Published in final edited form as:

Science. 2011 September 16; 333(6049): 1593–1602. doi:10.1126/science.1207532.

Focused Evolution of HIV-1 Neutralizing Antibodies Revealed by Structures and Deep Sequencing

Xueling Wu^{1,*}, Tongqing Zhou^{1,*}, Jiang Zhu^{1,*}, Baoshan Zhang¹, Ivelin Georgiev¹, Charlene Wang¹, Xuejun Chen¹, Nancy S. Longo¹, Mark Louder¹, Krisha McKee¹, Sijy O'Dell¹, Stephen Peretto¹, Stephen D. Schmidt¹, Wei Shi¹, Lan Wu¹, Yongping Yang¹, Zhi-Yong Yang¹, Zhongjia Yang¹, Zhenhai Zhang^{1,2}, Mattia Bonsignori³, John A. Crump^{4,5}, Saidi H. Kapiga⁶, Noel E. Sam^{5,6}, Barton F. Haynes³, Melissa Simek⁷, Dennis R. Burton^{8,9}, Wayne C. Koff⁷, Nicole A. Doria-Rose¹⁰, Mark Connors¹⁰, NISC Comparative Sequencing Program¹¹, James C. Mullikin¹¹, Gary J. Nabel¹, Mario Roederer¹, Lawrence Shapiro^{1,2}, Peter D. Kwong^{1,†}, and John R. Mascola^{1,†}

¹Vaccine Research Center, National Institutes of Health, Bethesda, MD 20892, USA

²Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

³Duke Human Vaccine Institute, Duke University School of Medicine, and Duke University Medical Center, Durham, NC 27710, USA

⁴Division of Infectious Diseases and International Health, Department of Medicine, and Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA

⁵Kilimanjaro Christian Medical Centre and Kilimanjaro Christian Medical College, Tumaini University, Moshi, Tanzania

⁶Kilimanjaro Reproductive Health Programme, Moshi, Tanzania

⁷International AIDS Vaccine Initiative (IAVI), New York, NY 10038, USA

⁸Department of Immunology and Microbial Science and IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA 92037, USA

⁹Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard, Cambridge, MA 02129, USA

¹⁰Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

¹¹NIH Intramural Sequencing Center (NISC), National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

Antibody VRC01 is a human immunoglobulin that neutralizes about 90% of HIV-1 isolates. To understand how such broadly neutralizing antibodies develop, we used x-ray crystallography and 454 pyrosequencing to characterize additional VRC01-like antibodies from HIV-1-infected individuals. Crystal structures revealed a convergent mode of binding for diverse antibodies to the same CD4-binding-site epitope. A functional genomics analysis of expressed heavy and light

[†]To whom correspondence should be addressed. pdkwong@nih.gov (P.D.K.); jmascola@nih.gov (J.R.M.).

*These authors contributed equally to this work.

chains revealed common pathways of antibody-heavy chain maturation, confined to the IGHV1-2*02 lineage, involving dozens of somatic changes, and capable of pairing with different light chains. Broadly neutralizing HIV-1 immunity associated with VRC01-like antibodies thus involves the evolution of antibodies to a highly affinity-matured state required to recognize an invariant viral structure, with lineages defined from thousands of sequences providing a genetic roadmap of their development.

HIV-1 exhibits extraordinary genetic diversity and has evolved multiple mechanisms of resistance to evade the humoral immune response (1–3). Despite these obstacles, 10 to 25% of HIV-1-infected individuals develop cross-reactive neutralizing antibodies after several years of infection (4–9). Elicitation of such antibodies could form the basis for an effective HIV-1 vaccine, and intense effort has focused on identifying responsible antibodies and delineating their characteristics. A variety of monoclonal antibodies (mAbs) have been isolated that recognize a range of epitopes on the functional HIV-1 viral spike, which is composed of three highly glycosylated gp120 exterior envelope glycoproteins and three transmembrane gp41 molecules. Some broadly neutralizing antibodies are directed against the membrane-proximal external region of gp41 (10, 11), but the majority recognize gp120. These include the quaternary structure–preferring antibodies PG9, PG16, and CH01-04 (12, 13); the glycan-reactive antibodies 2G12 and PGT121-137 (14, 15); and antibodies b12, HJ16, and VRC01-03, which are directed against the region of HIV-1 gp120 involved in initial contact with the CD4 receptor (16–19).

One unusual characteristic of all these gp120-reactive broadly neutralizing antibodies is a high level of somatic mutation. Antibodies typically accumulate 5 to 15% changes in variable domain–amino acid sequence during the affinity maturation process (20), but for these gp120 reactive neutralizing antibodies, the degree of heavy chain–somatic mutation is markedly increased, ranging from 19% for the quaternary structure–preferring antibodies (12), to 31% for antibody 2G12 (21, 22), and to 40 to 46% for the CD4-binding-site antibodies, HJ16 (17), VRC01, VRC02, and VRC03 (18) (table S1).

In the case of VRC01, the mature antibody accumulates roughly 70 total changes in amino acid sequence during the maturation process. The mature VRC01 can neutralize ~90% of HIV-1 isolates at a geometric mean inhibitory concentration (IC_{50}) of 0.3 μ g/ml (18), and structural studies show that it achieves this neutralization by precisely recognizing the initial site of CD4 attachment on HIV-1 gp120 (19). By contrast, the predicted unmutated germline ancestor of VRC01 has weak affinity for typical strains of gp120 (in the millimolar range) (19). Moreover, with only three VRC01-like antibodies identified in a single individual (donor 45), it has been unclear whether the VRC01 mode of recognition, genetic origin, and pathway of affinity maturation represent general features of the B cell response to the CD4-binding site of HIV-1 gp120. Here, we explore how broadly neutralizing HIV-1 immunity associated with VRC01-like antibodies develops, with an analysis of dozens of neutralizers from additional donors to answer questions of generality and to trace pathways of affinity maturation with thousands of VRC01-like antibody sequences.

Isolation of neutralizing antibodies from donors 74 and 0219 with a CD4-binding-site probe

We previously used structure-guided resurfacing to alter the antigenic surfaces on HIV-1 gp120 while preserving the initial site of attachment to the CD4 receptor (18). With the resurfaced stabilized core 3 probe (RSC3), over 30% of the surface residues of core gp120 were altered and the conformation stabilized by the addition of interdomain-disulfide bonds and cavity-filling point mutations (18). We used RSC3 and a mutant version containing a

single amino acid deletion in the CD4-binding loop (Δ RSC3) to interrogate a panel of 12 broadly neutralizing sera derived from the IAVI protocol G cohort of HIV-1-infected individuals (6, 23) (Fig. 1A). A substantial fraction of neutralization of three sera (23, 57, and 74) was specifically blocked by RSC3 compared with Δ RSC3, indicating the presence of CD4-binding-site-directed neutralizing antibodies. RSC3-neutralization competition assays also confirmed the presence of CD4-binding-site antibodies in the previously characterized sera 0219, identified in the Center for HIVAIDS Vaccine Immunology (CHAVI) 001 cohort (8) (Fig. 1A). Peripheral blood mononuclear cells (PBMCs) from protocol G donor 74 (infected with A/D recombinant) and from CHAVI donor 0219 (infected with clade A) were used for antigen-specific B cell sorting and antibody isolation. For donors 74 and 0219, respectively, a total of 0.13% and 0.15% of IgG⁺ (immunoglobulin G) B cells were identified (Fig. 1B and fig. S1). The heavy- and light-chain immunoglobulin genes from individual B cells were amplified and cloned into IgG1 expression vectors that reconstituted the full IgG (18, 24). From donor 74, two somatically related antibodies named VRC-PG04 and VRC-PG04b demonstrated strong binding to several versions of gp120 and to RSC3 but >100-fold less binding to Δ RSC3 (fig. S2 and table S2). From donor 0219, five somatically related antibodies named VRC-CH30, 31, 32, 33, and 34 displayed a similar pattern of RSC3/ Δ RSC3 reactivity (fig. S2 and table S2). Sequence analysis of these two sets of antibody clones (Fig. 1C and table S3) revealed that they originated from the same inferred immunoglobulin heavy-chain variable (IGHV) precursor gene allele IGHV1-2*02. Despite this similarity in heavy-chain V-gene origin, the two clone sets originated from different heavy-chain J segment genes and contained different light chains. The light chains of the VRC-PG04 and 04b somatic variants originated from an IG κ V3 allele, whereas the VRC-CH30-34 somatic variants derived from an IG κ V1 allele. Of note, all seven antibodies were highly affinity matured: VRC-PG04 and 04b displayed a heavy-chain-variable gene (V_H) mutation frequency of 30% relative to the germline IGHV1-2*02 allele, a level of affinity maturation similar to that previously observed with VRC01-03; the VRC-CH30-34 antibodies were also highly affinity matured, with a V_H mutation frequency of 23 to 25%.

To define the reactivities of these new antibodies on gp120, we performed competition enzyme-linked immunosorbent assays with a panel of well-characterized mAbs. Binding by each of the new antibodies was competed by VRC01-03, by other CD4-binding-site antibodies and by CD4-Ig, but not by antibodies known to bind gp120 at other sites (fig. S3). Despite similarities in gp120 reactivity and V_H -genomic origin, sequence similarities of heavy- and light-chain gene regions did not readily account for their mode of gp120 recognition (table S4). Finally, assessment of VRC-PG04 and VRC-CH31 neutralization on a panel of Env pseudoviruses revealed their ability to potently neutralize a majority of diverse HIV-1 isolates (Fig. 1D and table S5).

Structural definition of gp120 recognition by RSC3-identified antibodies from different donors: A remarkable convergence

To define the mode of gp120 recognition employed by donor 74-derived VRC-PG04, we crystallized its antigen-binding fragment (Fab) in complex with a gp120 core from the clade A/E recombinant 93TH057 that was previously crystallized with VRC01 (19). Diffraction data to 2.1 Å resolution were collected from orthorhombic crystals, and the structure was solved by molecular replacement and refined to a crystallographic *R* value of 19.0% (Fig. 2A and tables S6 and S7). The structure of VRC-PG04 in complex with HIV-1 gp120 showed striking similarity with the previously determined complex with VRC01, despite different donor origins and only 50% amino acid identity in the heavy-chain-variable region (Fig. 2). When gp120s were superimposed, the resultant heavy-chain positions of VRC-PG04 and VRC01 differed by a root-mean-square deviation (RMSD) of 2.1 Å in C α atoms, with even more precise alignment of the heavy-chain-second complementarity

determining (CDR H2) region (1.5 Å RMSD). Critical interactions such as the Asp368gp120 salt bridge to Arg71VRC01 were maintained in VRC-PG04 (Fig. 2B).

We also crystallized the gp120-Fab complex of the donor 45–derived VRC03 mAb, the isolation and initial characterization of which were previously described (18). VRC03 and VRC-PG04 share only 51% heavy-chain–variable protein sequence identity (table S4), and the heavy chain of VRC03 contains an unusual insertion in the framework 3 region (18). Diffraction data to 1.9 Å resolution were collected from orthorhombic crystals, and the structure was solved by molecular replacement and refined to a crystallographic *R* value of 18.7% (Fig. 2 and tables S6 and S8). VRC03 also showed recognition of gp120 that was strikingly similar to that of VRC-PG04 and VRC01, with similar interface residues (fig. S4) and pairwise RMSDs in Cα atoms of 2.4 and 1.9 Å, respectively. In particular, gp120–interactive surfaces of CDR H2 and CDR L3 showed similar recognition (pairwise Cα RMSDs of these regions of the antibodies ranged from 0.5 to 1.4 Å after superposition of gp120) (fig. S5).

In general, the repertoire of possible immunoglobulin products is very large, and highly similar modes of antibody recognition are expected to occur infrequently (25). To assess how atypical the VRC01-like antibody convergence was, we analyzed other families of HIV-1–specific antibodies that share common IGHV-gene origins (26–29), including the CD4-induced antibodies 17b, 412d, and X5, all of which derive from a common IGHV1-69 allele. Analysis of the recognition of gp120 by these antibodies indicated substantial variation, with angular difference in heavy-chain orientation between 17b, 412d, and X5 of over 90°, or roughly 10 times as much as among the VRC01-like antibodies (table S9). Also, the RSC3 probe may select for a particular mode of recognition, so we analyzed other CD4-binding-site antibodies that bind strongly to the RSC3 probe, including antibodies b12 and b13 (16, 30); these other RSC3-reactive antibodies also showed dramatic differences in heavy-chain orientation relative to the VRC01-like antibodies (table S10).

The remarkable convergence in recognition observed with VRC01, VRC03, and VRC-PG04 suggested a common mode of HIV-1 gp120 recognition, conserved between donors infected with a clade B (donor 45) or a clade A/D (donor 74) strain of HIV-1. The precision required for this mode of recognition likely arises as a consequence of the multiple mechanisms of immune evasion that protect the site of CD4 attachment on HIV-1 gp120 (30). We analyzed paratope surface properties and found that the average energy of antibody hydrophobic interactions (Δ^iG) correlated with the convergence in antibody recognition ($P = 0.0427$) (Fig. 3A) (31). Thus, although precise H-bonding is required for this mode of recognition (Fig. 2C), the convergence in structure appears to optimize regions with hydrophobic interactions (fig. S6 and table S11). Another important feature of this mode of recognition is its ability to focus precisely on the initial site of CD4 receptor attachment (19,32). Indeed, the breadth of HIV-1 neutralization among CD4-binding-site ligands correlated with targeting onto this site ($P = 0.0405$) (Fig. 3B).

This convergence in epitope recognition is accompanied by a divergence in antibody sequence identity (Fig. 1C, Fig. 3C, and table S4). All 10 antibodies isolated by RSC3 binding use the IGHV1-2*02 germline and accrue 70 to 90 nucleotide changes. Despite this similarity in the epitope recognized by these mature antibodies, only two residues from the germline IGHV1-2*02 allele mature to the same amino acids (Fig. 1C). Both of these changes occur at a hydrophobic contact in the critical CDR H2 region (56: Gly → Ala and 57: Thr → Val). The light chains for donors 45 and 74 antibodies arise from either IGVκ3-11*01 or IGVκ3-20*01, whereas the light chains of donor 0219 antibodies are derived from IGVκ1-33*01 (comparative contributions of different light chains to the interaction with gp120 are shown in figs. S4 and S7). For these light chains, no maturational

changes are identical. Despite this diversity in maturation, comparison of the VRC01, VRC03, and VRC-PG04 paratopes shows that many of these changes are of conserved chemical character (Fig. 3C); a hydrophobic patch in the CDR L3, for example, is preserved. These observations suggest that divergent amino acid changes among VRC01-like antibodies nevertheless afford convergent recognition when guided by affinity maturation.

Functional complementation of heavy and light chains among VRC01-like antibodies

Although the identification and sorting of antigen-specific B cells with resurfaced probes has resulted in the isolation of several broadly neutralizing antibodies, genomic analysis of B cell cDNA libraries provides thousands of sequences for analysis. These sequences specify the functional “antibodyome,” the repertoire of expressed antibody heavy- and light-chain sequences in each individual. High-throughput sequencing methods provide heavy-chain and light-chain sequences but do not retain information about their pairings. For VRC01-like antibodies, the structural convergence revealed by the crystallographic analysis indicated a potential solution: Different heavy and light chains might achieve functional complementation within this antibody family.

Heavy- and light-chain chimeras of VRC01, VRC03, VRC-PG04, and VRC-CH31 were produced by transient transfection (table S12) and tested for HIV-1 neutralization (table S13). VRC01 (donor 45) and VRC-PG04 (donor 74) light chains were functionally compatible with VRC01, VRC03, and VRC-PG04 heavy chains, although the VRC03 light chain was compatible only with the VRC03 heavy chain (Fig. 4A and table S13). Similarly, despite ~50% differences in sequence identity (table S4), the VRC-CH31 (donor 0219) heavy and light chains were able to functionally complement most of the other antibodies (Fig. 4A and table S13).

Identification of VRC01-like antibodies by deep sequencing of donors 45 and 74

To study the antibody repertoire in these individuals, we performed deep sequencing of cDNA from donor 45 PBMC (33). Because the variable regions of heavy and light chains are roughly 400 nucleotides in length, 454 pyrosequencing methods, which allow read lengths of 500 nucleotides, were used for deep sequencing. We first assessed heavy-chain sequences from a 2008 PBMC sample from donor 45, the same time point from which antibodies VRC01, VRC02, and VRC03 were isolated by RSC3 probing of the memory B cell population (18). mRNA from 5 million PBMC was used as the template for polymerase chain reaction (PCR) to preferentially amplify the IgG and IgM genes from the IGHV1 family. The 454 pyrosequencing provided 221,104 sequences, of which 33,386 encoded heavy-chain variable domains that encompassed the entire V(D)J region (Appendix 1). To categorize the donor 45 heavy-chain sequence information, we chose characteristics particular to the heavy chains of VRC01 and VRC03 as filters: (i) sequence identity, (ii) IGHV gene allele origin, and (iii) sequence divergence from the germline IGHV gene as a result of affinity maturation (Fig. 4B). Specifically, we divided sequences in- to IGHV1-2*02 allelic origin (4597 sequences) and non-IGHV1-2*02 origin (28,789 sequences) and analyzed divergence from inferred germline genes and sequence identity to the template antibodies VRC01 and VRC03 (Fig. 4B). Interestingly, no sequence of higher than 75% identity to the VRC01 or VRC02 heavy chain was found (Fig. 4B and fig. S8), although 109 sequences of greater than 90% sequence identity to VRC03 were found, and all were of IGHV1-2*02 origin. These heavy-chain sequences formed a well-segregated cluster on a contour plot (Fig. 4B, top middle panel). We next assessed the biological

function of two randomly selected heavy-chain sequences from this cluster. Chimeric antibodies were made by pairing each of the two heavy-chain sequences with the VRC03 light chain (table S14). In both cases, potent neutralization was observed, with neutralization similar to the original VRC03 antibody (Fig. 4E and table S15) (34).

A similar heavy-chain deep-sequencing analysis was performed with donor 74 PBMC from the same 2008 time point from which VRC-PG04 and VRC-PG04b were isolated. In the initial analysis, despite obtaining 263,764 sequences, of which 85,851 encompassed the full V(D)J regions of the heavy chain and 93,112 were unique, no sequences of greater than 75% nucleotide identity to VRC-PG04 were found (fig. S10 and appendix 2). Because the number of unique heavy-chain mRNAs present in the PBMC sample was likely much larger than the number of unique sequences obtained in the initial analysis, we repeated the deep sequencing of this sample with an increased number of 454 pyrosequencing reads and with protocols that optimized read length. In this analysis, 110,386 sequences of IGHV1-2*02 origin and 606,047 sequences of non-IGHV1-2*02-origin were found to encompass the V(D)J region of the heavy chain, a 10-fold increase in sequencing depth. Among these sequences, 4920 displayed greater than 75% nucleotide identity to VRC-PG04 (Fig. 4B and appendix 3). Heavy-chain sequences of the IGHV1-2*02 allelic origin segregated into several clusters, one at ~25% divergence and ~85% identity to the VRC-PG04 heavy chain, and several at 25 to 35% divergence and 65%, 85%, and 95% identity to VRC-PG04 (Fig. 4B, top right). To assess the biological function of these numerous 454-identified heavy-chain sequences, we selected 56 representative sequences from the quadrant defined by high divergence (16 to 38%) and high sequence similarity (60 to 100%) to VRC-PG04 (fig. S11). The 56 sequences were synthesized and expressed with the VRC-PG04 light chain (table S19). Remarkably, many of these antibodies displayed potent HIV-1 neutralization (35), confirming that these were functional VRC-PG04-like heavy chains (Fig. 4E and table S20).

We next performed a similar analysis of the antibody light chain. Because VRC01-03 and VRC-PG04 derive from IGκV3 alleles, we used primers designed to amplify the IGκV3 gene family. We chose a donor 45 2001 time point to maximize the likelihood of obtaining light-chain sequences capable of functional complementation (36). A total of 305,475 sequences were determined, of which 87,658 sequences encompassed the V-J region of the light chain (Appendix 4). To classify the donor 45 light-chain sequences into useful subsets, we again chose biologically specific characteristics: A distinctive 2-amino acid deletion in CDR L1 and high affinity maturation (17% and 19% for VRC01 and VRC-PG04, respectively). Two such sequences with ~90% sequence identity to the VRC01 and VRC03 light chains were identified (Fig. 4C). We assessed the biological function of these two light chains after synthesis and expression in combination with the VRC01, VRC03, and VRC-PG04 heavy chains (table S21). When paired with their respective matching wild-type heavy chain to produce a full IgG, both chimeric antibodies displayed neutralization similar to the wild-type antibody (Fig. 4D and table S22).

Maturation similarities of VRC01-like antibodies in different donors revealed by phylogenetic tools

The structural convergence in gp120 recognition and the functional complementation between VRC01-like antibodies from different donors suggested similarities in their maturation processes. We therefore used well-established phylogenetic tools to assess the evolutionary relationship among sequences derived from the same precursor germline gene (37). We hypothesized that if known VRC01-like sequences from one donor were added to the analysis of sequences of another donor, the resultant “cross-donor phylogenetic” analysis might reveal similarities in antibody maturation pathways. Specifically, with such an analysis, the exogenous sequences would be expected to interpose between dendrogram

branches containing VRC01-like antibodies from the original donor's antibodyome. We performed this analysis with heavy chains because all of the probe-identified VRC01-like antibodies derived from the same heavy chain IGHV1-2*02 allele. We added the donor 74-derived VRC-PG04 and 4b and donor 0219-derived VRC-CH30-32 heavy-chain sequences to the donor 45 heavy-chain sequences of IGHV1-2*02 genomic origin and constructed a tree rooted by the predicted VRC01 unmutated germline ancestor (18). This analysis revealed that sequences of high identity to VRC03 clustered as a subtree of a common node that was also the parent to donor 74 and 0219 VRC01-like heavy-chain sequences (Fig. 5A, left). Two donor 45 sequences chosen at random from the subtree derived from this common node were shown to neutralize HIV-1, whereas 11 heavy-chain sequences from outside this node did not neutralize ($P < 0.0001$) (fig. S9).

We also assessed the donor 74-derived IGHV1-2*02 heavy-chain sequences by including probe-identified VRC01-like antibodies from donor 45 and donor 0219 in the cross-donor phylogenetic analysis. In the tree rooted by the predicted VRC-PG04 unmutated germline ancestor, 5047 sequences segregated within the donor 45 and 0219-identified subtree (Fig. 5A, right). This subtree included the actual VRC-PG04 and 04b heavy-chain sequences, 4693 sequences of $>85\%$ identity to VRC-PG04, and several hundred sequences with identities as low as 68% to VRC-PG04. To test the functional activity of heavy-chain sequences identified by this analysis, we first assessed the tree location of the 56 heavy-chain sequences that were identified and expressed from the previously described identity/divergence grid (Fig. 6A and fig. S11). To these 56 sequences, we added 7 additional sequences from the donor 74 tree and 7 non-IGHV1-2*02 sequences to enhance coverage of the cross-donor segregated sequences (Fig. 6B and fig. S12). These 70 sequences were synthesized and expressed with the VRC-PG04 light chain (Fig. 6C and table S19). Among these 70 synthesized heavy-chain sequences, 25 did not express. Of the remaining 45 reconstituted antibodies, 24 were able to neutralize HIV-1 (Fig. 6B and table S20). Remarkably, all of the neutralizing sequences segregated into the subtree identified by the exogenously added donor 45 and 0219 VRC01-like antibodies ($P = 0.0067$) (Fig. 6D).

We also applied this cross-donor segregation method to the light chains antibodyome of donor 45. The light chains from donors 74 and 0219 did not segregate with known VRC01-like light chains from donor 45 (fig. S13), likely because these three light chains do not arise from the same inferred germline sequences. This difference may also reflect the dissimilarities in focused maturation of the two chains (Fig. 3A): In the heavy chain, focused maturation occurs in the CDR H2 region (encompassed solely within the IGHV1-2*02 V_H gene from which all VRC01-like heavy chains derive), and, in the light chain, selection pressures occur in the CDR L3 region (which is a product of different types of V-J recombination).

CDR H3 lineage analysis

The 37 heavy-chain sequences that both segregated into the VRC01 neutralizing subtree and expressed when reconstituted with the VRC-PG04 light chain could be clustered into nine CDR H3 classes (Fig. 6B), with sequences in each class containing no more than five nucleotide differences in CDR H3 from other sequences in the same class (fig. S14). A detailed junction analysis of the V(D)J recombination origins of these classes suggested that eight of the nine classes arose by separate recombination events (fig. S15); two of the classes (7 and 8) differed primarily by a single three-residue insertion/deletion, Arg-Tyr-Ser, and may have arisen from a single V(D)J recombination event (fig. S15b). Three of these classes (CDR H3-1, -2, and -9) were represented only by non-neutralizing antibodies, three by a single neutralizing antibody (CDR H3-4, -5 and -6), and three by a mixtures of neutralizing and non-neutralizing antibodies (CDR H3-3, -7 and -8) (38). Although it was

not clear whether the non-neutralizing heavy-chain sequences truly lacked neutralization function or whether this phenotype was due to incompatibilities in light-chain pairing, we chose to analyze CDR H3 classes only for those in which neutralization had been confirmed.

We further analyzed donor 74 IGHV1-2*02 heavy-chain sequences to provide an overview of CDR H3 diversity relative to sequence identity and divergence (Fig. 6E), and to identify those with CDR H3 sequences identical to the CDR H3s in each of the neutralizing classes. This analysis identified four clonal lineages (CDR H3 classes 3, 6, 7, and 8), with sequences that extended to 15% or less affinity maturation (figs. S16 and S17). CDR H3 class 7 included the probe-identified antibodies VRC-PG04 and 04b (Fig. 6B). In each case, a steady accumulation of changes in both framework and CDR regions led to increased neutralization activity (39), and changes at positions 48, 52, 58, 69, 74, 82, and 94 in the V gene, among others, appeared to be selected in several lineages (fig. S16). Overall, more than 1500 unique sequences could be classified into these four CDR H3 lineages (fig. S16). Although these CDR H3 lineages were inferred from a single time point, they likely provide insight into the specific maturation pathways by which the heavy chains of VRC01-like antibodies evolve from an initial unmutated recombinant to a broadly neutralizing antibody.

J chain analysis and maturation complexities

Among the heavy-chain VRC01-like sequences identified in donors 45 and 74, a significant skewing of J chain usage was observed (Fig. 5A): In donor 45, over 87% of the cross-donor-segregated sequences use the IGHJ1*01 allele, and in donor 74, 99% of the segregated sequences use the IGHJ2*01 allele. This preferential heavy J chain usage does not appear to be a requirement for binding specificity; indeed, the use of the J1 allele in VRC01-03, the J2 allele in VRC-PG04, and the J4 allele in VRC-CH31 provide examples for the functional compatibility of at least three different IGHJ alleles in VRC01-like antibodies. In addition to preferential J chain usage, other complexities in the maturation process could be inferred from similarities in mature heavy-chain genes and differences in CDR H3 sequence. In the absence of information on the natural pairing of heavy and light chains, the antibody maturation processes underlying these complexities is difficult to infer. Nevertheless, the deep sequencing data, with thousands of CDR H3-defined maturation intermediates (fig. S16), provide sufficient information to suggest that the maturation process may involve heavy-chain revision or other mechanisms of B cell diversification (40, 41).

Antibody genomics, HIV-1 immunity, and vaccine implications

Affinity maturation that focuses a developing antibody onto a conserved site of HIV-1 vulnerability provides a mechanism to achieve broad recognition of HIV-1 gp120. Such focused evolution may be common to broadly neutralizing antibodies that succeed in overcoming the immune evasion that protects HIV-1 gp120 from humoral recognition; the multiple layers of evasion may constrain or focus the development of nascent antibodies to particular pathways during maturation.

The structure-based genomics approach described provides tools for understanding antibody maturation. We show how deep sequencing can be used to determine the repertoire of specific families of heavy- and light-chain sequences in HIV-1-infected individuals. These partial anti-bodyomes can then be interrogated for unusual properties in sequence, or in maturation, to identify antibodies for functional characterization. We demonstrate three means of sieving a large database of antibody sequences: (i) by identity to a known mAb sequence and by divergence from putative germline (identity/divergence grid analysis), (ii) by cross-donor phylogenetic analysis of maturation pathway relationships, and (iii) by CDR H3 lineage analysis. These three means of sieving can be deployed either iteratively or in

combination (Fig. 6). An important aspect of our analyses was the functional characterization of selected sequences achieved through expression and reconstitution with known VRC01-like heavy or light chains, although other means of pairing, such as by frequency analysis (42), are possible. Although neutralization has been assessed on less than 100 reconstituted antibodies, the thousands of identified heavy- and light-chain sequences provide a large data set for analysis, which should enhance our understanding of the critical features of VRC01-like antibodies. For example, the correlation of sequence variation at particular positions with neutralization should provide insight into the allowed diversity and required elements of neutralization by this family of antibodies (fig. S18).

The deep sequencing and structural bioinformatics methodologies presented here facilitate analysis of the human antibodyome (fig. S19). This genomics technology allows interrogation of the antibody responses from infected donors, uninfected individuals, or even vaccine recipients and has several implications. For example, a genomic rooted analysis of the VRC01 antibodyome with standard phylogenetic tools may reveal a general B cell maturation pathway for the production of VRC01-like antibodies. Indeed, cross-donor phylogenetic analysis (Fig. 5B) suggests that common maturation intermediates with 20 to 30 affinity maturation changes from the IGHV1-2*02 genomic precursor are found in different individuals. These intermediates give rise to mature, broadly neutralizing VRC01-like antibodies, which have about 70 to 90 changes from the IGHV1-2*02 precursor (Fig. 5B and fig. S20). If modified gp120s with affinity to the maturation intermediates represented by the nodes of the tree were to stimulate the elicitation of these intermediates, then the analysis presented here can help guide the vaccine-induced elicitation of VRC01-like antibodies. Deep sequencing provides not only a means to identify such intermediates but also a means to facilitate their detection. Overall, the application of genomic technologies to analysis of antibodies facilitates both highly sensitive feedback and an unprecedented opportunity to understand the response of the human antibodyome to infection and vaccination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

X.W., T.Z., J. Z., G.J.N., M.R., L.S., P.D.K., and J.R.M. designed research; B.Z., C.W., X.C., M.L., K.M., S.O.D., S.P., S.D.S., W.S., L.W., Y.Y., Z.Y.Y., Z.Y., NISC, and J.M. performed experiments; X.W. isolated and characterized VRC01-like antibodies by RSC3 probe, devised and prepared samples for 454 pyrosequencing, and assisted with functional characterization; T.Z. determined and analyzed structures of VRC-PG04 and VRC03 with gp120 and assisted with functional characterization; J.Z. devised and carried out computational bioinformatics on the antibodyome; M.B., J.A.C., S.H.K., N.E.S., and B.F.H. contributed donor 0219 materials; M.S., D.R.B., and W.C.K. contributed Protocol G materials, including donor 74; N.D.R. and M.C. contributed donor 45 materials; X.W., T.Z., J.Z., I.G., N.S.L., Z.Z., L.S., P.D.K., and J.R.M. analyzed the data; and L.S., G.J.N., P.D.K., and J.R.M. wrote the paper, on which all authors commented. We thank J. Almeida and D. Douek for protocols of PBMC cDNA preparation and for helpful discussions; J. Stuckey for assistance with figures; T. Wrin for sequence information on the donor 74 virus; J. Binley, D. Montefiori, L. Morris, and G. Tomaras for donor 0219 serum characterization; and all of the IAVI Protocol G team members and the Protocol G clinical investigators, specifically, G. Miir, A. Pozniak, D. McPhee, O. Manigart, E. Karita, A. Inwoley, W. Jaoko, J. DeHovitz, L.-G. Bekker, P. Pitisuttithum, R. Paris, J. Serwanga, and S. Allen. We also thank H. Sato, I. Wilson, and members of the Structural Biology Section and Structural Bioinformatics Core, Vaccine Research Center, for discussions or comments on the manuscript. Support for this work was provided by the Intramural Research Program of the Vaccine Research Center, National Institute of Allergy and Infectious Diseases, and the National Human Genome Research Institute, NIH; by grants from the International AIDS Vaccine Initiative's Neutralizing Antibody Consortium; and by the Center for HIV AIDS Vaccine Immunology grant AI 5U19 AI 067854-06 from NIH. Use of sector 22 (Southeast Region Collaborative Access Team) at the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38. Structure factors and coordinates for antibodies VRC03 and VRC-PG04 in complex with HIV-1 gp120 have been deposited with the Protein Data Bank under accession codes 3SE8 and 3SE9, respectively. We have also deposited deep

sequencing data for donors 45 and 74 (Appendices 1 to 4) used in this study to National Center for Biotechnology Information Short Reads Archives (SRA) under accession no. SRP006992. Information deposited with GenBank includes the heavy- and light-chain variable region sequences of probe-identified antibodies VRC-PG04 and VRC-PG04b (accession nos. JN159464 to JN159467), VRC-CH30, VRC-CH31, and VRC-CH32 (JN159434 to JN159439), and VRC-CH33 and VRC-CH34 (JN159470 to 159473), as well as the sequences of genomically identified neutralizers: 24 heavy chains from donor 74, 2008 (JN159440 to JN159463), two heavy chains from donor 45, 2008 (JN159474 and JN159475), two light chains from donor 45, 2001 (JN159468 and JN159469), and 1561 unique sequences associated with neutralizing CDR H3 distributions with at least one low divergent member shown in Fig. 6B and fig. S16 (JN157873 to JN159433).

References and Notes

1. Burton DR, et al. *Nat. Immunol.* 2004; 5:233. [PubMed: 14985706]
2. Pantophlet R, Burton DR. *Annu. Rev. Immunol.* 2006; 24:739. [PubMed: 16551265]
3. Kwong PD, et al. *Nature.* 2002; 420:678. [PubMed: 12478295]
4. Stamatatos L, Morris L, Burton DR, Mascola JR. *Nat. Med.* 2009; 15:866. [PubMed: 19525964]
5. Sather DN, et al. *J. Virol.* 2009; 83:757. [PubMed: 18987148]
6. Simek MD, et al. *J. Virol.* 2009; 83:7337. [PubMed: 19439467]
7. Doria-Rose NA, et al. *J. Virol.* 2010; 84:1631. [PubMed: 19923174]
8. Gnanakaran S, et al. *PLOS Comput. Biol.* 2010; 6:e1000955. [PubMed: 20949103]
9. Gray ES, et al. *J. Virol.* 2009; 83:8925. [PubMed: 19553335]
10. Muster T, et al. *J. Virol.* 1993; 67:6642. [PubMed: 7692082]
11. Zwick MB, et al. *J. Virol.* 2001; 75:10892. [PubMed: 11602729]
12. Walker LM, et al. Protocol G Principal Investigators. *Science.* 2009; 326:285. [PubMed: 19729618]
13. Bonsignori M, et al. *J. Virol.* 2011; 85 10.1128/JVI.05045-11.
14. Trkola A, et al. *J. Virol.* 1996; 70:1100. [PubMed: 8551569]
15. Walker LM, et al. *Nature.* 2011 10.1038/nature10373.
16. Burton DR, et al. *Science.* 1994; 266:1024. [PubMed: 7973652]
17. Corti D, et al. *PLoS ONE.* 2010; 5:e8805. [PubMed: 20098712]
18. Wu X, et al. *Science.* 2010; 329:856. [PubMed: 20616233]
19. Zhou T, et al. *Science.* 2010; 329:811. [PubMed: 20616231]
20. Wrammert J, et al. *J. Exp. Med.* 2011; 208:181. [PubMed: 21220454]
21. Huber M, et al. *J. Virol.* 2010; 84:10700. [PubMed: 20702640]
22. Xiao X, et al. *Biochem. Biophys. Res. Commun.* 2009; 390:404. [PubMed: 19748484]
23. Materials and methods are available as supporting material on *Science* Online.
24. Scheid JF, et al. *Nature.* 2009; 458:636. [PubMed: 19287373]
25. Lerner RA. *Mol. Biosyst.* 2011; 7:1004. [PubMed: 21298133]
26. Huang CC, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101:2706. [PubMed: 14981267]
27. Sabin C, et al. *PLoS Pathog.* 2010; 6:e1001195. [PubMed: 21124990]
28. Breden F, et al. *PLoS ONE.* 2011; 6:e16857. [PubMed: 21479208]
29. Gorny MK, et al. *Mol. Immunol.* 2009; 46:917. [PubMed: 18952295]
30. Chen L, et al. *Science.* 2009; 326:1123. [PubMed: 19965434]
31. Significant correlations were observed between RMSDs of VRC01-like antibody interaction with gp120 and size of CDR interaction but not of surface area in general (fig. S6).
32. Zhou T, et al. *Nature.* 2007; 445:732. [PubMed: 17301785]
33. The mRNA was extracted from 20 million PBMC, reverse transcribed with oligo (dT)12-18, and a quarter of the resultant cDNA (equivalent to the transcripts of 5 million PBMC) was used as the template for PCR to preferentially amplify the IGHV1 gene family from both the IgG- and IgM-expressing cells. PCR products were gel purified and analyzed by 454 pyrosequencing.
34. We also assessed 454-derived sequences for structural compatibility with the VRC01, VRC03, and VRC-PG04 gp120-complex crystal structures using a threading algorithm that assessed structural

compatibility using the statistical potential based on distance-scaled, finite ideal-gas reference state (DFIRE) (43). None of the 10 sequences with optimal DFIRE scores (table S16), nor those with high germline divergence of non-IGHV1-2*02 genomic origin (table S17) displayed neutralization when reconstituted with the VRC01 light chain (Fig. 4E, fig. S9, and table S18). Thus, sequence similarity, IGHV1-2*02 origin, and divergence all correlate with neutralization potential, but other factors such as predicted structural compatibility failed to identify VRC01-like antibodies.

35. Over half of the reconstituted antibodies displayed a mean IC₅₀ of <0.5 µg/ml, a level of potency similar to that observed with antibodies reconstituted with the probe-identified mAb (VRC01, VRC03, or VRC-CH31) heavy chain paired with the VRC-PG04 light chain.
36. (i) VRC03L does not complement well with other heavy chains; (ii) VRC03 H was readily found among donor 45 2008 sequences; (iii) VRC01 and VRC02 H were not found among donor 45 2008 sequences; (iv) VRC01-03 were isolated from the memory B cell population. Results (i) to (iv) suggest that VRC03 came after VRC01; we therefore chose a pre-2008 time point to maximize chances of obtaining light chains that allowed for functional complementation with known VRC01-like heavy chains.
37. Although phylogenetic analysis is often used to study the evolution of a family of sequences and to understand the relationships between ancestral sequences and their descendants, we appreciate that there are some unique aspects to antibody evolution. Due to the nature of activation-induced cytidine deaminase (AID) activity, antibodies accumulate mutations at hot spots (CDRs) and thus do not occur in a stochastic manner throughout the antibody genome. Also, the process of V(D)J recombination introduces nucleotide insertions and deletions that alter germline DNA sequence. Our goal here was to elucidate the ontogeny of recombined antibody sequences in order to identify intermediate sequences related to mature neutralizing antibodies. We therefore used well-established maximum likelihood phylogenetic algorithms to analyze antibody sequence data and to build rooted trees of antibody sequences that are derived from a common ancestor (i.e., same VH-germline gene).
38. Several of the non-neutralizing heavy-chain sequences shown in the CDR H3 distribution of Fig. 6 are likely the result of PCR template switching. The single heavy chain depicted in the CDR H3 class 1 contour plot contains a unique CDR H3 sequence (fig. S15a), but with a V gene that displays high similarity to class 3 sequences (table S23). The same observation occurs for the two sequences in the class 2 contour plot. Also, the highly divergent (and outlier) sequence on the CDR H3 class 9 distribution plot contains the same CDR H3 as the other 140 class 9 sequences, but with a V gene that closely matches sequences found in class 8 (table S23b). Because only a few of more than 1500 unique sequences identified by CDR H3 analysis showed dissimilar V genes, and all of these appeared as single or double outliers, template switching can occur but appears to be rare. This rarity is also suggested by an analysis of 606,047 non-IGHV1-2*02 from donor 74 for sequences with the CDR H3s identified in Fig. 6B, which finds less than 100 sequences, of which the majority corresponds to the likely misassigned cluster in the non-IGHV1-2*02 sequence of donor 74 in Fig. 4, as described in (44).
39. A similar accumulation of somatic mutations was shown (45) with the broadly neutralizing antibodies PG9 and PG16 to correlate with an increase in neutralization breadth and potency.
40. Nemazee D, Weigert M. J. Exp. Med. 2000; 191:1813. [PubMed: 10839798]
41. Edry E, Melamed D. J. Immunol. 2004; 173:4265. [PubMed: 15383554]
42. Glanville J, et al. Proc. Natl. Acad. Sci. U.S.A. 2009; 106:20216. [PubMed: 19875695]
43. Zhou H, Zhou Y. Protein Sci. 2002; 11:2714. [PubMed: 12381853]
44. The peak at ~25% IGHV1-2*02 divergence and 88% identity was also seen in the sequence plot for sequences of non-IGHV1-2*02 origin. Cross-donor and CDR H3 analyses shows that these putative non-IGHV1-2*02– derived sequences segregate with VRC01-like antibodies in dendrograms and have CDR H3s that are identical to confirmed VRC01-like antibodies (fig. S16), indicating that sequences in the non-IGHV1-2*02 cluster are likely misassigned and actually of IGHV1-2*02 origin.
45. Pancera M, et al. J. Virol. 2010; 84:8098. [PubMed: 20538861]
46. Kabat, EA.; Wu, TT.; Gottesman, KS.; Foeller, C. Sequences of Proteins of Immunological Interest. 5th Edition. U.S. Department of Health and Human Services; Washington, DC: 1991.
47. Krissinel E, Henrick K. J. Mol. Biol. 2007; 372:774. [PubMed: 17681537]

Fig. 1.

Identification and characterization of mAbs from HIV-1– infected donors 74 and 0219. **(A)** RSC3 analysis of serum. Twelve sera from the IAVI Protocol G cohort and one serum from the CHAVI 001 cohort (donor 0219) were analyzed for RSC3 reduction in serum neutralization on HIV-1 strains JR-FL, PVO.4, YU2, and ZA12.29. Blue bars show the mean serum reduction in neutralization IC₅₀ resulting from RSC3 versus ΔRSC3 competition. Sera with the greatest reduction were further analyzed on HIV-1 strains Q168.a2, RW020.2, Du156.12, and ZM109.4. Red bars show the mean reduction on eight viruses. **(B)** Flow cytometric identification of RSC3-reactive IgG⁺ B cells from donors 74 and 0219. Gating and percentage of IgG⁺ B cells of interest (RSC3⁺ΔRSC3⁻) are indicated, with 40 and 26 sorted single B cells from donors 74 and 0219, respectively. Additional sorting details are shown in fig. S1. **(C)** Protein sequences of heavy- and light-chain variable regions of mAbs VRC-PG04 and VRC-PG04b, isolated from donor 74, and mAbs VRC-CH30-34 isolated from donor 0219. Sequences are aligned to the putative germline ancestral genes and to previously identified broadly neutralizing antibodies VRC01, VRC02, and VRC03. Framework regions (FR) and complementarity-determining regions (CDRs) are based on Kabat nomenclature (46). **(D)** Neutralization dendrograms. VRC-PG04 and VRC-CH31 were tested against genetically diverse Env-pseudoviruses representing the major HIV-1 clades. Neighbor-joining trees display the protein distance of gp160 sequences from 178 HIV-1 isolates tested against VRC-PG04 and a subset (80 isolates) tested against VRC-CH31. A scale bar denotes 1% distance in amino acid sequence. Tree branches are colored by the neutralization potencies of VRC-PG04 and VRC-CH31 against each particular virus.

Fig. 2.

Structure of antibodies VRC-PG04 and VRC03 in complex with HIV-1 gp120. **(A)** Overall structures. The liganded complex for the Fab of antibody VRC-PG04 from donor 74 and the HIV-1 gp120 envelope glycoprotein from isolate 93TH057 is depicted with polypeptide backbones in ribbon representation in the left image. The complex of Fab VRC03 from donor 45 is depicted in the right image, with surfaces of all variable domain residues that differ between VRC03 and VRC-PG04 colored according to their chemical characteristics. **(B and C)** Interaction close-ups. Critical interactions are shown between the CD4-binding loop of gp120 (purple) and the CDR H2 region of the broadly neutralizing mAbs VRC03 and VRC-PG04 (reported here) and VRC01 [reported previously (19)], with hydrogen bonds depicted as dotted lines. The 1.9 and 2.1 Å resolution structures of VRC03 and VRC-PG04, respectively, were sufficient to define interfacial waters shown in (C), which were unclear in the 2.9 Å structure of VRC01. The orientation shown in (C) is $\sim 180^\circ$ rotated about the vertical axis from the orientation shown in (B).

Fig. 3.

Focused evolution of VRC01-like antibodies. **(A)** Antibody convergence. The gp120 portions of liganded complexes with VRC01, VRC03, and VRC-PG04 were superimposed to determine the average antibody per residue C α deviation, and the per residue hydrophobic interaction (Δ^iG) was calculated (47). These two quantities were found to correlate ($P = 0.0427$), with antibody residues containing strong hydrophobic interactions (e.g., at positions 53 and 55 in the heavy chain, and 91 and 97 in the light chain, VRC-PG04—relative numbering) displaying high structural conservation. This correlation is visualized on VRC-PG04 in the left image, where the ribbon thickness is proportional to the corresponding per residue C α deviation and the paratope surface is colored according to hydrophobicity, from white (low) to red (high); notably, red surface patches map to thin ribbons. **(B)** Epitope convergence. The HIV-1 gp120 surface involved with CD4 binding contains conformationally invariant regions (e.g., associated with the outer domain) and conformationally variable regions (e.g., associated with the bridging sheet). We previously hypothesized that the conformationally invariant outer domain contact for CD4 represents a site of vulnerability (19). We analyzed the precision of CD4-binding-site ligand recognition (vertical axis) versus the IC₈₀ neutralization breadth (horizontal axis) and observed significant correlation ($R^2 = 0.6$, $P = 0.040$). **(C)** Divergences in sequence and convergences in recognition. The development of VRC01-like antibodies involves a heavy chain derived from the IGHV1-2*02 allele and selected light-chain V κ alleles. The far left image depicts a ribbon representation model of a putative germline antibody. Somatic hypermutation during the process of affinity maturation leads to a divergence in sequence, yet results in the convergent recognition of similar epitopes. Intersection of the epitope surfaces recognized by VRC01, VRC03, and VRC-PG04 (far right image), reveals a notable similarity to the site of vulnerability. The primary divergence of this intersection from the hypothesized site of vulnerability occurs in the region of HIV-1 gp120 recognized by the light chain of the VRC01-like antibodies. Although the separate epitopes on gp120 do show differences in recognition surface, these primarily involve the bridging sheet region, which is likely to adopt a different conformation in the functional viral spike before engagement of CD4.

Fig. 4.

Deep sequencing of expressed heavy and light chains from donors 45 and 74. **(A)** Heavy- and light-chain complementation. The neutralization profiles of VRC01 and VRC03 (donor 45), VRC-PG04 (donor 74), and VRC-CH31 (donor 0219) and their heavy- and light-chain chimeric swaps are depicted with 20-isolate neutralization dendrograms. Explicit neutralization IC_{50} s are provided in table S13. **(B)** The repertoire of heavy-chain sequences from donor 45 (2008 sample) and donor 74 (2008 sample). Heavy-chain sequences are plotted as a function of sequence identity to the heavy chain of VRC01 (left), VRC03 (middle), and VRC-PG04 (right) and of sequence divergence from putative genomic V_H alleles: Upper row plots show sequences of putative IGHV1-2*02 allelic origin; lower row plots show sequences from other allelic origins. Color coding indicates the number of sequences. **(C)** Repertoire of expressed light-chain sequences from donor 45 (2001 sample). Light-chain sequences are plotted as a function of sequence identity to VRC01 (left) and VRC03 (right) light chains, and of sequence divergence from putative genomic V-gene alleles. Sequences with two-residue deletions in the CDR L1 region (which is observed in VRC01 and VRC03) are shown as black dots. Two light-chain sequences, with 92.0% identity to VRC01 (sequence ID 181371) and with 90.3% identity to VRC03 (sequence ID 223454) are highlighted with red triangles. **(D)** Functional assessment of light-chain sequences identified by deep sequencing. The neutralization profiles of sequence 181371 reconstituted with the VRC01 heavy chain (named gVRC-L1_{d45}) and of sequence 223454 reconstituted with the VRC03 heavy chain (named gVRC-L2_{d45}) are depicted with 20-isolate neutralization dendrograms; explicit neutralization IC_{50} s are shown in table S22. **(E)** Functional assessment of heavy-chain sequences identified by deep sequencing. Heavy-chain sequences from donors 45 and 74 were synthesized and expressed with either the light chain of VRC01 or VRC03 (for donor 45) or the light chain of VRC-PG04 (for donor 74) and evaluated for neutralization. Neutralizing sequences are shown as red stars and are labeled. gVRC-H(n)_{d74} refers to the heavy chains with confirmed neutralization when reconstituted with the light chain of VRC-PG04, with controls as described in (34).

Fig. 5. Maturation similarities of VRC01-like antibodies in different donors revealed by cross-donor phylogenetic analysis. **(A)** Maximum-likelihood trees of heavy-chain sequences of the IGHV1-2*02 origin from donor 45 (left) and donor 74 (right). The subset of sequences shown was selected based on the germline divergence as described in (23). The donor 45 tree is rooted by the putative reverted unmutated ancestor of the heavy chain of VRC01 and also includes specific neutralizing sequences from donors 74 and 0219 (shown in red). Similarly, the donor 74 tree is rooted in the putative reverted unmutated ancestor of the heavy chain of VRC-PG04, and sequences from donors 45 and 0219 are included in the cross-donor phylogenetic analysis. Bars representing 0.1 changes per nucleotide site are shown. Insets show J chain assignments for all sequences within the neutralizing subtree identified by an iterative neighbor-joining tree analysis as described in (23). **(B)** Phylogenetically inferred maturation intermediates. Backbone ribbon representations are shown for HIV-1 gp120 (red) and the heavy-chain variable domains (green). Critical intermediates inferred from the phylogenetic tree in (A) are labeled I_{d45}, II_{d45}, III_{d45}, I_{d74}, and II_{d74}. The number of V_H-gene mutations is provided (e.g., for the 23 mutations associated with the first intermediation of donor 45, “I_{d45}: 23”), and the location of these is highlighted in the surface representation and colored according to their chemistry.

Fig. 6.

Analysis of the heavy-chain antibodyome of donor 74 and identification of heavy chains with HIV-1 neutralizing activity. Identity/divergence-grid analysis, cross-donor phylogenetic analysis, and CDR H3 analysis were coupled to functional characterization of selected heavy-chain sequences. This provides a means for identification of novel heavy chains with HIV-1 neutralizing activity. **(A)** Identity/divergence-grid analysis. The location of the 63 synthesized IGHV1-2*02 heavy chains from donor 74 is shown, including neutralizing (red stars) and non-neutralizing (black stars) sequences. **(B)** Cross-donor phylogenetic analysis and CDR H3 lineage analysis. A maximum-likelihood tree of the 70 synthesized heavy-chain sequences (including 7 non-IGHV1-2*02 sequences) is rooted at the putative reverted unmutated ancestor of VRC-PG04. The probe-identified VRC-PG and VRC-CH antibodies are shown in red text along with the 24 genomically identified heavy-chain sequences, gVRC-H(1-24)_{d74}, which were found to neutralize HIV-1 when reconstituted with the light chain of VRC-PG04. Grid locations and CDR H3 classes are specified for neutralizing and non-neutralizing sequences. Within each CDR H3 class, all sequences with identical CDR H3s are highlighted in orange in the far right grids (with the number of total sequences corresponding to each CDR H3 class shown). **(C)** Expression levels of selected heavy chains reconstituted with the light chain of VRC-PG04 versus breadth of neutralization. **(D)** Neutralization potency of reconstituted cross-donor phylogeny-predicted antibodies on seven HIV-1 isolates. **(E)** CDR H3 analysis of donor 74 heavy-chain sequences. For each of the 110,386 sequences derived from the IGHV1-2*02 allele, the CDR H3 was determined, and its percent identity to that of the VRC-PG04 heavy chain was color coded as shown and graphed. The sequences with high CDR H3 identity to VRC-PG04 reside in regions of high overall heavy-chain sequence identity, even for sequences with a low divergence from IGHV1-2*02.