

Supporting Information

Fera et al. 10.1073/pnas.1409954111

SI Methods

Compliance. All work related to human subjects complied with protocols approved by the Duke University Health System Institutional Review Board.

Expression and Purification of Fabs. The heavy- and light-chain variable and constant domains of the unmutated common ancestor (UCA), I3.2, I3.1, and I2 Fabs from the CH103 lineage were cloned into the pVRC-8400 expression vector using NotI and NheI restriction sites and the tissue plasminogen activator signal sequence. The C terminus of the heavy-chain constructs contained a noncleavable 6x histidine tag. Fabs were expressed using transient transfection of 293T cells using linear polyethylenimine (PEI) following the manufacturer's suggested protocol. After 5 d of expression, supernatants were clarified by centrifugation. His-tagged Fabs were loaded onto Ni-NTA superflow resin (Qiagen) preequilibrated with Buffer A (10mM Tris, pH 7.5, 100 mM NaCl), washed with Buffer A + 10 mM imidazole, and eluted with Buffer A + 350 mM imidazole. Fabs were then purified by gel filtration chromatography in Buffer A using a superdex 200 analytical column (GE Healthcare).

Plasmids containing the CMVR VRC01 Ig heavy and light chains were obtained from the NIH Aids Reagents Program. The protein was expressed in 293F cells as described for Fabs above. The clarified supernatant was diluted twofold using 1x PBS buffer and purified using protein A agarose resin (Pierce), according to manufacturer's protocols. Fractions containing the protein of interest were pooled, concentrated, and further purified by gel filtration chromatography in Buffer A using a superdex 200 analytical column (GE Healthcare). The VRC01 Fab was obtained by digesting the Ig using papain (Pierce) and by running the digested products through protein A resin (Pierce) according to manufacturer's protocol. The Fab was then purified by gel filtration chromatography in Buffer A using a superdex 200 analytical column (GE Healthcare).

Expression and Purification of HIV 92UG037.8 HIV-1 Subtype A Glycoproteins. The IRES-puro vector containing 92UG037.8 HIV-1 subtype A gp120 amino acid (a.a.) residues 1–492 (HXB2 numbering) Δ V123 (core) with a 6x-histidine tag inserted between residues 40 and 41 was kindly provided by Bing Chen (Boston Children's Hospital, Boston, MA). Site-directed mutagenesis was performed using manufacturer's protocols (Stratagene) for D-loop mutants of the CH505 gp120 core and for the V5 mutant of the 92UG037.8 gp120 core. Recombinant Env glycoproteins were expressed in 293T cells and purified as described for His-tagged Fabs.

Design, Expression, and Purification of pVRC8400-HIV-1 Clade C CH505 gp120 Δ V123 Constructs. The codon-optimized synthetic construct of the CH505 T/F HIV-1 subtype C gp120 containing a.a. 41–492 (HXB2 numbering) Δ V123 (core) was produced by GenScript with an N-terminal 6x-histidine tag and inserted into the pVRC-8400 expression vector as described for Fabs. The expression construct contained a leader sequence encoding the tissue plasminogen activator signal sequence

MDAMKRGLCCVLLLCGAVFVSPS

followed by a 6x-histidine tag and the codon-optimized sequence

⁴¹GVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVP-TDPNPQEMVLKNTENFNMWKNMVDQMHEDEVISLWDQSLKPCVKLTPLCVGAGNCNTSVITQACPQVSDPIPIH-

YCAPAGYAILKCNKNTFTGTGPCNNVSTVQCTHGKIPVV-STQLLLNGSLAEGEIIIRSENITNNVKTIIVHLNESVKIECTRP-GAGYCNINESKWNETLQRVSKKLKEYFPHKNITFQPSSG-GDLEITTHSFNCGGEFFYCNTSSLFNRTYMANSTDMANSTETNSTRITITIHCRIKQIINMWQEVGRAMYAPPIAGNITCISNITGLLLTRDGGKNNTEFRPGGGNMKNWRSLEYKYKVVVEVK⁴⁹²

between the 5' NheI and 3' NotI sites or by the codon-optimized sequence

⁴¹GVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVP-TDPNPQEMVLKNTENFNMWKNMVDQMHEDEVISLWDQSLKPCVKLTPLCVGAGNCNTSVITQACPQVSDPIPIHY-CAPAGYAILKCNKNTFTGTGPCNNVSTVQCTHGKIPVVST-QLLNGSLAEGEIIIRSENITNNVKTIIVHLNESVKIECTRP-GAGYCNINESKWNETLQRVSKKLKEYFPHKNITFQPSSG-GDLEITTHSFNCGGEFFYCNTSSLFNRTYMANSTDMANSTETNSTRITITIHCRIKQIINMWQEVGRAMYAPPIAGNITCISNITGLLLTRDGGKNNTEFRPGGGNMKNWRSLEYKYKVVVEVK⁴⁹²

between the 5' NheI and 3' NotI sites for the HIV-1 clade C CH505 gp120 Δ V123 expression construct with a V5 loop DT insertion or by the codon-optimized sequence

⁴¹GVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVP-TDPNPQEMVLKNTENFNMWKNMVDQMHEDEVISLWDQSLKPCVKLTPLCVGAGNCNTSVITQACPQVSDPIPIHY-CAPAGYAILKCNKNTFTGTGPCNNVSTVQCTHGKIPVV-STQLLLNGSLAEGEIIIRSENITNNVKTIIVHLNESVKIECTRP-GAGYCNINESKWNETLQRVSKKLKEYFPHKNITFQPSSGGDLEITTHSFNCGGEFFYCNTSSLFNRTYMANSTDMANSTETNSTRITITIHCRIKQIINMWQEVGRAMYAPPIAGNITCISNITGLLLTRDGGKNNTEFRPGGGNMKNWRSLEYKYKVVVEVK⁴⁹²

between the 5' NheI and 3' NotI sites for the HIV-1 clade C CH505 gp120 Δ V123 expression construct with a V5 loop ETF insertion or by the codon-optimized sequence

⁴¹GVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVP-TDPNPQEMVLKNTENFNMWKNMVDQMHEDEVISLWDQSLKPCVKLTPLCVGAGNCNTSVITQACPQVSDPIPIHY-CAPAGYAILKCNKNTFTGTGPCNNVSTVQCTHGKIPVVST-QLLNGSLAEGEIIIRSENITNNVKTIIVHLNESVKIECTRP-GAGYCNINESKWNETLQRVSKKLKEYFPHKNITFQPSSGGDLEITTHSFNCGGEFFYCNTSSLFNRTYMANSTDMANSTETNSTRITITIHCRIKQIINMWQEVGRAMYAPPIAGNITCISNITGLLLTRDGGKNNTEGKNNTEFRPGGGNMKNWRSLEYKYKVVVEVK⁴⁹²

between the 5' NheI and 3' NotI sites for the HIV-1 clade C CH505 gp120 Δ V123 expression construct with a V5 loop EGKNN insertion.

The insert was synthesized (GenScript) to replace the sequence encoding

¹²⁸TLNCTNATASNSSIIEGMKNCSFNITTELDRDKREKKNALFYKLDIVQLDGNSSQYRLI¹⁹⁴

from the V1/V2 loop with a sequence encoding a GAG and to replace the sequence encoding

³⁰⁰NNKTRTSIRIGPGQAFYATGQVIGDIREA³²⁹

from the V3 loop with a sequence encoding GAG.

Recombinant Env glycoproteins were expressed in 293T cells and purified as described for His-tagged Fabs.

Crystallization. All His-tagged Fabs were crystallized at ~10–12 mg/mL. Crystals were grown in 96-well format using hanging drop vapor diffusion and appeared after 24–48 h at 20 °C. UCA Fab crystals were obtained using microbatch seeding in a condition of 100 mM Hepes, pH 7.0, and 3 M NaCl from seeds of crystals obtained in a condition of 100 mM sodium acetate, pH 4.0, and 2 M ammonium sulfate. I3.2 crystals were grown over a reservoir of 100 mM Hepes, pH 7.0, 2 M ammonium sulfate, and 5% PEG 400; I3.1 crystals were grown over 100 mM sodium acetate, pH 5.0, and 2 M ammonium sulfate, and I2 crystals were grown over 100 mM Ches, pH 9.0, 1 M NaCl, and 20% PEG 4K. All crystals were harvested and cryoprotected by the addition of 20–25% glycerol to the reservoir solution and then flash-cooled in liquid nitrogen.

Structure Determination and Refinement. Diffraction data were obtained at 100°K from beam lines 24-ID-E at the Advanced Photon Source and 8.2.2 at the Advanced Light Source using a single wavelength. Datasets from individual crystals (1 crystal for each Fab) were processed with HKL2000 (1). Molecular replacement calculations were carried out with PHASER (2), using CH103 from the CH103/gp120 outer domain complex structure [Protein Data Bank (PDB) ID 4JAN] as the starting model. The CH103 model had to be separated into its variable and constant domains for the UCA and I3.2 Fab structure determinations. There were 4, 2, 4, and 1 molecules per asymmetric unit in the crystals of the UCA, I3.2, I3.1, and I2 Fabs, respectively.

The resulting electron density maps for the UCA and I3.1 were further improved by solvent flattening, histogram matching, and noncrystallographic symmetry averaging using the program DM (3). Phase combination was disabled in these calculations. After density modification, restrained refinement was performed for the UCA and I3.1 Fabs using Refmac in Coot. For all Fabs, subsequent refinement steps were carried out with Refmac (4) and PHENIX (5), and all model modifications were carried out with Coot (6). During refinement, maps were generated from combinations of positional, group B-factor, and TLS (translation/libration/screw) refinement algorithms. Secondary-structure restraints were included at all stages for all Fabs; noncrystallographic symmetry restraints were applied to the UCA and I3.1 Fabs throughout refinement.

The CDRL1 loop, which was three residues longer in all these Fabs than that in the CH103 Fab, was built manually using Coot.

The final refined structures contain residues 2–214 (heavy chain) and 2–209 (light chain) for the UCA, residues 1–215 (heavy chain) and 2–209 (light chain) for I3.2, residues 1–215 (heavy chain) and 2–209 (light chain) for I3.1, and residues 2–215 (heavy chain) and 3–209 (light chain) for I2. Density missing elsewhere is discussed in *Results*. Structure validations were performed periodically during refinement for each of the Fabs using the MolProbity server (5). The final refinement statistics are summarized in Table S1. The UCA, I3.2, I3.1, and I2 structures had, respectively, 87.4%, 92.7%, 89.7%, and 90.6% of their residues in favored regions of the Ramachandran plot, and 0.7%, 1.5%, 3.0%, and 2.6% in outlying regions. The C α rmsd of three of the molecules in the UCA asymmetric unit from the fourth are 0.276 Å, 0.459 Å, and 0.448 Å, and the last two have an rmsd of 0.292 Å relative to each other; the rmsd of the two I3.2 molecules in an asymmetric unit is 0.356 Å; the rmsd of three of the four molecules of I3.1 in an asymmetric unit from the fourth are 0.233 Å, 0.187 Å, and 0.198 Å.

Protein Structure Analysis and Graphical Representations. The heavy chains of the UCA, I3.2, I3.1, and I2 were superposed on that of the CH103 Fab by least squares fitting in Coot. The rmsds of different molecules in the asymmetric unit of a particular Fab were determined using PyMol by superposing the entire Fab molecules. To calculate C α rmsds between different Fabs, we used only the framework regions of the Fabs, as defined in Chothia et al. (7). All graphical representations with protein crystal structures were made using PyMol.

Biolayer Interferometry Measurements. Kinetic measurements of Fab binding to the autologous CH505 gp120 core and heterologous 92ug037.8 gp120 core and their corresponding mutants were carried out using the Octet QK^c system (ForteBio); 0.2 mg/mL of each His-tagged Fab was immobilized onto an anti-Human Fab-CH1 biosensor until it reached saturation. The gp120 core mutants were tested at concentrations of 1 μ M to 36 μ M, depending on the mutant. A reference sample of buffer alone was used to account for any signal drift that was observed during the experiment. Association and dissociation were each monitored for 5 min. All experiments were conducted in the Octet instrument at 30 °C in a buffer of 10 mM Tris, pH 7.5, and 100 mM NaCl with agitation at 1,000 rpm. Analyses were performed using a global fit of at least three measurements using nonlinear regression curve fitting using the Graphpad Prism software, version 6.

- Otwinowski Z, Minor W (1997) *Processing of X-Ray Diffraction Data Collected in Oscillation Mode* (Academic, New York).
- McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63(Pt 1):32–41.
- Cowtan K (1994) "dm": An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31:34–38.
- Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.

- Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
- Chothia C, Lesk AM (1987) Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196(4):901–917.
- Liao HX, et al.; NISC Comparative Sequencing Program (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496(7446):469–476.

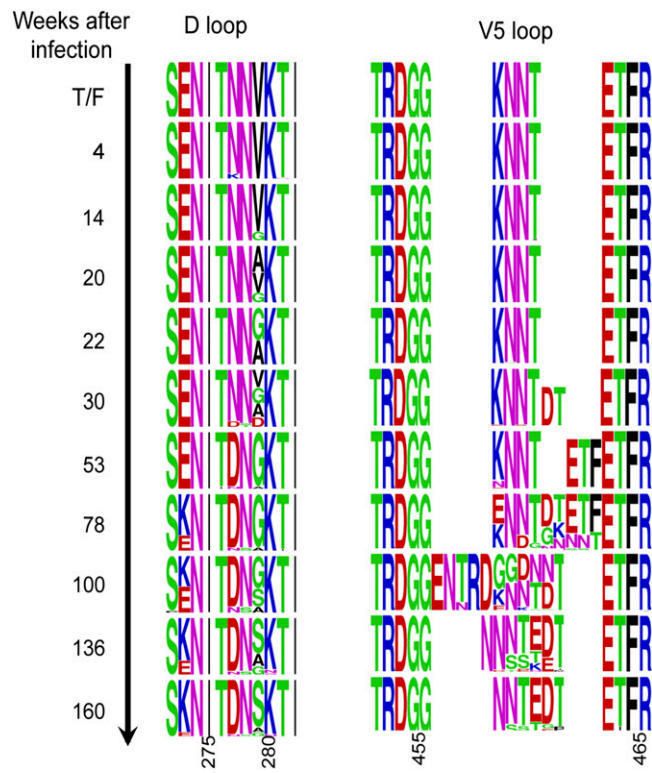


Fig. S1. Sequence logo of the D and V5 loops at the time points indicated on the left. The frequency of each amino acid is shown by its height; spaces allow for deletions and insertions. Amino acid position numbers are based on HIV-1 HXB2 numbering. GSTYC polar residues are green; NQ neutral residues, purple; KRH basic residues, blue; DE acidic residues, red; and PAWFLIMV hydrophobic residues, black.

Table S1. Data collection and refinement statistics

	UCA	I3.2	I3.1	I2
Data collection				
Space group	P12 ₁ 1	P2 ₁ 2 ₁ 2 ₁	P3 ₁ 12	P2 ₁ 2 ₁ 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.7, 71.1, 184.4	72.1, 99.1, 163.3	132.3, 132.3, 209.6	55.9, 67.0, 128.8
α , β , γ (°)	90, 93.8, 90	90, 90, 90	90, 90, 120	90, 90, 90
Resolution (Å)	46.85–3.49 (3.56–3.49)*	49.57–3.15 (3.20–3.15)	48.04–3.23 (3.29–3.23)	46.42–3.00 (3.05–3.00)
<i>R</i> _{sym} or <i>R</i> _{merge}	14.4 (37.3)	12.3 (38.6)	11.4 (65.0)	10.5 (80.6)
<i>I</i> / σ (<i>I</i>)	7.1 (3.0)	7.7 (2.1)	6.7 (1.1)	9.6 (1.3)
Completeness (%)	96.5 (95.1)	99.0 (98.1)	98.9 (97.6)	97.6 (97.5)
Redundancy	3.0 (3.0)	2.2 (2.2)	1.9 (2.0)	2.3 (2.3)
Refinement				
Resolution (Å)	46.85–3.49	49.57–3.15	48.04–3.23	46.42–3.00
No. reflections	23,779	20,706	33,905	10,152
<i>R</i> _{work} / <i>R</i> _{free} (%)	26.9/28.5	23.7/28.2	22.4/25.1	22.6/27.2
No. atoms				
Protein	12,800	6,454	12,840	3,220
Ligand/ion				
Water				
<i>B</i> -factors				
Protein	66.2	59.4	85.5	87.6
Ligand/ion				
Water				
rms deviations				
Bond lengths (Å)	0.005	0.003	0.004	0.004
Bond angles (°)	1.18	0.787	0.978	0.986

*Values in parentheses are for highest-resolution shell. One crystal of each Fab was used for data collection.

Table S2. C α rmsd (Å) of the framework variable domains of CH103 lineage Fabs

	UCA	I3.2	I3.1	I2	CH103
UCA	0				
I3.2	0.41	0			
I3.1	0.94	1.05	0		
I2	0.92	0.94	0.54	0	
CH103	1.07	1.01	0.73	0.74	0

Table S3. C α rmsd (Å) of the individual V_H (white) and V_L (gray) framework domains of CH103 lineage Fabs

	UCA	I3.2	I3.1	I2	CH103
UCA	0	0.38	0.43	0.44	0.46
I3.2	0.45	0	0.27	0.33	0.33
I3.1	0.54	0.41	0	0.36	0.34
I2	0.68	0.45	0.45	0	0.42
CH103	0.57	0.38	0.31	0.41	0

Values shaded in gray were determined by comparing the V_L framework domains of the indicated Fabs; values in white regions were determined by comparing V_H framework domains.

