Rational HIV immunogen design to target specific germline B cell receptors

Authors:
Joseph Jardine, Jean-Philippe Julien, Sergey Menis, Takayuki Ota, Oleksandr Kalyuzhniy, Andrew McGuire, Devin Sok, Po-Ssu Huang, Skye MacPherson, Meaghan Jones, Travis Nieusma, John Mathison, David Baker, Andrew B. Ward, Dennis R. Burton, Leonidas Stamatatos, David Nemazee, Ian A. Wilson, William R. Schief*

*To whom correspondence should be addressed: schief@scripps.edu

This Word file includes:

Materials and Methods
Figures S1-S26
Tables S1-S9
References

Supplementary Material:

Germline Antibody Analysis

Ab GL Predictions. The naïve heavy and light chain precursors of VRC01-class bNAbs 12A12 (*11*), 3BNC60 (*11*), NIH45-46 (*11*), PGV04 (*12*), PGV19, PGV20, VRC-CHA31 (*12*), VRC01 (*10*), (figs. S1-2) were calculated from the mature Ab sequences using JOINSOLVER® (*31*) to select probable variable (V), diversity (D) and joining (J) genes. In nearly all cases, the genes of highest probability predicted by JOINSOLVER® were selected, but in some cases (such as J gene selection), we chose the second most likely gene because it contained elements we believed made the selection more biologically relevant (figs. S20-21). For most antibodies (Abs), mature CDR3 loops were used, as it was not possible to unambiguously determine the original identity of the junction region. For some of the Abs (VRC01, NIH45-46, VRC-CH31 and PGV04), several variations of the CDRH3 or light chain (LC) pairings were constructed.

Homology Modeling. Initially, homology modeling of VRC01 and PGV04 germline (GL) Abs based on the mature Ab structures was used to generate starting structures for design. The coordinates of the mature VRC01 and PGV04 were extracted from the co-crystal structures bound to gp120 (PDBID 3NGB and 3SE9, respectively). The two amino-acid deletion in CDRL1 was modeled transferring the CDRL1 from a minimally mutated LC from the same VL gene (PDBID: 1GC1 and PDBID: 3F12) using RosettaRemodel (*32*). After correcting the length change, the GL sequence was threaded onto the coordinates of the mature Ab using RosettaFixBB (*33*) design, and the resulting

GL model was relaxed using RosettaRelax (*33*). In these calculations, the constant region of the Fab was omitted to reduce computational time.

Protein Production

eOD and gp120 Production. Unless otherwise stated, gp120 and eOD proteins were produced in FreeStyleTM 293F (Invitrogen) suspension cultures by transient transfection using 293Fectin (Invitrogen) of a pHLSec plasmid containing either mammalian codon-optimized eOD or gp120 with a C-terminal His_{6x} affinity tag. Protein was harvested from the supernatant after 96 h and purified by affinity chromatography with a HIS-TRAP column (GE) followed by Superdex 75 (for eOD) or Superdex 200 (for gp120) size exclusion chromatography (GE Healthcare) using an AKTA Express system (GE Healthcare).

***E. coli* Fab Production.** DNA segments encoding the Fab heavy and light chains containing His_{6x} and AviTag affinity tags were synthesized. DNA was codon optimized for *E. coli* expression and RNA structure, subcloned into pBAD-DEST49 Gateway® Destination Vector (Invitrogen) and transformed into TOP10 E. coli (Invitrogen). Single colonies were grown overnight at 37 °C in 10 mL Luria Broth (LB) plus Ampicillin (100 mg/mL). The starter cultures were expanded into 1 L of Terrific Broth (Fisher Scientific) plus Ampicillin and incubated at 37°C; when cells reached an OD600 of 1.0, Arabinose was added to the cultures to a final concentration of 0.05% to induce protein expression and the cells were then incubated overnight at 18°C. Cultures were pelleted and re-suspended in Start Buffer (20 mM imidazole, 500 mM sodium chloride, 20 mM sodium

phosphate), 1 tablet of protease inhibitor (Novagen). The cell suspension was thawed and lysed by the addition of 10x Bugbuster (Novagen) to a final concentration of 1X, 20 μL of Benzonase Nucleases (Novagen) and 1.0 μL of rLysozyme (Novagen) and gently rocked on a rotating mixer for 60 minutes. Lysed cells were pelleted and the supernatant was filtered through a 0.22 μm filter (Millipore). Supernatants were then passed over a 5 mL HisTrap FF crude affinity column (GE Healthcare). The resin was washed with 50 mL Wash Buffer (50 mM imidazole, 500 mM sodium chloride and 20 mM sodium phosphate, pH 7.4) and eluted with 20 mL of Elution Buffer (500 mM imidazole, 500 mM sodium chloride and 20 mM sodium phosphate, pH 7.4). Fractions containing the construct of interest were combined and further purified by affinity chromatography with a KappaSelect resin (GE Healthcare) at room temperature and eluted with 0.1 M glycine (pH 3.0) and immediately neutralized by adding an equal amount of PBS + 0.1M Tris (pH 8.0). The sample was then dialyzed into PBS (pH 7.4) overnight. Collected fractions were analyzed on a 12% SDS denaturing gel (Invitrogen) and positive fractions were combined and concentrated by ultrafiltration (Vivaspin, Bioexpress).

Mammalian Fab Production. Recombinant Fabs were produced in FreeStyleTM 293F (Invitrogen) suspension cultures by co-transfection of pHLSec plasmids containing expression constructs for light chain and Fab heavy chain using 293Fectin (Invitrogen). Protein was harvested after 96 h and the supernatant was concentrated to <200 mL by tangential flow concentration. Fabs were purified by Kappa Select (GE Healthcare) affinity chromatography followed by size exclusion chromatography (SEC) using Superdex 200 (GE Healthcare) in HBS.

Mammalian IgG Production. IgG Ab variants were produced in FreeStyleTM 293F (Invitrogen) suspension cultures by transfection of pFUSEss expression vectors (Invivogen) containing the light chain and heavy chain using 293Fectin (Invitrogen). Supernatants were harvested 96 h after transfection and Abs were purified using ProteinA SepharoseTM (GE Healthcare) and dialyzed overnight into PBS (0.01 M sodium phosphate, pH 7.4, 0.137 M sodium chloride).

Affinity Determination

Surface Plasmon Resonance (SPR). All binding experiments were carried out on a Biacore 2000 instrument (GE Healthcare) at 25˚C with HBSEP+BSA (0.01 M HEPES pH 7.4, 0.15 M sodium chloride, 3 mM EDTA and 0.005% (v/v) Surfactant P20) (GE Healthcare) + 1mg/mL BSA as running buffer. For binding analysis, ~300 response units (RUs) of IgG were captured on a CM5 sensor chip containing 8000-9000 RUs of amine-linked mouse anti-human IgG (Human Ab Capture kit, GE Healthcare). Samples of different protein concentrations were injected in duplicates over this surface at a flow rate of 50 µL/min. If necessary, surface regeneration was performed with two 60 s injections of 3 M magnesium chloride at a flow rate of 10µL/min. One flow cell contained anti-human IgG only and its interaction with the analyte was used as reference.

Data preparation and analysis were performed using Scrubber 2.0 (BioLogic Software). For kinetic analysis, biosensor data were globally fit to a mass transport limited simple bimolecular binding model:

$$A_0 \underset{k_m}{\overset{k_m}{\rightleftharpoons}} A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB$$

where A_0 represents injected analyte. For equilibrium analysis, each data set was fitted to a single site interaction model:

$$R_{eq} = \frac{R_{max} * K_1 * C_1}{K_1 * C_1 + 1}$$

where R_{eq} is the response value at equilibrium, C_A is the concentration of the analyte, R_{max} is the maximum response obtained when all binding sites are occupied by the analyte and K_D is the dissociation constant.

Preliminary Screen of gp120s. Binding for VRC01 and GL-PGV04 Abs was measured by SPR for YU2, HXB2, Bal, JRFL, JRCSF and DU179 gp120. BaL was the only construct that showed detectable binding to the GL of VRC01, although it did not show binding to the GL of PGV04. BaL core gp120 (based on the sequence http://www.ncbi.nlm.nih.gov/protein/AFJ93245.1) and a D368R knockout variant were displayed on the surface of yeast (*19*) to confirm specificity for the GL-VRC01 Ab. Surprisingly, we were unable to detect binding to either GL Ab on the surface of yeast, possibly suggesting a false positive in the original SPR binding assays. Nonetheless, the BaL sequence was used as one of our starting backgrounds.

Comprehensive Screening of HIV Env interactions with GL VH1-2*02 Abs. Ninety-six-well ELISA plates were coated overnight at 4°C with 50 uL PBS containing 5 ng/uL

of sheep D7324 anti-gp120 antibody (Aalto Bio reagents). The wells were washed four times with PBS containing 0.05% Tween-20 and blocked with 3% BSA at room temperature for 1 h. Virus supernatants from 63 different isolates were harvested after transfection in 293T cells and were lysed using a final concentration of 1% NP-40 and 50 uL of lysed virus were added to the wells. After incubating at 37°C for 2 h, the wells were washed four times with PBS containing 0.05% Tween-20. Serial dilutions of mAbs were then added to the wells, and the plates were incubated at room temperature for 1 h. After washing four times, goat anti-human IgG F(ab')2 conjugated to alkaline phosphatase (Pierce) was diluted 1:1000 in PBS containing 1% BSA and 0.025% Tween-20 and added to the wells. The plate was incubated at room temperature for 1 h, washed four times, and the plate was developed by adding 50 uL of alkaline phosphatase substrate (Sigma) to 5 mL alkaline phosphatase staining buffer (pH 9.8), according to the manufacturer's instructions. The optical density at 405 nm was read on a microplate reader (Molecular Devices).

Engineering GL Immunogens

Computational Design. A homology model of the GL-VRC01 was aligned onto the coordinates for the mature VRC01/gp120 complex (PDBID: 3NGB) using a backbone alignment to the mature Ab. Subsequently, a crystal structure of unliganded Gl-VRC01 confirmed the initial homology models. An evident clash between the GL-VRC01 CDRL1 insert and the glycan at position N276 was observed. The glycan was, therefore, removed with an N276D mutation and the redocked complex was minimized to reduce side-chain clashes.

Computational design to improve affinity was carried out using RosettaScripts (*34*) to call sequential modeling tasks in the Rosetta software suite. Initially, small variations in the rigid-body orientation were generated using RosettaDock (*35*) followed by RosettaBackrub (*36*) to sample slight conformations in the protein backbone. The initial modeling was done using low resolution side-chain representations (centroid mode) with the intention of creating diversity in the starting structure since the computational interface design was being carried out with a docked homology model rather than a high-resolution structure, therefore confidence in the input structure was limited. Following initial perturbation of the modeled complex, sequence design was performed using RosettaDesign with full atom representation. All side chains within 8 Å of the interface on both partners were allowed to sample alternative rotameric conformations, and interface mutations were sampled on gp120 to minimize the energy of the complex. Five rounds of interface optimization were carried out to generate sufficient diversity. Because small modifications in backbone position can cause significant bias in side-chain design, a backbone conformational variant was generated prior to each design step, by passing the pose through RosettaBackrub followed by RosettaDesign. After the pose was passed through five iterations of backrub followed by design, a final output structure was generated.

This procedure was repeated ~100,000 times with each run producing a unique model. Output decoys were filtered based on Rosetta calculated total score and binding energy (Ecomplex –Epartner1 – Epartner2) and, then loosely filtered on unsatisfied polar

residues and RMSD from the starting structure. Sequences from the top ~100 decoys were aligned, and the predicted mutations were manually inspected. We specifically selected mutations that made backbone contacts with the Ab or targeted side chains that were identical between both the GL and the mature Ab. Mutations that passed manual inspection were used to create directed libraries of gp120 mutants predicted to have increased affinity for the GL-VRC01.

Surface Display and Selection Overview. Positions on gp120 identified by Rosetta to improve GL Ab binding were screened experimentally using yeast surface display and fluorescence-activated cell sorting (FACS). Directed libraries were constructed using PCR assembly with partially degenerate primers (*37*) encoding the gp120 or eOD of interest. Mutations at positions that were computationally predicted to be beneficial for GL binding, as well as the wild-type amino acid were used to generate a library using degenerate codons so that all possible combinations of desired amino acids would be sampled at that position. The PCR assembly product was gel purified and cloned into pCTCON2 for yeast surface display via homologous recombination in yeast (*38*). gp120 variants were expressed on the surface of EBY100 as a fusion protein between Aga2p and a C-Myc tag (*19*). Biotinylated (Biotin Protein Ligase, Avidity) GL-VRC01 Ab was pre-incubated with phycoerythrin-conjugated streptavidin (SAPE) (Invitrogen) to create tetrameric complexes to maximize avidity. SAPE/Ab complexes were purified by Superdex 200 size exclusion chromatography (GE Healthcare) using an AKTA Avant system (GE Healthcare).

The yeast library was induced at an optical density of $1x10^7$ for ~12 h in C-Trp-Ura + 2% galactose dropout media at 30˚C. Induced cells were pelleted, washed and suspended in PBSA buffer (0.01 M sodium phosphate, pH 7.4, 0.137 M sodium chloride, 1 g/L bovine serum albumin) and were duel-labeled with fluorescein isothiocyanate labeled α-cMyc Ab (Immunology Consultants Laboratory) and the SAPE/GL-VRC01 Fab complex at 4˚C for 1 h.

Cells were analyzed by FACS on a BD Influx (BD Biosciences) and double positive clones were collected, expanded and re-induced for additional rounds of selection. After modest binding was achieved, the above protocol was modified such that cells were first labeled with biotinylated Fab, washed in PBSA, then labeled with SAPE for fluorescence.

After several rounds of enriching for binding to GL Abs, sequences were recovered. Cells were plated at a density of ~200 clones per plate on C-Trp-Ura + 2% glucose agarose plates, plasmid DNA was extracted and sequenced (Genewiz, La Jolla).

Development of germline-targeting Core gp120. After 3 rounds of sorting using libraries on core BaL [based on http://www.ncbi.nlm.nih.gov/protein/AFJ93245.1] and 93TH057, clones had enriched for the core BaL gp120 library that bound to GL-VRC01, but not to GL-PGV04. No clones enriched from the 93TH057-based library when sorted against either antibody. After 5 rounds of sorting, sequences were recovered from the core BaL library that bound to GL-VRC01. A representative sequence was produced as

soluble gp120 (referred to as CoreBal-GT1) in 293F cells, and was found to bind GL-VRC01 (Table 1). As a control, we also produced core gp120 BaL [based on http://www.ncbi.nlm.nih.gov/protein/AFJ93245.1] on which the library was initially based, as well as another quasi-species of BaL [http://www.ncbi.nlm.nih.gov/protein/AAR05834.1] that lacked a glycan at position N276 due to a T278A mutation. The latter protein bound to GL-VRC01 with a K_D of 4.6 μM, however, a D368R knockout mutation only reduced binding to GL-VRC01 to 6.9 μM and improved binding to GL-PGV19 to 10 nM, so we hypothesized that this interaction may not represent specific binding.

Development of germline-targeting eOD. We had separately developed an engineered outer domain lacking the inner domain, the V1, V2, and V3 regions, and β20/21, to serve as a minimal antigen for the CD4bs and other epitopes (*17*) and Huang et al, in preparation. To take advantage of these properties for germline-targeting, we sought to develop germline-targeting eODs. This was further encouraged during our development of germline-targeting gp120, by our finding that one of the gp120 modifications that enriched was a shortening of the β20/21 by 6 amino acids.

Using eOD instead of core gp120, new homology models were generated and the Rosetta interface design protocol was used to predict eOD specific mutations that would improve GL Ab binding. Rosetta-directed libraries were combined with mutations that had enriched from BaL core gp120 library sorts (and including two loops from BaL) and sorted as described above on the surface of yeast for GL-VRC01 and GL-PGV04

binding. Clones that showed binding were sequenced, the resulting protein (eOD-GT1) was expressed, and the K_D for GL-VRC01 binding was measured to be 44 μM by SPR (table S2). Mutations from that clone were modeled, and computational affinity maturation was carried out a second time, with a starting structure containing the previously identified mutations. This second computational library was screened on the surface of yeast as described above and resulted in eOD-GT2, which bound GL-VRC01 with a K_D of 15 μM by SPR (table S2).

Further optimization was carried out on a library containing error prone (EP) PCR-generated random mutations [GeneMorph II Random Mutagenesis Kit, Agilent Technologies] on the DNA recovered from eOD-GT2. The resulting eOD (eOD-GT3) had a 220 nM affinity for GL-VRC01 (table S2). A second round of error prone mutagenesis was conducted on the DNA that produced eOD-GT3, and screening of that error prone library enriched for an L260F mutation as well as a K464N/E mutation. Rather than testing those mutations individually, a larger computational library was constructed that allowed both of those positions to vary among the original and EP mutations as well as additional Rosetta directed mutations. The resulting computational/random mutagenesis library produced eOD-GT4 that bound GL-VRC01 with a K_D of 34 nM (table S2).

During the course of eOD-GT development, other VH1-2 Abs became available, including NIH45-46 and PGV19. GL-NIH45-46 bound to eOD-GT4 with a K_D of 1 μM and GL-PGV19 bound with a K_D of 27 nM as measured by SPR (table S2). We then used

a multistate sorting protocol **(**fig. S22**)** to optimize against multiple targets to identify the minimal number of mutations necessary for binding to diverse GL and mature VH1-2 Abs. A library was generated in which the original residue as well as others that had previously enriched in our sorts was sampled. The resulting library was divided in two and each was screened against GL-VRC01 or GL-NIH45-46. After two sorts, each library was split in half again, and one half was sorted against the opposite Ab while the other half was further sorted on the original Ab. 50 clones from each of the four libraries were sequenced. We found that some of the mutations that enriched for one of the GL Abs did not result in improved binding by the other Ab. From these data, only mutations that were present in all four libraries were selected and used to express eODs for binding assays with both Abs. By SPR, eOD-GT5 bound to GL-VRC01 with an affinity of 530 nM and bound to GL-NIH45-46 with an affinity of 4.8 μM (table S2)**.** We reverted the A281S mutation on loop D, as well as the S365L mutation on the CD4 binding loop, back to their original positions on eOD-base. While these mutations improved GL affinity, they also significantly modified the interface.

To further minimize mutations, a second reversion library was created and that library was divided into seven sub-libraries and sorted for binding to the GL Abs of VRC01, NIH45-46, PGV19, CHA31, PGV04, as well as mature VRC01 and PGV04 Abs. 50 clones from each the GL-VRC01, GL-NIH45-46, GL-PGV19, mature VRC01 and mature PGV04 libraries were sequenced. Common mutations that enriched in all libraries were selected and combined to create eOD-GT6. This final procedure removed four additional mutations while improving affinity for germline antibodies compared to eOD-

GT5 (table S2). We suspect the affinity improvements may be due largely to reverting the D457G mutation that had removed a salt bridge between D457 and R469.

eOD-GT6 differed from the starting eOD-base construct by 8 mutations, including the 6 shown in Table 2 along with S465T and N386D, as well as a point deletion at position 356 (del356). The S465T mutation and del356 were relics from the use of two BaL loops in eOD-GT1. The N386D mutation is discussed in the main text. We have tested reversion of the S465T mutation and it has no significant effect on GL-VRC01 affinity (table S7). del356, present in 84% of HIV-1 strains, was considered part of the scaffold.

Point Reversions. Point reversions of the mutations that contributed to improved GL Ab binding in eOD-GT6 were generated and tested for binding by SPR **(**Table 2**)**

Generation of Mouse Abs. The mouse VH gene collection was downloaded from IMGT, translated and aligned. Based on our sorting criteria, we were unable to find any Abs that contained all of the critical contact residues. Specifically, there were no Abs that had both Arg^{H71} and Trp^{H50}, which we believe to be very important for the initial germline binding interaction. We produced the 5 closest mouse VH genes with Arg^{H71} as GL-VRC01 hybrids, in which we replaced the human VH1-2*02 gene of GL-VRC01 with different mouse VH genes **(**fig S18**)**. Binding to these Abs was assessed for eOD-GT6 , as well as Core BaL-GT1. No construct showed detectable binding when flowed as analyte at 100 μM concentration.

Generation of Rhesus Abs. To investigate whether a macaque germline repertoire could potentially bind eOD-GT6, we looked for VH1-2-like Ab VH genes in the un-annotated genome using both BLAST and Ensembl to look for potential VH genes in the sequenced macaque genome. Several were identified **(**fig. S20**)** and GL-VRC01 VH gene replacements were generated. We assessed binding to eOD-GT6, and the results are described in the main text.

Crystallization and Structure Determination

Crystallography. For crystallization, GL-VRC01 Fab was produced in *E. coli* with a C-terminal His_{6x} tag on the heavy chain. Fab was first purified with a HisTrap column (GE Healthcare) followed by KappaSelect (GE Healthcare). Subsequently, to ensure a predominant monomeric population, the Fab was reduced and alkylated with DTT and iodoacetamide, as previously described (*39*). The resulting sample was purified by MonoS cation exchange chromatography (Sigma) followed by Superdex 200 size exclusion chromatography (GE Healthcare) using an AKTA Avant system (GE Healthcare). Purified GL-VRC01 Fab in a buffer containing 20 mMTris, 150 mM NaCl, pH 8.0 was concentrated to 8.5 mg/ml and setup for crystallization trials using the automated CrystalMation robotic system (Rigaku) at the Joint Center for Structural Genomics (www.jcsg.org). Crystals were obtained from a solution containing 20% PEG 3350 w/v, 0.2 M lithium citrate.

For crystallization of eOD-GT6, a minimal glycan (mglyc) construct possessing only two glycosylation sites (N18 and N65, eOD numbering) was designed and subsequently

transfected in lab-adapted 293S ($GnTI^{-/-}$) suspension cells. The secreted His_{6x}-tagged protein was isolated from the supernatant by affinity chromatography using HisTrap nickel columns (GE Healthcare). Before purification to homogeneity by Superdex 200 size exclusion chromatography (GE Healthcare), eOD-GT6mglyc was treated with EndoH (NEB) using a protocol similar to that previously described (*40*). The deglycosylated sample was concentrated to ~10 mg/ml and subjected to crystallization trials using the automated Rigaku CrystalMation robotic system at the Joint Center for Structural Genomics (www.jcsg.org). Initial crystal hits were obtained in a solution containing 20% PEG 3350 w/v, 0.2 M magnesium acetate at 4°C.

To obtain complex crystals of GL-VRC01+eOD-GT6, it was first necessary to resurface eOD-GT6 to mutate aspartic acid residues initially introduced to remove sites of glycosylation (NXT/S to DXT/S) to alanine residues (AXT/S). Proteins were expressed and purified as described above. After incubation in molar excess of eOD-GT6, the complex was Endo-H (NEB) treated and subsequently purified to homogeneity by Superdex 200 size exclusion chromatography (GE Healthcare). The deglycosylated sample was concentrated to ~10 mg/ml and subjected to crystallization trials using the automated Rigaku CrystalMation robotic system at the Joint Center for Structural Genomics (www.jcsg.org). Initial crystal hits were obtained in a solution containing 20% PEG 4000 w/v, 0.1 M sodium citrate, pH 5.6, 20% 2-propanol v/v. All crystal were prepared for X-ray diffraction analysis by first cryoprotecting them in the mother liquor supplemented with 20% glycerol followed by fast plunging into liquid nitrogen.

X-ray Data Collection, Data Processing and Structural Determination. Flash-cooled crystals of unliganded GL-VRC01, EndoH-treated eOD-GT6mglyc and EndoH-treated eOD-T6+GL-VRC01 were subjected to high energy X-ray radiation at the Advanced Photon Source (APS) or Stanford Synchrotron Radiation Lightsource (SSRL) and diffraction images were collected with strategies leading to datasets of high completeness and redundancy. Data processing was performed using XDS (*41*). Statistics for data collection and processing are reported in table S4. The GL-VRC01 Fab crystal structure was solved using VRC01 Fab coordinates from the PDBID 3NGB as a search model for molecular replacement in PHASER (*42*). To solve the unliganded EndoH-treated eOD-GT6mglyc crystal structure, coordinates of eOD from PDBID 3TYG were used as a search model for molecular replacement using PHASER (*42*). Finally, individual components were used as search models for molecular replacement of the EndoH-treated eOD-GT6+GL-VRC01 dataset. For all structures, refinement was performed using a combination of CCP4 (*43*), PHENIX (*44*) and COOT (*45*). The eOD-GT6mglyc crystals belonged to space group $P2_1$ with unit cell dimensions of a=44.96 Å, b=217.73 Å, c=44.99Å, $\alpha=\gamma=90°$, $\beta=119.97$. These unit cell dimensions are also consistent with hexagonal, C-centered orthorhombic or C-centered monoclinic lattices, however data processed in those higher-symmetry lattices had very high R-merge values (>40%). Significant peaks (~70% of the origin height) were observed in the self-rotation function for orthogonal non-crystallographic 2-fold axes in the XZ plane, -30° and 60° from Z. Significant pseudo merohedral-twinning was detected from analysis with phenix.xtriage (*44*). Molecular replacement was successful in placing four copies of eOD-GT6mglyc in the $P2_1$ cell. Amplitude-based twin refinement in Refmac5 (*45*) was applied throughout

refinement, and all six pseudo-merohedral twin operators possible for this lattice were tested (with default R_{merge}< 0.44 and twin domains > 0.07 parameters) resulting in final twin operations (twin fraction) of: 1) h, k, l (0.489); 2) h+l, -k, -l (0.408); and 3) l, -k, h (0.102). Refinement statistics for the final models are reported in table S4.

Multimerization and B Cell Activation

Multimeric Particles. Two different eOD-GT6 multimers were created with the objective to assess their ability to activate B cells *in vitro*. eOD trimers were generated by fusing the C-terminal end of eOD-GT6 to the N-terminal end of a GCN4 trimer (*15*) (PDBID: 1GCM) with a 12 amino-acid GGSGGSGGSGGG linker. The resulting protein was analyzed by SEC-MALS (Wyatt Corporation, Santa Barbara, CA) to confirm trimer formation (fig. S15). Self-assembling 60-mer particles were generated by fusing the C-terminus of lumuzine synthase (*16*) (PDBID: 1HQK) to the N-terminus of eOD-GT6 via a GGSGGSGGSGGSGGG linker (fig. S23). Multimerized particles were purified by lectin affinity. The resulting particles were analyzed by SEC-MALS (fig. S15) and Electron Microscopy (EM). YU2 gp140 trimers were provided by Richard Wyatt. These were confirmed to be predominantly trimeric by SEC-MALS (fig. S15) and confirmed to bind with high affinity to mature VRC01 by SPR (data not shown).

Electron microscopy. For EM, a 3 μL aliquot of the multimerized particles (~0.05 mg/mL) was applied for 5 sec onto a carbon coated 400 Cu mesh grid that had been glow discharged at 20 mA for 30 sec, then negatively stained with 2% uranyl formate for 30 sec. Data were collected using a FEI Tecnai F20 electron microscope operating at 120

keV using an electron dose of 30 $e^-/Å^2$ and a magnification of 100,000x. Images were acquired with a Gatan 4k x 4k CCD camera using a nominal defocus range of 500 to 900 nM (fig. S14).

VRC01 B Cell Activation. B cell lines were used as previously reported (*23*). Briefly, cells were suspended at 4 million cells/mL in Advanced DMEM, labeled with 1.5 μM Indo-1 (Invitrogen) for 30 min at 37°C and washed with 2 mM $CaCl_2$ HBSS, followed by another 30 min at 37°C. Aliquots of 2×10^6 cells in 0.5 mL were then stimulated at room temperature with BCR ligands. Ca^{2+} signals were recorded for 180 s, measuring the 405/485-nm emission ratio of Indo-1 fluorescence upon UV excitation. Calcium flux analysis was performed on an LSR II cytometer (BD Biosciences). BCR ligand concentrations was tested at 1 μM, 100 nM, 10 nM and 1 nM. Kinetic analyses were performed using FlowJo (Tree Star).

NIH45-46 B Cell Activation. B cell lines were used as previously reported (*24*). Briefly, the DG-75 human Burkitt's lymphoma (ATCC # CRL-2625) cell line was maintained in RPMI-1640 supplemented with 10% FBS. For electoroporation, 2 $X10^6$ DG-75 cells were suspended in 100 μl of cell line nucleofector solution V (Lonza, Cologne Germany), containing 5 μg of BCR-expressing plasmids and electroporated with the Amaxa Nucleofector II, program O-006 (Lonza, Cologne Germany). At 24 h post-transfection, cells were loaded with Fluo-4 Direct calcium indicator (Invitrogen, Carlsbad CA), in RPMI-1640 medium containing 10% FBS at 37 °C for 45 min. Cells were pelleted and stained with APC-conjugated mouse monoclonal anti-human IgG (BD Pharmingen Cat. #

550931) at a 1/10 dilution in 100 µl of RPMI-1640 with 10% FBS and Fluo-4 Direct for 15 min. The cells were washed with 5 mL in RPMI-1640 containing 10% FBS, pelleted, and resuspended at ~$1x10^6$ cells/mL in RPMI-1640 and subjected to Ca^{2+} flux analysis at a medium flow rate on an LSR II cytometer (Beckton Dickinson). For all the experiments, we gated on B cells expressing comparable numbers of BCRs.

Minimum levels of background fluorescence (Min_{FL}) were determined by averaging the background Fluo-4 absorbance in cells for 30 s, and then activation of exogenous BCRs with various immunogens was determined by monitoring changes in Fluo-4 fluorescence associated with cells expressing the exogenous BCRs (APC positive cells) for 210 s. Ionomycin was added to a final concentration of 6.5 nM for 60 s and maximum Fluo-4 fluorescence (Max_{FL}) was established by averaging changes in Fluo-4 fluorescence recorded during the last 10 s.

The percent of maximum Fluo-4 fluorescence at each time point, t, was determined using the formula: (Fluorescence at t-Min_{FL}) / (Max_{FL}-Min_{FL}) x 100. This analysis was performed on both the BCR positive (anti-IgG-APC positive) and BCR negative cells (anti-IgG-APC negative) simultaneously. The background Fluo-4 fluorescence signal from the BCR negative cells was subtracted from that of the BCR positive population at each time point.

Supplementary Figures:

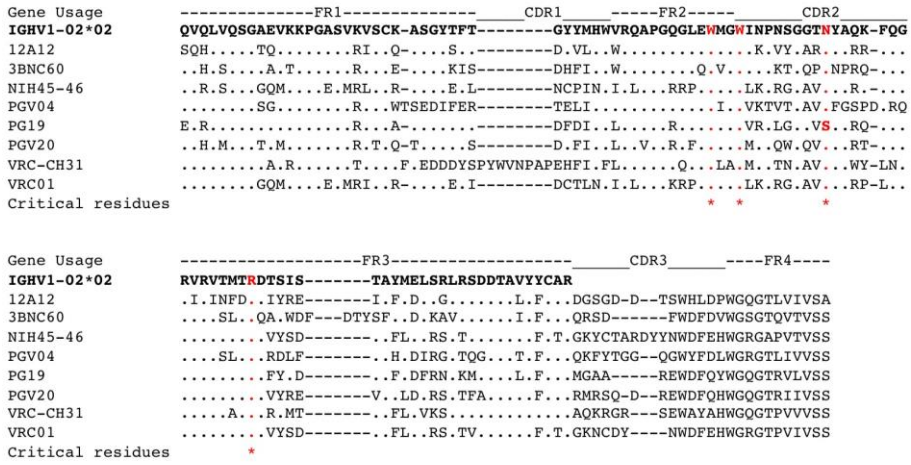


Fig. S1. Sequence alignment of the heavy chain of the 8 VRC01-class Abs used in this study aligned to their predicted GL VH precursor. Positions of GL encoded

residues important for binding to gp120 highlighted in red.

```
IGKV1-33*01       DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPSRFSGSGSG
12A12             ........................G.G.GSS.Q.............VHG....HR..........FH
3BNC60            .............R...T.......N----G........RR..........G.K..R...A....RRW.
VRC-CH31          .............L...........RG.GKD.......A.......VS...T..G..........FH

IGKV3-11*01       EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSG
VRC01             ........G........T.II...T..YG.--.......R.......V..SG.T..A...D.....RW.
NIH45-46          ................T.II...T...G.--.......R.......V..SG.T..A...D.....RW.

IGKV3-20*01       EIVLTQSPGTLSLSPGERATLSCRASQSVSSSYLAWYQQKPGQAPRLLIYGASSRATGIPDRFSGSGS
PGV04             ................T.S...T.ASYGH---MT...K....P.K...FAT.K..S.........QF

IGLV2-14*01       QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYEVSNRPSGVSNRFSGSK
PGV19             .......................-A...FR.FSS......V..R....L.FS.NR....I.H......
PGV20             .......P...........L....A.T------S.A....YAD...R.IVFDGNK...DI.S.....Q


IGKV1-33*01       TDFTFTISSLQPEDIATYYCQQYDNLP
12A12             .T.SL...G..RD.F...F.AVLEFFGPGTKVEIK
3BNC60            QEYNL..NN.....V...F..V.EFIVPGTRLDLK
VRC-CH31          QN.SL......A..V...F....ETFGQGTKVDIK

IGKV3-11*01       TDFTLTISSLEPEDFAVYYCQQRSNWP
VRC01             P.YN....N..SG..G......YEFFGQGTKVQVDIK
NIH45-46          A.YN.S..N..SG..G......YEFFGQGTKVQVDIK

IGKV3-20*01       GTDFTLTISRLEPEDFAVYYCQQYGSSP
PGV04             .KQY....T.M......R.....LEFFGQGTRLEIR

IGLV2-14*01       SGNTASLTISGLQAEDEADYYCSSYTSSSTL
PGV19             .............I......H.NA.EFFGGGTKVFVLG
PGV20             ..G..........S....Y.H.NAFEFFGGGTKLTVLS
```

Fig. S2. Sequence alignment of the light chain of the 8 VRC01-class Abs used in this study aligned to their predicted GL VL precursor.

```
eOD Base   DTITLPCRPAPPPHCSSNITGLILTRDGGNSNDESEIFRPGGGDMRDIARCQIAGTVVSTQLLLNGSLAEEEVVIRSVDFTDNAKSICVQLNTSVEINC
eOD-GT1    ............................T.D.KT.........................................E..R..................
eOD-GT2    ........................G..I.D.KT.........................................E..R..S...............
eOD-GT3    ........................G..I.D.KT.....S...................................E..R..S...............
eOD-GT4    ........................G..I.D.DT.....S.....................F..............E..R..S...............
eOD-GT5    ........................G..V.D.DT.....S.....................F.................R..................
eOD-GT6    ............................V....T.....S.....................F.................R..................
```

```
eOD Base   TGAGHCNISRAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNST
eOD-GT1    ............................-.............F..............D.....D....D..
eOD-GT2    ............................-......S..L.....F..............D.....D....D..
eOD-GT3    ............................-.R....S..L.....F..............D.....D....D..
eOD-GT4    ............................-.R....S..L.....F..............D.....D....D..
eOD-GT5    ............................-.R....S........F..............D.....D....D..
eOD-GT6    ............................-.R.............F..............D.............
```

Fig. S3. Sequence alignment of the eOD-Base and eOD-GT variants.

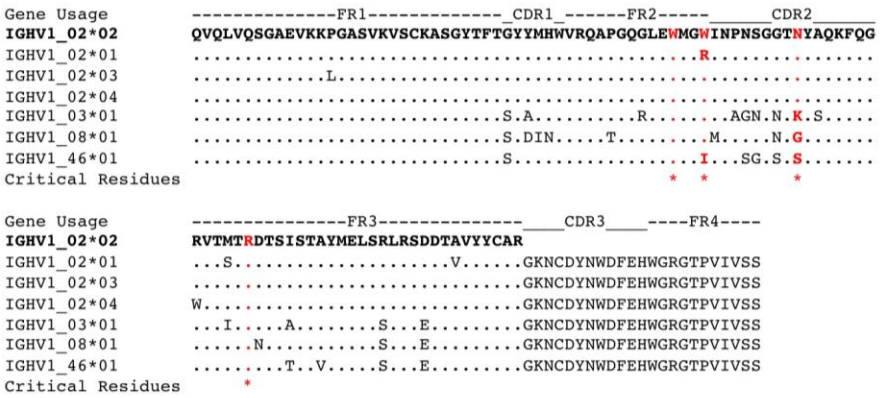


Fig. S4. Sequence alignment of human VH1-2 alleles and closely related human VH genes, with positions of GL encoded residues important for binding to gp120 highlighted in red.

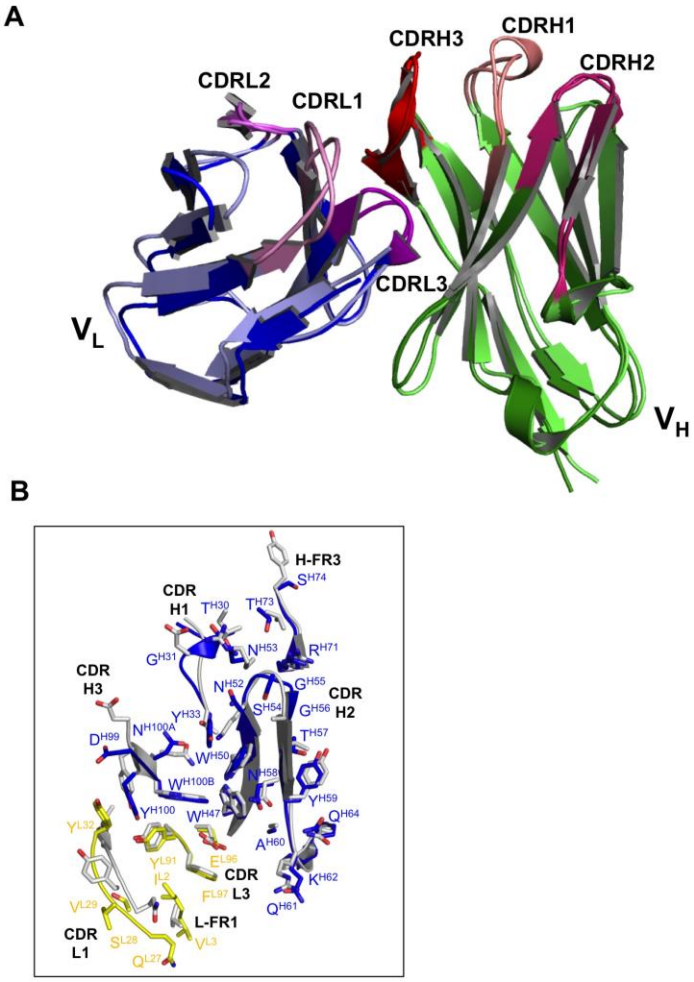


Fig. S5. (A) Comparison between the crystal structures of an unliganded GL-VRC01 Ab (heavy and light chains are colored light green and light blue, respectively) and a gp120-bound mature VRC01 Ab (PDBID: 3NGB, heavy and light chains are colored green and blue, respectively). Only the variable regions are shown for clarity, and structures are rendered as secondary structure cartoons. The CDR loops have been colored differently and are labeled. The overall conformation of the GL and mature VRC01 Fv is similar (core RMSD of 1.2 Å, despite a 66% sequence identity). The elbow angle varies significantly between mature VRC01 (223.1°) and GL-VRC01 (161.4°), but this difference can probably be attributed primarily to crystal packing (*46*). (**B**) Sidechains that directly mediate gp120 contacts are mostly structurally conserved between the GL and mature VRC01. GL (heavy and light chain colored blue and yellow, respectively) and mature VRC01 (gray) are rendered as sticks and ribbon.

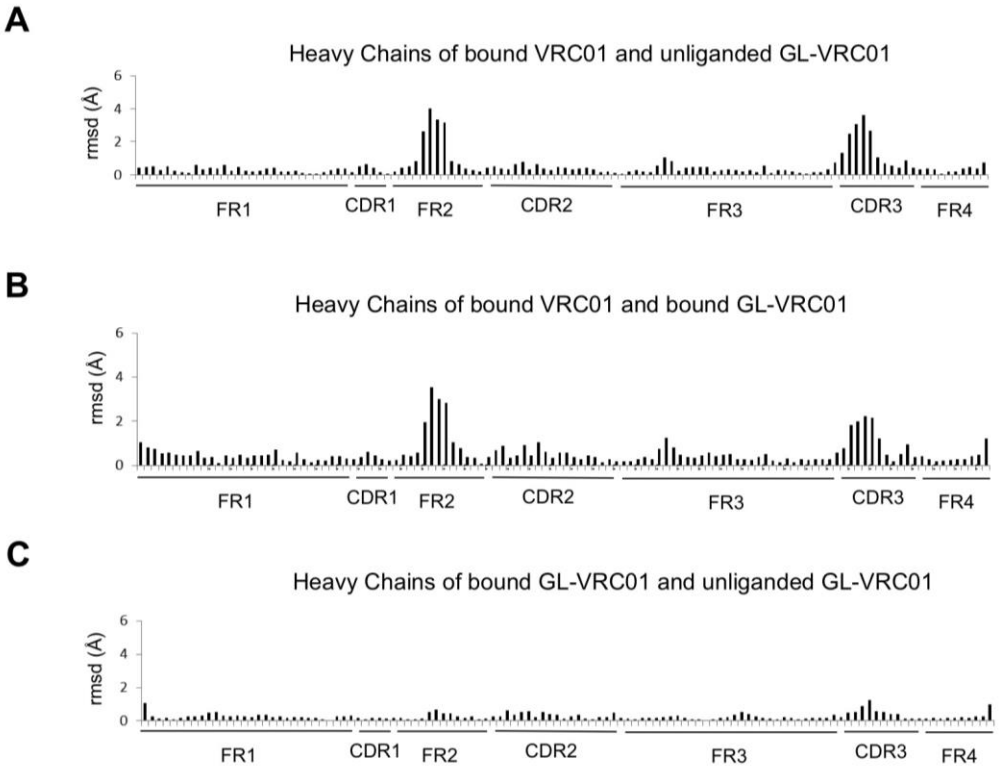


Fig. S6. RMSD comparison between the heavy chains of VRC01 and GL-VRC01 from different crystal structures. The gp120-bound VRC01 crystal structure has PDBID: 3NGB. In (A) and in (B), we observe that the bound-VRC01 varies significantly from GL-VRC01 in the FR2 and CDR3 regions. However, we note that FR2 does not significantly take part in interactions with gp120. Furthermore, analysis of the GL-VRC01 CDR3 loop is limited by the fact that the exact VDJ recombination event in the germline antibody is not predicted in this construct. (C) No large conformational changes are observed in the heavy chain of GL-VRC01 upon

binding eOD-GT6. Cα rmsd were calculated using Chimera (*30*).

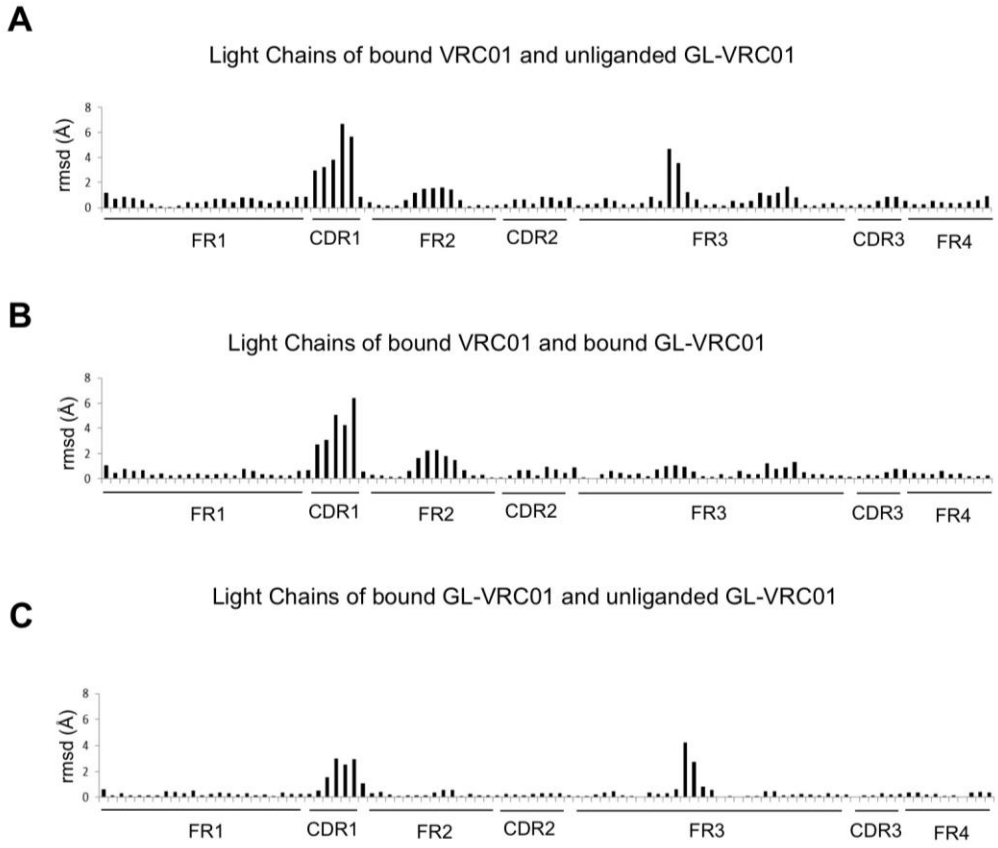


Fig. S7. RMSD comparison between the light chains of VRC01 and GL-VRC01 from different crystal structures. The gp120-bound VRC01 crystal structure has PDBID: 3NGB. In (A) and in (B), we observe that the bound-VRC01 varies significantly from GL-VRC01 in CDR1. A two residue deletion in this region results in conformational changes. (C) Slight conformational changes are observed in the light chain of GL-VRC01 upon binding eOD-GT6, particularly in CDR1 and FR3.

Cα rmsd were calculated using Chimera (*30*).

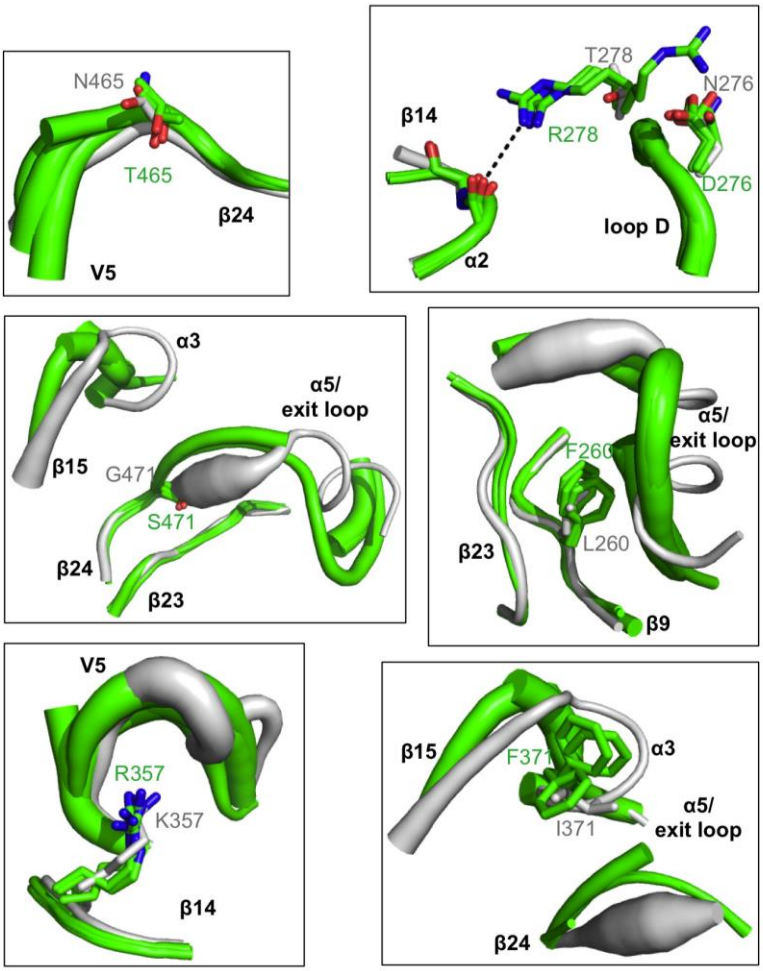


Fig. S8. Detailed view of the mutations in eOD-GT6 that allow binding of GL-VRC-class GL Abs. The crystal structures of unliganded eOD-GT6 (green) and unliganded gp120 core (PDBID: 3TGT, gray) have been superimposed to allow for direct comparison. Also, the four copies present in the asymmetric unit of eOD-GT6 crystal structure have been superimposed and are shown. Structures are rendered according to B-values, with thin and thick lines representing areas of low and high flexibility, respectively. Mutated residues, along with key neighboring residues are shown as sticks. Overall, the unliganded eOD-GT6 crystal structure indicates that key structural features are largely maintained with gp120, despite inner domain truncation and introduction of several mutations in the eOD-GT6 construct.

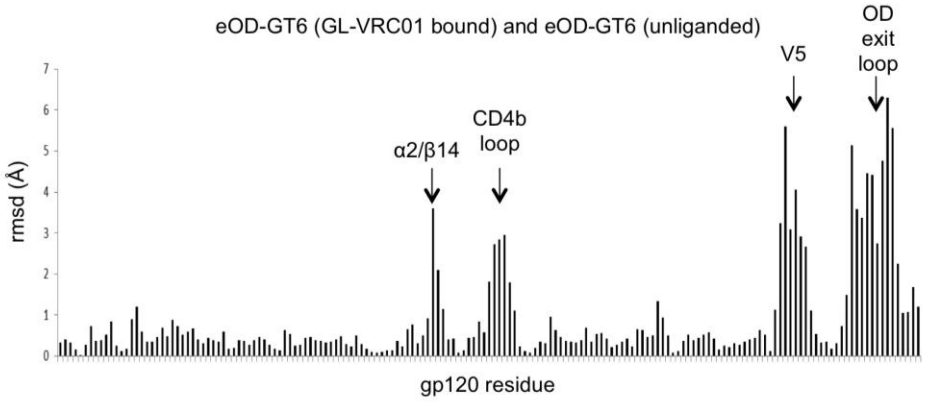


Fig. S9. RMSD comparison between unliganded eOD-GT6 and GL-VRC01 liganded eOD-GT6. Several conformational changes are observed in the eOD-GT6 upon GL-VRC01 binding, particularly in α2/β14, the CD4b loop, V5 and the eOD exit loop.

Cα rmsd were calculated using Chimera (*30*).

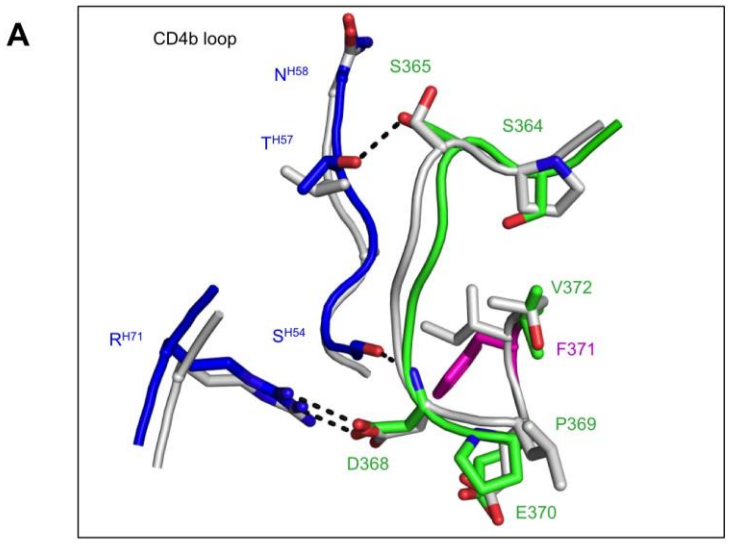


Fig. S10. Comparison between the crystal structures of GL-VRC01+eOD-GT6 and VRC01+gp120 core in the CD4b loop. GL-VRC01 heavy chain is shown in blue. eOD-GT6 is shown in green, with the mutated F371 and V460 shown in magenta. VRC01+gp120 core components are superposed and shown as gray cartoon. In the CD4b loop, GL-VRC01+eOD-GT6 reveals a mode of recognition nearly identical

to that observed in the VRC01+gp120 core (PDBID: 3NGB) structure.

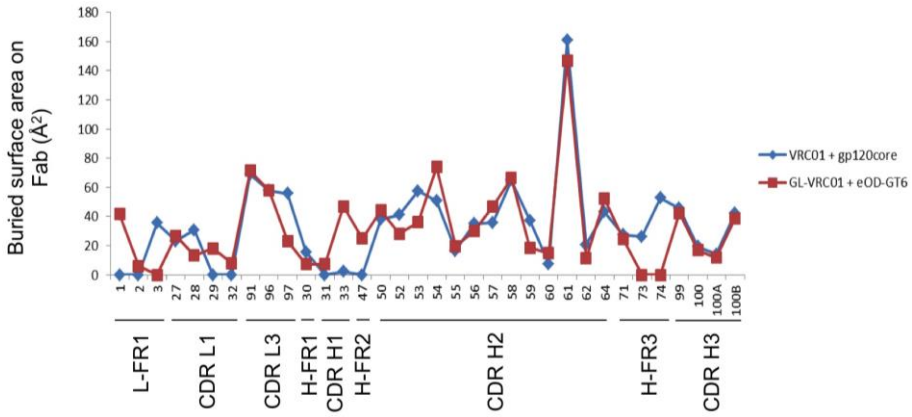


Fig. S11. Graphical rendering of the analysis of the buried surface area and H-bonds on VRC01 and GL-VRC01 in the complexes of VRC01+gp120core and GL-VRC01+eOD-GT6 (Table S5). Interfaces were calculated using PDBePISA (*29*).

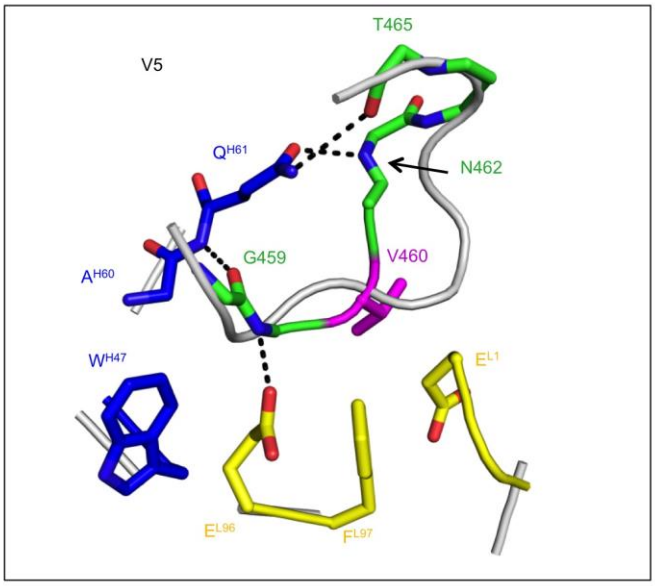


Fig. S12. Comparison between the crystal structures of GL-VRC01+eOD-GT6 and VRC01+gp120 core in the V5 region. GL-VRC01 heavy and light chains are shown in blue and yellow, respectively. eOD-GT6 is shown in green, with the mutated F371 and V460 shown in magenta. VRC01+gp120 core components are superposed and shown as gray cartoon. The GL-VRC01+eOD-GT6 reveals a different mode of recognition of V5 than that observed in the VRC01+gp120 core (PDBID: 3NGB) in that it forms H-bonds with N462 and T465, as opposed to H-bonds with N465 and T467. The conformation of the V5 loop is also significantly different in the two

structures.

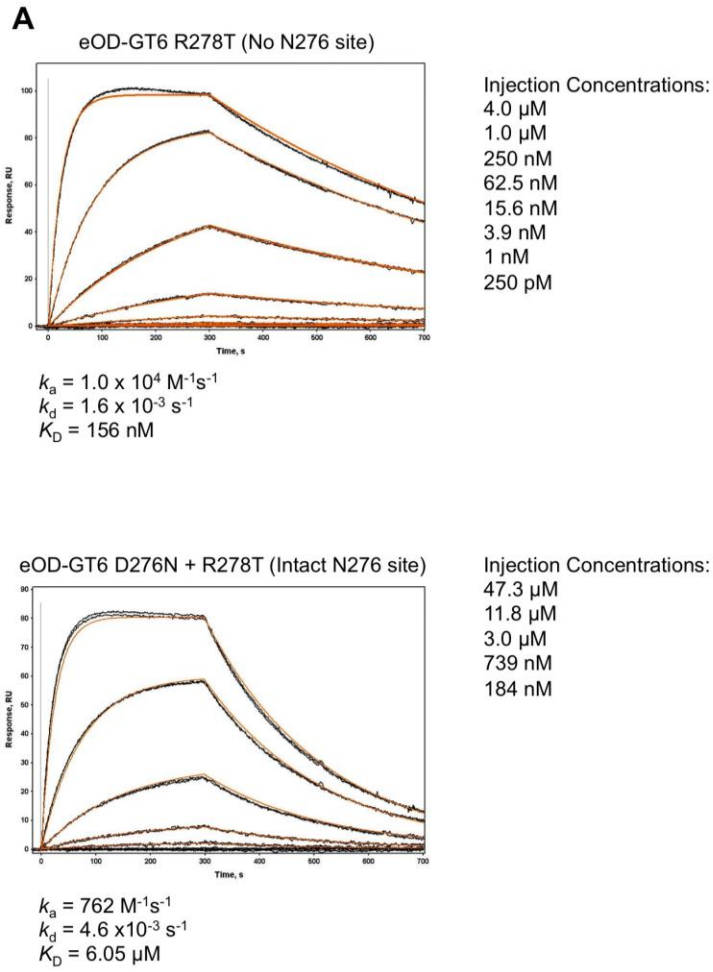
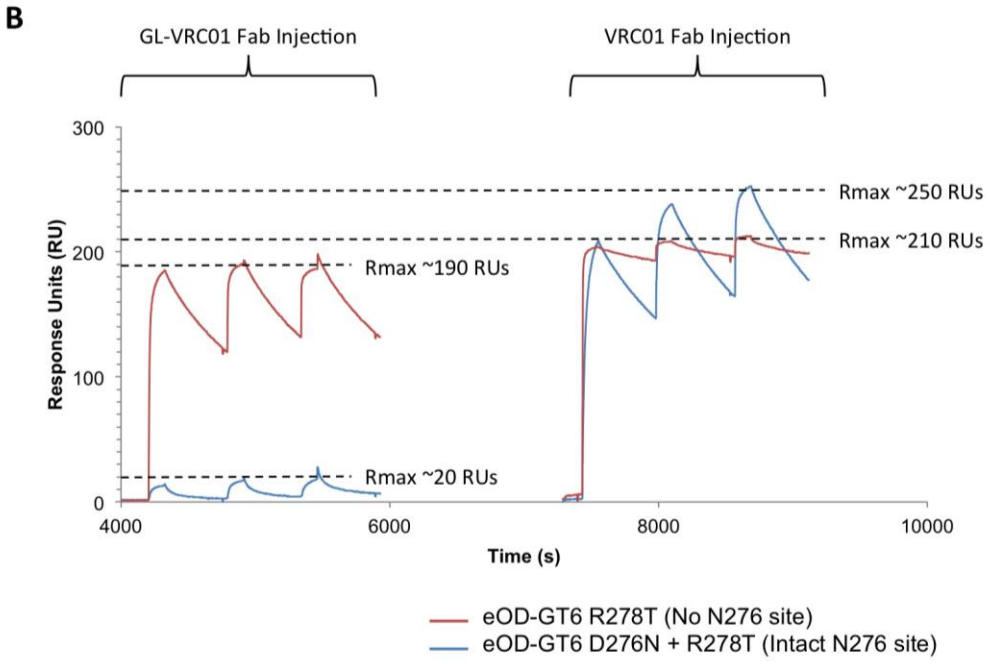
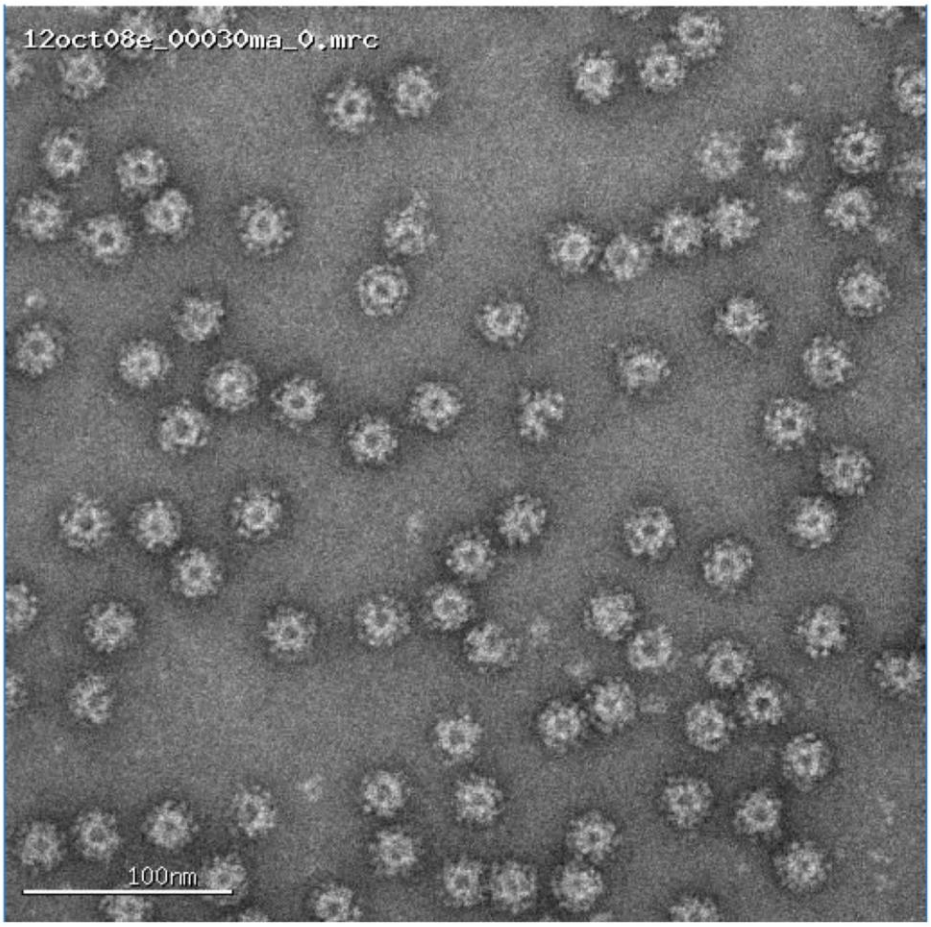


Fig. S13. Surface plasmon resonance experiments showed that only ~10% of the eOD-GT6 D276N/R278T double mutant that bound mature VRC01 was able to bind GL-VRC01, suggesting a model accounting for glycan under-utilization in which ~10% of the eOD-GT6 double mutant lacked the 276 glycan and had high affinity for GL-VRC01 whereas ~90% utilized the glycan at 276 and had no detectable affinity. (A) SPR traces of GL VRC01 immobilized while flowing eOD-GT6 with and without the N276 glycan. eOD preps containing the N276 glycan binding site have a significantly reduced on-rate, suggesting that the signal may be due to a small fraction of the sample underutilizing the N276 glycosylation position. (B) eOD-GT6 with and without the glycan immobilized while GL VRC01 and VRC01 Fab are flown as analyte. The significantly different Rmax for the germline antibody for the two constructs suggest that the germline is only capable of binding to a small fraction of eOD, presumably that which lacks the N276 glycan.

Fig. S13. Cont.


B
GL-VRC01 Fab Injection
VRC01 Fab Injection
300
200
100
0
Response Units (RU)
Rmax ~250 RUs
Rmax ~210 RUs
Rmax ~190 RUs
Rmax ~20 RUs
4000
6000
8000
10000
Time (s)
eOD-GT6 R278T (No N276 site)
eOD-GT6 D276N + R278T (Intact N276 site)

A



B

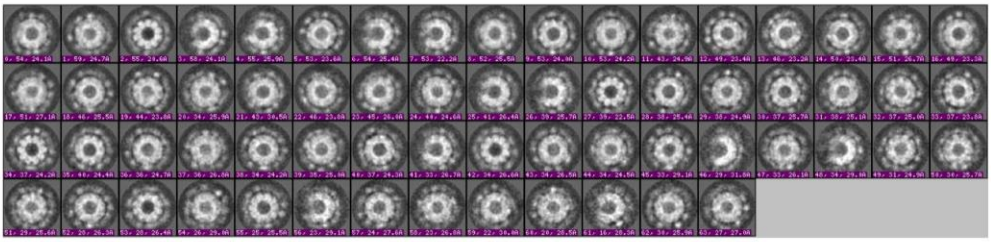


Fig. S14. Negative stain electron microscopy of eOD-GT6 60mer nanoparticles (A) Raw image and (B) 2D-class averages.

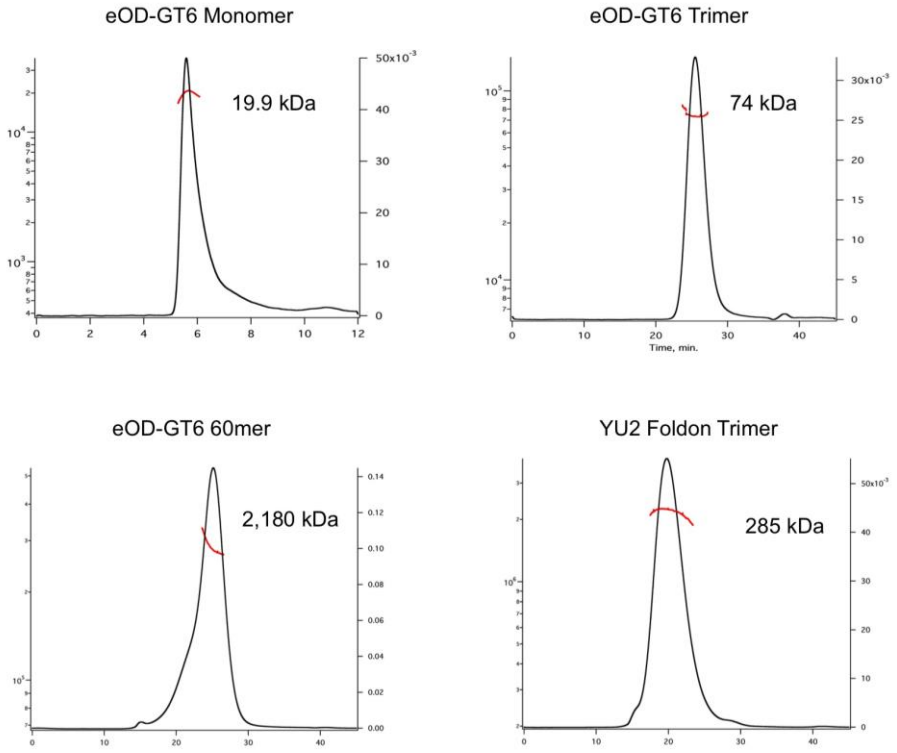


Fig. S15. Purification and characterization of multimeric eOD-GT6 and recombinant soluble YU2 gp140 trimer by size exclusion chromatography coupled in-line with multi-angle light scattering detectors (SEC-MALS). Traces show UV absorbance during the elution and lines under the peak the calculated molar mass of the eluting species. (A) eOD-GT6.0 monomer. Calculated MW: 19.9 kDa, measured MW: 19.9 kDa. (B) eOD-GT6 1gcm trimer. Calculated MW: 71.0, measured MW: 74.0 kDa. (C) eOD-GT6 60mer particle Calculated MW: 2,190 kDa, measured MW: 2,180 kDa.(D) YU2 gp140 trimer. Calculated MW: 237 kDa,

measured MW: 285 kDa.

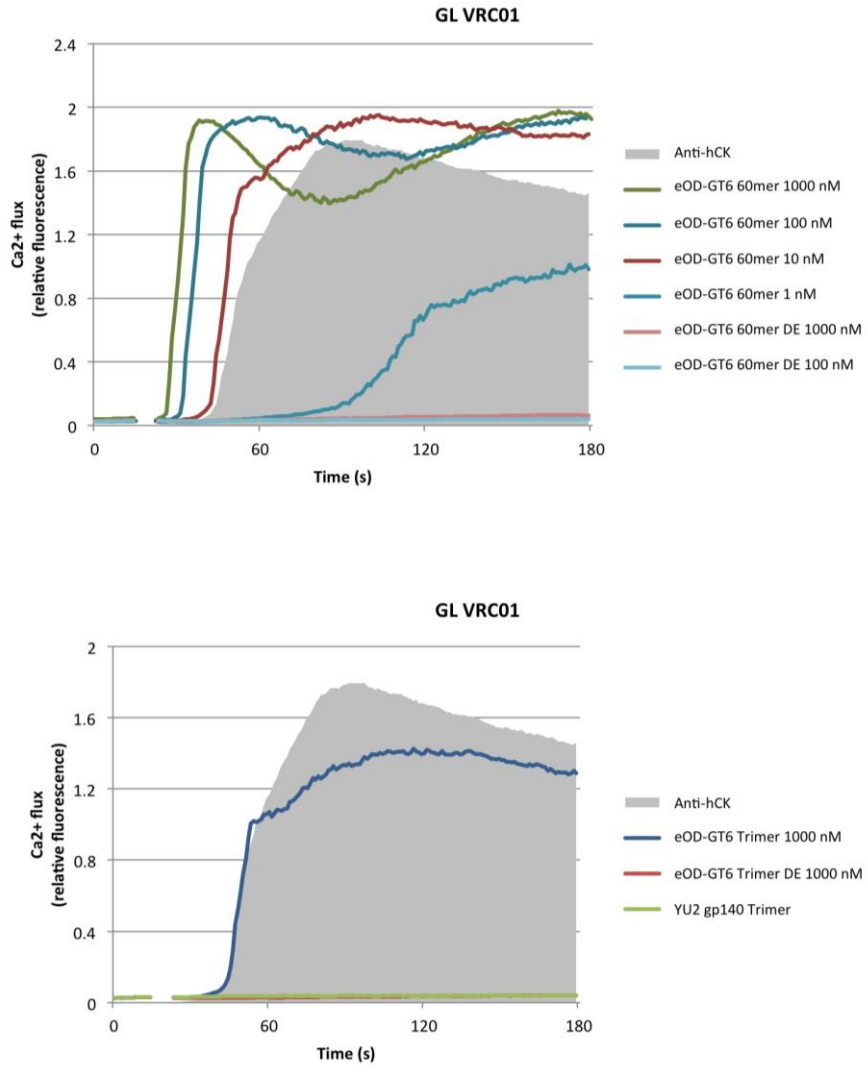


Fig. S16. Ca^{2+} flux activation of B cell lines expressing GL and Mature VRC01 (IgM), 12A12 (IgM and IgG), and NIH45/46 (IgG) with different constructs at varying concentrations. DE = dead epitope, in which a D368R mutation was made in gp120. Absence of activation for this negative control indicates specific activation for the

WT construct.

Fig. S16. Cont.

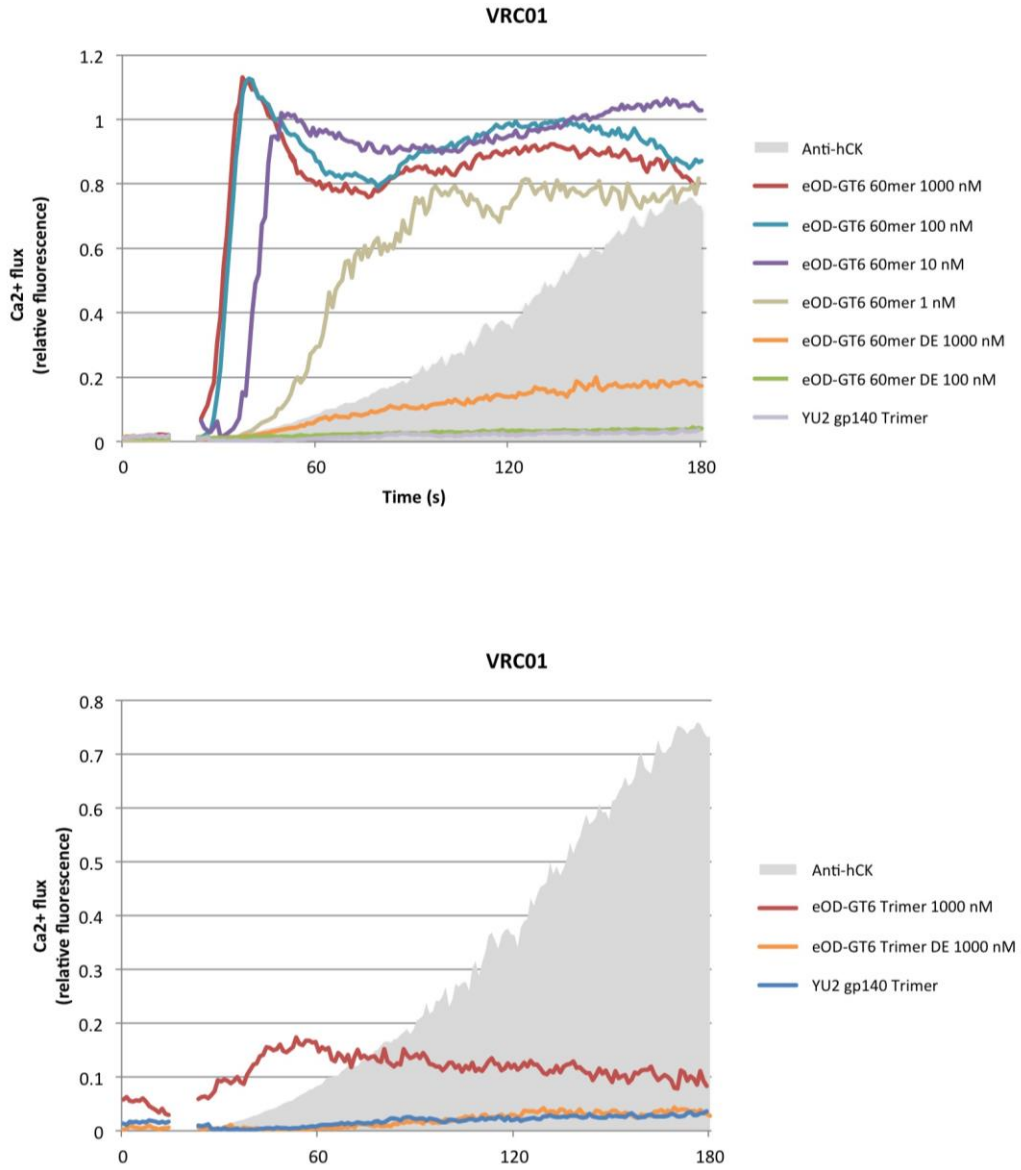

VRC01
Ca2+ flux
(relative fluorescence)
Time (s)
Anti-hCK
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
eOD-GT6 60mer DE 1000 nM
eOD-GT6 60mer DE 100 nM
YU2 gp140 Trimer
VRC01
Ca2+ flux
(relative fluorescence)
Anti-hCK
eOD-GT6 Trimer 1000 nM
eOD-GT6 Trimer DE 1000 nM
YU2 gp140 Trimer

Fig. S16. Cont.

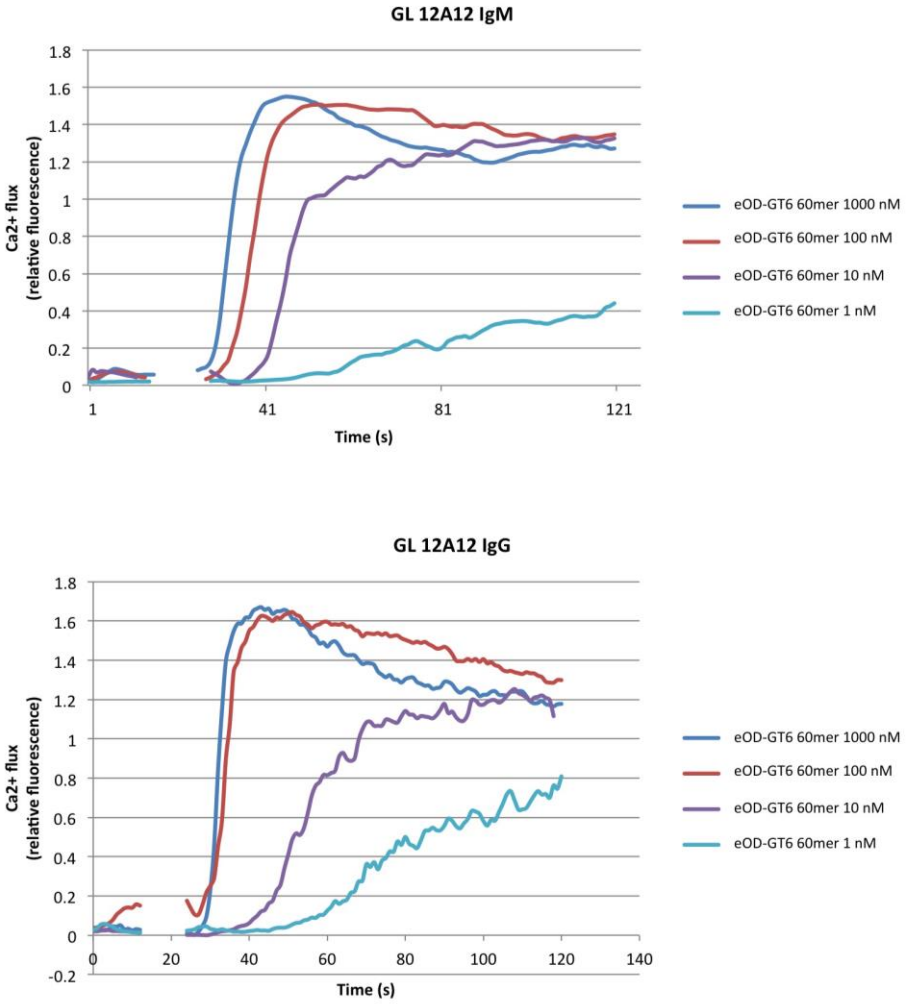
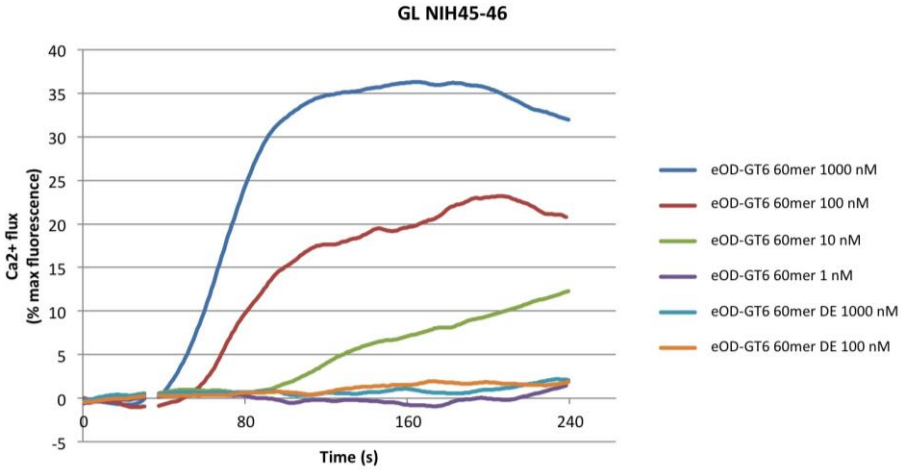

GL 12A12 IgM
Ca2+ flux (relative fluorescence)
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
GL 12A12 IgG
Ca2+ flux (relative fluorescence)
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM

Fig. S16. Cont.


GL NIH45-46
Ca2+ flux
(% max fluorescence)
40
35
30
25
20
15
10
5
0
-5
0
80
160
240
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
eOD-GT6 60mer DE 1000 nM
eOD-GT6 60mer DE 100 nM

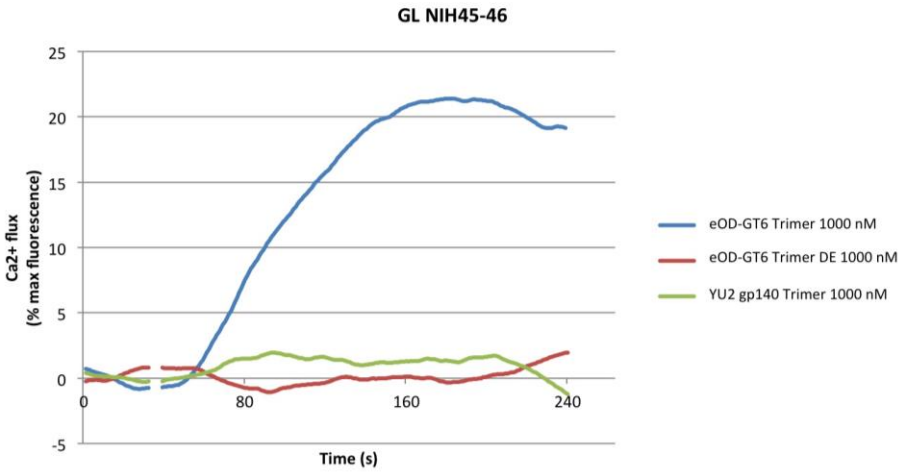
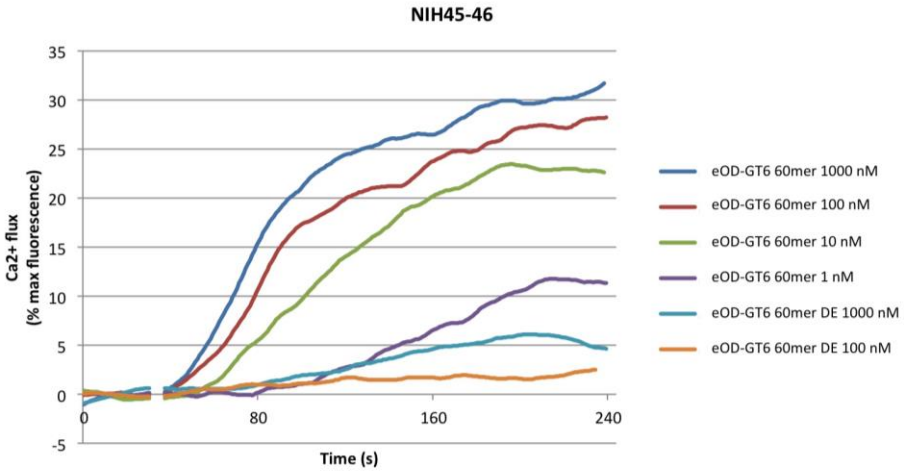
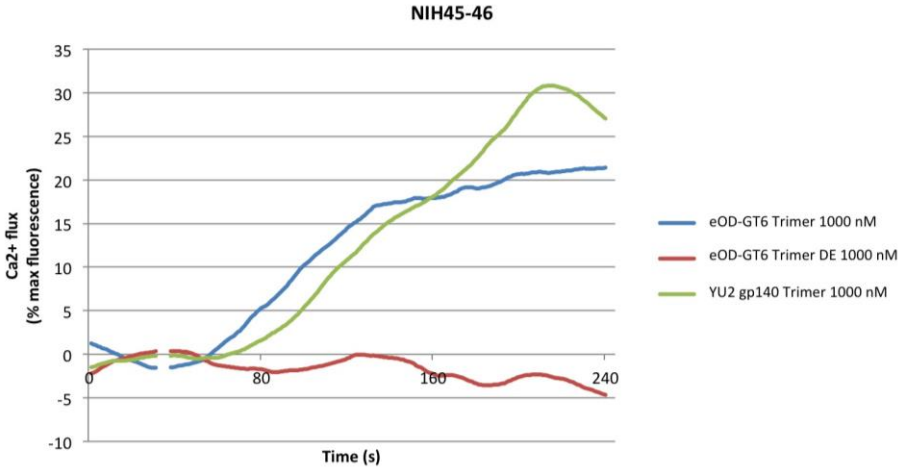

GL NIH45-46
Ca2+ flux
(% max fluorescence)
25
20
15
10
5
0
-5
0
80
160
240
Time (s)
eOD-GT6 Trimer 1000 nM
eOD-GT6 Trimer DE 1000 nM
YU2 gp140 Trimer 1000 nM

Fig. S16. Cont.


NIH45-46
Ca2+ flux
(% max fluorescence)
Time (s)
eOD-GT6 60mer 1000 nM
eOD-GT6 60mer 100 nM
eOD-GT6 60mer 10 nM
eOD-GT6 60mer 1 nM
eOD-GT6 60mer DE 1000 nM
eOD-GT6 60mer DE 100 nM


NIH45-46
Ca2+ flux
(% max fluorescence)
Time (s)
eOD-GT6 Trimer 1000 nM
eOD-GT6 Trimer DE 1000 nM
YU2 gp140 Trimer 1000 nM

A

```
Gene Usage          ---------------FR1-------------_CDR1_------FR2-----_____CDR2_____
IGHV1-02*02         QVQLV-QSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQKFQG
Rabbit VH1-a1       .C.S.EE..GRLVT..TPLTLT.TV..FSLSS.A.S.......K....I.I.S-S..S.Y..SWAK.
Critical Residues                                                 *  *        *

Gene Usage          ---------------FR3--------------____CDR3____----FR4----
IGHV1-02*02         RVTMTRDTSISTAYMELSRLRSDDTAVYYCAR
Rabbit VH1-a1       .F.ISKTST--.VDLKITSPTTE...T.F...GKNCDYNWDFEHWGRGTPVIVSS
Critical Residues       *
```

B

```
Gene Usage          --------------FR1-------------_CDR1_------FR2-----_____CDR2_____
IGHV1-02*02         QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQKFQG
IGHV1S7*01          -Q..KE..GGLV...G.L.LC.....F..SS...C.......K....I.C.YAG..S.H..SWVN.
IGHV1S47*01         .E...E..GGLVQ.EG.LTLT.....FD.SSNA.C.......K.P..IAC.YNGD.S.Y..SWVN.
Critical Residues                                                 *  *        *

Gene Usage          ---------------FR3--------------____CDR3____----FR4----
IGHV1-02*02         RVTMTRDTSISTAYMELSRLRSDDTAVYYCAR
IGHV1S7*01          .F.LS..NAQ..VCLQ.NS.TAA...T.F...GKNCDYNWDFEHWGRGTPVIVSS
IGHV1S47*01         .F.IS.S..LN.VTLQMTS.TAV...T.F...GKNCDYNWDFEHWGRGTPVIVSS
Critical Residues       *
```

Fig. S17. (A) Sequence alignment of the human VH1-2*02 gene with the VH-1 gene reported to be utilized by the majority of rabbit Abs (*47*). This gene only contains 1 of 4 critical residues and is 3 residues shorter, suggesting that the conformation of the backbone might also be different. (**B)** Alignment of the human VH1-2*02 gene with the closest rabbit VH gene. It is unclear if rabbits produce Abs possessing this gene with any useful frequency. Positions of human GL encoded residues important for binding to gp120 highlighted in red.

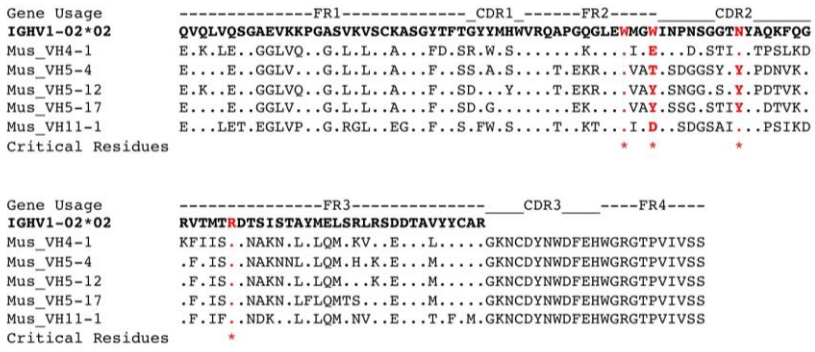


Fig. S18. Sequence alignment of the closest 5 mouse VH genes containing Arg^{H71} to VH1-2*02. Residues colored red represent major incompatibilities with gp120 binding. Positions of human GL encoded residues important for binding to gp120 highlighted in red.

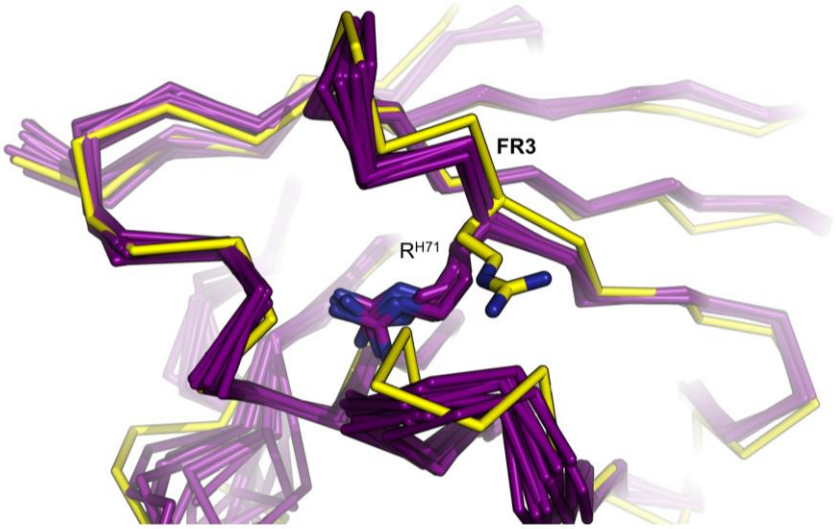


Fig. S19. Structural superposition of 10 Abs from the PDB that arose from the mouse VH5-17*01 gene (PDBIDs 2aab, 2zuq, 3hr5, 2ffd, 1seq, 2qhr, 3ra7, 2vh5, 2xkn, 3s88 and colored purple) and the gp120-bound VRC01 crystal structure (PDBID: 3GNB and colored yellow). There was no density for Arg^{H71} in PDBID: 2AAB, but the remaining 9 structures all have the Arg^{H71} side chain buried into the interior of the Ab structure, while this residue in the VRC01 structure is exposed. In the conformation found in the mouse Abs, Arg^{H71} would not be accessible to form the critical salt bridge with D368 on gp120. Additionally, the conformation of the backbone is significantly modified, further reducing the structural similarities with human VH1-2 Abs.

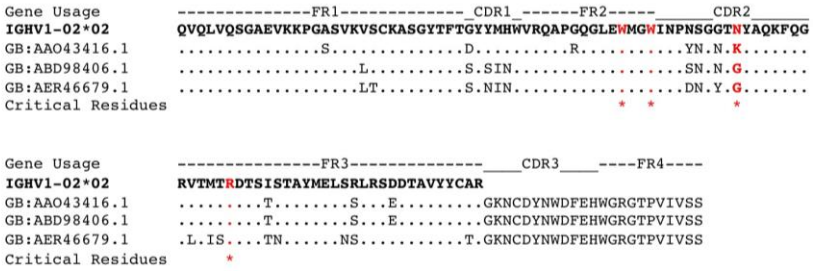


Fig. S20. Sequence alignment of the closest 3 macaque VH genes containing Arg^{H71} to VH1-2*02. While none of the VH genes identified contain all 4 critical contacts, our model suggests that the mutations at position Asn^{H57} do not result in obvious clashes with GP120 and as such, might allow binding to eOD-GT6. Positions of human GL encoded

residues important for binding to gp120 highlighted in red.

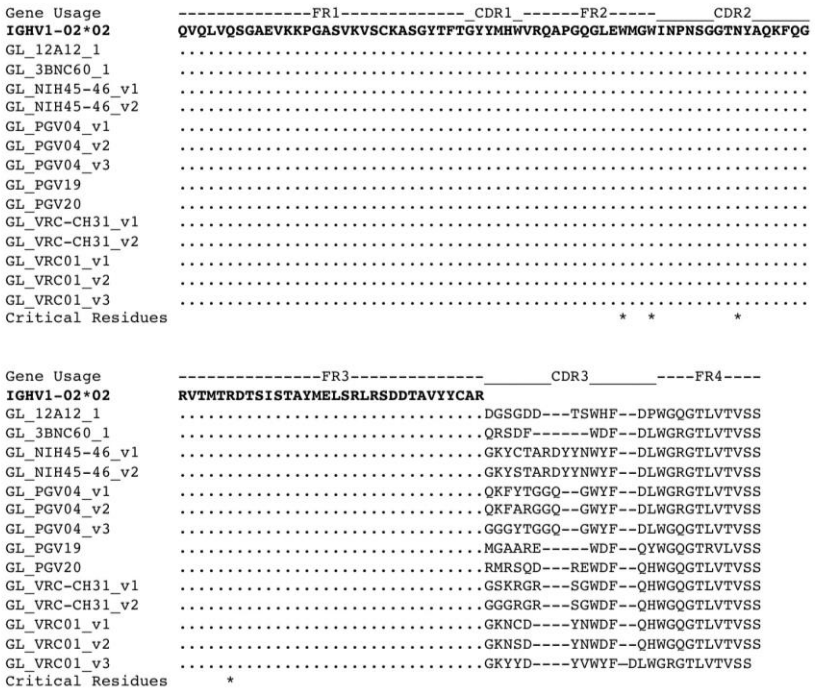


Fig. S21. Sequence alignment of the GL calculated heavy chain of the VRC01-class Abs used in this study.

```
IGKV1-33*01      DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPSRFSGSGSG
GL_12A12         ............................................................
GL_3BNC60        ............................................................
GL_VRC-CH31      ............................................................

IGKV3-11*01      EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSG
GL_NIH45-46_v1   ............................................................
GL_VRC01_v1      ............................................................

IGKV3D-15*02     EIVMMQSPATLSVSPGERATLSCRASQSVSSNLAWYQQKPGQAPRLLIYGASTRATGIPARFSGSGSG
GL_NIH45-46_v2   ....T.......................................................
GL_VRC01_v2      ............................................................

IGKV3-20*01      EIVLTQSPGTLSLSPGERATLSCRASQSVSSSYLAWYQQKPGQAPRLLIYGASSRATGIPDRFSGSGS
GL_PGV04         ............................................................

IGLV2-14*01      QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYEVSNRPSGVSNRFSGSK
GL_PGV19         E...........................................................
GL_PGV20         E...........................................................


IGKV1-33*01      TDFTFTISSLQPEDIATYYCQQYDNLP
GL_12A12         ...................AVLEFFGPGTKVEIK
GL_3BNC60        .....................EFIGPGTKVDIK
GL_VRC-CH31      .....................ETFGQGTKLEIK

IGKV3-11*01      TDFTLTISSLEPEDFAVYYCQQRSNWP
GL_NIH45-46_v1   .....................YEFFGQGTKLEIK
GL_VRC01_v1      .....................YEFFGQGTKLEIK

IGKV3D-15*02     TEFTLTISSLQSEDFAVYYCQQYNN*P
GL_NIH45-46_v2   .....................EFFGQGTKLEIK
GL_VRC01_v2      .....................EFFGQGTKLEIK

IGKV3-20*01      GTDFTLTISRLEPEDFAVYYCQQYGSSP
GL_PGV04         .....................LEFFGQGTRLEIK

IGLV2-14*01      SGNTASLTISGLQAEDEADYYCSSYTSSSTL
GL_PGV19         ......................EFFGGGTKVFVLG
GL_PGV20         ......................EFFGGGTKVFVLG
```

Fig. S22. Sequence alignment of the GL calculated light chain of the VRC01-class Abs used in this study.

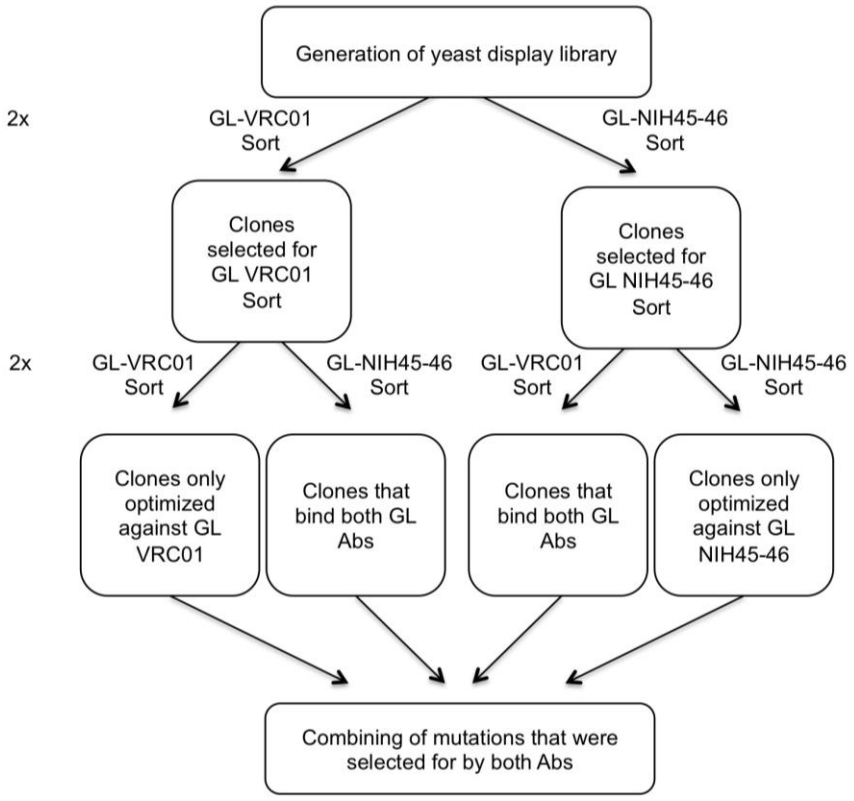

Generation of yeast display library
2x
GL-VRC01 Sort
GL-NIH45-46 Sort
Clones selected for GL VRC01 Sort
Clones selected for GL NIH45-46 Sort
2x
GL-VRC01 Sort
GL-NIH45-46 Sort
GL-VRC01 Sort
GL-NIH45-46 Sort
Clones only optimized against GL VRC01
Clones that bind both GL Abs
Clones that bind both GL Abs
Clones only optimized against GL NIH45-46
Combining of mutations that were selected for by both Abs

Fig. S23. Flow chart representing the multistate sorting protocol.

MQIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAIDAIVRHGGREEDITLVRVPGSWEIPVAAGE
LARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGLADLSLELRKPITFGVITADTLEQAIERAGT
KHGNKGWEAALSAIEMANLFKSLRGGSGGSGGSGGSGGGDTITLPCRPAPPPHCSSNITGLILTR
DGGVSNDETEIFRPSGGDMRDIARCQIAGTVVSTQLFLNGSLAEEEVVIRSVDFRDNAKSICVQL
NTSVEINCTGAGHCNISRAKWNNTLKQIASKLREQFGNRTIIFKQSSGGDPEFVTHSFNCGGEFF
YCDSTQLFNSTWFNST**

Fig. S24. Sequence of eOD-GT6 60mers. Particle base in red, GS linker in black, eOD-GT6 in blue.

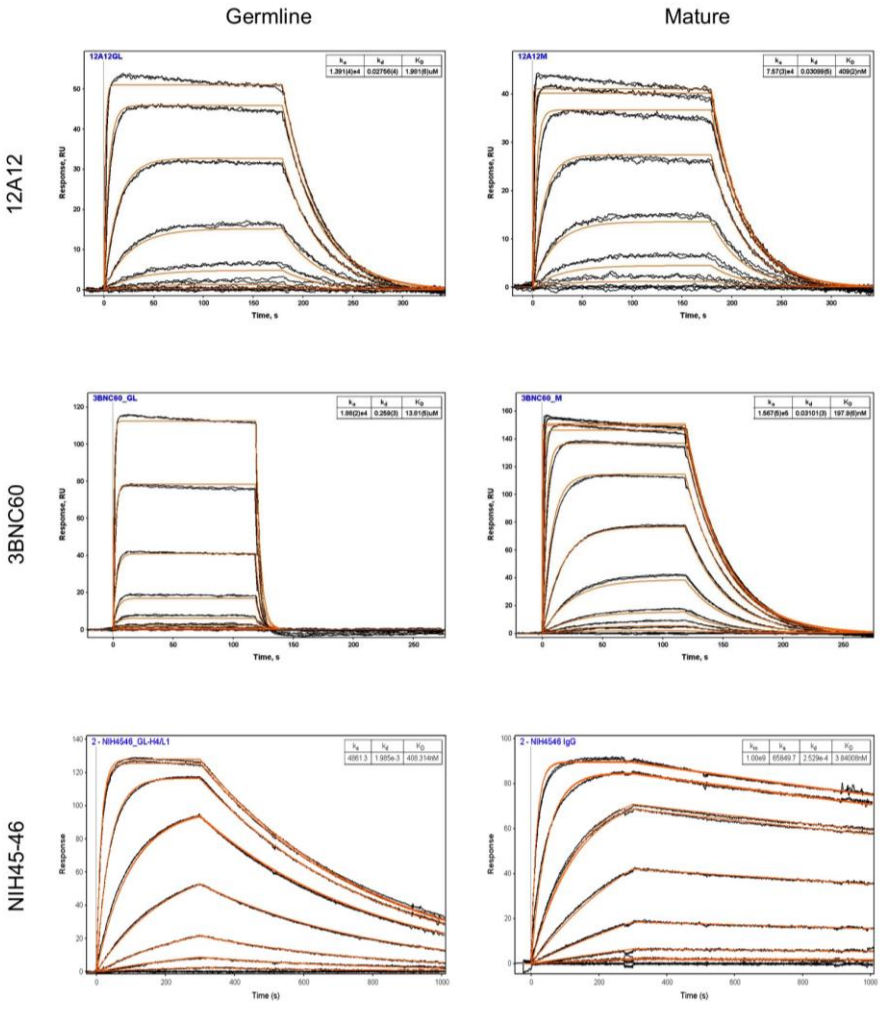

Germline
Mature
12A12
3BNC60
NIH45-46
12A12GL
12A12M
3BNC60_GL
3BNC60_M
2 - NIH4546_GL-H4/L1
2 - NIH4546 IgG
Response, RU
Response
Time, s
Time (s)

Fig. S25. Raw SPR data for eOD-GT6 against various GL and mature VRC01-class Abs binding to eOD-GT6.

Fig. S25. Cont.

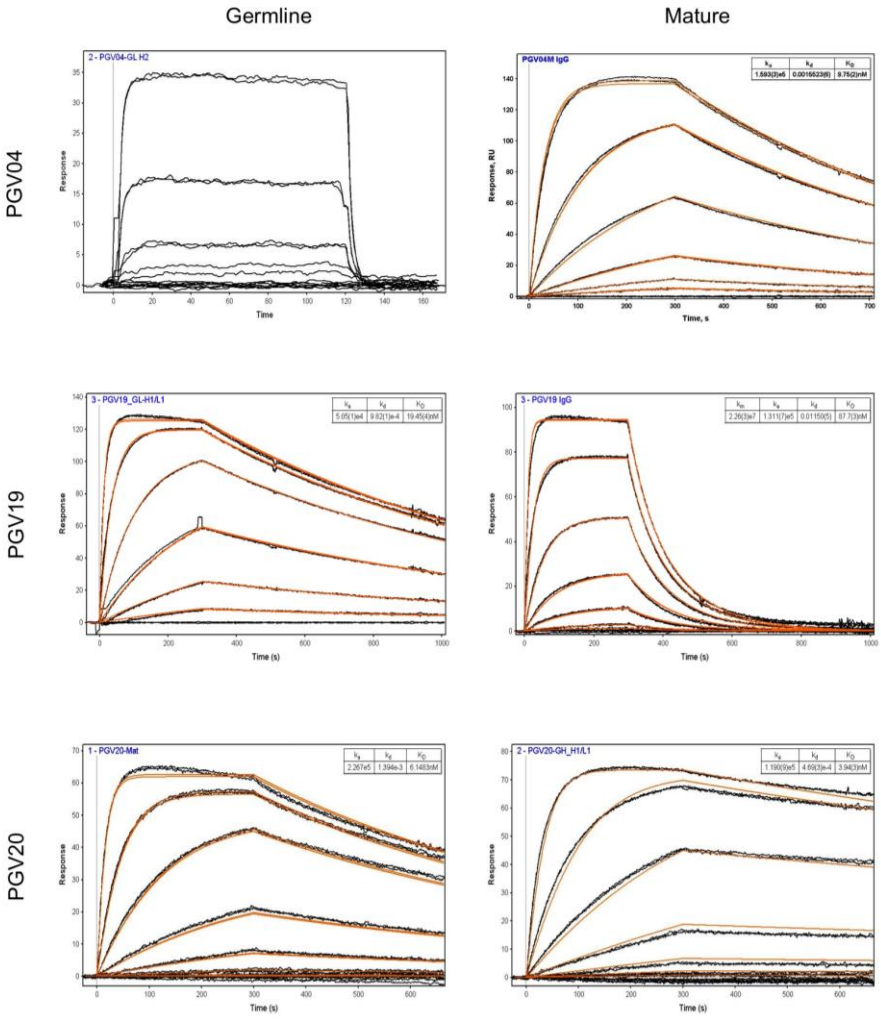
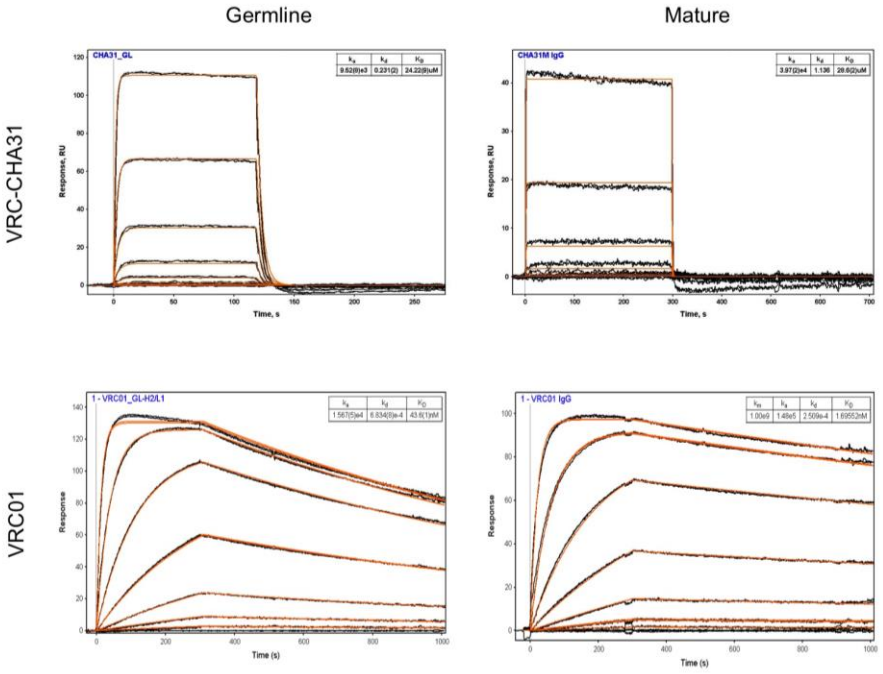

Germline
Mature
PGV04
PGV19
PGV20
2 - PGV04-GL H2
PGV04M IgG
3 - PGV19_GL-H1/L1
3 - PGV19 IgG
1 - PGV20-Mat
2 - PGV20-GH_H1/L1
Response
Response, RU
Time
Time, s
Time (s)

Fig. S25. Cont


Germline
Mature
VRC-CHA31
VRC01
CHA31_GL
CHA31M IgG
1 - VRC01_GL-H2/L1
1 - VRC01 IgG
Response, RU
Response
Time, s
Time (s)

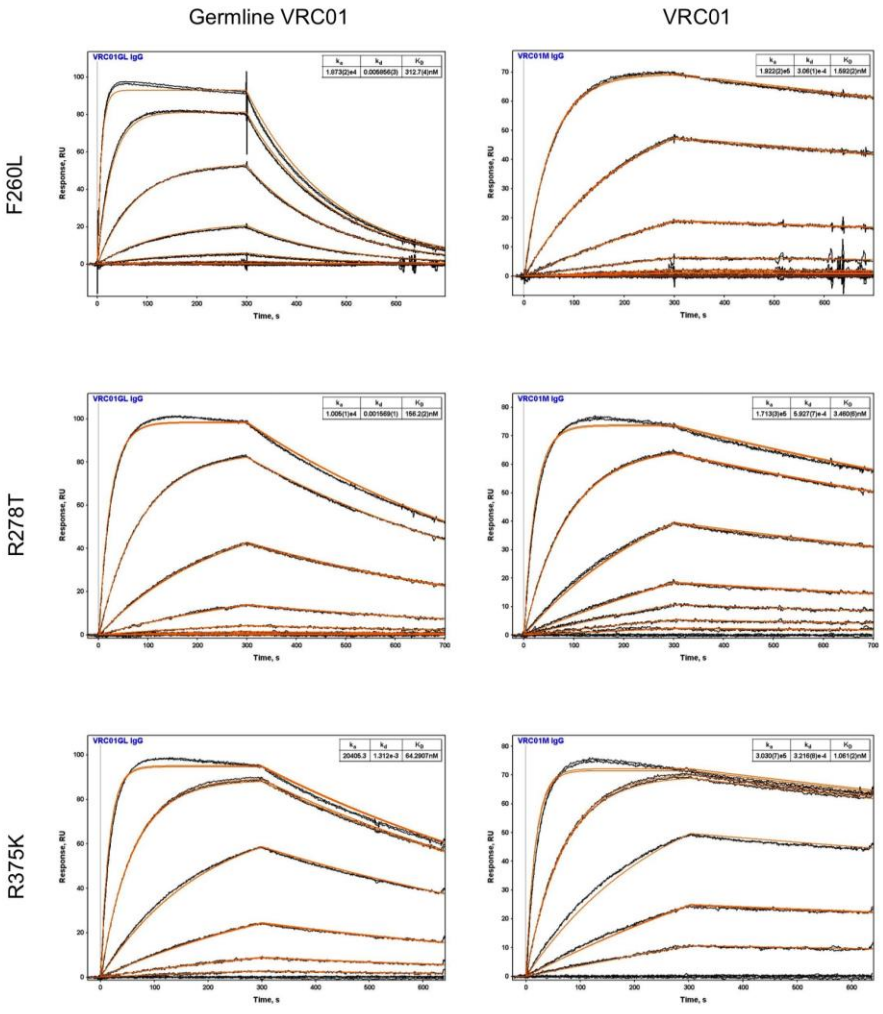
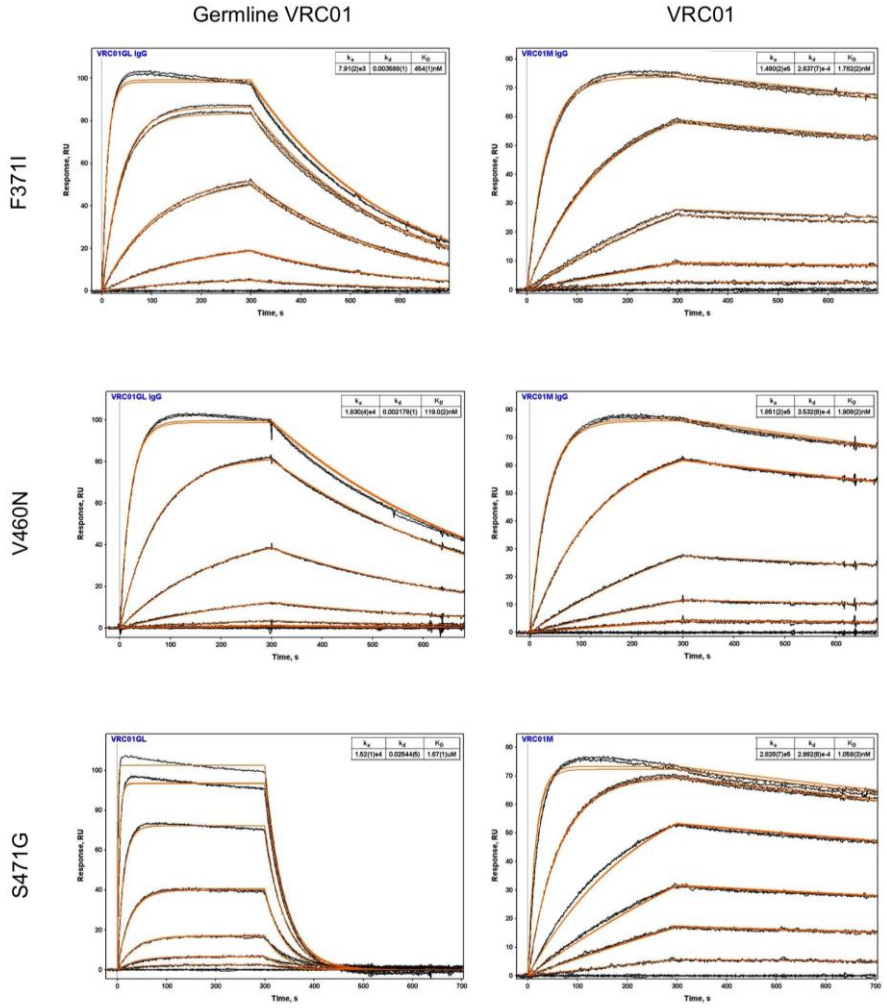

Germline VRC01
VRC01
F260L
R278T
R375K
VRC01GL IgG
VRC01M IgG
Response, RU
Time, s

Fig. S26. Raw SPR data for GL and mature VRC01 binding to eOD-GT6 reversion mutants.

Fig. S26. Cont.


Germline VRC01
VRC01
F371I
V460N
S471G
VRC01GL IgG
VRC01M IgG
VRC01GL
VRC01M
Response, RU
Time, s

Supplementary Tables:

Table S1. Germline Binding

Clade	Virus Strain	VRC01	3BNC60	PGV19	GL VRC01	GL 3BNC60	GL PGV19
A	BL274.W6M.ENV.A3	1.31	0.23	0.78	>100	>100	>100
	Q23	2.17	0.38	0.85	>100	>100	>100
	Q769env.h5	1.37	0.29	0.27	>100	>100	>100
	92RW020	2.65	0.49	20.81	>100	>100	>100
	94UG103	3.46	1.82	1.30	>100	>100	>100
	QB726.70M.ENV.C4	0.55	0.40	>100	>100	>100	>100
	QF495.23M.ENV.A1	1.28	0.13	>100	>100	>100	>100
	QF495.23M.ENV.A3	1.11	0.11	>100	>100	>100	>100
	QF495.23M.ENV.B2	1.12	0.11	>100	>100	>100	>100
	QF495.23M.ENV.D1	0.77	0.06	>100	>100	>100	0.00
	QH209.14M.ENV.A2	0.81	1.83	>100	>100	>100	>100
	QH359.21M.ENV.D1	3.40	0.85	>100	>100	>100	>100
	QG984.21M.ENV.A3	0.75	>100	0.12	>100	>100	>100
	QH343.21M.ENV.A10	>100	>100	>100	>100	>100	>100
	BG505.W6M.ENV.C2	>100	>100	>100	>100	>100	>100
B	YU-2	1.02	0.09	0.44	>100	>100	>100
	6535.5	7.17	0.98	1.90	>100	>100	>100
	CAAN5342A2	5.24	0.94	3.07	>100	>100	>100
	TRO.11	3.39	0.56	7.76	>100	>100	>100
	TRJO4551.58	5.12	3.68	5.32	>100	>100	>100
	SC422661.8	1.66	0.79	1.33	>100	>100	>100
	QH0692	2.85	0.54	3.54	>100	>100	>100
	SF162	5.77	6.84	2.19	>100	>100	>100
	ADA	2.41	0.80	1.38	>100	>100	>100
	JR-FL	1.45	0.38	1.71	>100	>100	>100
	JR-CSF	1.20	0.13	1.45	>100	>100	>100
	PVO.4	3.03	0.31	>100	>100	>100	>100
	RHPA4259.7	0.94	0.16	>100	>100	>100	>100
	92BR020	2.63	0.97	>100	>100	>100	>100
	AC10	11.39	>100	>100	>100	>100	>100
	WITO4160.33	>100	>100	>100	>100	>100	>100
B (T/F)	62357_14_D3_4589	2.93	0.16	1.18	>100	>100	>100
BC	CNE53	1.14	0.40	>100	>100	>100	>100
	CNE20	>100	>100	>100	>100	>100	>100
C	ZM53M.PB12	1.42	0.22	0.28	>100	>100	>100
	IAVI C22	7.88	1.18	3.44	>100	>100	>100
	ZM109	3.21	0.50	>100	>100	>100	>100
	ZM197	1.52	4.40	>100	>100	>100	>100
	ZM214	2.15	0.22	>100	>100	>100	>100
	QB099.391M.ENV.B1	1.30	>100	>100	>100	>100	>100
	DU172	>100	>100	>100	>100	>100	>100
	DU422	>100	>100	>100	>100	>100	>100
	ZM233	>100	>100	>100	>100	>100	>100
C (T/F)	BF1266	0.62	0.12	0.07	>100	>100	>100
D	QA013.70I.ENV.H1	5.40	22.61	3.98	>100	>100	>100
	QA013.70I.ENV.M12	0.88	0.33	>100	>100	>100	>100
G	X1193_C1	1.99	0.39	0.81	>100	>100	>100
	X1254_C3	1.05	0.18	2.94	>100	>100	>100
	X2131_C1_B5	6.02	3.02	4.67	>100	>100	>100
	P1981_C5_3	2.90	0.97	2.17	>100	>100	>100
	X2088_C9	>100	>100	>100	>100	>100	>100
A/D	MF535.W0M.ENV.D11	1.13	1.90	7.85	>100	>100	>100
	QA790.204I.ENV.A4	0.42	0.08	0.06	>100	>100	>100
	QA790.204I.ENV.C1	0.62	0.21	0.10	>100	>100	>100
	QA790.204I.ENV.C8	0.91	0.47	0.16	>100	>100	>100
A2/D	QG393.60M.ENV.A1	1.45	0.49	0.86	>100	>100	>100
	QG393.60M.ENV.B7	1.02	2.40	0.83	>100	>100	>100
AE	CRF01_AE clone 269	3.88	1.28	>100	>100	>100	>100
AG	263-8	2.17	0.35	1.33	>100	>100	>100
	T250-4	>100	>100	>100	>100	>100	>100
	235-47	>100	>100	>100	>100	>100	>100
C/D	BK184.W6M.ENV.D2	4.16	1.46	>100	>100	>100	>100

ELISA

Table S2: Characteristics of germline-targeting eOD Generations

Characteristics of each generation of germline-targeting eOD. K_Ds were measured by SPR kinetic fits except where indicated by (*) in which case K_Ds were measured by fits to equilibrium association data. ND, not done.

		K_D (nM) measured by SPR for Antibody-antigen interaction						
Antigen	Mutations from eOD Base	Antibody						
		VRC01		NIH45-46			PGV19	
		GL	Mat	GL	GL_v2	Mat	GL	Mat
eOD Base	0	$>10^5$	5	$>10^5$	54,000*	14	$>10^5$	3,100
eOD-GT1	10	44,000*	1	$>10^5$	ND	3	7,800*	1,000*
eOD-GT2	14	15,000*	3	$>10^5$	ND	4	13,000*	6,600*
eOD-GT3	15	220	3	4,600	ND	4	390	1800
eOD-GT4	17	34	2	1,000	ND	3	27	1800
eOD-GT5	14	530	57	4,800	ND	110	300	19,000*
eOD-GT6	8	44	2	410	9	4	19	88

Table S3: K_Ds for eOD-GT6 interactions with GL VRC01-class Abs with alternate human VH genes. Values in nM.

	eOD-GT6
VH1-2*01	$>10^5$
VH1-2*02	44
VH1-2*03	39
VH1-2*04	39
VH1-3*01	$>10^5$
VH1-8*01	32,000
VH1-46*01	$>10^5$

Table S4. X-Ray Crystallography Table and Refinement statistics

Crystal	GL-VRC01 Fab	eOD-GT6	eOD-GT6 + GL-VRC01 Fab
Data Collection	APS 23-ID	APS 23-ID	SSRL 12-2
Wavelength, Å	1.03320	1.03320	0.9795
Space group	H3	$P2_1$	C2
Unit cell a, b, c (Å)	172.44, 172.44, 92.43	44.96, 217.73, 44.99	179.26, 63.3, 61.3
α, β, γ (°)	90, 90, 120	90, 119.97, 90	90, 90.43, 90
Molecules per ASU	2	4	1 Fab + 1 eOD
Resolution (Å) *	40.0 – 2.1 (2.2 – 2.1)	40.0 – 2.5 (2.6 – 2.5)	40.0 – 2.4 (2.5 – 2.4)
Completeness*	100 (100)	99.9 (99.9)	93.6 (96.8)
Redundancy*	5.8 (5.8)	3.9 (3.9)	2.7 (3.0)
No. total reflections	347,661	99,546	74,140
No. unique reflections	59,809	25,751	25,368
I/σ*	13.6 (2.8)	8.1 (2.9)	9.4 (2.1)
R_{sym} †,*	8.6 (52.1)	14.2 (43.0)	8.5 (43.2)
Refinement statistics			
Resolution (Å)	40.0 – 2.1	40.0 – 2.5	40.0 – 2.4
No. reflections total/R_{free}	58,400/2,914	24,371/1,336	25,310/1,253
R_{cryst}/R_{free} ‡, §	17.3/21.2	20.9/25.9	20.0/26.3
RMSD bond length (Å)	0.007	0.006	0.005
RMSD bond angles (°)	1.13	1.16	1.00
Protein atoms/ solvent atoms	6,518/387	5,183/45	4,588/173
Wilson B-value ($Å^2$)	30.0	44.9	57.3
Overall average B-value ($Å^2$)	38.1	25.9	47.6
Average B-value protein ($Å^2$)	35.5	31.6	47.8
Average B-value solvent ($Å^2$)	59.5	20.4	42.2
Ramachandran Preferred %	98.1	90.3	93.8
Allowed %	1.7	8.3	5.4
PDB ID	XXX	XXX	XXX

Comment [JJ1]: JP

* Values in parentheses are for the highest resolution shell.

† $R_{sym}= \Sigma|I-\langle I\rangle|/\Sigma\langle I\rangle$, where I is the observed intensity, and $\langle I\rangle$ is the average intensity of multiple observations of symmetry related reflections.

Table S5: Analysis of the buried surface area and H-bonds on VRC01/GL-VRC01 in the complexes of VRC01+gp120core and GL-VRC01+eOD-GT6.

Interfaces were calculated using PDBePISA (*29*).

	gp120 - VRC01		eOD-GT – GL-VRC01	
	Residue #	BSA (Å²)	Residue #	BSA (Å²)
FR1	E1	0	E1	41.6
	I2	0	I2	5.9
	V3	35.6	V3	0
CDR L1	Q27	23.1	Q27	26.6
	Y28 (H)	30.7	S28 (H)	13.4
	G29	0	V29	18.2
	A32	0	Y32	8.1
CDR L3	Y91 (H)	68.8	Y91 (H)	71.5
	E96 (H)	57.7	E96 (H)	57.8
	F97	55.7	F97	23.2
FR1	I30	15.6	T30	7.2
CDR H1	D31	0	G31	7.3
	T33	2.3	Y33	46.8
FR2	W47	0	W47	24.8
CDR H2	W50 (H)	38.1	W50 (H)	44.1
	K52 (H)	41.0	N52 (H)	28.3
	R53	57.5	N53	36.0
	G54 (H)	50.8	S54 (H)	74.0
	G55	16.5	G55	19.7
	A56	35.0	G56	30.0
	V57	35.9	T57 (H)	47.0
	N58 (H)	64.7	N58 (H)	66.3
	Y59 (H)	37.3	Y59	18.6
	A60	7.5	A60	14.8
	R61 (H)	160.8	Q61 (H)	146.7
	P62	20.5	K62	11.2
	Q64 (H)	43.1	Q64 (H)	52.2
FR3	R71 (H)	27.5	R71 (H)	24.6
	V73	26.1	T73	0
	Y74	52.7	S74	0
CDR H3	D99 (H)	45.6	D99 (H)	42.4
	Y100	19.3	Y100	16.8
	N100A	14.0	N100A	11.7
	W100B (H)	42.3	W100B (H)	38.8
Total BSA (Å²)		1125.7		1075.6

Table S6: Analysis of the buried surface area and H-bonds on gp120 core and eOD-GT6. Blue shaded areas show residues present in eOD-GT6.

	gp120core + VRC01		eOD-GT6 – GL-VRC01	
	Residue #	BSA (Å²)	Residue #	BSA (Å²)
	K97 (HS)	26.5		
β3	G124	38.8		
	G198	13.2		
Loop D	E275		V78	5.9
	N276	22.6	D79 (H)	31.4
	T278 (H)	123.0	R81 (H)	130.0
	N279 (H)	55.6	D82 (H)	54.1
	N280 (H)	68.4	N83 (H)	70.9
	A281 (H)	70.1	A84	66.6
	K282 (H)	30.9	K85 (H)	39.4
	T283	12.1	S86	0
	S365 (H)	62.2	S138 (H)	66.4
β15	G366	22.0	G139	17.7
	G367	24.0	G140	27.6
	D368 (H)	49.1	D141 (H)	54.1
α3	I371	44.2	F144	22.5
	W427	7.7		
β21	T430	57.8		
β23	T455	31.6	T25	24.5
	R456 (H)	5.7	R26 (H)	5.8
	D457 (H)	47.2	D27 (H)	47.4
	G458 (H)	49.8	G28 (H)	43.2
	G459 (H)	68.7	G29 (H)	61.4
	A460	37.8	V30	64.3
V5	N461	68.4	S31	17.5
	N462		N32 (H)	11.4
	T463	13.6	D33	0
	N465 (H)	9.7	T35 (H)	0.2
β24	E466	5.5	E36	2.9
	T467 (H)	15.4	I37	14.6
	R469 (H)	22.8	R39 (H)	24.2
	G472	8.3	G42 (H)	61.4
	G473	28.4	G43 (H)	22.1
	N474	16.8	D44 (H)	36.1
	K476	0.3	R46	47.4
α5	D477		D47	6.4
Total BSA (Å²)		1158.2		1102.0

Table S7: SPR data for all experiments.

ND, not determined. NB, no binding. SPR traces for eOD-GT6 bound to GL and Mat Abs as well as the 6 point mutations are shown in fig S24-25.

Construct	Antibody	*k*a	*k*d	K_D	Equlibrium
eOD-GT1	NIH45-46 GL	NB	NB	-	NB
eOD-GT1	NIH45-46 Mature	1.42E+05	4.24E-04	2.98E-09	ND
eOD-GT1	PGV19 GL	ND	ND	-	7.80E-06
eOD-GT1	PGV19 Mature	1.17E+05	1.26E-01	1.08E-06	1.04E-06
eOD-GT1	VRC01 GL	ND	ND	-	4.40E-05
eOD-GT1	VRC01 Mature	2.60E+05	3.67E-04	1.41E-09	ND
eOD-GT2	NIH45-46 GL	NB	NB	-	NB
eOD-GT2	NIH45-46 Mature	1.44E+05	5.34E-04	3.71E-09	ND
eOD-GT2	PGV19 GL	ND	ND	-	1.25E-05
eOD-GT2	PGV19 Mature	ND	ND	-	6.59E-06
eOD-GT2	VRC01 GL	ND	ND	-	1.50E-05
eOD-GT2	VRC01 Mature	2.21E+05	5.99E-04	2.71E-09	ND
eOD-GT3	NIH45-46 GL	5.33E+03	2.48E-02	4.65E-06	ND
eOD-GT3	NIH45-46 Mature	1.18E+05	4.38E-04	3.71E-09	ND
eOD-GT3	PGV19 GL	4.50E+04	1.73E-02	3.85E-07	ND
eOD-GT3	PGV19 Mature	7.00E+04	1.28E-01	1.83E-06	1.81E-06
eOD-GT3	VRC01 GL	1.30E+04	2.87E-03	2.21E-07	ND
eOD-GT3	VRC01 Mature	2.00E+05	5.41E-04	2.70E-09	ND

eOD-GT4	NIH45-46 GL	3.43E+03	3.42E-03	9.99E-07	ND
eOD-GT4	NIH45-46 Mature	1.86E+05	5.18E-04	2.79E-09	ND
eOD-GT4	PGV19 GL	5.02E+04	1.39E-03	2.77E-08	ND
eOD-GT4	PGV19 Mature	4.86E+04	8.90E-02	1.83E-06	1.79E-06
eOD-GT4	VRC01 GL	1.39E+04	4.73E-04	3.41E-08	ND
eOD-GT4	VRC01 Mature	2.59E+05	5.55E-04	2.15E-09	ND
eOD-GT5	NIH45-46 GL	5.81E+02	2.79E-03	4.81E-06	ND
eOD-GT5	NIH45-46 Mature	1.05E+04	1.16E-03	1.11E-07	ND
eOD-GT5	PGV19 GL	5.20E+03	1.56E-03	3.01E-07	ND
eOD-GT5	PGV19 Mature	3.90E+03	7.60E-02	1.95E-05	1.92E-05
eOD-GT5	VRC01 GL	1.51E+03	7.99E-04	5.29E-07	ND
eOD-GT5	VRC01 Mature	2.39E+04	1.36E-03	5.70E-08	ND
eOD-GT6	12A12 GL	1.39E+04	2.76E-02	1.98E-06	1.86E-06
eOD-GT6	12A12 Mature	7.57E+04	3.10E-02	4.09E-07	3.71E-07
eOD-GT6	3BNC60 GL	1.88E+04	2.59E-01	1.38E-05	1.36E-05
eOD-GT6	3BNC60 Mature	1.57E+05	3.10E-02	1.98E-07	1.85E-07
eOD-GT6	NIH45-46 GL	4.86E+03	1.99E-03	4.08E-07	ND
eOD-GT6	NIH45-46 Mature	6.58E+04	2.53E-04	3.84E-09	ND
eOD-GT6	PGV04 GL	ND	ND	-	5.20E-05
eOD-GT6	PGV04 Mature	1.59E+05	1.55E-03	9.73E-09	ND
eOD-GT6	PGV19 GL	5.05E+04	9.82E-04	1.94E-08	ND
eOD-GT6	PGV19 Mature	1.31E+05	1.15E-02	8.78E-08	ND
eOD-GT6	PGV20 GL	1.19E+05	4.69E-04	3.94E-09	ND
eOD-GT6	PGV20 Mature	2.27E+05	1.39E-03	6.15E-09	ND
eOD-GT6	VRC-CH31 GL	9.52E+03	2.31E-01	2.43E-05	2.38E-05
eOD-GT6	VRC-CH31 Mature	3.97E+04	1.14E+00	2.86E-05	2.84E-05

eOD-GT6	VRC01 GL	1.57E+04	6.83E-04	4.36E-08	ND
eOD-GT6	VRC01 Mature	1.48E+05	2.51E-04	1.69E-09	ND
eOD-GT6	VRC03 GL	NB	NB	-	NB
eOD-GT6	VRC03 Mature	1.79E+05	1.59E-01	8.92E-07	8.72E-07
eOD-GT6	VRC06 GL	NB	NB	-	NB
eOD-GT6	VRC06 Mature	5.53E+04	1.21E-01	2.19E-06	2.10E-06
eOD-GT6 V460N	VRC01 GL	1.83E+04	2.18E-03	1.19E-07	ND
eOD-GT6 V460N	VRC01 Mature	1.85E+05	3.53E-04	1.91E-09	ND
eOD-GT6 S471G	VRC01 GL	1.54E+04	2.54E-02	1.65E-06	1.37E-06
eOD-GT6 S471G	VRC01 Mature	2.83E+05	2.99E-04	1.06E-09	ND
eOD-GT6 F260L	VRC01 GL	1.87E+04	5.86E-03	3.13E-07	ND
eOD-GT6 F260L	VRC01 Mature	1.92E+05	3.06E-04	1.59E-09	ND
eOD-GT6 R278T	VRC01 GL	1.00E+04	1.57E-03	1.57E-07	ND
eOD-GT6 R278T	VRC01 Mature	1.71E+05	5.93E-04	3.46E-09	ND
eOD-GT6 R357K	VRC01 GL	2.04E+04	1.31E-03	6.43E-08	ND
eOD-GT6 R357K	VRC01 Mature	3.03E+05	3.22E-04	1.06E-09	ND
eOD-GT6 F371I	VRC01 GL	7.91E+03	3.59E-03	4.54E-07	ND
eOD-GT6 F371I	VRC01 Mature	1.48E+05	2.64E-04	1.78E-09	ND
eOD-GT6 T465S	VRC01 GL	1.13E+03	4.65E-04	4.13E-08	ND
eOD-GT6 T465S	VRC01 Mature	9.90E+04	2.46E-04	2.50E-09	ND
eOD-GT6 D368R	NIH45-46 GL	NB	NB	-	NB
eOD-GT6 D368R	NIH45-46 Mature	NB	NB	-	NB
eOD-GT6 D368R	PGV19 GL	1.39E+04	1.93E-01	1.39E-05	ND
eOD-GT6 D368R	PGV19 Mature	ND	ND	-	1.93E-05
eOD-GT6 D368R	VRC01 GL	2.92E+03	8.59E-02	2.94E-05	ND
eOD-GT6 D368R	VRC01 Mature	ND	ND	-	4.00E-05

eOD-GT6 D279A	NIH45-46 GL	NB	NB	-	NB
eOD-GT6 D279A	NIH45-46 Mature	NB	NB	-	NB
eOD-GT6 D279A	PGV19 GL	NB	NB	-	NB
eOD-GT6 D279A	PGV19 Mature	NB	NB	-	NB
eOD-GT6 D279A	VRC01 GL	NB	NB	-	NB
eOD-GT6 D279A	VRC01 Mature	ND	ND	-	1.20E-04
eOD-GT6 D386N + D463N	VRC01 GL	6.47E+03	7.99E-04	1.24E-07	ND
eOD-GT6 D386N + D463N	VRC01 Mature	5.55E+04	4.44E-04	7.99E-09	ND
eOD-GT6 D386N + D463N	NIH45-46 GL	2.20E+03	2.38E-03	1.08E-06	ND
eOD-GT6 D386N + D463N	NIH45-46 Mature	2.19E+04	7.71E-04	3.52E-08	ND
eOD-GT6 D386N + D463N	PGV19 GL	1.93E+04	1.07E-03	5.54E-08	ND
eOD-GT6 D386N + D463N	PGV19 Mature	2.00E+06	2.00E+00	1.00E-06	ND
eOD-GT6	VRC01_VH1-2*01	NB	NB	-	NB
eOD-GT6	VRC01_VH1-2*03	1.59E+04	6.13E-04	3.86E-08	ND
eOD-GT6	VRC01_VH1-2*04	1.50E+04	5.77E-04	3.85E-08	ND
eOD-GT6	VRC01_VH1-3*01	NB	NB	-	NB
eOD-GT6	VRC01_VH1-8*01	ND	ND	-	3.20E-04
eOD-GT6	VRC01_VH1-46*01	NB	NB	-	NB
eOD-GT6	VRC01_Rhe1	NB	NB	-	NB
eOD-GT6	VRC01_Rhe2	ND	ND	-	2.74E-05
eOD-GT6	VRC01_Rhe4	ND	ND	-	3.78E-05
eOD Base	12A12 GL	NB	NB	-	NB
eOD Base	12A12 Mature	5.28E+04	2.00E-02	3.79E-07	ND
eOD Base	3BNC60 GL	NB	NB	-	NB
eOD Base	3BNC60 Mature	-	-	-	-
eOD Base	NIH45-46 GL	NB	NB	-	NB

eOD Base	NIH45-46 Mature	4.34E+04	6.18E-04	1.43E-08	ND
eOD Base	PGV04 GL	NB	NB	-	NB
eOD Base	PGV04 Mature	4.57E+04	5.18E-03	1.13E-07	ND
eOD Base	PGV19 GL	NB	NB	-	NB
eOD Base	PGV19 Mature	5.07E+04	1.58E-01	3.12E-06	2.99E-06
eOD Base	PGV20 GL	ND	ND	-	1.70E-05
eOD Base	PGV20 Mature	1.89E+05	2.54E-03	1.34E-08	
eOD Base	VRC-CH31 GL	NB	NB	-	NB
eOD Base	VRC-CH31 Mature	ND	ND	-	3.02E-05
eOD Base	VRC01 GL	NB	NB	-	NB
eOD Base	VRC01 Mature	7.05E+04	3.51E-04	4.98E-09	ND
HXB2 Core GP120	12A12 GL	NB	NB	-	NB
HXB2 Core GP120	12A12 Mature	2.60E+04	1.67E-04	6.42E-09	ND
HXB2 Core GP120	3BNC60 GL	NB	NB	-	NB
HXB2 Core GP120	3BNC60 Mature	1.17E+04	4.22E-04	3.62E-08	ND
HXB2 Core GP120	NIH45-46 GL	NB	NB	-	NB
HXB2 Core GP120	NIH45-46 Mature	1.30E+04	4.56E-04	3.50E-08	ND
HXB2 Core GP120	PGV04 GL	NB	NB	-	NB
HXB2 Core GP120	PGV04 Mature	8.86E+03	4.28E-04	4.83E-08	ND
HXB2 Core GP120	PGV19 GL	NB	NB	-	NB
HXB2 Core GP120	PGV19 Mature	5.62E+03	1.13E-04	2.02E-08	ND
HXB2 Core GP120	PGV20 GL	NB	NB	-	NB
HXB2 Core GP120	PGV20 Mature	8.97E+04	1.71E-03	1.90E-08	ND
HXB2 Core GP120	VRC-CH31 GL	NB	NB	-	NB
HXB2 Core GP120	VRC-CH31 Mature	4.23E+03	1.99E-04	4.71E-08	ND
HXB2 Core GP120	VRC01 GL	NB	NB	-	ND
HXB2 Core GP120	VRC01 Mature	1.71E+04	7.74E-05	4.54E-09	ND

Core.Bal-GT1	12A12 GL	2.61E+04	8.27E-02	3.17E-06	3.14E-06
Core.Bal-GT1	12A12 Mature	1.56E+05	1.62E-04	1.04E-09	ND
Core.Bal-GT1	3BNC60 GL	4.65E+04	7.64E-01	1.64E-05	1.59E-05
Core.Bal-GT1	3BNC60 Mature	2.57E+05	1.01E-03	3.94E-09	ND
Core.Bal-GT1	NIH45-46 GL	NB	NB	-	NB
Core.Bal-GT1	NIH45-46 Mature	1.10E+05	7.01E-05	6.37E-10	ND
Core.Bal-GT1	PGV19 GL	1.18E+05	3.00E-03	2.55E-08	ND
Core.Bal-GT1	PGV19 Mature	7.40E+05	1.01E-02	1.36E-08	ND
Core.Bal-GT1	PGV20 GL	7.58E+03	4.21E-02	5.55E-06	ND
Core.Bal-GT1	PGV20 Mature	2.66E+05	9.93E-04	3.73E-09	ND
Core.Bal-GT1	VRC-CH31 GL	1.36E+04	4.81E-01	3.54E-05	3.30E-05
Core.Bal-GT1	VRC-CH31 Mature	3.12E+04	5.58E-02	1.79E-06	2.00E-06
Core.Bal-GT1	VRC01 GL	3.00E+05	4.00E-01	1.33E-06	1.80E-06
Core.Bal-GT1	VRC01 Mature	1.96E+05	9.66E-05	4.93E-10	ND

Table S8: K_Ds for eOD-GT6 interactions with GL VRC01-class Abs with selected non-human VH genes. Values in nM.

	eOD-GT6
Mus VH4-1	$>10^5$
Mus VH5-4	$>10^5$
Mus VH5-12	$>10^5$
Mus VH5-17	$>10^5$
Mus VH11-1	$>10^5$
Rhe1	$>10^5$
Rhe2	30,000
Rhe4	40,000

Table S9. Strains in LANL Lacking the N276 Glycan

HIV strains lacking glycan at 276 out of 2867 different sequences from the LANL database:

Strain name	276	277	278
A.CM.97.97CM_MP640.AM279366	S	L	T
A.UG.07.191955_A11.HM215272	N	I	N
A1.KE.00.KNH1211.AF457070	N	I	A
A1.KE.04.QG984_21M_ENV_A3.FJ866117	D	I	S
A1.TZ.01.A341.AY253314	N	I	A
B.AU.96.MBCD36.AF042105	N	F	M
B.CH.x.NAB8pre_cl_11.EU023927	D	F	S
B.CN.06.CNHLJSM06048.EU131794	N	F	M
B.CN.x.B05.EU363829	N	F	M
B.CO.01.PCM039.AY561239	N	F	A
B.GB.04.MM39d11p.HM586193	N	F	M
B.GB.08.F455b_B4.HQ595778	N	V	E
B.GB.x.MB314.Y13719	N	F	M
B.JP.05.-426.AB428556	N	F	X
B.US.01.108051_006.HM769944	D	F	K
B.US.06.YOMI_A6.EU578667	N	F	A
B.US.87.SFMHS5.AF025753	D	F	T
B.US.97.1013_03.AY331287	N	F	X
B.US.97.62357_14_D3_4589.EU289189	K	F	A
B.US.97.M02_3_SW.U84854	D	F	S
B.US.x.F1540TOB8U.GU728134	S	F	T
B.US.x.H0002GH.DQ222211	D	F	K
B.US.x.L3516TOB8U.GU728392	D	F	M
C.BW.00.00BW076820.AF443089	A	L	T
C.BW.00.00BW087421.AF443090	N	L	A
C.BW.00.00BW18595.AF443099	N	I	E
C.BW.00.00BW20361.AF443102	N	L	A
C.BW.96.96BW11B01.AF110971	I	I	T
C.BW.99.99BW47547.AF443086	S	L	T
C.GE.03.03GEMZ033.DQ207941	D	I	N
C.IN.00.HIV_00836_2.EF117265	K	L	D

C.IN.x.CALCMANDAL.AJ276221	N	L	I
C.MW.03.CHV0011247_0478_H2.FJ444251	E	L	G
C.MW.x.BF1266_431a.HM215360	N	L	E
C.TZ.00.390_F1_B7.HQ697983	N	L	K
C.TZ.03.6471_v1_c16.HM215328	D	L	N
C.TZ.x.346_F4_D2_12.HM215302	N	I	N
C.ZA.00.1184MB.AY838566	N	L	A
C.ZA.00.1189MB.AY838565	N	L	A
C.ZA.00.1225MB.AY463227	N	L	I
C.ZA.00.J112MA.AY838568	N	L	K
C.ZA.02.02ZAPS005MB1.DQ351235	D	L	T
C.ZA.02.02ZAPS013MB1.DQ351222	N	L	A
C.ZA.02.02ZAPS014MB1.DQ351218	N	L	A
C.ZA.03.03ZAPS032MB1.DQ445633	N	L	A
C.ZA.03.03ZAPS056MB1.DQ396374	N	L	K
C.ZA.03.03ZAPS066MB2.DQ396375	N	M	K
C.ZA.03.03ZAPS067MB2.DQ396389	N	L	E
C.ZA.03.03ZAPS116MB1.DQ445635	N	L	A
C.ZA.03.03ZAPS136MB1.DQ351231	N	L	A
C.ZA.03.03ZAPS151MB1.DQ396392	K	L	T
C.ZA.03.03ZASK034B1.AY878065	N	P	T
C.ZA.03.03ZASK078B1.AY901971	K	L	T
C.ZA.03.03ZASK098B1.AY878061	N	L	A
C.ZA.03.03ZASK224MB1.DQ275664	K	L	T
C.ZA.03.03ZASK232B1.DQ093589	D	L	T
C.ZA.04.04ZAPS157MB1.DQ351232	N	L	A
C.ZA.04.04ZASK139B1.AY878072	D	L	T
C.ZA.04.04ZASK168B1.AY878058	D	L	T
C.ZA.04.04ZASK181B1.AY878062	S	L	T
C.ZA.04.04ZASK204B1.DQ056414	N	L	A
C.ZA.04.SK144B1.AY703911	N	L	A
C.ZA.05.05ZAFV5.DQ382363	N	L	A
C.ZA.05.CAP174_4w.GQ999981	N	L	A
C.ZA.05.CAP239_5w_F1.GQ999991	N	I	L
C.ZA.06.2833264.HQ595757	N	L	A
C.ZA.06.CHV0005989_CAP129.1.15B2.FJ443417	K	L	E

C.ZA.06.CHV0006091_CAP222.1.11A6.FJ443515	D	L	E
C.ZA.07.3514597.HQ595759	N	M	A
C.ZA.07.CHV0008598_CAP40.2.01F2.FJ443865	N	L	K
C.ZA.09.704MC004N.GU080162	S	L	T
C.ZA.09.704MC016N.GU080173	D	L	E
C.ZA.98.TV013B.AF391246	S	L	E
C.ZA.99.COT9.DQ447272	N	I	K
C.ZA.99.LT36.AY522729	N	L	A
C.ZM.02.14M_BML_1012.HM036760	N	L	A
C.ZM.02.21M_BML_1012.GU939143	N	L	A
C.ZM.04.Z205FPB5NOV04ENV5.2.GQ485436	N	L	A
D.TZ.87.87TZ4622.U65075	S	L	T
D.UG.97.9FPC2.EU853126	N	L	A
D.UG.97.pt632.EU281996	S	L	T
D.UG.98.98UG57131.AF484505	N	I	K
D.UG.99.99UGA07412.AF484477	N	X	X
D.UG.99.99UGD23550.AF484485	K	L	E
D.YE.01.01YE386.AY795903	N	L	A
F1.RU.08.D88_845.GQ290462	N	I	K
K.CD.97.97ZR_EQTB11.AJ249235	D	I	T
U.NL.01.U_NL_01_H10986_C11.EF029069	N	I	I
01_AE.CF.90.90CF402.U51188	D	L	T
01_AE.CN.06.CNE3.HM215410	D	L	T
01_AE.CN.07.BJX4_6.GU475020	N	L	A
01_AE.TH.00.00TH_C3347.AY945721	N	L	E
01_AE.TH.01.OUR414I.AY358050	N	L	A
01_AE.TH.06.99PL2.EU743787	N	I	K
01_AE.TH.95.NI1144.AF070703	N	L	A
01_AE.TH.97.97TH6_107.AY125894	N	P	T
02_AG.CM.01.01CM_0119MA.GU201498	D	I	T
02_AG.CM.01.01CM_0186BA.GU201500	D	I	T
02_AG.CM.01.01CM_4410HAL.AY371142	S	I	T
02_AG.CM.02.02CM_4082STN.AY371141	S	I	T
02_AG.CM.04.250.EU513189	N	T	I
02_AG.US.99.99US_MSC1134.AY444809	S	L	T
05_DF.BE.93.VI961.AF076998	N	I	L

07_BC.CN.x.CH110.EF117257	K	L	T
14_BG.ES.00.X477.AF423759	N	F	X
14_BG.ES.05.X772_8.FJ670528	N	F	X
17_BF.PE.02.PE02_PCR0155.EU581828	N	I	F
24_BG.ES.08.X2456_2.FJ670526	N	F	A
35_AD.AF.06.047H.GQ477443	N	I	L
01B.MY.04.04MYKL016_1.DQ366663	X	L	T
02A1.GH.97.AG2.AB052867	N	L	A
02A1U.CM.05.280.EU513183	D	I	T
02G.CM.02.02CM_3228MN.AY371147	N	I	L
A1C.TZ.97.97TZ06.AF361876	D	L	T
A1D.DK.96.FSA.DQ912822	D	I	L
A1D.UG.98.98UG57129.AF484503	N	I	I
AC.TZ.04.6540_v4_c1.HM215330	H	I	G
AC.TZ.x.605_F4_12_5.HM215322	D	L	E
BF1.BR.99.BREPM107.AY771588	N	I	K
BF1.ES.02.X1241.AY536238	N	I	A
BF1.IT.05.83166.GU595152	N	I	X
CD.TZ.x.89_F1_2_25.HM215349	E	I	T
CF1.MW.07.CH010180_w12_p1.HM204593	N	L	K
DF.CM.93.CA4.AJ277819	N	I	R
DU.CM.x.247.EU683891	D	F	T
GKU.SE.95.SE9010.AY352655	D	M	T
O.BE.87.ANT70.L20587	D	I	L
O.CM.96.96ABB009.AF383231	D	I	K
O.CM.98.98CMA124.AF383245	N	I	M
O.CM.98.98CMA307.AF383246	D	I	R
O.CM.98.98CMA323.AF383248	S	I	T
O.CM.98.98CMA407.AF383250	N	I	F
O.FR.92.VAU.AF407418	D	I	S
N.CM.02.DJO0131.AY532635	-	-	N
N.CM.02.SJGddd.GQ324959	-	-	N
N.CM.04.04CM_1015_04.DQ017382	-	-	X
N.CM.06.U14296.GQ324962	-	-	-
N.CM.06.U14842.GQ324958	S	D	S
N.CM.95.YBF30.AJ006022	-	-	N

N.CM.97.YBF106.AJ271370	-	-	-
CPZ.CD.90.ANT.U42720	R	K	N
CPZ.CM.05.SIVcpzEK505.DQ373065	-	-	N
CPZ.CM.05.SIVcpzLB7.DQ373064	-	-	-
CPZ.CM.98.CAM3.AF115393	-	-	-
CPZ.CM.98.CAM5.AJ271369	D	L	R

References:

1. S. A. Plotkin, Correlates of protection induced by vaccination. *Clin Vaccine Immunol* **17**, 1055 (2010).
2. D. R. Burton, P. Poignard, R. L. Stanfield, I. A. Wilson, Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337**, 183 (2012).
3. W. R. Schief, Y.-E. A. Ban, L. Stamatatos, Challenges for structure-based HIV vaccine design. *Curr Opin HIV AIDS* **4**, 431 (2009).
4. D. R. Burton *et al.*, A Blueprint for HIV Vaccine Discovery. *Cell Host Microbe* **12**, 396 (2012).
5. D. C. Ekiert *et al.*, Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246 (2009).
6. P. Zhang *et al.*, Depletion of interfering antibodies in chronic hepatitis C patients and vaccinated chimpanzees reveals broad cross-genotype neutralizing activity. *Proc Natl Acad Sci U S A* **106**, 7537 (2009).
7. J. P. Julien, P. S. Lee, I. A. Wilson, Structural insights into key sites of vulnerability on HIV-1 Env and influenza HA. *Immunol Rev* **250**, 180 (2012).
8. L. Kong *et al.*, Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proc Natl Acad Sci U S A* **109**, 9499 (2012).
9. D. R. Burton, Antibodies, viruses and vaccines. *Nat Rev Immunol* **2**, 706 (2002).
10. X. L. Wu *et al.*, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856 (2010).

11. J. F. Scheid *et al.*, Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633 (2011).
12. X. Wu *et al.*, Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593 (2011).
13. T. Q. Zhou *et al.*, Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811 (2010).
14. R. Diskin *et al.*, Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **334**, 1289 (2011).
15. A. P. West, Jr., R. Diskin, M. C. Nussenzweig, P. J. Bjorkman, Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc Natl Acad Sci U S A* **109**, E2083 (2012).
16. R. Arnaout *et al.*, High-resolution description of antibody heavy-chain repertoires in humans. *Plos One* **6**, (2011).
17. R. Pejchal *et al.*, A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334**, 1097 (2011).
18. A. Leaver-Fay *et al.*, ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487**, 545 (2011).
19. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nature Protocols* **1**, 755 (2006).
20. C. Genomes Project *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56 (2012).
21. M. F. Bachmann, G. T. Jennings, Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol* **10**, 787 (2010).

22. X. Zhang, W. Meining, M. Fischer, A. Bacher, R. Ladenstein, X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *J Mol Biol* **306**, 1099 (2001).

23. T. Ota *et al.*, Anti-HIV B cell lines as candidate vaccine biosensors. *J Immunol*, (2012).

24. S. Hoot *et al.*, Recombinant HIV Envelope Proteins Fail to Engage Germline Versions of Anti-CD4bs bNAbs. *PLoS Pathog* **9**, e1003106 (2013).

25. C. Sundling *et al.*, Soluble HIV-1 Env trimers in adjuvant elicit potent and diverse functional B cell responses in primates. *J Exp Med* **207**, 2003 (2010).

26. K. L. Knight, C. R. Winstead, Generation of antibody diversity in rabbits. *Curr Opin Immunol* **9**, 228 (1997).

27. Several mouse VH genes were identified that contained the ArgH71 important for eOD-GT6 binding, but crystal structures of mouse Abs indicated that ArgH71 would not be accessible for binding, as it forms internal H-bonds with backbone carbonyl groups (fig. S15).

28. B. F. Haynes, G. Kelsoe, S. C. Harrison, T. B. Kepler, B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat Biotechnol* **30**, 423 (2012).

29. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774 (2007).

30. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605 (2004).

31. M. M. Souto-Carneiro, N. S. Longo, D. E. Russ, H. W. Sun, P. E. Lipsky, Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *J Immunol* **172**, 6790 (2004).

32. P.-S. Huang *et al.*, RosettaRemodel: a generalized framework for flexible backbone protein design. *Plos One* **6**, (2011).

33. B. Kuhlman *et al.*, Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364 (2003).

34. S. J. Fleishman *et al.*, RosettaScripts: a scripting language interface to the rosetta macromolecular modeling suite. *Plos One* **6**, (2011).

35. J. J. Gray *et al.*, Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* **331**, 281 (2003).

36. I. W. Davis, W. B. Arendall, D. C. Richardson, J. S. Richardson, The backrub motion: How protein backbone shrugs when a sidechain dances. *Structure* **14**, 265 (2006).

37. W. P. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49 (1995).

38. L. Benatuil, J. M. Perez, J. Belk, C.-M. Hsieh, An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* **23**, 155 (2010).

39. S. Bryson *et al.*, Cross-neutralizing human monoclonal anti-HIV-1 antibody 2F5: Preparation and crystallographic analysis of the free and epitope-complexed forms of its F-ab ' fragment. *Protein Peptide Lett* **8**, 413 (2001).

40. R. S. Depetris *et al.*, Partial enzymatic deglycosylation preserves the structure of cleaved recombinant HIV-1 envelope glycoprotein trimers. *J Biol Chem* **287**, 24239 (2012).

41. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125 (2010).

42. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658 (2007).

43. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235 (2011).

44. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213 (2010).

45. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126 (2004).