Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus

Hua–Xin Liao1,2*, Rebecca Lynch3, Tongting Zhou4,5* Feng Gao4–6, S. Munir Alam1,2, Scott D. Boyd1, Andrew Z. Fure4, Krishna M. Roskin7, Chaim A. Schramm7, Zhenhai Zhang5, Jian Zhu5, Lawrence Shapiro3,1,5, NICe Comparative Sequencing Program, James C. Mullikin,1,2, David C. Montefiori1,2, Robert Parks1,2, Krissey E. Lloyd1,2, Richard M. Soerence1,2, Kelly A. Soderberg2,2, Myron Cohen8, Gift Kamanga9, Mark K. Louder1, Lillian M. Tran1, Yue Chen1,1,2, Fangping Cai1,2, Sherry Chen1,2, Stephanie Moquin1, Xiulan Du1, M. Gordon Joyce1, Sanjay Srivatsan3, Baoshan Zhang8, Anqi Zheng8, George M. Shaw11, Beatrice H. Hahn11, Thomas B. Kepler12, Bette T. M. Korber8, Peter D. Kwong3, John R. Mascola1 & Barton F. Haynes1,2

Current human immunodeficiency virus-1 (HIV-1) vaccines elicit strain–specific neutralizing antibodies. However, cross-reactive neutralizing antibodies arise in approximately 20% of HIV-1–infected individuals, and details of their generation could provide a blueprint for effective vaccination. Here we report the isolation, evolution and structure of a broadly neutralizing antibody from an African donor followed from the time of infection. The mature antibody, CH103, neutralized approximately 55% of HIV-1 isolates, and its co-crystal structure with the HIV-1 envelope protein gp120 revealed a new loop-based mechanism of CD4–binding-site recognition. Virus and antibody gene sequencing revealed concomitant virus evolution and antibody maturation. Notably, the unmutated common ancestor of the CH103 lineage avidly bound the transmitted/founder HIV-1 envelope glycoprotein, and evolution of antibody neutralization breadth was preceded by extensive viral diversification in and near the CH103 epitope. These data determine the viral and antibody evolution leading to induction of a lineage of HIV-1 broadly neutralizing antibodies, and provide insights into strategies to elicit similar antibodies by vaccination.

Induction of HIV-1 envelope (Env) broadly neutralizing antibodies (BnAbs) is a key goal of HIV-1 vaccine development. BnAbs can target conserved regions that include conformational glycans, the gp41 membrane proximal region, the V1/V2 region, glycan-associated C3/V3 on gp120, and the CD4-binding site in gp120. Most mature BnAbs have one or more unusual features (long third complementarity-determining region of the heavy chain (HCDR), polyreactivity for non-HIV-1 antigens, and high levels of somatic mutations), suggesting substantial barriers to their elicitation. However, in particular, CD4-binding site BnAbs have extremely high levels of somatic mutation, suggesting complex or prolonged maturation pathways. Moreover, it has been difficult to find Env proteins that bind with high affinity to BnAb germline or unmutated common ancestors (UCAs), a trait that would be desirable for candidate immunogens for induction of BnAbs. Although it has been shown that Env proteins bind to UCAs of BnAbs targeting the gp41 membrane proximal region, and to UCAs of some V1/V2 BnAbs, so far, heterologous Env proteins have not been identified that bind the UCAs of CD4-binding site BnAb lineages, although they should exist.

Eighty per cent of heterologous HIV-1 infections are established by one transmitted/founder virus. The initial neutralizing antibody response to this virus arises approximately 3 months after transmission and is strain-specific. The antibody response to the transmitted/founder virus drives viral escape, such that virus mutants become resistant to neutralization by autologous plasma. This antibody–virus race leads to poor or restricted specificities of neutralizing antibodies in ~80% of patients; however in ~20% of patients, evolved variants of the transmitted/founder virus induce antibodies with considerable neutralization breadth, such as BnAbs.

There are several potential molecular routes by which antibodies to HIV-1 may evolve, and indeed, types of antibody with different neutralizing specificities may follow different routes. Because the initial autologous neutralizing antibody response is specific for the transmitted/founder virus, some transmitted/founder Env proteins might be predisposed to binding the germ line or UCA of the observed BnAb in those rare patients that make BnAbs. Thus, although neutralizing breadth generally is not observed until chronic infection, a precise understanding of the interaction between virus evolution and maturing BnAb lineages in early infection may provide insight into events that ultimately lead to BnAb development. BnAbs studied so far have only been isolated from individuals who were sampled during chronic infection. Thus, the evolutionary trajectories of virus and antibody from the time of virus transmission to the development of broad neutralization remain unknown.

We and others have proposed vaccine strategies that begin by targeting UCAs, the putative naive B-cell receptors of BnAbs with relevant Env immunogens to trigger antibody lineages with potential ultimately to develop breadth.
with Env proteins specifically selected to stimulate somatic mutation pathways that give rise to BnAbs. Both aspects of this strategy have proved challenging owing to a lack of knowledge of specific Env proteins capable of interacting with UCAs and early intermediate antibodies of BnAbs.

Here we report the isolation of the CH103 CD4-binding site BnAb clonal lineage from an African patient, CH505, who was followed from acute HIV-1 infection to BnAb development. We show that the CH103 BnAb lineage is less mutated than most other CD4-binding site BnAbs, and may be first detectable as early as 14 weeks after HIV-1 infection. Early autologous neutralization by antibodies in this lineage triggered virus escape, but rapid and extensive Env evolution in and near the epitope region preceded the acquisition of plasma antibody neutralization breadth defined as neutralization of heterologous viruses. Analysis of the co-crystal structure of the CH103 Fab fragment and a gp120 core demonstrated a new loop-binding mode of antibody neutralization.

Isolation of the CH103 BnAb lineage

The CH505 donor was enrolled in the CHAVI001 acute HIV-1 infection cohort93 approximately 4 weeks after HIV-1 infection (Supplementary Fig. 1) and followed for more than 3 years. Single genome amplification of 53 plasma viral Env gp160 RNAs24 from 4 weeks after transmission identified a single clade C transmitted/founder virus. Serological analysis demonstrated the development of autologous neutralizing antibodies at 14 weeks, CD4-binding site antibodies that bound to a recombinant Env protein (resurfaced stabilized core 3 (RSC3)) at 53 weeks, and evolution of plasma cross-reactive neutralizing activity from 41–92 weeks after transmission96 (Fig. 1, Supplementary Table 1 and Supplementary Fig. 2). The natural variable regions of heavy-chain (VH) and light-chain (VL) gene pairs of antibodies CH103, CH104 and CH106 were isolated from peripheral blood mononuclear cells (PBMCs) at 136 weeks after transmission by flow sorting of memory B cells that bound RSC3 Env protein5,13,36 (Fig. 1b). The V(D)J gene of antibody CH105 was similarly isolated, but no VL gene was identified from the same cell. Analysis of characteristics of V(D)J (VH4–59, posterior probability (PP) = 0.99; D3–16, PP = 0.74; JI4, PP = 1.00) and VL (VJ2–3, PP = 1.00; J1, PP = 1.00) rearrangements in monoclonal antibodies CH103, CH104, CH105 and CH106 demonstrated that these antibodies were representatives of a single clonal lineage that we designated as the CH103 clonal lineage (Fig. 2 and Supplementary Table 2).

Neutralization assays using a previously described37 panel of 196 geographically and genetically diverse Env-pseudoviruses representing the major circulating genetic subtypes and circulating recombinant forms demonstrated that CH103 neutralized 55% of viral isolates, with a geometric mean half-maximum inhibitory concentration

![Figure 2](image-url)
(IC_{50}) of 4.54 mg ml^{-1} among sensitive isolates (Fig. 1c and Supplementary Table 3). Enzyme-linked immunosorbent assay (ELISA) cross-competition analysis demonstrated that CH103 binding to gp120 was competed by known CD4-binding site ligands such as monoclonal antibody VRC01 and the chimeric protein CD4-Ig (Fig. 1d). CH103 binding to RSC3 Env was also substantially diminished by gp120, with Pro363Asn and Δ371lle mutations known to reduce the binding of most CD4-binding site monoclonal antibodies\(^5,30\) (Supplementary Fig. 3).

Molecular characterization of the CH103 BnAb lineage

The RSC3 probe isolated CH103, CH104, and CH106 BnAbs by single-cell flow sorting. The CH103 clonal lineage was enriched by V_{iD}J_{H} and V_{i}J_{L} sequences identified by pyrosequencing PBMC DNA\(^{34,38}\) obtained 66 and 140 weeks after transmission, and complementary DNA antibody transcripts\(^6\) obtained 6, 14, 53, 92 and 144 weeks after transmission. From pyrosequencing of antibody gene transcripts, we found 457 unique heavy- and 171 unique light-chain clonal members (Fig. 2a, b). For comprehensive study, a representative 14-member BnAb pathway was reconstructed from V_{iD}J_{H} sequences (1AH92U, IAJZET and 1A102R) recovered by pyrosequencing, and V_{iD}J_{H} genes of the inferred intermediate (I) antibodies (I_{1–4, 17})\(^{1,16,34}\) (T. B. Kepler, manuscript submitted; http://arxiv.org/abs/1303.0424) that were paired and expressed with either the UCA or I2 V_{i}J_{L} depending on the genetic distance of the V_{iD}J_{H} to either the UCA or mature antibodies (Fig. 2c and Supplementary Table 2). The mature CH103, CH104 and CH106 antibodies were paired with their natural V_{i}J_{L}. The CH105 natural V_{i}J_{L} isolated from RSC3 memory B-cell sorting was paired with the V_{i}J_{L} of I2.

Whereas the V_{iD}J_{H} mutation frequencies (calculated as described in the Methods) of the published CD4-binding site BnAbs VRC01, CH31 and NIH45–46 are 30–36% (refs 5–7, 22, 39), the V_{iD}J_{H} frequencies of CH103 lineage CH103, CH104 and CH105 and CH106 are 13–17% (Fig. 2c). Furthermore, antibodies in the CH103 clonal lineage do not contain the large (>3 nucleotides) insertion or deletion mutations common in the VRC01 class of BnAbs\(^1\),\(^3\) with the exception of the V_{i}J_{L} of CH103, which contained a three amino-acid light-chain complementarity-determining region 1 (LCDR1) deletion.

It has been proposed that one reason that CD4-binding site BnAbs are difficult to induce is because heterologous HIV-1 Env proteins do not bind their UCAs\(^1\),\(^18\),\(^22\). We wondered, however, whether the CH505 transmitted/founder Env, the initial driving antigen for the CH103 BnAb lineage, would preferentially bind to early CH103 clonal lineage members and the UCA compared to heterologous Env proteins. Indeed, a heterologous gp120 transmitted/founder Env, subtype B B.63521 (B.63521), did not bind to the CH103 UCA (Fig. 2d) but did bind to later members of the clonal lineage. Affinity for this heterologous Env protein increased four orders of magnitude during somatic evolution of the CH103 lineage, with maximal dissociation constant (K_{D}) values of 2.4–7.0 nM in the mature CH103–CH106 monoclonal antibodies (Fig. 2d). The CH103 UCA monoclonal antibody did not bind to heterologous transmitted/founder Env proteins AE.427299, B.9021 and C.1086 (Supplementary Table 4), confirming lack of heterologous Env binding to CD4-binding site UCAs. Moreover, the gp120 Env RSC3 protein was also not bound by the CH103 UCA and earlier members of the clonal lineage (Supplementary Fig. 3a), and no binding was seen with RSC3 mutant proteins known to disrupt CD4-binding site BnAb binding (Supplementary Fig. 3b).

In contrast to heterologous Env proteins, the CH505 transmitted/founder Env gp140 bound well to all of the candidate UCAs (Supplementary Table 5), with the highest UCA affinity of K_{D} = 37.5 nM. In addition, the CH505 transmitted/founder Env gp140 was recognized by all members of the CH103 clonal lineage (Fig. 2d). Whereas affinity to the heterologous transmitted/founder Env B.63521 increased by more than four orders of magnitude as the CH103 lineage matured, affinity for the CH505 transmitted/founder Env increased by more than tenfold (Fig. 2d). To demonstrate Env escape from CH103 lineage members directly, autologous recombinant gp140 Env proteins isolated at weeks 30, 53 and 78 after infection were expressed and compared with the CH505 transmitted/founder Env for binding to the BnAb arm of the CH103 clonal lineage (Supplementary Table 6 and Supplementary Fig. 4). Escape-mutantEnv proteins could be isolated that were progressively less reactive with the CH103 clonal lineage members. Env proteins isolated at weeks 50, 53 and 78 lost UCA reactivity and only bound intermediate antibodies 3, 2 and 1, as well as BnAbs CH103, CH104, CH105 and CH106 (Supplementary Table 6). In addition, two Env escape mutants from week-78 viruses also lost either strong reactivity to all intermediate antibodies or all lineage members (Supplementary Table 6).

To quantify CH103 clonal variants from initial generation to induction of broad and potent neutralization, we used pyrosequencing of antibody cDNA transcripts from five time points, weeks 6, 14, 53, 92 and 144 after transmission (Supplementary Table 7). We found two V_{iD}J_{H} chains closely related to, and possibly members of, the CH103 clonal lineage (Fig. 2a, Supplementary Table 7). Moreover, one of these V_{iD}J_{H} chains when reconstructed in a full IgG1 backbone and expressed with the UCA V_{i}J_{L} weakly bound the CH505 transmitted/founder Env gp140 at an end-point titre of 11 μg ml^{-1} (Fig. 2a). These reconstructed antibodies were present concomitant with CH505 plasma autologous neutralizing activity at 14 weeks after transmission (Supplementary Fig. 2). Antibodies that bound the CH505 transmitted/founder Env were present in plasma as early as 4 weeks after transmission (data not shown). Both CH103 lineage V_{iD}J_{H} and V_{i}J_{L} sequences peaked at week 53, with 230 and 83 unique transcripts, respectively. V_{i}J_{L} clonal members fell to 46 at week 144, and V_{i}J_{L} members dropped to 76 at week 144.

Polyreactivity is a common trait of BnAbs, suggesting that the generation of some BnAbs may be controlled by tolerance mechanisms\(^{10,21,40}\). Conversely, polyreactivity can arise during the somatic evolution of B cells in germinal centres as a normal component of B-cell development\(^41\). The CH103 clonal lineage was evaluated for polyreactivity as measured by HEp-2 cell reactivity and binding to a panel of autoantigens\(^4\). Although earlier members of the CH103 clonal lineage were not polyreactive by these measures, polyreactivity was acquired together with BnAb activity by the intermediate antibody I2, I1 and clonal members CH103, CH104 and CH106 (Supplementary Fig. 5a, b). The BnAbs CH106 and intermediate antibody I1 also demonstrated polyreactivity in protein arrays with specific reactivity to several human autoantigens, including elongation factor-2 kinase and ubiquitin-protein ligase E3A (Supplementary Fig. 5c, d).

Structure of CH103 in complex with HIV-1 gp120

Crystals of the complex between the CH103 Fab fragment and the ZM176.66 strain of HIV diffracted to 3.25 Å resolution, and molecular replacement identified solutions for CH103 Fab and for the outer domain of gp120 (Fig. 3a). Inspection of the CH103–gp120 crystal lattice (Supplementary Fig. 6) indicated that the absence of the gp120 inner domain was probably related to proteolytic degradation of the extended gp120 core to an outer domain fragment. Refinement to a R_{work}/R_{free} ratio of 19.6%/25.6% (Supplementary Table 8) confirmed a lack of electron density for gp120 residues amino-terminal to gp120 residue Val 255 or carboxy-terminal to Gly 472 (gp120 residues are numbered according to standard HXB2 nomenclature), and no electron density was observed for gp120 residues 301–324 (V3), 398–411 (V4) and 421–439 (B20–21). Superposition of the ordered portions of gp120 in complex with CH103 with the fully extended gp120 core bound by antibody VRC01 (ref. 7) indicated a highly similar structure (Cz root mean squared deviation (r.m.s.d.) 1.16 Å) (Fig. 3b). Despite missing portions of core gp120, the entire CH103 epitope seemed to be present in the electron density for the experimentally observed gp120 outer domain.

The surface bound by CH103 formed an elongated patch with dimensions of ~40 × 10 Å, which stretched across the site of initial
CD4 contact on the outer domain of gp120 (Fig. 3c). The gp120 surface recognized by CH103 correlated well with the initial site of CD4 contact; of the residues contacted by CH103, only eight were not predicted to interact with CD4. CH103 interacted with these gp120 residues through side-chain contact with Ser 256 in loop D, main- and side-chain contacts with His 364 and Leu 369 in the CD4-binding loop, and main- and side-chain contacts with Asn 463 and Asp 464 in the V5 loop (Fig. 3d). Notably, residue 463 is a predicted site of N-linked glycosylation in strain ZM176.66 as well as in the autologous CH505 virus, but electron density for an N-linked glycan was not observed. Overall, of the 22 residues that monoclonal antibody CH103 was observed to contact on gp120, 14 were expected to interact with CD4 (16 of these residues with antibody VRC01), providing a structural basis for the CD4-epitope specificity of CH103 and its broad recognition (Supplementary Table 9).

Residues 1–215 on the antibody heavy chain and 1–209 on the light chain showed well-defined backbone densities. Overall, CH103 uses a CDR H3 dominated mode of interaction, although all six of the complementary-determining regions (CDRs) interacted with gp120 as well as the light-chain framework region 3 (FWR3) (Supplementary Table 9). Of the residues predicted to contact on gp120, 14 were expected to interact with CD4. CH103 interacted with these gp120 residues through side-chain contact with Ser 256 in loop D, main- and side-chain contacts with His 364 and Leu 369 in the CD4-binding loop, and main- and side-chain contacts with Asn 463 and Asp 464 in the V5 loop (Fig. 3d). Notably, residue 463 is a predicted site of N-linked glycosylation in strain ZM176.66 as well as in the autologous CH505 virus, but electron density for an N-linked glycan was not observed. Overall, of the 22 residues that monoclonal antibody CH103 was observed to contact on gp120, 14 were expected to interact with CD4 (16 of these residues with antibody VRC01), providing a structural basis for the CD4-epitope specificity of CH103 and its broad recognition (Supplementary Table 9).

Residues 1–215 on the antibody heavy chain and 1–209 on the light chain showed well-defined backbone densities. Overall, CH103 uses a CDR H3 dominated mode of interaction, although all six of the complementary-determining regions (CDRs) interacted with gp120 as well as the light-chain framework region 3 (FWR3) (Supplementary Table 7). Of the residues predicted to contact on gp120, 14 were expected to interact with CD4. CH103 interacted with these gp120 residues through side-chain contact with Ser 256 in loop D, main- and side-chain contacts with His 364 and Leu 369 in the CD4-binding loop, and main- and side-chain contacts with Asn 463 and Asp 464 in the V5 loop (Fig. 3d). Notably, residue 463 is a predicted site of N-linked glycosylation in strain ZM176.66 as well as in the autologous CH505 virus, but electron density for an N-linked glycan was not observed. Overall, of the 22 residues that monoclonal antibody CH103 was observed to contact on gp120, 14 were expected to interact with CD4 (16 of these residues with antibody VRC01), providing a structural basis for the CD4-epitope specificity of CH103 and its broad recognition (Supplementary Table 9).
diversification in these regions preceded neutralization breadth. Expanding diversification early in viral evolution (4–22 weeks after transmission; Supplementary Figs 8 and 9) coincided with autologous neutralizing antibody development, consistent with autologous neutralizing antibody escape mutations. Mutations that accumulated from weeks 41 to 78 in CH505 Env contact regions immediately preceded development of neutralizing antibody breadth (Fig. 4 and Supplementary Figs 8 and 9). By weeks 30–53, extensive within-sample diversity resulted from both point mutations in and around CH103 contact residues, and to several insertions and deletions in V1 and V5 (Supplementary Fig. 9). A strong selective pressure seems to have come into play between weeks 30 and 53, perhaps due to autologous neutralization escape, and neutralization breadth developed after this point (Fig. 4 and Supplementary Figs 8 and 9). Importantly, owing to apparent strong positive selective pressure between weeks 30 and 53, there was a marked shift in the viral population that is evident in the phylogenetic tree, such that only viruses carrying multiple mutations relative to the transmitted/founder, particularly in CH103 contact regions, persisted after week 30. This was followed by extreme and increasing within-time-point diversification in key epitope regions, beginning at week 53 (Supplementary Fig. 9). Emergence of antibodies with neutralization breadth occurred during this time (Supplementary Fig. 2 and Supplementary Table 1). Thus, plasma breadth evolved in the presence of highly diverse forms of the CH103 epitope contact regions (Fig. 4 and Supplementary Fig. 2).

To evaluate and compare the immune pressure on amino acids in the region of CH103 and CD4 contacts, we compared the frequency of mutations in evolving transmitted/founder sequences of patient CH505 during the first year of infection and in 16 other acutely infected subjects followed over time (Supplementary Fig. 10). The accumulation of mutations in the CH505 viral population was concentrated in regions likely to be associated with escape from the CH103 lineage (Supplementary Fig. 10a), and diversification of these regions was far more extensive during the first six months of infection in CH505 than in other subjects (Supplementary Fig. 10b). However, by one year into their infections, viruses from the other subjects had also begun to acquire mutations in these regions. Thus, the early and continuing accumulation of mutations in CH103 contact regions may have potentiated the early development of neutralizing antibody breadth in patient CH505.

Neutralization of viruses and the CH103 lineage

Heterologous BnAb activity was confined to the later members (I3 and later) of the BnAb arm of the CH103 lineage, as manifested by their neutralization capacity of pseudoviruses carrying tier 2 Env proteins A.Q842 and B.BG1168 (Fig. 5a). Similar results were seen with Env proteins A.Q168, B.JRFL, B.SF162 and C.ZM106 (Supplementary Tables 14 and 15). By contrast, neutralizing activity of clonal lineage members against the autologous transmitted/founder Env pseudovirus appeared earlier, with measurable neutralization of the CH505 transmitted/founder virus by all members of the lineage after the UCA except monoclonal antibody 1AH92U (Fig. 5a). Thus, within the CH103 lineage, early intermediate antibodies only neutralized the transmitted/founder virus, whereas later intermediate antibodies gained neutralization breadth, indicating evolution of neutralization breadth with affinity maturation, and CH103–CH1106 BnAbs evolved from an early autologous neutralizing antibody response. Moreover, the clonal lineage was heterogeneous, with an arm of the lineage represented in Fig. 5a evolving neutralization breadth and another antibody arm capable of mediating only autologous transmitted/founder virus neutralization. Although some escape-mutant viruses are clearly emerging over time (Supplementary Table 4), it is important to point out that, although the escape-mutant viruses are driving BnAb evolution, the BnAbs remained capable of neutralizing the CH505 transmitted/founder virus by all members of the lineage (Fig. 5a). Of note, the earliest mutations in the heavy-chain lineage clustered near the contact points with gp120, and these remained fixed throughout the period of study, whereas mutations that accumulated later tended to be further from the binding site and may be affecting binding less directly (Supplementary Fig. 11). Thus, stimulation of the CH103 BnAbs occurs in a manner to retain reactivity with the core CD4-binding site epitope present on the transmitted/founder Env. One possibility that might explain this is that the footprint of UCA binding contracts to the central core binding site of the CH103 mature antibody. Obtaining a crystal structure of the UCA with the transmitted/founder Env should inform this notion. Another possibility is that because affinity maturation is occurring in the presence of highly diverse forms of the CD4-binding site epitope, antibodies that favour tolerance of variation in and near the epitope are selected instead of those antibodies that acquire increased affinity for particular escape Env proteins. In both scenarios, persistence of activity to the transmitted/founder form and early viral variants
would be expected. Figure 5b and Supplementary Fig. 11 show views of accumulations of mutations or entropy during the parallel evolution of the antibody paratope and the Env epitope bound by monoclonal antibody CH103.

**Vaccine implications**

In this study, we demonstrate that the binding of a transmitted/founder Env to a UCA B-cell receptor of a BnAb lineage was responsible for the induction of broad neutralizing antibodies, thus providing a logical starting place for vaccine-induced CD4-binding site BnAb clonal activation and expansion. Importantly, the number of mutations required to achieve neutralization breadth was reduced in the CH103 lineage compared to most CD4-binding site BnAbs, although the CH103 lineage had reduced neutralization breadth compared to more mutated CD4-binding site BnAbs. Thus, this type of BnAb lineage may be less challenging to attempt to recapitulate by vaccination. By tracking viral evolution through early infection we found that intense selection and epitope diversification in the transmitted/founder virus preceded the acquisition of neutralizing antibody breadth in this individual—thus demonstrating the viral variants associated with development of BnAbs directly from autologous neutralizing antibodies and illuminating a pathway for induction of similar B-cell lineages.

These data have implications for understanding the B-cell maturation pathways of the CH103 lineage and for replicating similar pathways in a vaccine setting. First, we demonstrate in CH505 that BnAbs interact with extensive escape-generated epitope diversification may be an evolutionary force that also drives incidental acquisition of polyreactivity. Polyreactivity to host molecules in the CH103 lineage arose during affinity maturation in the periphery coincident with BnAb activity. This finding is compatible with the hypothesis that BnAbs may be derived from an inherently polyreactive pool of B cells, with polyreactivity providing a neutralization advantage via heteroligation of Env and host molecules. Alternatively, as CH103 affinity maturation involves adapting to the simultaneous presence of diverse co-circulating forms of the epitope, the selection of antibodies that can interact with extensive escape-generated epitope diversification may be an evolutionary force that also drives incidental acquisition of polyreactivity.

Thus, a candidate vaccine concept could be to use the CH505 transmitted/founder Env or Env subunits (to avoid dominant Env non-neutralizing epitopes) to initially activate an appropriate naive B-cell response, followed by boosting with subsequently evolved

---

**Figure 5** Development of neutralization breadth in the CH103 clonal lineage. a, Phylogenetic CH103 clonal lineage tree showing the IC_{50} (μg ml^{-1}) of neutralization of the autologous transmitted/founder (C.CH505), heterologous tier clades A (A.Q842) and B (B.BG1168) viruses as indicated. b, Interaction between evolving virus and developing clonal lineage mapped on to models of CH103 developmental variants and contemporaneous virus. The outer domain of HIV gp120 is depicted in worm representation, with worm thickness and colour (white to red) mapping the degree of per-site sequence diversity at each time point. Models of antibody intermediates are shown in cartoon diagram, with somatic mutations at each time point highlighted in spheres and coloured red for mutations carried over from I8 to mature antibody, cyan for mutations carried over from I4 to mature antibody, green for mutations carried over from I3 to mature antibody, blue for mutations carried over from I2 to mature antibody, orange for mutations carried over from I1 to mature antibody, and magenta for CH103 mutations from I1. Transient mutations that did not carry all the way to mature antibody are coloured in deep olive. The antibody (paratope) residues are shown in surface representation and coloured by their chemical types as indicated.
CH505 Env variants either given in combination, to mimic the high diversity observed in vivo during affinity maturation, or in series, using vaccine immunogens specifically selected to trigger the appropriate maturation pathway by high-affinity binding to UCA and antibody intermediates. These data demonstrate the power of studying subjects followed from the transmission event to the development of plasma BnAb activity for concomitant isolation of both transmitted/founder viruses and their evolved quasispecies along with the clonal lineage of induced BnAbs. The finding that the transmitted/founder Env can be the stimulator of a potent BnAb and bind optimally to that BnAb UCA is a crucial insight for vaccine design, and could allow the induction of BnAbs by targeting UCAs and intermediate ancestors of BnAb clonal lineage trees.

METHODS SUMMARY

Serial blood samples were collected from an HIV-1-infected subject CH505 from 4 to 236 weeks after infection. Monoclonal antibodies CH103, CH104 and CH106 were generated by the isolation, amplification and cloning of single RSC3-specific memory B cells as described. VDJ and Vβ 454 pyrophoresis was performed on samples from five time points after transmission. Inference of UCA, and identification and production of clone members were performed as described in the Methods (see also Kepler, T. B., manuscript submitted; http://archive.org/abs/1303.0424). Additional VDJ and Vβ genes were identified by 454 pyrophoresing and select VDJ and Vβ genes were used to produce recombinant antibodies as reported previously and described in the Methods. Binding of patient plasma antibodies and CH103 clonal lineage antibodies to autologous and heterologous HIV-1 Env proteins was measured by ELISA and surface plasmon resonance, and neutralizing activity of patient plasma and CH103 antibody clonal lineage members was determined in a T2M-bl-based pseudovirus neutralization assay. Crystallographic analysis of CH103 bound to the HIV-1 outer domain was previously reported, and as described in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 2 December 2012; accepted 7 March 2013.

Published online 3 April 2013.
and C. Andrews for project management. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-Eng-38. The opinions herein are those of the authors and should not be construed as official or representing the views of the US Department of Health and Human Services, National Institute for Allergy and Infectious Diseases.

Author Contributions H.-X.L., R.L., T.Z. and F.G. contributed equally to this work. H.-X.L. led production of antibodies and Env proteins, designed assays, analysed data and edited the paper; R.L. generated antibodies and performed assays; T.Z. co-led the structural biology team, performed structural studies, analysed data, and edited the paper; F.G. generated autologous Env sequences and viruses; S.M.A. performed surface plasmon resonance analysis; S.D.B., A.Z.F., J.C.M. and K.M.R. performed pyrosequencing; C.A.S., Z.Z., J.Z. and L.S. analysed pyrosequences; S.G., P.H., B.T. and M.K. performed antibody and Env sequence analysis, and edited the paper; G.K. and G.Y. performed polyreactivity assays and analysis; S.-M.X. and D.C.M. performed neutralization assays and analysis; R.P., K.E.L. and R.M.S developed and performed ELISAs; K.A.S., M.C. and G.K. performed cohort development, patient recruitment, management and sampling; M.K.L. and L.M.T. performed neutralization assays; Y.C., F.C. and S.C. performed Env cloning and sequencing; S.M., X.D., M.G.J., S.S., B.Z. and A.Z. performed experiments related to crystallization, structure determination, and structural analysis; G.M.S. and B.H.H. generated autologous Env sequences and edited the paper; T.B.K. performed antibody gene sequence analysis and inferred ancestor and intermediate antibodies and edited the paper; P.D.K. co-led the structural biology team and collected and analysed data, and edited the paper; J.R.M. isolated antibodies, designed assays, analysed data, and edited the paper; B.F.H. designed and directed the study, read and interpreted antinuclear antibody assays, analysed data, and wrote and edited the paper.

Author Information The GenBank accession numbers for 292 CH505 Env proteins are KC247375–KC247667, and accessions for 459 V(H)DJ(H) and 174 V(L)J(L) sequences of antibody members in the CH103 clonal lineage are KC575845–KC576303 and KC576304–KC576477, respectively. Atomic coordinates and structure factors for unbound CH103 Fab as well as CH103 Fab in complex with the ZM176.66 outer domain have been deposited with the Protein Data Bank under accession codes 4JAM for CH103 Fab, and 4JAN for the CH103–gp120 complex. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests; details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-X.L. (hliao@duke.edu) or B.F.H. (barton.haynes@duke.edu).

NISC Comparative Sequencing Program

Jesse Becker1, Betty Benjamin1, Robert Blakesley1, Gerry Bouffard1, Sheila Brooks3, Holly Coleman1, Mia Dekhtyar2, Michael Gregory1, Xiaobin Guan1, Jyoti Gupta1, Joel Han1, April Hargrove1, Shi-ling Ho1, Taccara Johnson1, Richelle Legatzi1, Sean Lovett1, Quino Maduro1, Cathy Masiello1, Baishali Maskeri1, Jenny McDowell1, Casandra Montemayor1, James Mullikin1, Morgan Park1, Nancy Riebow1, Karen Schandler2, Brian Schmidt1, Christina Sison1, Mal Stanttripop2, James Thomas1, Pam Thomas3, Meg Vemulapalli1 & Alice Young1

1NISC Comparative Sequencing Program, NIH, Bethesda, Maryland 20892, USA.
METHODS

Study subject. Plasma and PBMCs were isolated from serial blood samples that were collected from an HIV-1-infected subject CH505 starting 6 weeks after infection up to 236 weeks after infection (Supplementary Table 1) and frozen at ~80 °C and liquid nitrogen tanks, respectively. During this time, no antiretroviral therapy was administered. All work related to human subjects was in compliance with Institutional Review Board protocols approved by the Duke University Health System Institutional Review Board. Antibodies isolated from PBMCs were tested in binding and neutralization assays.

Inference of UCA and identification of clone members. The inference of the UCA from a set of clonally related genes is described elsewhere (Kepler, T. B., manuscript submitted; http://arxiv.org/abs/1303.0424). In brief, we parameterize the VDJ rearrangement process in terms of its gene segments, recombination points, and n-regions sequences (non-templated nucleotides polymerized in the recombination junctions by the action of terminal deoxynucleotidyl transferase). Given any multiple sequence alignment (A) for the set of clonally related genes and any tree (T) describing a purported history, we can compute the likelihood for all parameter values, and thus the posterior probabilities on the rearrangement parameters conditional on A and T. We can then find the unmutated ancestor with the greatest posterior probability, and compute the maximum likelihood alignment A* and tree T* given this unmutated ancestor, and then recompute the posterior probabilities on rearrangement parameters conditional on A* and T*. We iterate the alternating conditional maximizations until convergence is reached. We use ClustalW for the multiple sequence alignment, dnaml (PHYLIP) to infer the maximum likelihood tree, and our own software for the computation of the likelihood over the rearrangement parameters. The variable regions of heavy- and light-chain (V\textsubscript{H}D\textsubscript{J} and V\textsubscript{L}J) gene segments were inferred from the natural pairs themselves. The posterior probabilities for these two gene segments are 0.999 and 0.993, respectively. We first inferred the unmutated ancestor from the natural pairs as described above. We identified additional clonally related variable region sequences from deep sequencing and refine the estimate of the UCA iteratively. We identified all variable region segments inferred to have been rearranged to the same V\textsubscript{H}D\textsubscript{J} and J\textsubscript{L}, and to have the correct CDR3 length. For each sequence, we counted the number of mismatches between the sequence and the presumed V\textsubscript{H}D\textsubscript{J} gene up to the codon for the second invariant cysteine. Each iteration was based on the CDR3 of the current posterior modal unmutated ancestor. For each candidate sequence, we computed the number of nucleotide mismatches between its CDR3 and the unmutated ancestor CDR3. The sequence was rejected as a potential clone member if the z-statistic in a test for difference between proportion is greater than 2 (ref. 48). Once the set of candidates has been thus filtered by CDR3 distance, the unmutated ancestor was inferred on that larger set of sequences as described above. If the new posterior modal unmutated ancestor differed from the previous one, the process was repeated until convergence was reached. Owing to the inherent uncertainty in unmutated ancestor inference, we inferred the six most likely V\textsubscript{H}UC A sequences resulting in four unique amino acid sequences that were all of Cys 2 is omitted because the unmutated form of the ancestral and , V\textsubscript{H} for expression as IgG1 * V\textsubscript{H}V\textsubscript{L} given this unmutated ancestor, and then and J\textsubscript{V} and 5–7,22,36 variants to 119 9 45 The of CH105 were paired with either genes. }
VitDIII to either the UCA or mature antibodies for expressing as full-length IgG1 antibodies as described\(^\text{31}\) (Supplementary Table 2). Recombinant HIV-1 proteins. HIV-1 Env genes for subtype B, 63521, subtype C, 1086, and subtype CRF_01, 427299, as well as subtype C, CH505 autologous transfomed/finder Env were obtained from acutely infected HIV-1 subjects by single genome amplification\(^\text{31}\), codon-optimized by using the codon usage of highly expressed human housekeeping genes\(^\text{52}\), de novo synthesized (Gene-Script) as gp140 or gp120 (AE.427299) and cloned into a mammalian expression plasmid (GEMinG, Invitrogen). Recombinant Env glycoproteins were produced in 293F cells cultured in serum-free medium and transfected with the HIV-1 gp140- or gp120-expressing pcDNA3.1 plasmids, purified from the supernatants of transfected 293F cells by using Galanthus nivalis lectin-agarse (Vector Labs) column chromatography\(^\text{44,51}\) and stored at −80 °C. Select Env proteins made as CH505 transfomed/finder Env were further purified by superose 6 column chromatography to trimeric forms, and in binding assays that showed similar results as with the lectin-purified oligomers. ELISA. Binding of patient plasma antibodies and CH103 clonal lineage antibodies to autologous and heterologous HIV-1 Env proteins was measured by ELISA as described previously\(^\text{13,53}\). Plasma samples in serial threefold dilutions starting at 1:30 to 1:52,4470 or purified monoclonal antibodies in serial threefold dilutions starting at 100 µg ml\(^{-1}\) to 0.000 µg ml\(^{-1}\) diluted in PBS were assayed for binding to autologous and heterologous HIV-1 Env proteins. Binding of biotin-labelled CH103 at the subsaturating concentration was assayed for cross-competition by unlabelled HIV-1 antibodies and soluble CD4-Ig in serial fourfold dilutions starting at 10 µg ml\(^{-1}\). The half-maximal effective concentration (EC\(_{50}\)) of plasma samples and monoclonal antibodies to HIV-1 Env proteins were determined and expressed as either the reciprocal dilution of the plasma samples or concentration of monoclonal antibodies. Surface plasmon resonance affinity and kinetics measurements. Binding Km and rate constant (association rate (k\(_{a}\)) measurements of monoclonal antibodies and all candidate UCAs to the autologous Env C. CH505 gp140 and/or the heterologous Env B.63521 gp120 were carried out on BiAcore 3000 instruments as described previously\(^\text{34,42}\). Anti-human IgG Fc antibody (Sigma Chemicals) was immobilized on a CM5 sensor chip to about 15,000 response units and each antibody was captured to about 50–200 response units on three individual flow cells for replicate analysis, in addition to having one flow cell captured with the control Synagis (anti-RSV) monoclonal antibody on the same sensor chip. Double referencing for each monoclonal antibody–HIV-1 Env binding interactions was used to subtract nonspecific binding and signal drift of the Env proteins to the control surface and blank buffer flow, respectively. Antibody capture level on the sensor surface was optimized for each monoclonal antibody to minimize rebinding and any associated avidity effects. C.CH505 Env gp140 protein was injected at concentrations ranging from 2 to 25 µg ml\(^{-1}\), and B.63521 gp120 was injected at 50–400 µg ml\(^{-1}\) for UCAs and early intermediates IA8 and IA4, 10–100 µg ml\(^{-1}\) for intermediate IA3, and 1–25 µg ml\(^{-1}\) for the distal and mature monoclonal antibodies. All curve-fitting analyses were performed using global fit of to the 1:1 Langmuir model and are representative of at least three measurements. All data analysis was performed using the BiAevaluation 4.1 analysis software (GE Healthcare). Neutralization assays. Neutralizing antibody assays in TZM-bl cells were performed as described previously\(^\text{40}\). Neutralizing activity of plasma samples in eight serial threefold dilutions starting at 1:20 dilution and for recombinant monoclonal antibodies in eight serial threefold dilutions starting at 50 µg ml\(^{-1}\) were tested against autologous and heterologous HIV-1 Env-pseudotyped viruses in TZM-bl-based neutralization assays using the methods as described\(^\text{34,52}\). Neutralization breadth of CH103 was determined using a previously described\(^\text{52}\) panel of 196 of geographically and genetically diverse Env-pseudotyped viruses representing the major geographically and genetically diverse Env-pseudotyped viruses. The data were calculated as a reduction in luminescence units compared with control wells, and reported as IC\(_{50}\) in either reciprocal dilution for plasma samples or in micrograms per microlitre for monoclonal antibodies. Crystallization of antibody CH103 and its gp120 complex. The antigen binding fragment (Fab) of CH103 was generated by LyS-C (Roche) digestion of IgG1 CH103 and purified as previously described\(^\text{13}\). The extended gp120 core of HIV-1 clade C ZM176.66 was used to form a complex with Fab CH103 by using recombinantly expressed methods\(^\text{53}\). In brief, for CH103 crystals appeared in a condition from the JCSG space group with cell dimensions at a = 99.9, b = 146.4, c = 66.3, \(\gamma = 90.8\), \(\beta = 97.7\) and \(\alpha = 90.0\), and contained two Fab molecules per asymmetric unit (Supplementary Table 8). The crystal structures of Fab CH103 were solved by molecular replacement using Phaser\(^\text{54}\) in the CCP4 program suite\(^\text{48}\) with published antibody structures as searching models. The gp120–CH103 crystal also belonged to the P2\(_2\) space group with cell dimensions at \(a = 48.9, b = 208.7, c = 69.4, g = 90.0, \beta = 107.2\) and \(\gamma = 90.0\), and contained two gp120–CH103 complexes per asymmetric unit (Supplementary Table 8). The high-resolution CH103 Fab structure was used as an initial model to place the CH103 Fab component in the complex. With the CH103 Fab position fixed, searching with the extended gp120 core of ZM176.66 in the VRC01-bound form as an initial model failed to place the gp105 component in the complex. After trimming the inner domain and bridging sheet regions from the gp120 search model, Phaser was able to place correctly the remaining outer domain of gp120 into the complex without considerable clashes. Analysis of the packing of the crystallographic lattice indicated a lack of space to accommodate the inner domain of gp120, suggesting possible protease cleavage of gp120 by the containing fungi during crystallization\(^\text{54}\). Structural refinements were carried out with PHENIX\(^\text{55}\). Starting with torsion-angle simulated annealing with slow cooling, iterative manual model building was carried out on COOT\(^\text{56}\) with maps generated from combinations of standard positional and individual positional TLS (translation/libration/screw) refinement algorithms and non-crystallographic symmetry (NCS) restraints. Ordered solvent were added during each macro cycle. Throughout the refinement processes, a cross validation (\(R_{	ext{cv}}\)) test set consisting of 5% of the data was used and hydrogen atoms were included in the refinement model. Structure validations were performed periodically during the model building/refinement process with MolProbity\(^\text{57}\) and pdb-care\(^\text{4}\). X-ray crystallographic data and refinement statistics are summarized in Supplementary Table 8. The Kabat nomenclature\(^\text{58}\) was used for numbering of amino acid residues in amino acid sequences in antibodies. Protein structure analysis and graphical representations. PISA\(^\text{59}\) was used to perform secondary structure prediction interfaces analysis, CCP4 (ref. 66) was used for structural alignments. All graphical representation with protein crystal structures were made with PyMol\(^\text{60}\). Polyelectrolyte analysis of antibodies. All antibodies in CH103 clonal lineage were assayed at 50 µg ml\(^{-1}\) for autoreactivity to HEP-2 cells (Innverness Medical Professional Diagnostics) by indirect immunofluorescence staining and a panel of autogens by anionic antibody assays using the methods as reported previously\(^\text{56}\). The intermediate antibody IA1 and CH106 were identified as reactive with HEP-2 cells and then selected for further testing for reactivity with human host cellular antigens using ProteinArray 5 microchip (Invitrogen) according to the manufacturer's instructions. In brief, the ProteinArray 5 microchips were blocked and exposed to 2 µg ml\(^{-1}\) IA1, CH106 or an isotype-matched (IgG1, k) human myeloma protein, 151 K (Southern Biotech) for 90 min at 4 °C. Protein–antibody interactions were detected by 1 µg ml\(^{-1}\) Alexa Fluor 647-conjugated anti-human IgG. The arrays were scanned at 635 nm with 10-µm resolution, using

©2013 Macmillan Publishers Limited. All rights reserved
100% power and 600 gain (GenePix 4000B scanner, Molecular Devices). Fluorescence intensities were quantified using GenePix Pro 5.0 (Molecular Devices). Lot-specific protein spot definitions were provided by the microchip manufacturer and aligned to the image.

47. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
48. Zar, J. H. *Biostatistical Analysis* (Pearson, 1974).
49. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 (2010).
50. Moir, S. et al. Normalization of B cell counts and subpopulations after antiretroviral therapy in chronic HIV disease. *J. Infect. Dis.* **197**, 572–579 (2008).
51. Liao, H. X. et al. High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. *J. Virol. Methods* **158**, 171–179 (2009).
52. Liao, H. X. et al. Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity* **38**, 176–186 (2013).
53. André, S. et al. Increased immune response elicited by DNA vaccination with a synthetic gp120 sequence with optimized codon usage. *J. Virol.* **72**, 1497–1503 (1998).
54. Liao, H. X. et al. A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses. *Virology* **353**, 268–282 (2006).
55. Montefiori, D. C. et al. Magnitude and breadth of the neutralizing antibody response in the RV144 and Vax003 HIV-1 vaccine efficacy trials. *J. Infect. Dis.* **206**, 431–441 (2012).
56. Huang, J. et al. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012).
57. Wolinsky, S. M. et al. Response: HIV-1 evolution and disease progression. *Science* **274**, 1010–1011 (1996).
58. Kwong, P. D. et al. Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem.* **274**, 4115–4123 (1999).
59. Kwon, Y. D. et al. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proc. Natl Acad. Sci. USA* **109**, 5663–5668 (2012).
60. Bai, Y., Auperin, T. C. & Tong, L. The use of in situ proteolysis in the crystallization of murine CsF-77. *Acta Crystallogr. F* **63**, 139–138 (2007).
61. Mandel, C. R., Gebauer, D., Zhang, H. & Tong, L. A serendipitous discovery that in situ proteolysis is essential for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Crystallogr. F* **62**, 1041–1045 (2006).
62. Otrubowski, Z. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307 (1997).
63. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
64. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
65. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
66. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
67. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
68. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
69. DeLano, W. L. The PyMOL Molecular Graphics System http://www.pymol.org (DeLano Scientific, 2002).