18**, e1010624 (2022).
64. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
65. Mirdita, M. et al. ColabFold: Making Protein folding accessible to all. *Nat. Methods* <https://doi.org/10.1038/s41592-022-01488-1> (2022).
66. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* <https://doi.org/10.1101/2021.10.04.463034v1> (2021).
67. Mirdita, M., Steinegger, M. & Soding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
68. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform* **20**, 473 (2019).
69. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
70. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980 (2003). vol.
71. Mirdita, M. et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* **45**, D170–D176 (2017).
72. Mitchell, A. L. et al. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res* <https://doi.org/10.1093/nar/gkz1035> (2019).
73. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
74. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371 (1973).
75. Rost, B. & Sander, C. Conservation and prediction of solvent accessibility in protein families. *Proteins* **20**, 216–226 (1994).
76. Stewart-Jones, G. B. E. et al. Trimeric HIV-1-Env structures define Glycan shields from Clades A, B, and G. *Cell* **165**, 813–826 (2016).
77. Hamelryck, T. & Manderick, B. PDB file parser and structure class implemented in Python. *Bioinforma. Oxf. Engl.* **19**, 2308–2310 (2003).
78. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinforma. Oxf. Engl.* **25**, 1422–1423 (2009).
79. Schommers, P. et al. Restriction of HIV-1 escape by a highly broad and potent neutralizing antibody. *Cell* **180**, 471–489.e22 (2020).
80. Henderson, R. et al. Disruption of the HIV-1 Envelope allosteric network blocks CD4-induced rearrangements. *Nat. Commun.* **11**, 520 (2020).
81. Lorin, V. et al. Epitope convergence of broadly HIV-1 neutralizing IgA and IgG antibody lineages in a viremic controller. *J. Exp. Med.* **219**, e20212045 (2022).

82. Schrödinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.8. (2015). <https://pymol.org/>
83. Schoofs, T. et al. HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1. *Science* **352**, 997–1001 (2016).
84. Arthos, J. et al. Biochemical and biological characterization of a dodecameric CD4-Ig fusion protein: implications for therapeutic and vaccine strategies. *J. Biol. Chem.* **277**, 11456–11464 (2002).
85. Trkola, A. et al. Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG. *J. Virol.* **69**, 6609–6617 (1995).
86. Huang, J. et al. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012).
87. Zwick, M. B. et al. Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *J. Virol.* **75**, 10892–10905 (2001).
88. Gorny, M. K. et al. Neutralization of diverse human immunodeficiency virus type 1 variants by an anti-V3 human monoclonal antibody. *J. Virol.* **66**, 7538–7542 (1992).
89. Krebs, S. J. et al. Longitudinal analysis reveals early development of three MPER-directed neutralizing antibody lineages from an HIV-1-infected individual. *Immunity* **50**, 677–691.e13 (2019).
90. Tomaras, G. D. et al. Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82**, 12449–12463 (2008).
91. Brown, E. P. et al. High-throughput, multiplexed IgG subclassing of antigen-specific antibodies from clinical samples. *J. Immunol. Methods* **386**, 117–123 (2012).
92. Dussupt, V. et al. Potent Zika and dengue cross-neutralizing antibodies induced by Zika vaccination in a dengue-experienced donor. *Nat. Med.* **26**, 228–235 (2020).
93. Falkowska, E. et al. Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity* **40**, 657–668 (2014).
94. Calarese, D. A. et al. Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**, 2065–2071 (2003).
95. 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).
96. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci. Publ. Protein Soc.* **30**, 70–82 (2021).
97. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
98. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
99. Hothorn, T., Hornik, K., van de Wiel, M. A. & Zeileis, A. A Lego system for conditional inference. *Am. Stat.* **60**, 257–263 (2006).

## Acknowledgements

We thank Johannes PM Langedijk and Lucy Rutten for helpful discussion, Indie Showell-DeLeon and Diana Wasson for technical assistance. We thank Phuc Pham for checking the code. This work was supported by a

cooperative agreement between The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the U.S. Department of the Army [WW81XWH-18-2-0040] and by funds from the NIH (1R01AI176520-01) to SV.

## Author contributions

H.B. and M.Ro. conceived the study, analyzed the results and wrote the manuscript. H.B. and P.V.T. designed and performed analyses. E.L. and Y.L. curated sequence alignments and designed consensus sequences. P.V.T. and M.G.J. characterized redesigned Env in vitro. M.Z., M.M., C.E.P., T.T., P.A.R., A.H., V.D., B.S., L.M.-R., A.S., E.K., M.Ra., G.S., J.F., A.S., L.W., V.P., S.J.K, D.L.B., S.T. produced Env proteins and mapped antibody responses. J.A.A. and S.V. provided support and funding.

## Competing interests

The views expressed are those of the authors and should not be construed to represent the positions of the U.S. Army, the Department of Defense, or the Department of Health and Human Services. A patent application on the invention disclosed in this publication is filed. Hongjun Bai, Eric Lewitus and Morgane Rolland are the co-inventors. Other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-48139-x>.

**Correspondence** and requests for materials should be addressed to Morgane Rolland.

**Peer review information** *Nature Communications* thanks Tongqing Zhou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024