Deep Mining of the Human Antibody Repertoire Identifies Frequent and Immunogenetically Diverse CDRH3 Topologies Targetable by Vaccination

Rumi Habib[1,2,3]*, Shahlo O. Solieva[1,4]*, Zi Jie Lin[1,5]*, Sukanya Ghosh[1]*, Kelly Bayruns[1], Maya Singh[1], Colby J. Agostino[1,4], Nicholas J. Tursi[1,6], Kirsten J. Sowers[2,3], Jinwei Huang[1], Ryan S. Roark[2,3,7,8], Mansi Purwar[1,] Younghoon Park[2,3], Kasirajan Ayyanathan[2,3], Hui Li[2,3], John W. Carey[2,3], Amber Kim[1], Joyce Park[1], Madison E. McCanna[1], Ashwin N. Skelly[2,3,6], Neethu Chokkalingam[1], Sinja Kriete[1], Nicholas Shupin[1], Alana Huynh[1], Susanne Walker[1], Niklas Laenger[1], Jianqiu Du[1], Jiayan Cui[1], Beatrice H. Hahn[2,3], Ami Patel[1], Amelia Escolano[1,6], Peter D. Kwong[7,9], Lawrence Shapiro[7,8], Gregory R. Bowman[4], George M. Shaw[2,3]#, David B. Weiner[1]#, Jesper Pallesen[1]#, Daniel W. Kulp[1,3,4,6]#

[1]Vaccine and Immunotherapy Center, The Wistar Institute, Philadelphia, PA 19104, USA
[2]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[3]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Department of Chemistry, Indiana University, Bloomington, IN 47405, USA
[6]Penn Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[7]Aaron Diamond AIDS Research Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA
[8]Department of Biochemistry and Molecular Biophysics and Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA
[9]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA
*These authors contributed equally
#correspondence

Abstract

Many vaccination strategies against highly variable pathogens such as HIV-1 aim to elicit broadly neutralizing antibodies (bnAbs) with particular immunogenetic or structural features. The V2 apex of the HIV-1 Env protein is a promising target for a class of bnAbs that contain conserved structural motifs in the heavy chain complementarity determining region 3 (CDRH3). Here, we show that these structural motifs are targetable by vaccination by characterizing V2 apex 'axe-like' CDRH3s in the human repertoire and developing new immunogens capable of selectively engaging them. We determined the frequency and diversity of axe-like CDHR3s in healthy human donors using a series of structural informatics approaches finding these precursors in 86.5% of donors. Axe-targeting immunogens based on the HIV-1 Env Q23.17 were developed and bound axe-like precursors in cryo-EM structures, induced V2 apex-specific antibody responses in humanized mice, and induced axe-like heterologous neutralizing antibodies in rhesus macaques. These results unveil a new structure-guided immunoinformatic vaccine design paradigm that can be employed to elicit immunogenetically diverse yet structurally conserved classes of antibodies.

Introduction

Following the identification of SARS-CoV-2, a highly efficacious vaccine for the virus was developed in less than a year. Yet, despite over 35 years of research, there is still no protective vaccine for HIV. The HIV envelope glycoprotein (Env), which is the only target for neutralizing antibodies, has several features that make it a challenge for vaccine development. These include a large and dynamic glycan shield, conformational masking of epitopes, and extensive sequence diversity[1, 2, 3, 4]. As a result of these properties, the antibody response to Env is usually strain-specific[5]. However, 10-20% of infected individuals develop broadly neutralizing antibodies (bnAbs) targeting conserved epitopes on Env[6, 7, 8, 9, 10, 11, 12]. BnAbs have been shown to be protective against infection by diverse heterologous viruses in humans and rhesus macaques[13, 14, 15, 16].

Among the various classes of bnAbs, antibodies targeting the V2 apex are some of the most frequently elicited during infection, are highly potent, and have relatively straightforward developmental pathways[17, 18, 19]. These features make the apex a promising target for vaccine development. However, to date, vaccination has not been able to elicit V2 apex bnAbs in outbred animal models, and there is a need for immunogens capable of doing so. V2 apex bnAbs primarily rely on their long heavy chain third complementarity determining regions (CDRH3s) for binding to Env, forming distinct microdomains that contact several glycans and the C-strand of the V1V2 region[19, 20, 21, 22, 23, 24, 25, 26, 27]. These CDRH3s can be roughly classified into three categories based on their structural topology: “needle-like” in which β-hairpin CDRH3s bearing sulfated tyrosines at their tips extend into the trimer apex hole[19, 20, 21], “axe/hammerhead-like” in which CDRH3s form extended β-sheet interactions with the C-strand[22, 23, 24, 25], and “combined” which share

features of both[26, 27]. Because of these unusual CDRH3s, B cells expressing V2 apex bnAb precursors are estimated to have low frequencies in the naïve human B cell repertoire, making priming a major hurdle to their elicitation[20]. This challenge is compounded by the fact that, unlike bnAb precursors to other epitopes, V2 apex bnAbs utilize diverse V, D, and J genes, making it difficult to identify V2 apex bnAb precursors based on sequence features alone[20]. Yet, despite their diverse gene usage, V2 apex bnAbs form reproducible structural classes[22, 27]. Given the diversity of the antibody repertoire[28], we hypothesize that a myriad of genetically divergent CDRH3s can converge on structurally homologous solutions to engage the V2 apex. Here, we propose that V2 apex bnAb precursors can be identified by the structural topology of their CDRH3s, and that by targeting structural rather than immunogenetic classes of antibodies, we can capture that diversity and maximize the chances of successfully engaging rare bnAb precursors (Fig 1a).

Axe-like V2 apex bnAb lineages are typified by the human bnAb lineages PG9 and CH01[29, 30]. To understand the structural plasticity and frequency of axe-like bnAb lineage precursor B cells, we first conducted large-scale sequence and structure prediction screens to identify the frequency of CH01-like, PG9-like, and rhesus axe-like antibodies in the human B cell repertoire, identifying this class of antibodies as a promising target for vaccination. Next, we next designed and characterized a stable recombinant Env trimer immunogen, Q23.MD39, based on the clade A Env Q23.17, which exhibits affinity for the inferred germline precursors (iGLs) of the PG9 and CH01 lineages[22, 29, 31, 32]. Because we had previously shown that immunization with DNA-encoded native-like HIV trimers can enhance antibody responses compared to conventional protein bolus vaccination in WT mice[33], we first delivered Q23.MD39 using DNA immunization and showed that it can assemble into well-formed trimers *in vivo*. To further enhance the antigenicity of this

construct, we next engineered self-assembling ferritin nanoparticles decorated with Q23.MD39. These nanoparticles showed enhanced engagement of axe-like iGLs and could be delivered by DNA vaccination to assemble *in vivo*. We took advantage of the high affinity of Q23.MD39 for the CH01 iGL to solve an atomic resolution structure of this antibody, finding that it retains the same unique axe-like CDRH3 conformation as the mature CH01 antibody. We further enhanced the affinity of Q23.MD39 for additional axe-like iGLs through a single-round mammalian display method and showed that it can elicit axe-like neutralizing antibodies to the V2 apex as determined by cryo-EM in SHIV-infected macaques and V2 apex-targeted responses in humanized mouse models.

Results

Estimation of axe-like B cell frequency in the human repertoire

Currently, four axe-like V2 apex bnAb lineages have been described. These include the rhesus lineages V033-a and 41328-a[27], as well as the prototypical human lineages CH01[29] and PG9[23]. The V033-a, CH01, and PG9 bnAb lineages are promising targets for vaccine development given their high breadth, potency, and relatively low levels of somatic hypermutation. Even though the CH01 lineage has the second-shortest CDRH3 length of canonical human V2 apex bnAbs, previous estimates of B cell precursor frequency have suggested that CH01-like B cells may be one of the rarer classes of V2 apex bnAbs[20]. This was primarily due to CH01's rare D-J gene pairing and its skewed D gene placement in its CDRH3. For the CH01 bnAb lineage, the major contribution of the D gene to the CDRH3 of the antibody is a "YYGSG" sequence motif that forms the C-terminal section of the CDRH3 axe. We hypothesized that CDRH3-dominated bnAb precursor searching should take into account two previously unconsidered criteria: 1) that antibodies with alternative

CDRH3 lengths, J genes, and D gene placement could represent a larger pool of potential bnAb precursors, and 2) that only a subset of these sequences can form the requisite CDRH3 shape present in target antibody class. Recent advancements in B cell receptor repertoire sequencing, structure prediction tools, and computational power make it possible to apply these criteria to identify potential bnAb precursors through this new structure informatic approach.

The Observed Antibody Space (OAS) database, updated in 2021 for increased accessibility, contains over one billion antibody sequences, allowing for a rich, in-depth antibody repertoire analysis at a previously unachievable level[34]. Furthermore, it is now possible to model the antibody sequences within the OAS database with AlphaFold2 (AF2), which represents a fundamental breakthrough in the structure prediction field[35]. We conducted sequence searches within the OAS database to estimate frequencies of potential bnAb precursors. Our first set of searches were focused on the CH01 bnAb lineage. We conducted a precursor search of OAS imposing the following CH01 criteria: CDHR3 length of 26 amino acids, V3 gene family, J2 gene and D gene amino acids in the same location in the CDRH3 ("motif register"), similar to previously published criteria[20]. This CH01 CDRH3 criteria resulted in an average frequency estimation of about 0.01 precursors per million B cells ('Search 1' in Fig. 1b). As the JH gene amino acids do not make direct Env contacts, we removed the JH gene restriction and observed 1 precursor per million B cells, and an increase in the number of donors with sequence matches ('Search 2' in Fig. 1b). This drastic increase relative to a previous study[20] was expected since the J2 gene utilized by CH01 is the least frequently used JH gene used in humans[36]. The frequency then increased significantly to ~100 precursors per million B cells when the CDRH3 length (20-40 residues in length), D gene motif position (+/- 2 to 4 residues) and V gene restrictions were relaxed ('Search 3-7' in Fig. 1b).

Using the most relaxed search criteria, 'Search 7', we observed a frequency of 228 precursors per million B cells encoding a diverse set of 48,713 non-redundant antibody sequences. We further observed axe-like CH01 sequences in 33 out of the 37 donors who have >1 million sequences in OAS (Extended Figure 1a). Taken together, these results suggest that CH01-like bnAb precursors may be more frequent in humans than previously estimated.

To test whether the sequences found from our OAS searches are predicted to have an axe-shaped CDRH3 fold we modeled them with AF2. First, we determined that AF2 could predict the known axe-like CDRH3 of CH01 with high structural precision (< 2Å RMSD) (Fig. 1c). To assess which mutations in CH01 were required for AF2 to predict the axe-like structure, we conducted an *in silico* saturation mutagenesis screen by predicting each point mutation within CH01's CDRH3. Interestingly we observed positions D105, D106, and to a lesser extent positions G113, S114, G115, led to high structural deviation when mutated, suggesting these amino acids are playing a key role in this axe-like micro-fold of the CDRH3 of CH01 (Fig. 1d). While AF2 can struggle to model most CDRH3 structures[37, 38, 39], our data demonstrates that the axe-like CDRH3 structure of CH01 can be predicted with high accuracy.

Next, we sought to discover which of the BCR sequences from search criteria 7 could fold into axe-like structures with AF2. To automate detection of axe-like structures, we developed a general axe-like CDRH3 feature pattern that we termed 'turn-beta-turn' from axe harboring antibodies with known structures (Fig. 1e, Extended Data Fig. 2a). The 'turn-beta-turn' pattern consists of at least two consecutive 'turn' residues defined using canonical turn metrics and types[40, 41], followed by two or more residues with 'beta' backbone geometry at least 20Å away from the start of the

CDRH3 and capped by at least two more consecutive 'turn' residues (see methods). We validated the specificity of this 'turn-beta-turn' pattern by searching 610 diverse antibody structures which target many different proteins from the SAbDab database[42] and found five known axe-like structures harboring a 'turn-beta-turn' pattern (CH01 lineage, PDB IDs 3TCL, 3U46, 3U4B, and CAP256-VRC26 lineage antibodies, PDB IDs: 4ORD1, 4ORG), and only three other antibodies which exhibit an axe-like fold (Extended Data Fig. 2b). Of the search 7 sequences, we found 5,146 of the sequences have axe-like structures. We found that, out of the donors with >1 million antibody sequences, 10.8% do not have any axe-like sequence or structure shaped antibodies, 18.9% have axe-like sequence signatures without an axe shape and 78.4% have a predicted CH01-like axe shaped CDRH3. To further support AF2 predictions, we used molecular dynamics (MD) simulations to assess the stability of the axe microfolds in our structural models[43]. We conducted 10 replicate MD simulations of 250 ns each (aggregate of 2.5 μs per antibody) for four AF2 models and CH04 (PDB ID: 3TCL) as a control. We applied our axe structural definition to the simulations to determine the frequency of axe-like structures within each simulation. We found that the CDRH3 maintained an axe-like structure for a majority of the time in the CH04 simulations (Fig. 1f). Three antibodies that matched the 'turn-beta-turn' structural motif (seq10941, seq15061, seq20621, shown in blue in Fig. 1f) also maintained an axe-like structure in a majority of the simulations. The antibody that did not match the 'turn-beta-turn' structural motif (seq10821, shown in red in Fig 1f) had a drastically lower axe-like frequency compared to CH04. These simulations give us insight into the dynamics of the CDRH3 loops and show that this approach could be a useful tool for verifying the stability of CDRH3s with predicted micro-folds, such as the axe shape.

The axe-like PG9 CDRH3 has a sequence signature that differs from CH01, and we sought to understand how frequently the PG9 sequence and structure were found in human repertoires. We employed our expanded sequence and structure modeling pipeline on PG9 as we have for CH01 (Extended Data Fig. 1b). We found PG9-like sequences in 35 out of 37 highly sequenced donors (>1M sequences) (Extended Data Fig. 1b). From the 38,427 PG9-like sequences, we found that 6,965 sequences had axe-like structures and that 81% of highly sequenced donors have axe-like antibodies. These searches show that, similarly to CH01, the human antibody repertoire likely contains an immunogenetically diverse set of PG9-like antibodies (Extended Data Fig. 1c).

Rhesus V2 apex bnAb motifs are frequent in the human repertoire

The recent discovery of rhesus-derived axe-like bnAb lineages (V033-a and 41238-a) gives us additional targets for vaccine design[27]. Interestingly, although human V2 apex bnAbs use diverse V, D, and J genes, a marked feature of rhesus V2 apex bnAbs is their shared usage of the IGHD3-15*01 gene[27]. These bnAbs invariably incorporate a D gene-templated "EDDYG" motif that is absent from D genes in the human repertoire. While humans do not contain a germline D gene that encodes "EDDYG", an outstanding question is how frequently V-D-J rearranged antibodies in the human repertoire contain this rhesus sequence motif. To explore this, we conducted repertoire and structural searches to understand how often an "EDDYG" motif is incorporated into human CDRH3s.

We performed three searches using either a larger rhesus D gene sequence of "EDDYGYYT", the core rhesus D gene motif of "EDDYG", and the motif found in the V033-a lineage "[EG]DDYG" with two or more N-terminal amino acids and six or more C-terminal residues relative to the motif

(Fig. 2a, Extended Data Fig. 1d). We observed two donors harbored antibodies with the larger rhesus D gene sequence (Fig. 2a, Search 1). CDRH3s found in donors Hu-CD1 and Hu-BD3[44] are strikingly similar to rhesus lineages V033-a and 41328-a, suggesting a subset of the human population may have rhesus-like V2 apex bnAb precursors (Fig. 2b). While the full D-gene motif was rare amongst donors, the core D gene motifs were found in 89% and 97% of donors (Fig. 2a, Search 2 and Search 3). For CDRH3 lengths of >23, the core D-gene motifs were found at 6 per million and 8 per million antibody sequences. To further probe the immunogenetics of antibodies with these core D-gene motifs, we cataloged the human D-genes used in antibodies found in Search 2. We observed that IGHD4-11*01, IGHD4-17*01 and IGHD4/OR15-4a*01 each contained the 'DYG' sequence (Fig. 2b) and were all heavily enriched in the core D-gene search relative the frequency of these D genes found across OAS (Fig. 2c). We folded Search 3 CDRH3s using AF2 and found that 267 sequences out of the 24,696 sequences form axe-like structures (three axe-like and one non-axe examples are shown in Fig. 2d). Indeed, MD simulations of rhesus-like axe CDRH3s demonstrate that this fold is highly stable (Fig. 2e). Therefore, despite the nearly universal usage of an IGHD3-15*01 D gene-derived "EDDYG" motif in rhesus V2 apex bnAbs, the absence of an orthologous human D gene may not limit the value of the rhesus model for V2 apex bnAb immunogen design and testing.

While only four axe-like V2 apex bnAbs have thus far been described, it is likely that far more exist with these sequence and structure features. Using the four axe-like bnAbs as a reference, we found 12,378 total axe-like antibody sequences. Of the donors with > 1 million antibody sequences in OAS, we found that 1 donor did not have any axe-like CDRH3s, but 97% of donors had antibodies with axe-like sequence motifs and 86% of donors had antibodies with axe-like structural

motifs (Fig. 2f). Collectively, these studies reveal that not only are axe-like sequence motifs identified at high frequency in the human antibody repertoire, but also that antibodies with predicted axe-like structures are observed in the overwhelming majority of donors, pointing to this structural class of bnAbs as a promising vaccine target.

Design, characterization, and nucleic acid delivery of Q23.17-based native-like trimer and nanoparticle immunogens

To design an immunogen that can engage axe-like V2 apex bnAb precursors, we sought to develop a stable trimeric Env protein that could easily multimerize on nanoparticles and be delivered by nucleic acid platforms while retaining its structure *in vivo*. To discover a starting immunogen sequence, we searched for Envs that have detectable affinity for inferred germline precursors of V2 apex bnAbs, as we reasoned that these may be more favorable for V2 apex priming. Several groups have conduct screens of Envs to identify such strains with "permissive" V2 epitopes that have natural propensities for V2 apex iGL engagement[22, 29, 31]. We selected a clade A Env, Q23.17[45], as our platform for immunogen development because of its relatively high binding affinity for the iGLs of the axe-like CH01 and PG9 bnAb lineages[22, 29, 31] and sensitivity to mature V2 apex bnAbs. We note that this Env also contains many of the engineered "Apex-GT5" residues recently reported to sensitize the Env BG505 to precursors of the PCT64 lineage[20].

To design a stabilized recombinant Q23.17 trimer, we tested several published Env trimer stabilization strategies including MD39[46], MD64[47], v5[48], 7S[49], RnS-DS[50, 51], and Olio6[47] mutations. We found that the Q23.MD39 and Q23.7S demonstrated superior expression and antigenicity (data not shown). The MD39 mutations are a subset of the 7S mutations, and thus we

concluded that the MD39 mutations were primarily responsible for the stable phenotype of the trimer, focusing on Q23.MD39 for further investigation. Q23.MD39 bound strongly by ELISA to the trimer-specific or trimer-preferring mAbs PGDM1400, PGT151, and VRC34 and bound poorly to V2p-, CD4i-, and linear V3-specific non-neutralizing antibodies, indicating an antigenic profile consistent with a stable prefusion trimer (Fig. 3a). In contrast, Q23.gp120.foldon, a construct consisting of three gp120 domains loosely held together by a foldon trimerization domain, showed high binding to non-nAbs and poor binding to quaternary structure-dependent bnAbs (Fig. 3a). We confirmed the size and homogeneity of Q23.MD39 by size exclusion chromatography multiple angle light scattering (SEC-MALS) which showed that Q23.MD39 formed trimers of the expected size (Extended Data Fig. 3b).

There is extensive evidence that the multimeric display of vaccine antigens significantly enhances their immunogenicity through avidity effects and enhanced trafficking, among other effects[52, 53, 54, 55]. We have previously shown that DNA-launched nanoparticles (DLNPs) are capable of self-assembly *in vivo* and are highly immunogenic[56]. To this end, we engineered Q23.Ferritin, a self-assembling ferritin nanoparticle decorated with our Q23.MD39 trimer using a short glycine-serine linker (Extended Data Fig. 7). Using NS-EM and antigenic profiling by ELISA, we showed that this nanoparticle retains the well-formed phenotype of the Q23.MD39 trimer, expresses primarily as a well-formed nanoparticle by SEC, and has significantly enhanced binding to the axe-like iGLs (Fig. 3b-c, Extended Data Fig. 3c). We measured the affinity of our design in trimeric, nanoparticle and gp120 forms for the CH01, PG9, V033, and 41328 iGLs which demonstrated that Q23-based immunogens have affinity for axe-like iGL precursor antibodies (Fig. 3c).

DNA delivery of HIV Env trimers is a promising approach for driving unique antibody responses in WT mice[33]. Importantly, SEC analysis showed that Q23.MD39 expressed primarily as a trimer (Extended Data Fig. 3a). Unlike purification of recombinant Env immunogens, nucleic acid delivery strategies are unable to purify out poorly folded or unassembled Env proteins. To show that the trimer was capable of homogenous assembly and expression *in vivo*, we immunized WT BALB/c mice with a DNA vaccine encoding the Q23.MD39 construct. Mice were immunized with 5ug, 10ug or 25ug of DNA encoded Q23.MD39 with or without adjuvanting with plasmid IL-12 at weeks 0, 3 and 6, following which antibody responses were tracked 2 weeks following each immunization (Fig. 3d, Extended Data Fig. 4a). The immunized animals developed strong binding antibody responses, demonstrating that the trimer was immunogenic when delivered as a DNA vaccine (Fig. 3d, Extended Data Fig. 4a). However, they failed to develop neutralizing antibody responses or V2 apex-targeted responses (data not shown). To determine if the trimer structure and concomitantly the antigenicity of Q23.MD39 produced *in vivo* was maintained, we employed our recently developed Antigen Conformation Tracing In Vivo by ELISA (ACTIVE) assay[33] (Fig. 3e). Mice were administered with DNA containing Q23.MD39 or BG505.MD39 intramuscularly. *In vivo* assembled trimers from harvested muscle tissue of mice were analyzed by ELISA for binding to various bnAbs, non-nAbs, and the CH01 iGL at several timepoints post-administration (Extended Data Fig. 4b). High binding of muscle homogenate expressing Q23.MD39 to trimer-specific bnAbs such as PGDM1400, and low binding to non-nAbs such as 17B, indicated that Q23.MD39 retains its favorable antigenic profile when expressed *in vivo* (Fig. 3e, Extended Data Fig. 4b). BG505.MD39, a similarly stabilized and closely related Env, showed a similar binding profile to Q23.MD39. However, only Q23.MD39 was able to bind to CH01 iGL *in vivo* (Fig. 3e, Extended Data Fig. 4b). Escalating-dosing immunization regimes delivered over the course of a

few weeks have been shown to significantly improve immune responses through an increase in durable germinal center responses[57, 58]. Here, with a single DNA immunization, we detected escalating amounts of intact prefusion Env trimer in muscle tissue for at least 3 weeks post-immunization (Fig 3e, Extended Data Fig. 4b). This data shows that Q23.MD39 retains its sensitivity for V2 bnAb iGLs *in vivo* and marks the first direct demonstration of the ability of an Env trimer produced *in vivo* to engage a bnAb precursor.

We next sought to understand whether Q23.Ferritin could also be delivered by DNA vaccines. We have previously demonstrated that DNA immunizations can be used to deliver self-assembling nanoparticle immunogens such as eOD-GT8 60mer[56], and others have shown that Env trimer nanoparticles can be delivered by mRNA[59]. The favorable expression profile of Q23.Ferritin combined with our use of a genetic fusion approach again allowed us to deliver the immunogen via DNA immunization (Extended Data Fig. 3c). DNA delivery of the Q23.Ferritin construct with or without plasmid IL-12 in WT mice elicited high-titer binding antibody responses, demonstrating that Q23.Ferritin is immunogenic when delivered by DNA vaccination (Extended Data Fig. 4a).

Cryo-EM structure of Q23.MD39 bound to the CH01 iGL reveals key molecular contacts of axe-like precursor antibodies

To gain structural insights into how Q23.MD39 binds to the axe-like CH01 iGL with high affinity, we determined atomic resolution cryo-EM structures of Q23.MD39 in complex with the Env gp120-gp41 interface bnAb 35O22 alone (3.19Å) and in complex with the CH01 iGL and 35O22 (3.35Å) (Fig. 4a). We observed a prefusion-closed Q23.MD39 Env trimer with high structural similarity to BG505.SOSIP.664, further indicating that Q23.MD39 is a well-formed native-like

trimer. Overlay of our CH01 iGL antibody structure with mature antibody structures of lineages members CH03 and CH04 showed identical conformations for all CDR loops except the CDRL1. This may be due to CH01 iGL lacking the somatic hypermutation of the mature antibodies or because the CH01 iGL is the only antibody structure of this lineage solved in complex with a full Env (Fig. 4b). The CH01 iGL antibody bound to Q23.MD39 Env trimer with a 1:1 stoichiometry (Fig. 4a) and engaged the C-strand using an intricate backbone hydrogen bond network similar to that found in CH03/CH04 structures in complex with a C-strand scaffolded protein (Fig. 4c)[22]. 4c)[22]. The sidechains of the C-strand on protomer A interact with positions within the axe shape: Y100g with R169 and Y100h with K168 (Fig. 4c, top middle and top right panels). The CDRH1 may interact with glycan at N156A, although the density for the full glycan could not be built. The sidechains of CDRH1 also interact with both the K171 and Y173 on the gp120A C-strand which are conserved residues among different HIV isolates at these positions[31] (Fig. 4c bottom left panel). The sidechains on the CDRH2 and CDRH3 form extensive contacts with the N160A glycan (Fig. 4c bottom middle panel). The structure of the CH01 iGL maintains the axe-like CDRH3 of the mature CH01-lineage bnAbs, indicating that the unique CDRH3 topology is likely a feature present in the unmutated ancestor of the bnAb lineage. This finding corroborates the structure-guided immunoinformatic approach to bnAb precursor identification conducted in this study, as it suggests that CH01-like germline precursors may also harbor this unique CDRH3 shape.

Structures of mature CH03 and CH04 were previously determined using trimeric scaffold mimics with CAP256.SU (CH03) and A244 (CH04) V1/V2 sequences[22]. The key heavy chain epitope-paratope interactions identified in their structures were very similar to our CH01 iGL structure determined using a trimeric Env. However, due to the unnatural trimeric scaffold protein used in

CH03/CH04 structures, the light chain interactions were unclear. One notable feature we discovered that was not revealed in previous structures was the close proximity between CDRL2 and the V2b hypervariable loop from gp120B (Fig 4c, bottom right panel). Furthermore, while the N139 glycan was previously proposed as a second N160 glycan mimic[22] we did not observe binding of the CH01 iGL to glycans on neighboring protomers (Fig 4d). However, we observed by glycan molecular modeling that N160B may be capable of making favorable interactions between CDRL1 and the back of CDRH3 (Fig. 4d).

Through this structural analysis, we hypothesize there are three major contributing factors to the high affinity of Q23.17 for the axe-like CH01 iGL: (i) C-strand residues that are optimal for interactions with the CH01 iGL CDRH3, (ii) a lack of disfavored glycans at the apex epitope such as N130 and loop V2b glycans as previously reported[31], and (iii) the length of loop V2b being within the threshold accommodated by the CH01 iGL, in particular by CDRL2 and CDRH3. This observation suggests V2b could be a promising location for incorporating designed loops or mutations, with the closest contact in our structure being CH01 iGL HC Y100h and gp120B E185 (4.82 Å). With this structure, we can now visualize an axe-like precursor antibody engaging with the C-strand and further rationalize potential mutations that may increase affinity for CH01-like axe shaped precursors.

Mammalian display mutagenesis to enhance axe-like iGL binding

While Q23.MD39 has modest binding for known axe-like iGLs, we hypothesized that further optimization of the Env is likely required to elicit true axe-like iGLs antibodies with high efficiency. The iGLs available for axe-like human bnAb lineages contain affinity matured non-

templated regions which comprise a significant portion of the epitope-paratope interface, and so true naïve B cell precursors are likely to have lower affinities for Env. To this end, we used a novel mammalian display mutagenesis approach to enhance the affinity of Q23.MD39 for axe-like iGLs. We optimized a previous mammalian display protocol[46], by performing single round sorting followed by next-generation sequencing. Briefly, Q23.MD39 was membrane anchored using a PDGFR transmembrane domain and a lentiviral scanning NNK library spanning Env residues 110-192 was created with silent barcoding mutations flanking each NNK codon. The top 5% of cells were sorted against axe-like iGLs and deep sequenced (Fig. 5a). To identify affinity-enhancing mutations, the frequency of particular mutants in the sorted population was compared to the frequency of the same mutation in unsorted cells. This method enabled the identification of a significantly higher number of affinity-enhancing mutations with greater efficiency and speed compared to traditional mammalian display mutagenesis approaches (Fig. 5b).

As a proof of principle, we sorted our Q23 library following staining with the V033 iGL. Sequencing of sorted cells led to the identification of mutations spanning the length of the library that were either enriched or selected against, providing an array of potential affinity enhancing mutations (Fig. 5b). From these enrichments, we filtered positions with low numbers of counts among either the selected or unselected cells, enrichments that arose from low numbers of a mutation appearing in the unsorted library, mutations that occurred at positions where the WT amino acid was also enriched, and mutations that altered or introduced cysteine or proline residues. A subset of 24 especially enriched mutations that passed these criteria were synthesized as soluble Q23.MD39 trimer and then expressed, purified and tested for enhanced binding against the V033 iGL. We observed that the majority of the mutations conferred increased binding to the antibody

by ELISA ranging from low (<1.5 fold) to high (>3 fold) increase in binding (Fig. 5c). We next asked whether these mutations have an additive enhancement of binding affinity when combined. We selected high binding mutations, M161T, L165F, and K168V and progressively incorporated them into Q23.MD39. ELISA binding to the V033 iGL showed increases in affinity as each mutation was incorporated (Fig. 5d). Subsequent addition of a K117A, identified through sorts against other axe-like antibodies, further increased the affinity of the V033 iGL, ultimately resulting in a construct named Q23.V033GT. We assessed binding against V033 iGL by SPR and found that the unmodified Q23.MD39 did not have measurable affinity (up to 1uM), however, Q23.V033GT bound at a K_D of 100nM (Extended Data Fig. 8).

Construction of and infection with a germline targeting SHIV

We next asked whether our Q23.V033GT iGL-targeted immunogen would be able to elicit structurally homologous antibodies in rhesus macaques. We introduced our germline targeting mutations into a replicating SHIV[60, 61] for the following reasons: (i) the high and sustained antigenic loads during viral infection maximize the probability of engaging bnAb precursor B cells, similar to extended dosing or using nucleic acid delivery; (ii) SHIVs coevolve with the bnAb lineage, selecting for affinity maturation and the acquisition of neutralization breadth, thereby allowing us to verify that primed B cells are genuine bnAb precursors; (iii) the pattern of Env escape can aid in the design of boosting immunogens, since escape variants select for affinity maturation and neutralization breadth; and (iv) sequencing of Env escape serves as a sensitive indicator of nAb-targeted epitopes, allowing identification of both on- and off-target nAb lineages[19, 62].

To construct this germline targeting SHIV, we first sought to incorporate our mammalian display-identified mutations into the WT SHIV-Q23[60]. We found that the V033-targeting mutations in Q23.V033GT could be incorporated into the virus without impacting viral infectivity (Extended Data Fig. 5a). A SHIV bearing all 4 of these mutations, SHIV-Q23.V033GT maintained infectivity and p27 antigen titers similar to WT (Extended Data Fig. 5a). Strikingly, neutralization assays showed an over 50-fold increase in neutralization potency by the V033 iGL against the resulting SHIV-Q23.V033GT, verifying that the germline targeting mutations retained their effect in the context of an infectious virion (Fig. 6a). Antigenic profiling of the germline targeting SHIV demonstrated that the Q23.V033GT Env retained a closed prefusion state, showing no sensitivity to a panel of CD4i and V2p antibodies consisting of CH58, CAP228-3D, 447-52D, 17b, A32, 697-D, and 1393. The SHIV does show a modest increase in neutralization sensitivity to the antibody 3074, suggesting a more exposed V3 loop (Fig. 6a).

Three rhesus macaques were infected with SHIV-Q23.V033GT, as previously described[19, 60]. All rhesus macaques were productively infected, with peak viral loads between 10^7-10^8 vRNA copies/ml of plasma 2 weeks post infection (Extended Data Fig. 5b). Viral loads for one rhesus macaque, CL60, were especially high during the first 12 weeks of infection (>$5x10^6$ vRNA copies/ml). This animal failed to develop any detectable antibody response to SHIV infection and succumbed to rapid clinical progression of AIDS[63] (Extended Data Fig. 5b). The other two macaques exhibited stable setpoint viral loads between $5x10^4$-$5x10^5$ vRNA copies/ml (Extended Data Fig. 5b). Each mounted autologous Q23.V033GT and tier 1A nAb responses by 12 weeks post infection (Fig. 6b). Interestingly, the entire nAb response of rhesus macaque CK52 at week 12 was immunofocused to the V2 apex, as the response was dependent on the introduced germline

targeting mutations (Fig. 6b). In both CK11 and CK52, we measured neutralization of V2 apex site-directed mutants (N160K and N187S) and observed striking changes in neutralization, thus confirming the presence of V2 apex targeted nAb responses (Fig. 6b). Single-genome sequencing of the circulating plasma virus quasispecies in CK52 and CK11 showed strong selection at the V2 apex at week 12 post-infection, with the addition of glycans to the V2b hypervariable loop and selection for neutralization escape at V2 apex C-strand residues 170-172 (Fig. 6c Extended Data Fig 5. d-e).

To understand how efficiently these C-strand targeted nAbs were being primed and affinity matured by SHIV-Q23.V033GT compared to SHIVs bearing wild-type Envs, we compared the kinetics of V2 apex targeted nAb elicitation in CK52 and CK11 with 6 macaques infected with WT SHIV-BG505.N332 and 4 macaques infected with WT SHIV-Q23.17 over the first 24 weeks of infection (Fig. 6c-d). We used kinetics of epitope escape as a measure of C-strand targeted nAb elicitation, since virus escape has been shown to precede or coincide with the first detection of epitope specific nAbs[64, 65]. To quantify escape, we calculated Hamming distance sequence similarity measurements in the V2 apex epitope relative to the starting SHIV sequence. At week 12 post infection, SHIV-Q23.V033GT infected macaques exhibited significantly more selection at the V2 apex than the SHIV-BG505.N332 ($p = 0.0102$) or SHIV-Q23.17 ($p = 0.0170$) infected macaques (Fig. 6c-d, p values determined by one-way ANOVA followed by Tukey's post-hoc test). The acceleration of selective pressure on the V2 apex in SHIV-Q23.V033GT infected macaques was even more apparent at week 24 (Fig. 6d, $p < 0.0001$ against both BG505.N332 and Q23.17 by one-way ANOVA followed by Tukey's post-hoc test). These data suggest that infection with SHIV-Q23.V033GT resulted in significantly faster and stronger selection at the V2 apex as

compared to infection with primary HIV Envs, highlighting the ability of this Env to immunofocus early responses to the V2 apex. These findings suggest that a germline targeted SHIV with enhanced affinity for V2 apex precursors can consistently elicit V2 apex C-strand targeted nAbs early in infection.

To determine if the macaques with V2 apex responses also generated neutralization breath, we measured neutralization activity against a panel of nine heterologous tier 2 viruses. Heterologous neutralization was detectable in both macaques CK11 and CK52 at 28 weeks post-infection. Both plasma samples neutralized BG505 at titers of 1:20 to 1:30. CK52 plasma also neutralized C1080 at a titer of 1:200 and CH505 at a titer of 1:20 (Fig. 6e). For both monkeys, heterologous plasma neutralization was mapped to the V2 apex C-strand using site-directed mutants (Extended Data Fig. 5f). Finally, we employed a powerful imaging technique, Epitope Mapping by Polyclonal Electron Microscopy (EMPEM), using both negative stain (nsEMPEM)[66] and cryo-EM (cryoEMPEM)[67] to confirm the presence of V2 apex directed antibodies in SHIV immune sera. The nsEMPEM data revealed strongly epitope focused responses, as we observed only V2 apex and gp41 binding antibodies at week 24 in macaque CK11 and only V2 apex antibodies in macaque CK52 (Fig 6f, Extended Data Fig 6a, b). Further, we observe multiple V2 apex targeting antibodies with an off-center angle of approach characteristic of axe-like bnAbs. By fitting high resolution Env-axe antibody complexes into the CK52 nsEMPEM density, we observed significant overlap with that of the V033, 41328 and CH01 iGL Fabs, indicating the presence of a structurally homologous antibody (Extended Data Fig 6c). Next, we performed cryo-EMPEM on the CK52 sample at week 24 and obtained a 4.2Å resolution structure (Fig 6g). Consistent with nsEMPEM results, the cryo-EMPEM confirmed V2 apex engagement targeting the C strand region similar to

the approach of axe-like bnAbs (Fig 6h). Further refinement of antibody CK52.1 revealed turn-beta-turn density likely contributed by the CDRH3 interacting with the C-Strand as expected for axe-like engagement of Env V2 apex (Fig 6h, 6i). Together, the escape mutations, neutralization sensitivity and structural data demonstrate that SHIV-Q23.V033GT re-elicited axe-like antibodies to the Env V2 apex early in infection.

Construction of and immunization with a pan-axe targeting immunogen

We next sought to simultaneously enhance the affinity of Q23.MD39 for all known axe-like iGL antibodies. We reasoned that an immunogen with increased affinity for multiple immunogenetically distinct iGLs of this class would be generally more efficient in binding naïve germline B cell precursors harboring this CDRH3 topology. To do this, we took advantage of the large number of enrichments identified by our mammalian display approach. By generating matrices of enrichments following sorting of our NNK library against several iGLs of interest and looking for enrichments in common, we were able to identify a set of mutations that simultaneously increased affinity for the V033, 41328, PG9 and CH01 iGLs (Fig. 7a-b, Extended Data Fig 7). The mutations in this trimer, Q23.RH-GT (for Rhesus-Human Germline Targeting), were all distal to the C-strand, the primary point of contact at the epitope-paratope interface, and instead may enhance affinity through second shell interactions or indirect restructuring of the Env apex, making it more amenable to engagement by axe-like antibodies.

To test the ability of this immunogen to elicit axe-like bnAb precursor lineages, we immunized hD3-3/JH6 mice[68]. This model has the mouse DQ52 and JH1-4 segments replaced by human IGHD3-3 and IGHJ6 segments, respectively, allowing the B cells to incorporate these genes during

VDJ recombination and produce diverse CDRH3s of lengths longer than WT mice and comparable to human V2 apex bnAbs that incorporate these same genes[68]. The hD3-3/JH6 mice (n=3) were immunized with 25µg of DNA encoding an RH-GT trimer along with IL-12 adjuvant, and boosted 3 weeks post prime (Fig. 7c). All mice showed strong binding to the Q23.RH-GT trimer but diminished binding to a V2 apex C-strand epitope knockout trimer (Q23.MD39 R169E/K71E), indicating the presence of robust C-strand directed responses (Fig. 7c). Control mice immunized with Q23.ferritin immunogen containing a WT V2 apex epitope. These mice failed to induce the same level of epitope-specific response. These data indicate that the Q23.RH-GT immunogen was successfully able to induce V2 apex responses in a stringent rearranging mouse model.

Discussion

Germline targeting approaches have successfully primed HIV bnAb precursors for lineages where the bnAb paratope consists of mostly germline-templated amino acids[69]. However, bnAbs with CDRH3-dominated modes of binding have proven more challenging, with V2 apex bnAbs being a particularly difficult target given their long CDRH3s containing extensive non-templated regions. Groups have previously described approaches to eliciting antibodies with predominantly CDRH3-mediated binding through the design of immunogens that bind B cells with strictly defined criteria, consisting of invariant V, D, and J gene templated regions but diverse non-templated junctions[70]. A major limitation of these approaches is that they are predicated on the selection of a single suitable template antibody lineage. Here we expand on this immunogen design approach by instead targeting a structural class of V2 apex bnAbs, leveraging immunogenetically diverse but structurally homologous antibodies to optimize a trimer immunogen for enhanced affinity of axe-like V2 apex precursors and show elicitation of axe-like antibodies by cryo-EM.

Our findings suggest that the frequency of antibodies with axe-like CDRH3 topologies is likely far higher than previous estimates based solely on the immunogenetic features of B cells[20]. Here we build on previous work by leveraging recent advancements in deep learning structure prediction to incorporate predicted a unique B cell CDRH3 topology into our immunoinformatic searches. We note three key observations that support our structurally focused approach to precursor targeting. First, our structure of the CH01 iGL retains its distinct axe-like CDRH3 topology, as was seen with PG9 iGL[20]. This suggests axe-like precursors acquire their distinct CDRH3 conformation during VDJ recombination, as opposed to through affinity maturation, and that additional precursors may be identified by the presence of this distinct CDRH3 conformation. Second, we demonstrate that AlphaFold2 is capable of folding axe-like CDRH3 structures and that we can use a structurally conserved 'turn-loop-turn' definition to identify new axe-like CDRH3s in an automated fashion. The third observation is the recent isolation of rhesus V2 apex bnAbs that recapitulate the axe shape found in C-strand targeting human V2 apex bnAbs despite using rhesus-specific D genes[27]. This structural homology underscores the ability of immunogenetically diverse CDRH3s to adopt similar solutions to C-strand recognition. Furthermore, the identification of B cells bearing the key rhesus "EDDYG" motif in the human repertoire supports the use of rhesus V2 apex bnAb precursors for human vaccinology. With the advent of more accurate protein structure prediction tools such as AF3[71] and Ig-Fold[38], we expect our new approach to provide an increasingly more powerful method for further precursor analyses.

Our data show that germline-targeting mutations can be engineered into Env Q23.17 to elicit V2 apex-specific responses in mouse models with long CDRH3s and induce antibodies that harbor

V2 apex targeting turn-beta-turn axe motif and neutralize heterologous tier-2 viruses in rhesus macaques. Our data suggest that the HuD3-3/J6 mouse model may be of utility to drive V2 apex-specific responses. However, it remains to be seen whether these antibodies harbor axe-like CDRH3s and can be effectively boosted to obtain broad neutralization. The SHIV infection model has been employed to study Env-antibody coevolution resulting in identification of bnAbs that recapitulate sequence and structural features of human bnAbs[19, 27]. We exploited this novel platform for evaluation of germline-targeting Env. One limitation on our interpretation is the limited number of monkeys available and that one exhibited high persisting plasma viral load and progression to AIDS and death. In both evaluable remaining monkeys, the GT Env was able to rapidly elicit escape mutations at the V2 apex and drive limited heterologous neutralization. However, due to the stochastic nature of SHIV infections, we cannot easily examine specific reversion mutations of the GT Env back toward a WT V2 apex epitope. Therefore, these studies should be followed up with rationally designed booster immunogens in heterologous vaccination experiments.

In summary, we propose that it is possible to design immunogens that target the broad axe-like structural class of CDRH3s. The Q23.RH-GT is such an immunogen that has improved affinity for all four axe-like inferred germline antibodies with the ability to activate similar antibodies in animal models. We demonstrate that electron microscopy can be employed as a critical tool for validating CDRH3 structural motifs induced by SHIV infection. We find it is also key to determine the frequency of B cells harboring a defined structural class of CDRH3s, as we showed here that axe-like precursor antibodies are present in most people. As the axe microdomain is not the only structural class of CDRH3s, we anticipate our structure-guided immunoinformatic approach will

open the door for many new studies on additional important epitopes on HIV-1 Env, as well as other glycoproteins with rich sets of antibody structures such as SARS-CoV-2, Influenza and RSV.

Acknowledgements

We would like to thank Drs. Frederick Alt and Ming Tian for provision of the human D3-3/J6 mouse model. In addition, we would like to thank the Wistar Animal, Molecular Screening & Protein Expression, and Flow Cytometry Facilities for experimental support. Funding was provided under NIH grants U19 AI166916 and R01 AI165080. R.H is supported by a Training Grant in HIV Pathogenesis (T32-AI007632). S.O.S. is supported by a Training Grant in Structural Biology and Molecular Biophysics (T32-GM132039).

Author contributions

S.O.S., C.J.A., G.R.B., and D.W.K. conducted in silico antibody and repertoire analyses. R.H., K.B., J.H., A.K., J.P., M.E.M., N.C., S.K., N.S., A.H., S.W., N.L., and D.W.K contributed to the design, expression, and in vitro characterization of immunogens. K.B., Y.P., A.K., J.P., M.E.M., N.C., S.K., N.S., and A.H. performed antibody expression and purification. K.J.S and H.L. cloned and characterized the SHIVs described in this study. R.H., K.J.S., Y.P., K.A., H.L., J.W.C., and A.N.S. collected and/or analyzed data from SHIV infected macaques. R.H., K.B., S.G., M.S., N.J.T., and M.P. conducted and/or analyzed murine immunization experiments. Z.J.L, J.D., and J.C. performed NS-EM, NS-EMPEM, cryo-EMPEM, and/or cryo-EM imaging and structure determination. R.S.R., A.E., B.H.H., P.D.K, and L.S. provided critical reagents or animal models. R.H., S.O.S., Z.J.L., S.G., and D.W.K. designed the figures and wrote the manuscript. A.P.,

G.R.B., G.M.S., D.B.W., J.P., and D.W.K provided experimental oversight and funding. All authors reviewed the manuscript.

Conflicts

R.H. and D.W.K. have pending intellectual property interests related to immunogens and approaches discussed in the manuscript. D.W.K. discloses a paid association with Inovio (SRA). D.B.W. has received grant funding, participates in industry collaborations, has received speaking honoraria, and has received fees for consulting, including serving on scientific review committees and board series. Remuneration received by D.B.W. includes direct payments and stock or stock options. D.B.W. also discloses the following paid associations with commercial partners: GeneOne (consultant), Geneos (advisory board), AstraZeneca (advisory board, speaker), Inovio (BOD, SRA, Stock), Sanofi (advisory board) and BBI (advisory board).

Methods

Inferred germline precursor (iGL) inference

iGLs for rhesus bnAb lineages 41328-a and V033-a were constructed by reverting V, D and J gene templated amino acids to their germline gene sequences. Non-templated CDRH3 residues were left unchanged from the mature bnAb sequence. Sequences for the iGLs inferred in this paper are available at Genbank accession numbers XXXX. For human bnAb lineages CH01 and PG9, previously reported iGLs were used (originally referred to as CH01 UA3 and PG9 UA respectively)[32].

Repertoire frequency analysis

Our antibody searches used the heavy chain only sequences from healthy, HIV-unexposed human donors. This resulted in over 1 billion sequences. We calculated the B cell frequencies within each donor for each search by taking the number of sequence hits and dividing that by the number of sequences the donor had in the OAS database. We then removed redundant sequences within each donor. For each donor, we first grouped the donor's sequences into groups based on CDRH3 length. Then, we compared CDRH3 sequences in a pairwise fashion for each length group. If the number of different residues between two CDRH3s was less than 15% of the CDRH3 length (greater than 85% similarity within CDRH3) then the sequence was considered redundant and was removed. For a CDRH3 of length 20, this would be a difference of 3 or fewer amino acids and for length 40, a difference of 6 or fewer amino acids. Our exact searches and example scripts are deposited in the GitHub repository: https://github.com/ssolieva/q23-manuscript.git

AlphaFold2 modeling

We generated 1 seed per model using ColabFold[72] (AlphaFold2-multimer[73]) with custom templates for our structure predictions. We created custom templates by removing the CDRH3 loop from the CH01, V033, and PG9 structures. The custom templates were used without a multiple sequence alignment and the CDRH3 region was untemplated. Each antibody was modeled in about 6 minutes. We paired our non-redundant variable heavy chain sequences from the OAS searches with the mature variable light chain from the antibody the search was based on (CH01, PG9, V033).

Structural axe definition

We created a structural definition of an axe using the four available axe antibodies. We defined an axe as having the following pattern: at least two turn residues, followed by at least two beta residues at least 20 A from the start of the CDRH3, followed by at least two turn residues. We defined a turn residue as having any of the following: (i) DSSP-defined hydrogen-bonded turn or bend, (ii) hydrogen bond between residues *i* and *i+3* or between residues *i* and *i+4, or* (iii) Phi/Psi angle-defined Type I, I', II, or III turn (see table below). We defined a beta residue as having any of the following: (i) DSSP-defined beta sheet (residue in isolated beta-bridge or extended strand), or (ii) Phi/Psi angle-defined beta sheet (phi < -50 and psi > 95 or phi < -50 and psi < -170). The beta residue was defined as a 'far beta' if the residue's alpha carbon was at least 20 A away from the alpha carbon of the first residue in the CDRH3.

Phi/Psi angle definitions for turn types I, I', II, and III.		
Turn Type	Residue i+1	Residue i+2
Type I	Phi: -100 to -25 Psi: -75 to 25	Phi: -150 to -50 Psi: -50 to 50
Type I'	Phi: 25 to 100 Psi: 0 to 75	Phi: 50 to 150 Psi: -50 to 60
Type II	Phi: -100 to -25 Psi: 75 to 180	Phi: 50 to 125 Psi: -50 to 50
Type III	Phi: -100 to -50 Psi: -60 to 10	Phi: -150 to -60 Psi: 75 to 160

To validate our structural axe definition, we applied it to 610 human antibody structures from the SAbDab[42]. The specific structures we used and their classifications using our definition can be found in our GitHub repository: https://github.com/ssolieva/q23-manuscript.git

MD simulations

Molecular dynamics simulations were seeded from AlphaFold2-generated, cryo-em, or x-ray crystallography structures using the GROMACS software[74] (version 2023.2) and the CHARMM36m force field[75]. The structures consisted of a variable heavy chain and the mature variable light chain for the respective bnAb (CH01, V033). The antibodies were each solvated using TIP3P waters[76] and 0.1 uM NaCl in a dodecahedron box. Energy minimization was performed using steepest descent minimization, followed by NVT equilibration for 200 ps, NPT equilibration for 1 ns, and 10 independent 250 ns production runs per antibody, with snapshots saved at 20 ps intervals. We calculated the frequency of axe-like structures throughout each trajectory by applying our axe structural definition at 200 ps intervals. This analysis was conducted using the MDTraj software[77]. All input files, starting structures, and analysis files are deposited in the following GitHub repository: https://github.com/ssolieva/q23-manuscript.git

***In-silico* saturated mutagenesis**

In-silico saturated mutagenesis was conducted using AlphaFold2 and mutating each residue on the CH01 CDRH3 to every possible amino acid. Each prediction had 5 models and 10 seeds per model, resulting in 50 structures per mutation. The RMSD values were computed between the CDRH3 regions of the CH01 AlphaFold2 structure and the AlphaFold2 predictions of the mutations. The

analysis files are deposited in the following GitHub repository: https://github.com/ssolieva/q23-manuscript.git

DNA design and plasmid synthesis

The amino acid sequence for Q23.17 was obtained from a previously published sequence (genbank accession number AF004885.1). Stabilized trimer constructs were developed by adding published stabilization mutations into the Q23.17, and all constructs included the T533A "repair" mutation. Constructs were then codon optimzed, and an optimized IgE leader sequence was added to the N terminus of the protein to provide efficient processing and secretion. All plasmids were synthesized and cloned (GenScript) into our modified pVax1 backbone (Inovio Pharmaceuticals).

Antibody expression and purification

Expi293F cells (ThermoFisher) were maintained in Expi293 expression medium (ThermoFisher). All cell lines were mycoplasma negative and tested on a regular basis. All proteins were produced by Expifectamine transfection of Expi293F cells following the manufacturer's protocol. Transfection enhancers were added 18 h after transfection and supernatants were harvested 6 days later. Antibodies were purified using Protein A agarose according to the manufacturer's protocol to purify the IgG. Purity was confirmed with Coomassie staining of SDS-page gels and concentration was determined using a nanodrop.

Pseudovirus production and purification

Pseudotyped viruses were produced using HEK 293 T cells transfected with 4 ug of a plasmid expressing the Env of interest and 8 ug of a plasmid expressing the HIV-1 backbone Δ Env (pSG3ΔEnv − NIH AIDS Reagents) using GeneJammer (Aglient). Forty-eight hours after

transfection, cell supernatant was harvested, filtered through a 45 µm filter, aliquoted, and stored at −80 °C.

Trimer production and purification

Env-based trimers were expressed in Expi293F cells maintained in Expi293 expression medium (ThermoFisher). Trimers were produced by Expifectamine transfection (Gibco, A14524) of Expi293F cells following the manufacturer's protocol. The trimer-containing supernatants were obtained by centrifuging (4000 × g, 25 mins) and filtering (0.2 um Nalgene Rapid-Flow Filter) the 293F cultures, following which trimers were purified from supernatants by lectin purification using lectin beads (Vector Laboratories) and lectin elution buffer (1M Methyl alpha-D-mannopyranoside). The trimers were then purified over a size-exclusion chromatography column (GE S200 Increase) in PBS. The molecular weight and homogeneity of the trimers were confirmed by protein conjugated analysis from ASTRA with data collected from a size-exclusion chromatography-multi-angle light scattering (SEC-MALS) experiment run in PBS using a GE S6 Increase column followed by DAWN HELEOS II and Optilab T-rEX detectors. The trimers were aliquoted at 1 mg/ml and flash frozen in thin-walled PCR tubes prior to use.

Antigen Conformation Tracing in vivo by ELISA (ACTIVE)

For Antigen Conformation Tracing *in vivo* by ELISA, BALB/c mice were administered with 100 µg DNA plasmid co-formulated with 12 U hyaluronidase in thetibialis anterior (TA) muscles of the mice as described previously[33]. At predetermined time points TA muscles were harvested and homogenized in T-PER extraction buffer (Thermo Fisher Scientific) containing protease inhibitor (Roche). Muscle homogenates were subsequently concentrated using a 3kDa Amicon Ultra 0.5 mL centrifugation kit(Millipore Sigma). Concentrations of total proteinswere estimated

using BCA assay kit (Thermo Fisher Scientific). In order to determine correct folding and binding of *in vivo* expressed antigens, 96 well ELISA plates (Corning, 3690) were coated with 4 ug/ml of recombinant PGT128 Fab fragments in PBS and incubated overnight at 4° C . After washing, plates were blocked with 5% skimmed milk in PBS containing 1% newborn calf serum (NBS) and 0.2% Tween for 1 hour at RT. Total protein concentrations were normalized and added to the plate followed by serial dilutions. Recombinant BG505.MD39, Q23.MD39, Q23.Ferritin and gp120-foldon were added as positive control standards. Plates were incubated at RT for 2 h, then followed by washing, antibodies of interest were added at 10 ug/ml except for PGT145 which was added at 50 ug/ml, for 1 h at 37 °C. Plates were further washed and incubated with Goat Anti-Human IgG Fc Fragment conjugated with HRP (Bethyl Laboratories Inc) for 1hour at RT. Plates were developed for 5 min with 1-step ultra TMB (ThermoFisher) and stopped with 1 N H2SO4. Absorbance at an optical density (OD) of 450 nm and 570 nm was measured using Synergy2 plate reader (BioTek Instrument). The background 570 nm OD was subtracted from the 450 nm reading. The data was analyzed and fitted using Graph Pad Prism 10.2.

Antibody Digestion for Complexation

Monoclonal and polyclonal antibodies were digested into antigen-binding fragments (Fabs) by adding antibodies into 6-12 mL of digestion buffer (100 mM sodium acetate 10 mM L-cysteine 0.3 mM EDTA pH 5.6 or 7) followed with the addition of papain (2%, w/w), pre-incubated for 15 mins in digestion buffer. Digestion reactions were quenched using 3 mM iodoacetamide and fabs were purified using protein A resin. Protein A resin was added into digestion mixture and incubated for 15 mins on ice. Digest antibody/Protein A resin mixture was filtered through a gravity column. Flow through was collected and buffer exchanged into 1X PBS using Amicon-Ultra concentrator

with molecular weight cutoff of 10 kDa. Protein A was resin washed with at least 10 CV of 1X PBS and eluted using protein A elution buffer. Flow through, wash, and elution fractions were analyzed by SDS-PAGE. CH01iGL Fabs were unable to be purified using Protein A resin due to unwanted binding of Fab to Protein A. Instead, digestion mixture mixed with 3mM iodoacetamide was buffer exchanged into 1X PBS. CH01 iGL Fabs/Fc/IgG mixture was separated using Superdex 200 increase 10/300 size exclusion column using 1X PBS as the running buffer to purify the CH01iGL Fabs/Fc mixture. Fractions containing CH01iGL Fabs/Fc were concentrated using Amicon-Ultra concentrator with a molecular weight cutoff 10kDa. All Fabs were stored at 4°C.

Negative stain electron microscopy sample preparation and data collection

SEC purified MD39 trimers were further dialyzed into Tris-buffered saline (TBS). A total of 4 µL of purified proteins (0.005 mg/mL) was adsorbed onto glow discharged carbon-coated Cu400 EM grids. The grids were then stained with 4 µL of 2% uranyl formate, blotted, and stained again with 4 µL of the stain followed by a final blot. Image collection was performed on a FEI Tecnai T12 microscope equipped with Oneview Gatan camera at 62,750x camera magnification resulting in pixel size of 2.356 Å.

For nsEMPEM, Q23.GT4 (50 µg) was mixed with 500 µg of polyclonal fabs and incubated at 4 degrees overnight. The resulting complex mixture was purified using Superose 6 increase 10/300 size exclusion column and complex fractions were collected. The collected fractions were diluted immediately and applied to glow discharged carbon coated 300 mesh Cu grids (Electron Microscopy Sciences (EMS) CFTH300-Cu-50) to incubate for 2 mins before blotting. Complexes were stained using 2% (w/v) uranyl formate. Micrographs were collected on a FEI Tecnai T12 microscope at 62,750x camera magnification resulting in pixel size of 2.356 Å/pixel.

Negative stain electron microscopy data processing

Micrographs were imported into RELION5.0[78] and performed CTF estimation using CtfFind4. Particles were picked using Laplacian of Gaussian (LoG) picker and extracted for 2D classifications. For negative stain EMPEM, after the final round of 2D classification, there were 209,484 particles in the CK52 dataset and 59,730 particles in the CK11 dataset. These particles were then 3D classified with 50 classes for CK52 and 60 classes for CK11 (T=4). Good 3d classes were individually 3D refined using a 15 Å ligand-free Env density as a reference.

Preparation of Q23.MD39 in complex with CH01 iGL and 35O22

Q23.MD39/CH01iGL/35O22 complex was prepared by mixing Fabs with a 1:40:40 molar ratio (Q23.MD39:CH01iGL:35O22) in 1X PBS. Q23.MD39 complex was purified by Size-Exclusion Chromatography using the ENrich™ 650 column (BioRad) with 1X PBS as the running buffer and fractions containing the complex were collected and concentrated to 0.1 mg/mL. The resulting protein complex was analyzed by negative stain EM immediately and aliquoted for storage in -80°C until Cryo-EM sample preparation.

Cryo Electron Microscopy Sample Preparation, Data Collection, and Data Processing

Q23.MD39/CH01-iGL/35O22 sample was diluted to 0.07 mg/mL and deposited on graphene-oxide (GO) coated Au-Flat grids (Protochips). Graphene oxide coating of Au-Flat grids was done in-house following a protocol from previous literature[79]. 4 µL of sample was added to the GO-coated grid at 4°C under 100% humidity in a Mark IV Vitrobot (FEI), blotted with Whatman #1 filter paper, and then plunged immediately into liquid ethane. Dose-fractionated data for Q23.MD39/CH01-iGL/35O22 was collected in Counting mode at a magnification of 81,000x

resulting in pixel size of 1.054 Å/pixel using aberration-free image shift (AFIS) protocol through the EPU software (ThermoFisher).

Data processing was performed employing a standard cryo-EM data processing workflow comprising motion correction, CTF estimation, reference-free LoG picking, 2D classification, manual inspection/selection of 2D class averages, and 3D refinement in RELION v3.1[78]. 3D classification was employed to remove junk particles as well as separating particles with or without CH01 iGL bound and with two and three 35O22-bound particles. Particles from two and three 35O22-bound Q23/CH01 iGL classes were individually 3D Refined and Bayesian Polished in RELION. Polished particles were pooled and transferred into CryoSPARC and focus 3D classified to further improve the CH01iGL density. The resulting 3D classes were individually refined, and the best class was manually inspected and used for another round of Non-Uniform Refinement to generate the final density maps.

Cryo EMPEM Sample Preparation and Data Collection

Q23.GT4 (100 µg) were added into 1 mg of CK52 week24 polyclonal Fabs and incubated at 4 degrees overnight. CK52wk24/Q23.GT4 complex was purified using S6i 10/300 size exclusion chromatography in 1X PBS and fractions were collected, pooled, and concentrated using Amicon-Ultra concentrator with a molecular weight cutoff of 10 kDa. Complexes were diluted to 0.075 mg/mL and deposited on GO-coated UltrAuFoil R1.2/1.3 grids (Quantifoil). Sample grid was loaded onto a Thermo Scientific Glacios (Thermo Fisher) equipped with a Falcon 4 detector. Dose-fractionated data was collected at 150,000x magnification resulting in a pixel size of 0.95 Å/pixel

using AFIS protocol through EPU software (Thermofisher). Data was collected with a total dose of 50 $e^-/Å^2$ divided over 45 frames.

Cryo EMPEM Data Processing

Micrographs were imported into cryoSPARC for motion correction and CTF estimation. Initial map reconstructed from blob picker was used for template picking. Particles from both picking strategies, with duplicated particles removed, were extracted (binned by 4), and 2D classified, and trimer particles were further filtered through two rounds of heterogenous refinement. Particles were re-extracted (binned by 2) and one round of Non-Uniform refinement was performed. Finally, particles were re-extracted with the original pixel size and were Non-Uniform refined using a 300 Å sphere mask. Particles used in the refinement were symmetry expanded through C3 symmetry. Symmetry expanded particles were 3D classified (K=10) using an 80 Å focus mask on the partial fab density and a 300 Å solvent mask. Particles from good classes were pooled then subjected to local refinement in CryoSPARC using a low-pass filtered map from one of the good 3D classes as a static mask. The resulting particles with further improved angular assignments were 3D-classified again using a 120 Å sphere focus mask around the Fab and the same solvent mask. A final round of local refinement was performed.

Model Building

A homology model of Q23.MD39 was generated from SWISS-MODEL[80] using a refined model of another high-resolution prefusion-closed Env trimer from the Pallesen Lab (unpublished). A homology model of CH01-iGL Fab was generated from SWISS-MODEL using CH03 Fab (PDB: 5ESV)[22]. The crystal structure of 35O22 Fab (PDB: 4TOY) was used. Homology models were

docked into the Q23.MD39/CH01 iGL/35O22 density map (UCSF ChimeraX)[81] and real-space refined using Coot[82]. In-house developed protocols using Rosetta was employed for model refinement. N-linked glycans were added manually in Coot. Q23/35O22 atomic model was refined using the same protocol. Model-to-map fit was validated using EMRinger[83], glycan geometry using Privateer31[84], and model geometry using MolProbity[85].

Serological trimer-binding ELISA

Binding titers to trimer were determined by coating plates with 2ug/ml of recombinant PGT128 antibody overnight in PBS. After washing, plates were blocked with 5% skim milk in PBS with 1% newborn calf serum (NBS) and 0.2% Tween for 1 h at RT. Recombinant trimer was added at 4 ug/ml for 2 h at RT. Serum was serially diluted, added to plates, and incubated at 37 °C for 1 h. Antigen and species-specific IgG was then detected across absorbed secondary anti-mouse HRP antibody (Bethyl Laboratories Inc). Plates were developed for 5 min with 1-step ultra TMB (ThermoFisher) and stopped with 1 N H_2SO_4. Absorbance at an optical density (OD) of 450 nm and 570 nm was measured on a Synergy2 plate reader (BioTek Instrument). The background 570 nm OD was subtracted from the 450 nm reading.

Recombinant antibody binding ELISA

Antibody affinity to Env trimers was determined by coating plates with 4 ug/ml PGT121 or PGT128 fab in 1×PBS for 3 h at RT. After washing with 1×PBS containing 0.05% Tween, plates were blocked overnight with 1×PBS containing 0.1% Tween and 5% skim milk. Plates were washed and trimer was incubated at 10 ug/ml for 1 h at RT. Plates were washed and recombinant antibody was serially diluted, added to plates, and incubated for 1 h at RT. After washing the plates,

goat anti-human IgG Fc (Bethyl Laboratories Inc) at a dilution of 1:10,000 was incubated for 1 h at RT. Plates were washed and developed for 10 min with 1-step ultra TMB (ThermoFisher) and stopped with 1 N H_2SO_4. Absorbance at an optical density (OD) of 450 nm and 570 nm was measured on a Synergy2 plate reader (BioTek Instrument). The background 570 nm OD was subtracted from the 450 nm reading.

DNA encoded immunogen immunization in WT (BALB/c) mice

Female BALB/c mice of 6-8 weeks old were immunized with 5ug, 10ug and 25ug of plasmid DNA encoding Q23.MD39 trimer or 5ug and 10ug of Q23.Ferritin nanoparticle. The animals received these immunizations either with or without plasmid DNA encoding IL-12 as intramuscular (IM) injections into the tibia anterior (TA) muscles, followed by *in vivo* electroporation (EP) using the CELLECTRA ® -3P device (Inovio Pharmaceuticals). Mice were immunized at 0,3 and 6 weeks and sera were collected 2 weeks post each immunization through the submandibular vein for assessment of humoral immune responses.

Neutralization Assays

Pseudotyped viruses and SHIVs were titered on TZM-bl cells to determine IU/ml. All sera samples were heat-inactivated for 60 min at 56°C. 96 well plates were seeded with 10^4 TZM-bl cells/well cultured in DMEM + 10% FBS one day prior to the assay. Serum was serially diluted in media supplemented with 10% normal human serum and incubated with pseudotyped virus and dextran (ThermoFisher) for 1 hour at 37°C, following which the virus/serum mixture was added to adherent TZM-bl cells. Forty-eight hours after incubation, media was removed and cells were lysed using PBS + 0.01% Triton-X (Promega). Luciferase luminescence was then measured using

the Synergy2 plate reader (BioTek Instruments). Serum titer was determined for 50% virus neutralization (ID50).

Mammalian display library design

For cell surface display, Q23.MD39 was genetically fused to a PDGFR transmembrane domain via a G/S linker as previously described[46]. This construct was synthesized in a pENTR backbone by Twist Biosciences. A scanning NNK library covering HXB2 residues 110-192 of the Q23.MD39 construct was produced using a BioXP 3250 instrument. NNK fragments were then pooled and assembled into the Q23.PDGFR.pENTR construct via Gibson assembly (NEB, E2611S). The Gibson reaction was then transformed into Stbl2 cells, and a small aliquot of the transformation culture was plated to ensure a transformation efficiency of >10X the library diversity. Library assembly and diversity was confirmed by sequencing on an Illumina MiSeq.

Mammalian display cell culture protocol

All 293t cells were grown in high glucose DMEM supplemented with Glutamax and pyruvate, 10% fetal bovine serum, and 1% penicillin/streptomycin. Cells were transduced at an MOI of approximately 0.1. 24h after transduction, 2µg/ml of puromycin was added to the cells, and cells were constantly grown in media containing 2µg/ml puromycin for subsequent steps. Once a sufficient number of cells had grown, approximately 10 million cells were physically dislodged from the cell culture flask. They were then washed in PBS, followed by staining with monoclonal antibodies of interest for 15 minutes at RT. Cells were then washed again with PBS, following which they were stained with anti-Myc FITC (Invitrogen, 13-2511) and anti-human IgG BV421 (BD Biosciences, 562581) antibodies for 15 minutes at RT. Cells were washed twice, resuspended

in PBS, and finally stained with 7-AAD (ThermoFisher, A1310) before sorting. Cells were collected by sorting on a BD FACS Melody sorter. Cells were gated on singlets, 7-AAD-, and Myc+, following which the top 1-5% of iGL binding cells were collected in DMEM with 50% FBS. In addition, control cells that were unstained with any iGL and only Myc+ were sorted.

Mammalian display sequencing

Genomic DNA was extracted from sorted cells using a GenElute mammalian genomic DNA miniprep kit (Millipore Sigma, G1N70) according to the manufacturers instructions. Forward and reverse primers containing the Illumina P5 and P7 adapters respectively were designed to target regions of the Q23.PDGFR construct outside the area of the scanning NNK library, and used to amplify the region of interest. Amplified libraries were then purified using AMPureXP beads, and checked on a BioAnalyzer instrument to confirm that amplicons were of the correct size. The libraries were then sequenced on an Illumina MiSeq instrument using a 600 cycle MiSeq v3 kit.

Mammalian display data analysis

Following sequencing, reads were adaptor trimmed and paired. Library mutations in each sequence were identified by the presence of synonymous barcoding mutations flanking each NNK residue. The total number of cells bearing each mutation across the length of the NNK library was determined, and amino acid frequencies were calculated by dividing the number of cells with a particular amino acid mutation by the total number of cells bearing any mutation at that residue. Relative enrichments of mutations in the libraries were calculated via the ratio of the frequency of a sequence in the sorted library over the frequency of the same sequence in the control cells. Enrichments resulting from a low number of mutation counts in the control library were discarded.

SHIV infection of rhesus macaques

All Indian rhesus macaques were housed at Bioqual Inc (Rockville MD), according to AAALAC standards. All experiments were approved by the University of Pennsylvania and Bioqual Institutional Animal Care and Use Committees. Three days prior to SHIV infection, rhesus macaques were subcutaneously injected with 25mg/kg anti-CD8α mAb (MT807R1). Macaques were inoculated with SHIV by intravenous infusion. Blood draws, processing, and storage was performed as previously described[19].

Env single genome sequencing

3' SHIV half genomes were sequenced as previously described[61]. Briefly, viral RNA was synthesized from plasma virions using the Qiagen BioRobot EZ1 Workstation with EZ1 Virus Mini Kit v2.0 (Qiagen). Viral RNA was then used to synthesize cDNA using SuperScript III reverse transcriptase (Invitrogen). cDNA was then serially diluted in 96 well plates and amplified using nested PCR such that <30% of wells were PCR-positive. Positive wells were sequenced using an Illumina Miseq sequencer. Sequences from wells that showed multiple unique Env sequences were excluded. All Env sequences reported in this study are available at Genbank under accession numbers XXXX.

1. Stephenson, K.E., Wagh, K., Korber, B. & Barouch, D.H. Vaccines and Broadly Neutralizing Antibodies for HIV-1 Prevention. *Annual Review of Immunology* **38**, 673-703 (2020).

2. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312 (2003).

3. Kwong, P.D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420** (12/12/2002).

4. Moore, J.P., Parren, P.W.H.I. & Burton, D.R. Genetic Subtypes, Humoral Immunity, and Human Immunodeficiency Virus Type 1 Vaccine Development. *Journal of Virology* **75** (2001-7-1).

5. Moore, P.L. The Neutralizing Antibody Response to the HIV-1 Env Protein. *Current HIV research* **16** (2018).

6. Gray, E.S. *et al.* The Neutralization Breadth of HIV-1 Develops Incrementally over Four Years and Is Associated with CD4+ T Cell Decline and High Viral Load during Acute Infection. *Journal of Virology* **85** (2011/05).

7. Doria-Rose, N.A. *et al.* Breadth of Human Immunodeficiency Virus-Specific Neutralizing Activity in Sera: Clustering Analysis and Association with Clinical Variables. *Journal of Virology* **84** (2010/02).

8. Haynes, B.F., Kelsoe, G., Harrison, S.C. & Kepler, T.B. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. Nat Biotechnol; 2012. pp. 423-433.

9. Kwong, P.D. & Mascola, J.R. HIV-1 Vaccines Based on Antibody Identification, B Cell Ontogeny, and Epitope Structure. Cell Press; 2018. pp. 855-871.

10. Landais, E. *et al.* Broadly Neutralizing Antibody Responses in a Large Longitudinal Sub-Saharan HIV Primary Infection Cohort. *PLOS Pathogens* **12** (Jan 14, 2016).

11. Hraber, P. *et al.* Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS (London, England)* **28** (2014/01/01).

12. Rusert, P. *et al.* Determinants of HIV-1 broadly neutralizing antibody induction. *Nature Medicine 2016 22:11* **22** (2016-09-26).

13. Gautam, R. *et al.* A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges. *Nature* **533**, 105-109 (2016).

14. Corey, L. *et al.* Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. *New England Journal of Medicine* **384**, 1003-1014 (2021).

15. Julg, B. *et al.* Broadly neutralizing antibodies targeting the HIV-1 envelope V2 apex confer protection against a clade C SHIV challenge. *Science Translational Medicine* **9** (2017-09-06).

16. Hessell, A.J. *et al.* Effective, low-titer antibody protection against low-dose repeated mucosal SHIV challenge in macaques. *Nature Medicine 2009 15:8* **15** (2009-06-07).

17. Landais, E. *et al.* HIV Envelope Glycoform Heterogeneity and Localized Diversity Govern the Initiation and Maturation of a V2 Apex Broadly Neutralizing Antibody Lineage. *Immunity* **47**, 990-1003.e1009 (2017).

18. Doria-Rose, N.A. *et al.* Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* **508**, 55-62 (2014).

19. Roark, R.S. *et al.* Recapitulation of HIV-1 Env-antibody coevolution in macaques leading to neutralization breadth. *Science* **371** (2021).

20. Willis, J.R. *et al.* Human immunoglobulin repertoire analysis guides design of vaccine priming immunogens targeting HIV V2-apex broadly neutralizing antibody precursors. *Immunity* **55** (2022/11/08).

21. Lee, J.H. *et al.* A Broadly Neutralizing Antibody Targets the Dynamic HIV Envelope Trimer Apex via a Long, Rigidified, and Anionic β-Hairpin Structure. *Immunity* **46**, 690-702 (2017).

22. Gorman, J. *et al.* Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. *Nature Structural and Molecular Biology* **23**, 81-90 (2016).

23. McLellan, J.S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336-343 (2011).

24. Pejchal, R. *et al.* Structure and function of broadly reactive antibody PG16 reveal an H3 subdomain that mediates potent neutralization of HIV-1. *Proceedings of the National Academy of Sciences* **107** (2010-6-2).

25. Julien, J.-P. *et al.* Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proceedings of the National Academy of Sciences of the United States of America* **110** (2013/03/03).

26. 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 Reports* **31** (2020/04/07).

27. Roark, R.S. *et al.* HIV-1 neutralizing antibodies in SHIV-infected macaques recapitulate structurally divergent modes of human V2 apex recognition with a single D gene. *bioRxiv* (2024-06-12).

28. Briney, B. *et al.* Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature 2019 566:7744* **566** (2019-01-21).

29. Bonsignori, M. *et al.* Analysis of a Clonal Lineage of HIV-1 Envelope V2/V3 Conformational Epitope-Specific Broadly Neutralizing Antibodies and Their Inferred Unmutated Common Ancestors. *Journal of Virology* **85**, 9998-10009 (2011).

30. Walker, L.M. *et al.* Broad and Potent Neutralizing Antibodies from an African Donor Reveal a New HIV-1 Vaccine Target. *Science (New York, N.y.)* **326** (2009/10/10).

31. Voss, J.E. *et al.* Elicitation of Neutralizing Antibodies Targeting the V2 Apex of the HIV Envelope Trimer in a Wild-Type Animal Model. *Cell Reports* **21**, 222-235 (2017).

32. Liao, H.-X. *et al.* Vaccine Induction of Antibodies Against a Structurally Heterogeneous Site of Immune Pressure within HIV-1 Envelope Protein Variable Regions 1 and 2. *Immunity* **38** (2013/01/01).

33. Xu, Z. *et al.* Induction of tier-2 neutralizing antibodies in mice with a DNA-encoded HIV envelope native like trimer. *Nature Communications 2022 13:1* **13** (2022-02-04).

34. Olsen, T.H., Boyles, F. & Deane, C.M. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Science : A Publication of the Protein Society* **31** (2022/01).

35. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596** (2021).

36. Volpe, J.M. & Kepler, T.B. Large-scale analysis of human heavy chain V(D)J recombination patterns. *Immunome Research* **4** (2008).

37. Valdés-Tresanco, M.S. *et al.* Structural Modeling of Nanobodies: A Benchmark of State-of-the-Art Artificial Intelligence Programs. *Molecules 2023, Vol. 28, Page 3991* **28** (2023-05-09).

38. Ruffolo, J.A. *et al.* Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications 2023 14:1* **14** (2023-04-25).

39. Yin, R. & Pierce, B.G. Evaluation of AlphaFold antibody–antigen modeling with implications for improving predictive accuracy. *Protein Science : A Publication of the Protein Society* **33** (2024/01).

40. Brevern, A.G.d. Extension of the classical classification of β-turns. *Scientific Reports* **6** (2016).

41. Lovell, S.C. *et al.* Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics* **50** (2003/02/15).

42. Dunbar, J. *et al.* SAbDab: the structural antibody database. *Nucleic Acids Research* **42** (2014/01/01).

43. Meller, A., Bhakat, S., Solieva, S. & Bowman, G.R. Accelerating Cryptic Pocket Discovery Using AlphaFold. *Journal of Chemical Theory and Computation* **19** (2023/07/07).

44. Waltari, E. *et al.* 5′ Rapid Amplification of cDNA Ends and Illumina MiSeq Reveals B Cell Receptor Features in Healthy Adults, Adults With Chronic HIV-1 Infection, Cord Blood, and Humanized Mice. *Frontiers in Immunology* **9** (2018).

45. Poss, M. & Overbaugh, J. Variants from the Diverse Virus Population Identified at Seroconversion of a Clade A Human Immunodeficiency Virus Type 1-Infected Woman Have Distinct Biological Properties. *Journal of Virology* **73**, 5255-5264 (1999).

46. Steichen, J.M. *et al.* HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* **45**, 483-496 (2016).

47. Kulp, D.W. *et al.* Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. *Nature Communications 2017 8:1* **8** (2017-11-21).

48. Peña, A.T.d.l. *et al.* Improving the Immunogenicity of Native-like HIV-1 Envelope Trimers by Hyperstabilization. *Cell Reports* **20** (2017/08/22).

49. Antanasijevic, A. *et al.* Structural and functional evaluation of de novo-designed, two-component nanoparticle carriers for HIV Env trimer immunogens. *PLOS Pathogens* **16** (Aug 11, 2020).

50. 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 Reports* **33**, 108432-108432 (2020).

51. Rutten, L. *et al.* A Universal Approach to Optimize the Folding and Stability of Prefusion-Closed HIV-1 Envelope Trimers. *Cell Reports* **23** (2018/04/10).

52. Kato, Y. *et al.* Multifaceted Effects of Antigen Valency on B Cell Response Composition and Differentiation In Vivo. *Immunity* **53** (2020/09/15).

53. Tokatlian, T. *et al.* Innate immune recognition of glycans targets HIV nanoparticle immunogens to germinal centers. *Science* **363** (2019-02-08).

54. Xu, Z. & Kulp, D.W. Protein engineering and particulate display of B-cell epitopes to facilitate development of novel vaccines. *Current Opinion in Immunology* **59** (2019/08/01).

55. Kelly, H.G., Kent, S.J. & Wheatley, A.K. Immunological basis for enhanced immunity of nanoparticle vaccines. *Expert Review of Vaccines* **18** (2019-3-4).

56. Xu, Z. *et al.* In Vivo Assembly of Nanoparticles Achieved through Synergy of Structure-Based Protein Engineering and Synthetic DNA Generates Enhanced Adaptive Immunity. *Advanced Science* **7** (2020/04).

57. Bhagchandani, S.H. *et al.* Two-dose "extended priming" immunization amplifies humoral immune responses by synchronizing vaccine delivery with the germinal center response. *bioRxiv* (2023).

58. Cirelli, K.M. *et al.* Slow Delivery Immunization Enhances HIV Neutralizing Antibody and Germinal Center Responses via Modulation of Immunodominance. *Cell* **177**, 1153-1171.e1128 (2019).

59. Mu, Z. *et al.* mRNA-encoded HIV-1 Env trimer ferritin nanoparticles induce monoclonal antibodies that neutralize heterologous HIV-1 isolates in mice. *Cell Reports* **38** (2022/03/15).

60. Li, H. *et al.* New SHIVs and Improved Design Strategy for Modeling HIV-1 Transmission, Immunopathogenesis, Prevention and Cure. *Journal of Virology* (2021).

61. Li, H. *et al.* Envelope residue 375 substitutions in simian-human immunodeficiency viruses enhance CD4 binding and replication in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E3413-E3422 (2016).

62. Bibollet-Ruche, F. *et al.* A Germline-Targeting Chimpanzee SIV Envelope Glycoprotein Elicits a New Class of V2-Apex Directed Cross-Neutralizing Antibodies. *mBio* **14** (Jan-Feb 2023).

63. Brown, C.R. *et al.* Unique Pathology in Simian Immunodeficiency Virus-Infected Rapid Progressor Macaques Is Consistent with a Pathogenesis Distinct from That of Classical AIDS. *Journal of Virology* **81** (2007/06).

64. Bar, K.J. *et al.* Early Low-Titer Neutralizing Antibodies Impede HIV-1 Replication and Select for Virus Escape. *PLoS Pathogens* **8** (2012/05).

65. Wang, H., Morris, D.J., Shaw, G.M. & Kwong, P.D. Potent cross-reactive HIV-1 neutralization in fusion peptide-primed SHIV-infected macaques (update once preprint is out). *Cell* (2024).

66. Bianchi, M. *et al.* Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization. *Immunity* **49** (2018/08/08).

67. Antanasijevic, A. *et al.* Polyclonal antibody responses to HIV Env immunogens resolved using cryoEM. *Nature Communications 2021 12:1* **12** (2021-08-10).

68. Schiffner, T. *et al.* Vaccination induces broadly neutralizing antibody precursors to HIV gp41. *Nature Immunology 2024 25:6* **25** (2024-05-30).

69. Leggat, D.J. *et al.* Vaccination induces HIV broadly neutralizing antibody precursors in humans. *Science* **378** (2022-12-02).

70. Steichen, J.M. *et al.* A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science* **366** (2019).

71. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature 2024 630:8016* **630** (2024-05-08).

72. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods 2022 19:6* **19** (2022-05-30).

73. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* (2022-03-10).

74. Abraham, M.J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).

75. Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* **14** (2017/01).

76. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79** (1983/07/15).

77. McGibbon, Robert T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109** (2015/10/10).

78. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7** (2018).

79. Patel, A., Toso, D., Litvak, A. & Nogales, E. Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from tilted grids. *bioRxiv* (2021-03-08).

80. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46** (2018/07/02).

81. Meng, E.C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Science : A Publication of the Protein Society* **32** (2023/11).

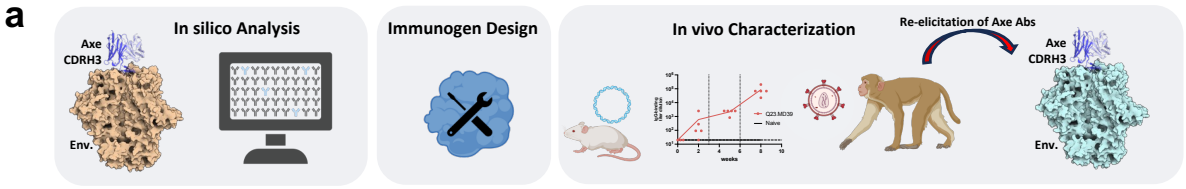
82. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* (2004).

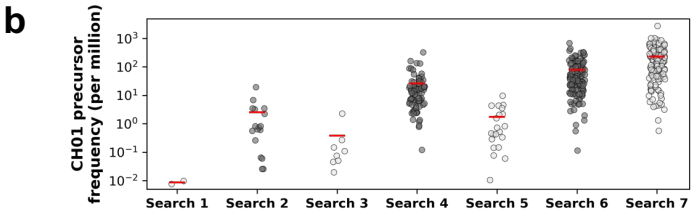
83. Barad, B.A. *et al.* EMRinger: side chain–directed model and map validation for 3D cryo-electron microscopy. *Nature Methods 2015 12:10* **12** (2015-08-17).

84. Agirre, J. *et al.* Privateer: software for the conformational validation of carbohydrate structures. *Nature Structural & Molecular Biology 2015 22:11* **22** (2015-11-04).

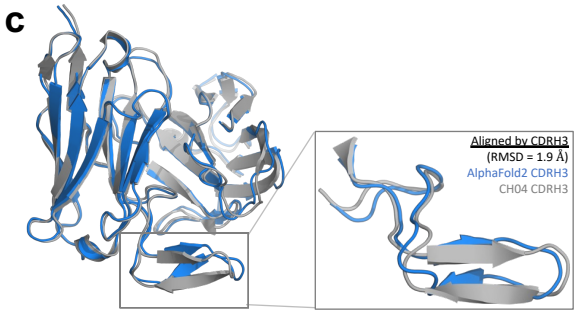
85. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**, 12-21 (2010).

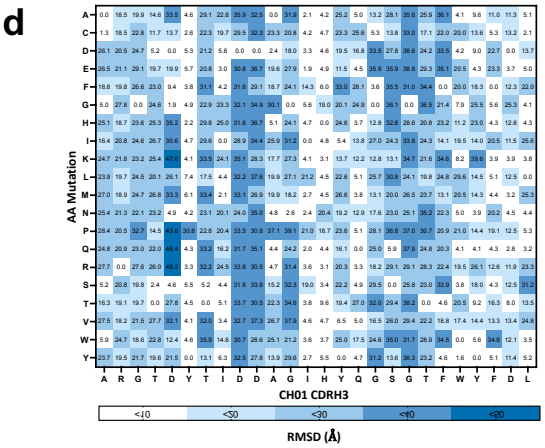
Figure 1

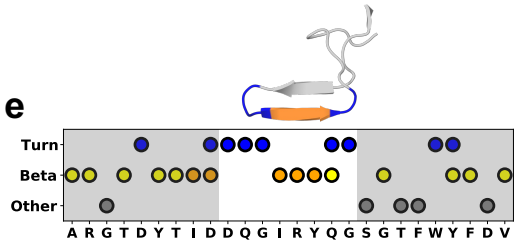


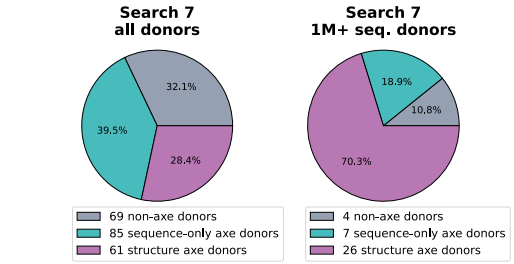


Search	Motif register	JH gene	VH gene	# donors	# donors with 1M+ sequences
1	+/- 0	J2	V3	3/215	3/37
2	+/- 0	Any	V3	17/215	13/37
3	+/- 2	J2	V3	8/215	7/37
4	+/- 2	Any	V3	76/215	28/37
5	+/- 4	J2	V3	20/215	16/37
6	+/- 4	Any	V3	121/215	32/37
7	+/- 4	Any	Any	146/215	33/37









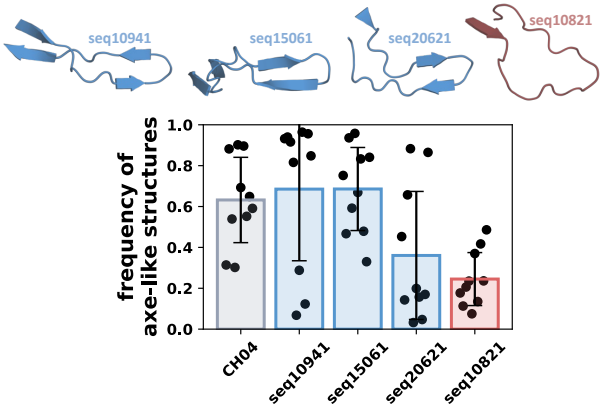
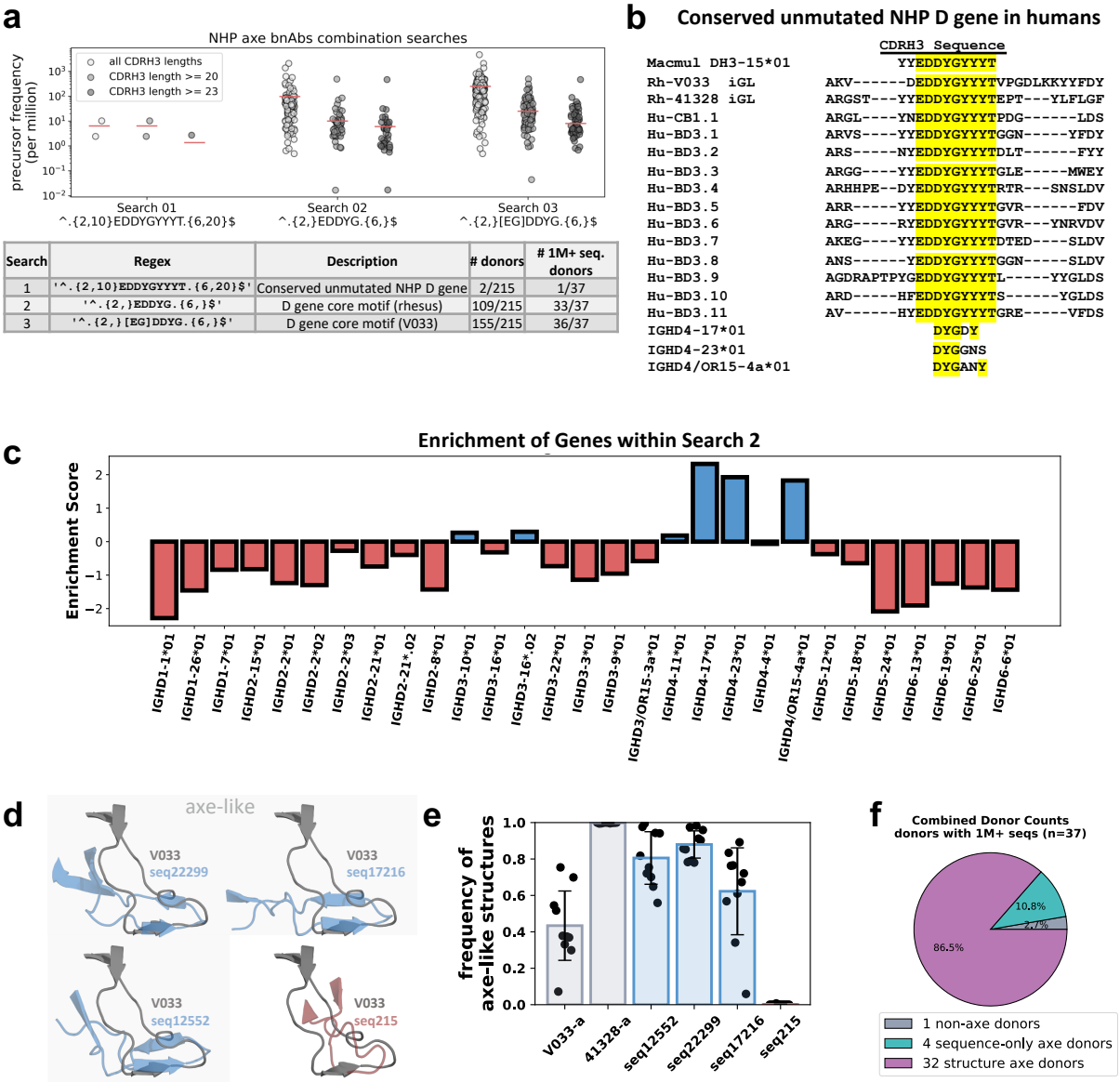


Fig. 1: Using structural characteristics of axe-like bnAbs to estimate axe-like B cell frequency in the human repertoire. a, An overview of our study, starting from in silico analysis to immunogen design to in vivo characterization and re-elicitation of Axe Abs in NHPs. **b,** CH01-like antibody search results from the OAS database. Each point on the graph represents a donor that had at least one sequence matching the search criteria. All searches are detailed in the table. The motif used was a DH gene motif, Y[YQK]GSG, but we did not include a specific DH gene. **c,** AlphaFold2 model of CH01 (blue) has a low CDRH3 CA RMSD to the published CH04 structure (gray, PDB ID 3tcl). **d,** AlphaFold models from CDRH3 saturated mutagenesis where each position along the CH01 CDHR3 was mutated to every possible residue. The x-axis shows the native CH01 CDRH3 sequence while the y-axis shows the mutations. The minimum CDRH3 CA RMSD (Å) to the AlphaFold-generated control structure is shown for each mutation. Small RMSD values (white, light blue) represent positions where the mutations did not greatly affect the fold while large RMSD values (darker blue) represent positions where the mutations greatly perturbed the CDRH3 fold. **e,** The scatter plot shows our structural definition of an axe CDRH3 fold applied to CH04 (PDB ID 3tcl), where blue shows residues that are involved in a turn, yellow shows beta-sheet residues, orange shows beta-sheet residues that are at least 20 Å away from the start of the CDRH3, and gray shows residues that did not fit our turn or beta criteria. The pie chart on the left shows the distribution of donors (n=215) that have no structural or sequence axe-like motif (gray), donors that have a sequence axe-like motif but were not predicted to have any axe-like structures (cyan), and donors that have a sequence axe-like motif that folded into an axe-like structure. The pie chart on the right is the same, but only shows donors which have at least 1M sequences in the OAS database. **f,** Structures of antibodies found from Search 7, 3 of which are predicted to be axe-like (blue) and 1 was predicted to not be axe-like (red). The plot below the structures shows the ratio of frames within each MD simulation where the structure was classified as axe-like. Error bars represent one standard deviation from the mean.

Figure 2



Search	Regex	Description	# donors	# 1M+ seq. donors
1	'^.{2,10}EDDYGYYYT.{6,20}$'	Conserved unmutated NHP D gene	2/215	1/37
2	'^.{2,}EDDYG.{6,}$'	D gene core motif (rhesus)	109/215	33/37
3	'^.{2,}[EG]DDYG.{6,}$'	D gene core motif (V033)	155/215	36/37

Fig. 2: Estimating the frequency rhesus-like axe antibodies within the human repertoire. a, Various rhesus-like antibody search results from the OAS database, broken down into CDRH3 lengths (all lengths, at least 20 AA in length, and at least 23 AA in length). Each point on the graph represents a donor that had at least one sequence matching the search criteria. The three searches are detailed in the table. **b,** Sequence alignments of the two rhesus antibodies with rhesus and human genes, and the sequences found in Search 1. **c,** The enrichment of IGHD genes within Search 2, with positive enrichment in blue and negative enrichment in red. **d,** Structures of antibodies found in Search 3, 3 of which are classified as axe-like and 1 is not classified as axe-like. **e,** The ratio of frames within each trajectory where the structure was classified as axe-like. Error bars represent one standard deviation from the mean. **f,** Distribution of highly sequenced donors (1M+ sequences, n=37) amongst the most expanded searches of CH01-like, rhesus-like, and PG9-like antibodies.

Figure 3

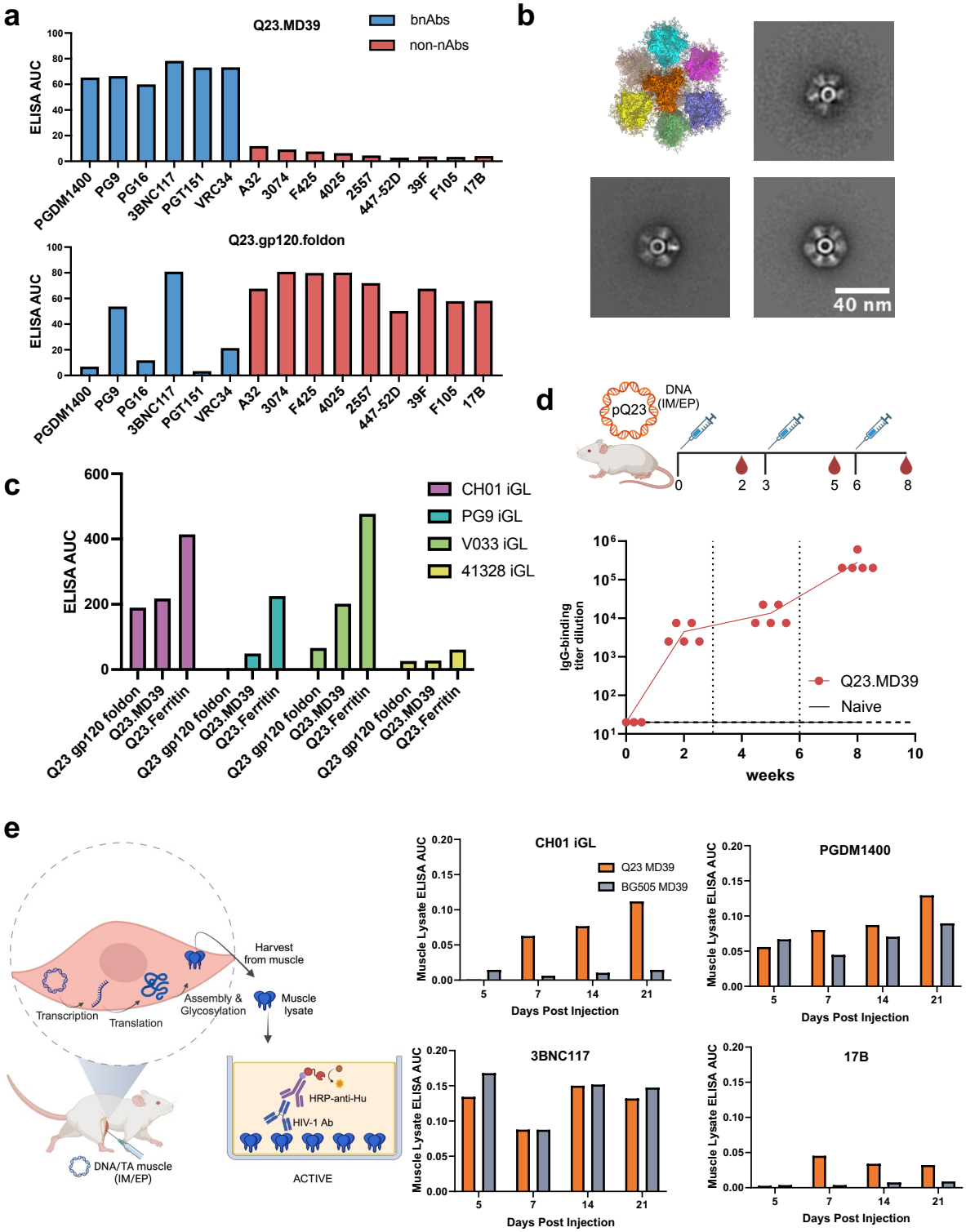

a
Q23.MD39
bnAbs
non-nAbs
ELISA AUC
PGDM1400
PG9
PG16
3BNC117
PGT151
VRC34
A32
3074
F425
4025
2557
447-52D
39F
F105
17B
Q23.gp120.foldon
b
40 nm
c
CH01 iGL
PG9 iGL
V033 iGL
41328 iGL
Q23 gp120 foldon
Q23.MD39
Q23.Ferritin
d
pQ23
DNA
(IM/EP)
IgG-binding
titer dilution
Q23.MD39
Naive
weeks
e
Harvest
from muscle
Transcription
Translation
Assembly &
Glycosylation
Muscle
lysate
DNA/TA muscle
(IM/EP)
HRP-anti-Hu
HIV-1 Ab
ACTIVE
CH01 iGL
Q23 MD39
BG505 MD39
Muscle Lysate ELISA AUC
Days Post Injection
PGDM1400
3BNC117
17B

Fig. 3: Characterization and DNA-delivery of Q23-based immunogens. a, antigenic profile of Q23.MD39 and Q23.gp120.foldon against a panel of broadly neutralizing and non-neutralizing antibodies. **b**, Model and NS-EM 2D class averages of Q23.Ferritin nanoparticles. **c,** ELISA area under the curve of binding of Q23.gp120.foldon, Q23.MD39, and Q23.Ferritin to axe-like V2 apex bnAb iGLs. **d**, Time course of vaccine-specific serum-IgG binding endpoint titers in WT BALB/c mice immunized with 25µg of DNA-delivered Q23.MD39. Circles indicate individual mice, and vertical dotted lines indicate immunization timepoints. **e,** Left, overview of the ACTIV assay. Right, ELISA area under the curve binding of muscle lysate from mice immunized with Q23.MD39 (orange) or BG505.MD39 (grey) to various antibodies.

Figure 4

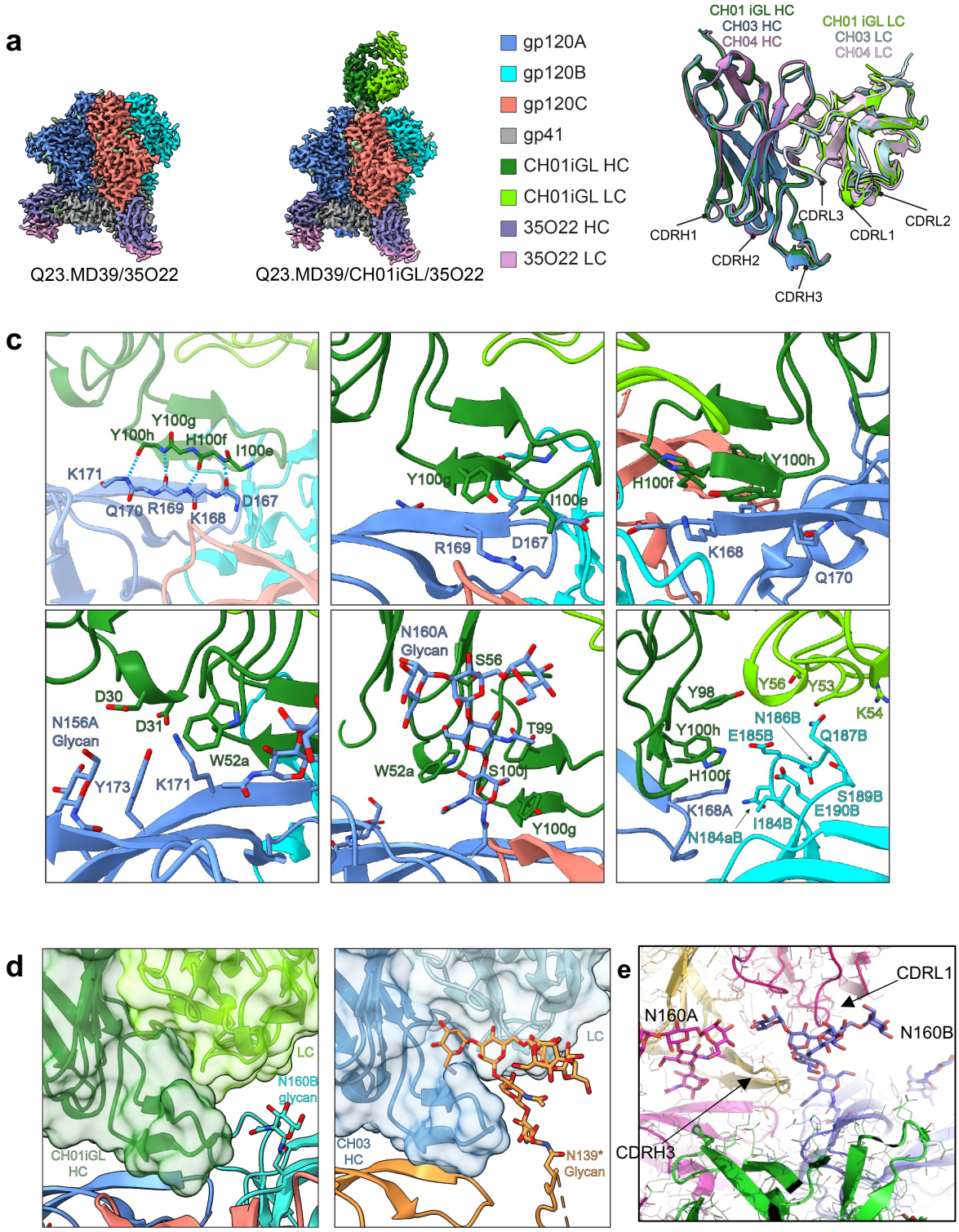


Fig 4: Cryo-EM structure of Q23.MD39 in complex with the CH01 iGL a, Colored density maps of Q23.MD39/35O22, Q23.MD39/CH01 iGL/35O22. **b**, Overlayed atomic models of CH01 iGL, CH03 (PDB: 5ESV), and CH04 (PDB: 5ESZ). **c**, Interactions at the epitope-paratope interface of CH01 iGL and Q23.MD39. **d**, Comparison of N160 glycan on gp120B of Q23.MD39 with N139 glycan in the atomic models of CH03 and CH04. **e,** N160B glycan modeled using Rosetta.

Figure 5

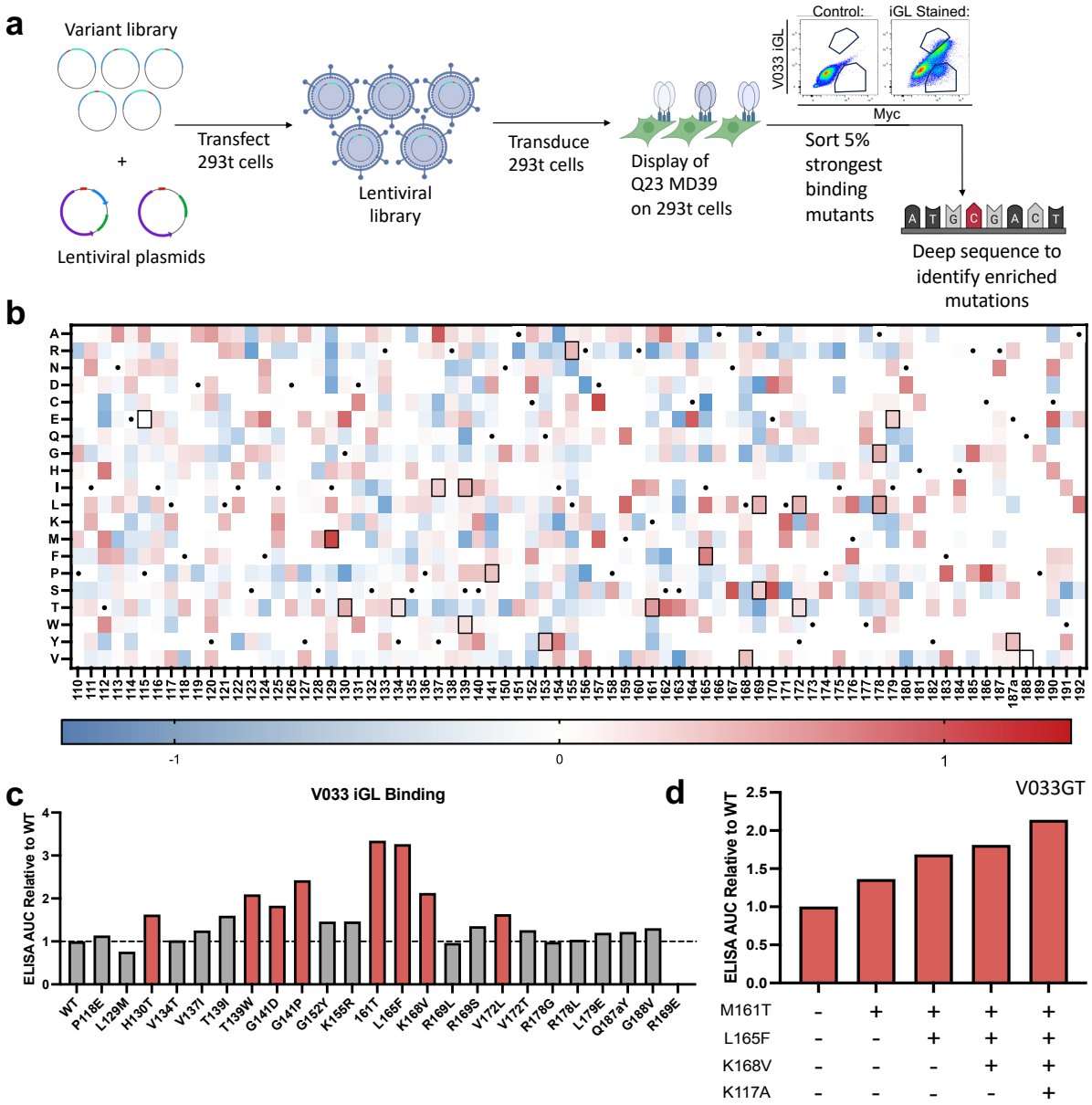


Fig. 5: A single-round mammalian display approach for the identification of V033 iGL affinity enhancing mutations a, Overview of our mammalian display approach. An immunogen variant library is packaged into lentiviral vectors, following which 293t cells are transduced with the library. Cells displaying the immunogen variant library are then stained with an antibody of interest, sorted, and deep sequenced to identify enriched mutations. **b**, Positive and negative enrichments following a single round of sorting against the V033 iGL. Black dots indicate the WT residue at each position. Boxes indicate mutations evaluated in (c) **c,** ELISA area under the curve binding values of the V033 iGL to Q23.MD39 site-directed mutants incorporating mammalian-display identified mutations. AUC values are normalized to WT. **d**, ELISA area under the curve binding values of the V033 iGL to Q23.MD39 mutants sequentially incorporating mammalian-display identified mutations. AUC values are normalized to WT.

Figure 6

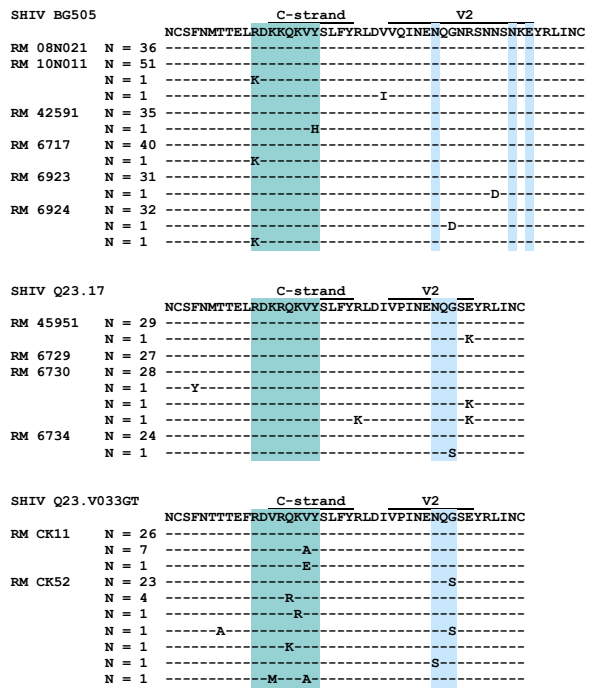
a

	CH58	CAP228-3D	3074	447-52D	17b	A32	697-D	1393	V033-a.01	V033 iGL
SHIV Q23 WT	>100	>100	82	>100	>100	>100	>100	>100	0.019	4.8
SHIV Q23 V033GT	>100	>100	9.9	>100	>100	>100	>100	>100	0.034	0.096

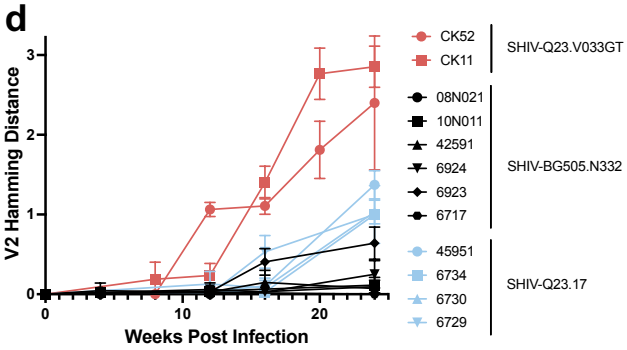
b

RM ID	WPI	MLV	Q23.17	Q23.V033GT	Q23.N160K	Q23.N187S	MN.3	MW965	TH023.6
			A	A	A	A	B	C	AE
CK52	12	<20	<20	106	<20	<20	482	819	100
	16	<20	<20	247	<20	<20	1775	4117	2023
	24	<20	1762	1039	9001	988	2597	6061	2511
CK11	12	<20	90	45	24	39	330	2340	164
	16	<20	146	628	88	48	1354	5858	944
	24	<20	574	2118	99	24	3133	9930	3438
CL60	12	<20	<20	<20	<20	<20	<20	<20	<20
	24	<20	<20	<20	<20	<20	<20	<20	<20

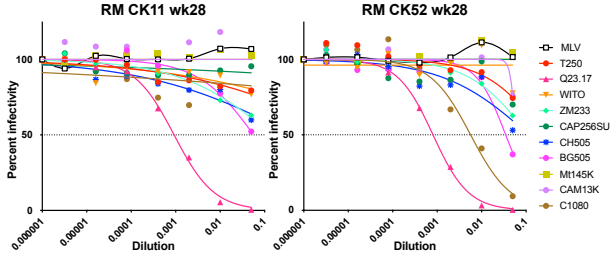
c



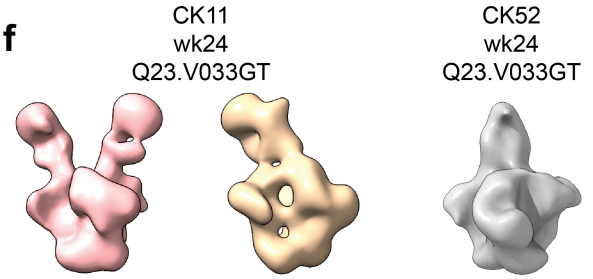
d



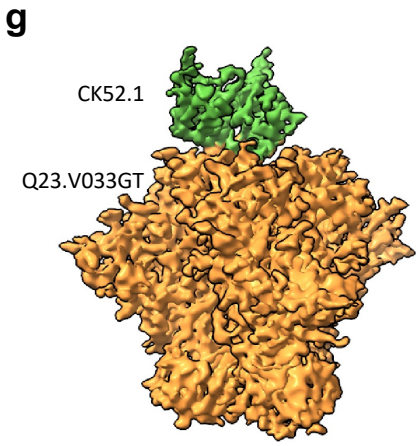
e



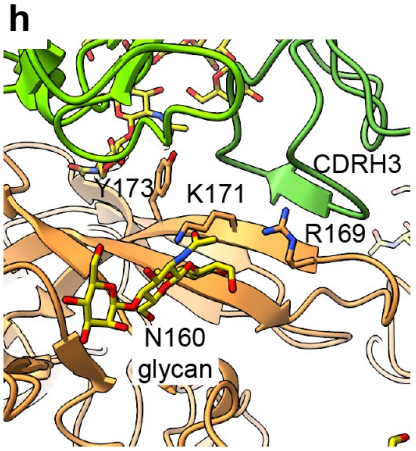
f



g



h



i

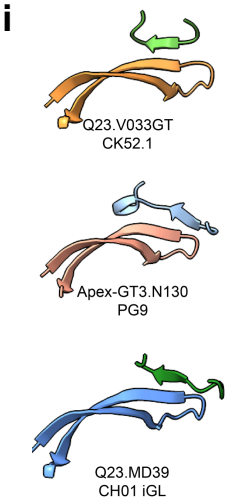


Fig. 6: Infection with a germline-targeting SHIV elicits axe-like antibodies to the V2 apex. a, Neutralization IC50 values of SHIV-Q23.17 and SHIV-V033GT against a panel of non-neutralizing antibodies, the mature V033-a.01 bnAb, and the V033 iGL. **b**, Longitudinal plasma ID50 titers against MLV, autologous mutants, and tier 1A viruses. MLV, Murine Leikemia Virus **c,** Single genome sequencing of circulating plasma viruses at week 12 post-infection in rhesus macaques infected with SHIV-BG505, SHIV-Q23.17, and SHIV Q23.V033GT. C-strand residues that are the primary site of Env escape (166-173) from V2 apex bnAbs are highlighted in teal. V2b residues that are potential sites of glycan additions are highlighted in blue. N = number of viruses. **d**, Longitudinal Hamming distances of V2 residues C131-C198 in the RMs from **c**. Error bars indicate a 95% confidence interval. **e**. Neutralization ID50 curves of week 28 plasma from macaques CK11 and CK52 against a panel of heterologous Tier 2 viruses. **f,** 3D reconstruction of selected NS-EM 3D classes that displayed V2 apex binding antibodies for macaques CK11 and CK52 at week 24. **g,** Cryo-EMPEM reconstructed density map of a V2 apex antibody (green) binding to Q23.V033GT (orange). **h,** Close-up view of the epitope recognized by the CK52 antibody CDRH3. **i,** Comparison of CK52.1 CDRH3 with PG9 (PDB:7T77) and CH01 iGL.

Figure 7

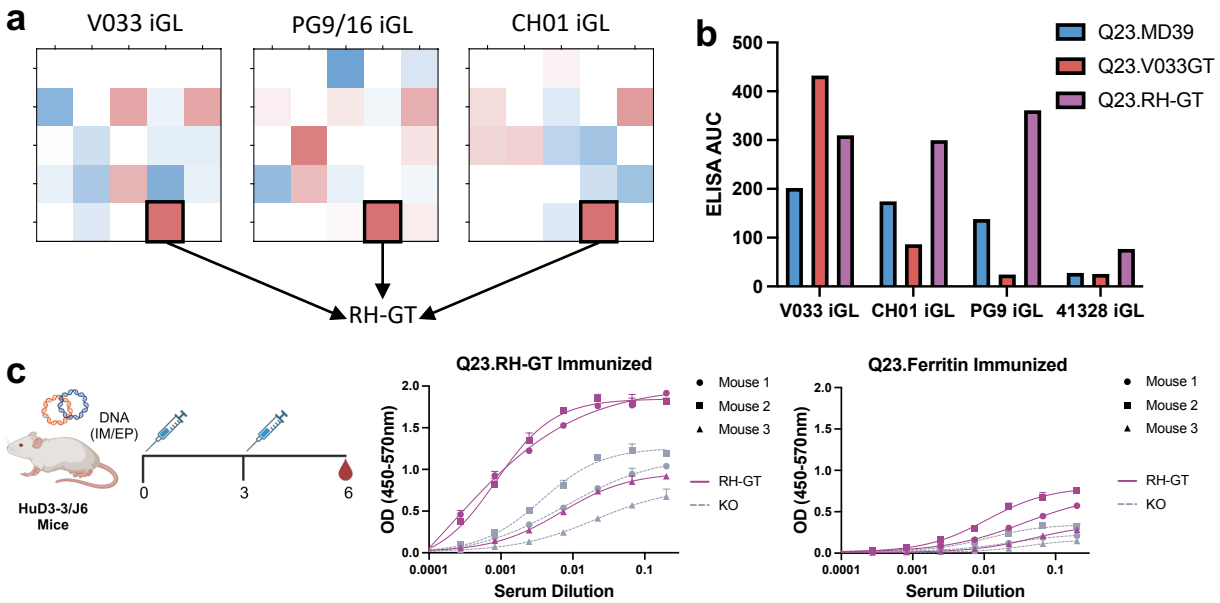


Fig. 7: Generation of a pan-axe targeting immunogen. a, Overview of our approach for immunogen design. Mutations enriched in multiple mammalian display mutagenesis sorts against a structural iGL class are identified and incorporated into an immunogen. **b,** ELISA area under the curve binding of Q23.MD39, Q23.V033GT, and Q23.RH-GT against the axe-like V2 apex bnAb precursors. **c,** Left, HuD3-3/J6 mice were primed and boosted with 25µg of DNA-encoded Q23.RH-GT adjuvanted with 0.5µg of IL-12 plasmid. Right, ELISA binding of serum from mice immunized with Q23.RH-GT or WT Q23.Ferritin to Q23.RH-GT or a Q23 epitope knockout. Shapes indicate individual mice. Solid purple lines indicate binding to Q23.RH-GT. Dashed grey lines indicate binding to Q23.KO.