ARTICLE

doi:10.1038/nature12053

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

Hua-Xin Liao[1,2]*, Rebecca Lynch[3]*, Tongqing Zhou[3]*, Feng Gao[1,2]*, S. Munir Alam[1,2], Scott D. Boyd[4], Andrew Z. Fire[4], Krishna M. Roskin[4], Chaim A. Schramm[5], Zhenhai Zhang[5], Jiang Zhu[3], Lawrence Shapiro[3,5], NISC Comparative Sequencing Program†, James C. Mullikin[6,7], S. Gnanakaran[8], Peter Hraber[8], Kevin Wiehe[1,2], Garnett Kelsoe[1,2], Guang Yang[1,2], Shi-Mao Xia[1,2], David C. Montefiori[1,2], Robert Parks[1,2], Krissey E. Lloyd[1,2], Richard M. Scearce[1,2], Kelly A. Soderberg[1,2], Myron Cohen[9], Gift Kamanga[10], Mark K. Louder[3], Lillian M. Tran[3], Yue Chen[1,2], Fangping Cai[1,2], Sheri Chen[1,2], Stephanie Moquin[3], Xiulian Du[3], M. Gordon Joyce[3], Sanjay Srivatsan[3], Baoshan Zhang[3], Anqi Zheng[3], George M. Shaw[11], Beatrice H. Hahn[11], Thomas B. Kepler[12], Bette T. M. Korber[8], Peter D. Kwong[3], John R. Mascola[3] & Barton F. Haynes[1,2]

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[1–9]. 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[4,10–13]. In particular, CD4-binding site BnAbs have extremely high levels of somatic mutation, suggesting complex or prolonged maturation pathways[4–7]. 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[7,14–18]. Although it has been shown that Env proteins bind to UCAs of BnAbs targeting the gp41 membrane proximal region[16,19], and to UCAs of some V1/V2 BnAbs[20], so far, heterologous Env proteins have not been identified that bind the UCAs of CD4-binding site BnAb lineages[7,18,21–23], although they should exist[21].

Eighty per cent of heterosexual HIV-1 infections are established by one transmitted/founder virus[24]. The initial neutralizing antibody response to this virus arises approximately 3 months after transmission and is strain-specific[25,26]. The antibody response to the transmitted/founder virus drives viral escape, such that virus mutants become resistant to neutralization by autologous plasma[25,26]. 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[2,20,27–33].

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[6,11,15,34]. Because the initial autologous neutralizing antibody response is specific for the transmitted/founder virus[31], 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[1,3–7,20,27,29]. 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[6,11,13–16,18,19,21]. This would be followed by vaccination

[1]Duke University Human Vaccine Institute, Departments of Medicine and Immunology, Duke University School of Medicine, Durham, North Carolina 27710, USA. [2]Duke Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, Durham, North Carolina 27710, USA. [3]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [4]Department of Pathology, Stanford University, Palo Alto, California 94305, USA. [5]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA. [6]NISC Comparative Sequencing Program, NIH, Bethesda, Maryland 20892, USA. [7]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [8]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA. [9]Departments of Medicine, Epidemiology and Microbiology and Immunology, University of North Carolina, North Carolina 27599, USA. [10]University of North Carolina Project, Kamuzu Central Hospital, Lilongwe, Malawi. [11]Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [12]Department of Microbiology, Boston University, Boston, Massachusetts 02215, USA.
*These authors contributed equally to this work.
†A full list of participants and their affiliations appears at the end of the paper.

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 cohort[35] 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 RNAs[24] 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))[5] at 53 weeks, and evolution of plasma cross-reactive neutralizing activity from 41–92 weeks after transmission[30] (Fig. 1, Supplementary Table 1 and Supplementary Fig. 2). The natural variable regions of heavy-chain (*VHDJH*) and light-chain (*VLJL*) 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 protein[5,13,36] (Fig. 1b). The *VHDJH* gene of antibody CH105 was similarly isolated, but no *VLJL* gene was identified from the same cell. Analysis of characteristics of VHDJH (VH4–59, posterior probability (PP) = 0.99; D3–16, PP = 0.74; JH4, PP = 1.00) and VLJL (Vλ3–1, PP = 1.00; Jλ1, 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 described[5,37] panel of 196 geographically and genetically diverse Env-pseudoviruses representing the major circulated genetic subtypes and circulating recombinant forms demonstrated that CH103 neutralized 55% of viral isolates, with a geometric mean half-maximum inhibitory concentration

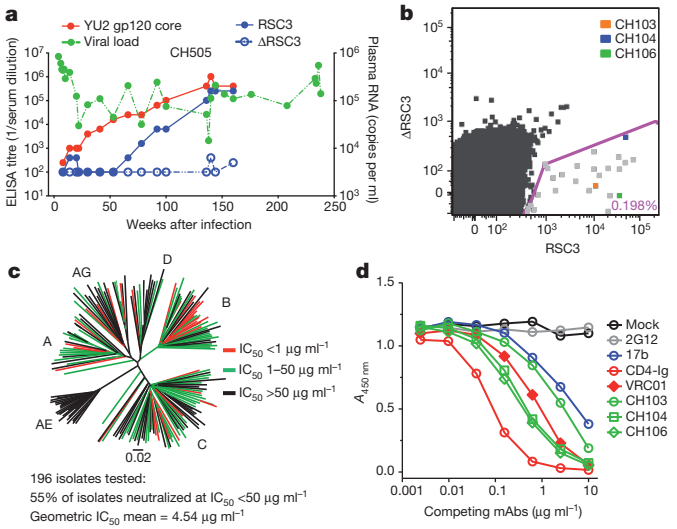

Figure 1 | Development of neutralization breadth in donor CH505 and isolation of antibodies. a, Shown are HIV-1 viral RNA copies and reactivity of longitudinal plasmas samples with HIV-1 YU2 gp120 core, RSC3 and negative control RSC3Δ371Ile (ΔRSC3) proteins. **b**, PBMCs from week 136 were used for sorting $CD19^+$, $CD20^+$, IgG^+, $RSC3^+$ and $\Delta RSC3^-$ memory B cells (0.198%). Individual cells indicated as orange, blue and green dots yielded monoclonal antibodies CH103, CH104 and CH106, respectively, as identified by index sorting. **c**, The neutralization potency and breadth of the CH103 antibody are displayed using a neighbour-joining tree created with the PHYLIP package. The individual tree branches for 196 HIV-1 Env proteins representing major circulating clades are coloured according to the neutralization IC_{50} values as indicated. **d**, Cross competition of CH103 binding to YU2 gp120 by the indicated HIV-1 antibodies, and soluble CD4-Ig was determined by ELISA. mAbs, monoclonal antibodies.

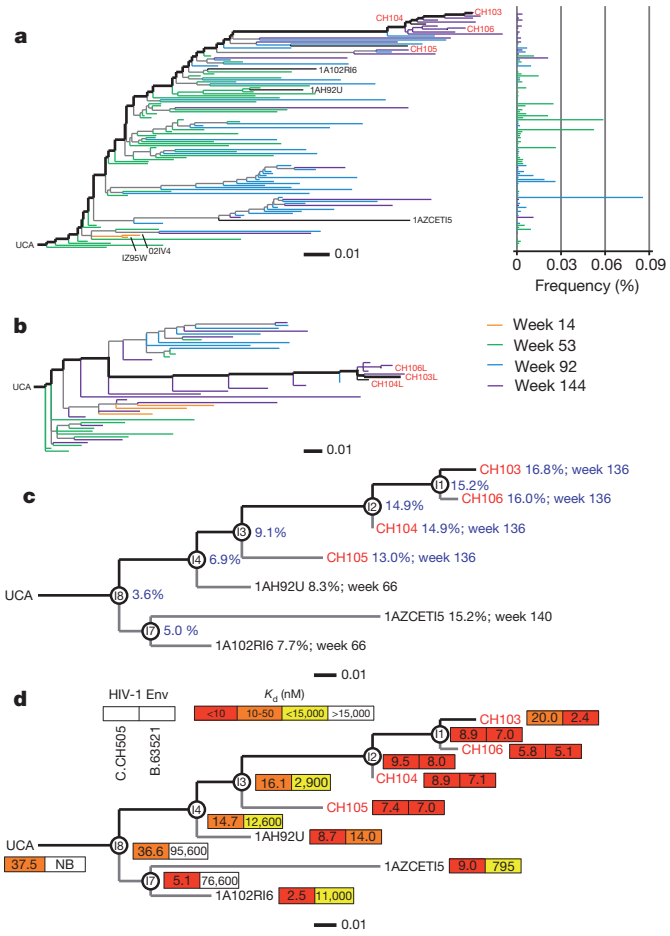

Figure 2 | CH103 clonal family with time of appearance, V_HDJ_H mutations and HIV-1 Env reactivity. a, **b**, Phylogenies of VHDJH (**a**) and VLJL (**b**) sequences from sorted single memory B cells and pyrosequencing. The ancestral reconstructions for each were performed as described in the Methods. The phylogenetic trees were subsequently computed using neighbour-joining on the complete set of DNA sequences (see Methods) to illustrate the correspondence of sampling date and read abundance in the context of the clonal history. Within time-point V_H monophyletic clades are collapsed to single branches; variant frequencies are indicated on the right. Isolated mature antibodies are red, pyrosequencing-derived sequences are black. The inferred evolutionary paths to observed matured antibodies are bold. **c**, Maximum-likelihood phylogram showing the CH103 lineage with the inferred intermediates (circles, I1–4, I7 and I8), and percentage mutated V_H sites and timing (blue), indicated. **d**, Binding affinities (K_d, nM) of antibodies to autologous subtype C CH505 (C.CH505; left box) and heterologous B.63521 (right box) were measured by surface plasmon reasonance.

(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 chimaeric protein CD4-Ig (Fig. 1d); CH103 binding to RSC3 Env was also substantially diminished by gp120, with Pro363Asn and Δ371Ile 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, CH105 and CH106 BnAbs by single-cell flow sorting. The CH103 clonal lineage was enriched by V_HDJ_H and V_LJ_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_HDJ_H sequences (1AH92U, 1AZCET and 1A102R) recovered by pyrosequencing, and V_HDJ_H genes of the inferred intermediate (I) antibodies (I1–I4, I7, I8)[11,16,34] (T. B. Kepler, manuscript submitted; http://arxiv.org/abs/1303.0424) that were paired and expressed with either the UCA or I2 V_LJ_L depending on the genetic distance of the V_HDJ_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_LJ_L. The CH105 natural V_HDJ_H isolated from RSC3 memory B-cell sorting was paired with the V_LJ_L of I2.

Whereas the V_HDJ_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_HDJ_H frequencies of CH103 lineage CH103, CH104, CH105 and CH106 are 13–17% (Fig. 2c). Furthermore, antibodies in 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_LJ_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[7,18,22]. 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 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 no 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-mutant Env proteins could be isolated that were progressively less reactive with the CH103 clonal lineage members. Env proteins isolated at weeks 30, 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_HDJ_H chains closely related to, and possibly members of, the CH103 clonal lineage (Fig. 2a, Supplementary Table 7). Moreover, one of these V_HDJ_H chains when reconstituted in a full IgG1 backbone and expressed with the UCA V_LJ_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_HDJ_H and V_LJ_L sequences peaked at week 53, with 230 and 83 unique transcripts, respectively. V_HDJ_H clonal members fell to 46 at week 144, and V_LJ_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[10]. 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 (β20–21). 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 (Cα 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

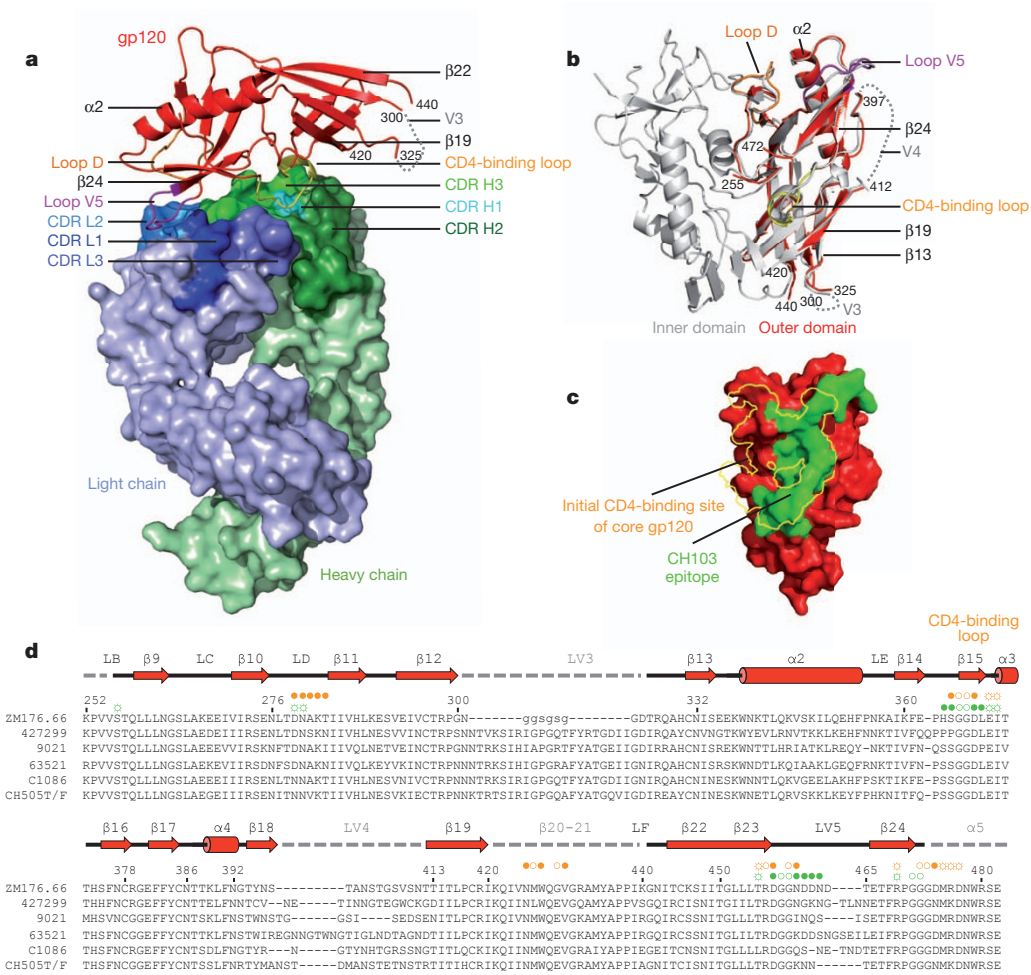

Figure 3 | Structure of antibody CH103 in complex with the outer domain of HIV-1 gp120. a, Overall structure of the CH103–gp120 complex, with gp120 polypeptide depicted in red ribbon and CH103 shown as a molecular surface (heavy chain in green, light chain in blue). Major CH103-binding regions on gp120 are coloured orange for loop D, yellow for the CD4-binding site, and purple for loop V5. **b**, Superposition of the outer domain of gp120 bound by CH103 (red), and core gp120 bound by VRC01 (grey), with polypeptide shown in ribbon representation. **c**, CH103 epitope (green) on gp120 outer domain (red), with the initial CD4-binding site superposed (yellow boundaries) in surface representation. **d**, Sequence alignment of outer domains of the crystallized gp120 shown on the first line, and diverse HIV-1 Env proteins recognized by CH103. Secondary structure elements are labelled above the alignment, with grey dashed lines indicating disordered regions. Symbols in yellow or green denote gp120 outer domain contacts for CD4 and CH103, respectively, with open circles representing main-chain contacts, open circles with rays representing side-chain contacts, and filled circles representing both main-chain and side-chain contacts.

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 complementarity-determining regions (CDRs) interacted with gp120 as well as the light-chain framework region 3 (FWR3) (Supplementary Fig. 7a, b and Supplementary Tables 10 and 11). It is important to note that ~40% of the antibody contact surface was altered by somatic mutation in the HCDR2, LCDR1, LCDR2 and FWR3. In particular, residues 56 on the heavy chain, and residues 50, 51 and 66 on the light chain are altered by somatic mutation to form hydrogen bonds with the CD4-binding loop, loop D and loop V5 of gp120. Nevertheless, 88% of the CH103 V_HDJ_H and 44% of the V_LJ_L contact areas were with amino acids unmutated in the CH103 germ line, potentially providing an explanation for the robust binding of the transmitted/founder Env to the CH103 UCA (Supplementary Fig. 7c, d and Supplementary Table 12).

Evolution of transmitted/founder Env sequences

Using single genome amplification and sequencing[24] we tracked the evolution of CH505 *env* genes longitudinally from the transmitted/founder virus to 160 weeks after transmission (Fig. 4 and Supplementary Fig. 8). The earliest recurrent mutation in Env, Asn279Lys (HIV-1 HXB2 numbering), was found 4 weeks after infection, and was in Env loop D in a CH103 contact residue. By week 14, additional mutations in loop D appeared, followed by mutations and insertions in the V1 loop at week 20. Insertions and mutations in the V5 loop began to accumulate by week 30 (Fig. 4). Thus, the transmitted/founder virus began to diversify in key CD4 contact regions starting within 3 months of infection (Supplementary Figs 8 and 9). Loop D and V5 mutations were directly in or adjacent to CH103/Env contact residues. Although the V1 region was not included in the CH103–Env co-crystal, the observed V1 CH505 Env mutations were adjacent to contact residues for CD4 and VRC01 so are likely to be relevant. It is also possible that early V1 insertions (Fig. 4) were selected by inhibiting access to the CD4-binding site in the trimer or that they arose in response to early T-cell pressure. CD4-binding-loop mutations were present by week 78. Once regions that could directly affect CH103-lineage binding began to evolve (loop D, V5, the CD4-binding loop, and possibly loop V1), they were under sustained positive selective pressure throughout the study period (Fig. 4, Supplementary Figs 8 and 9 and Supplementary Table 13).

Considerable within-sample virus variability was evident in Env regions that could affect CH103-lineage antibody binding, and

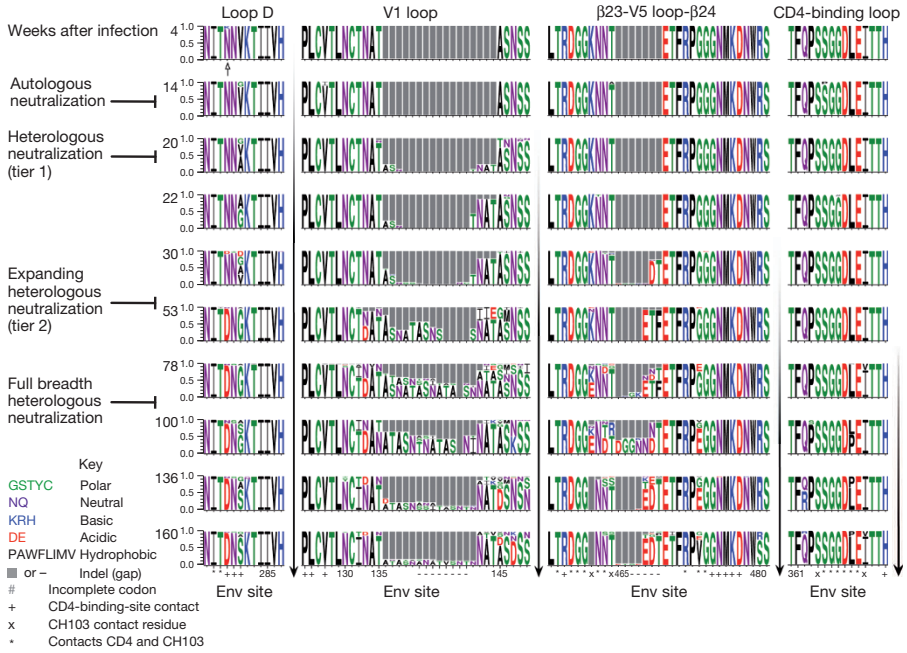

Figure 4 | Sequence logo displaying variation in key regions of CH505 Env proteins. The frequency of each amino acid variant per site is indicated by its height, deletions are indicated by grey bars. The first recurring mutation, Asn279Lys, appears at week 4 (open arrow). The timing of BnAb activity development (from Supplementary Fig. 2 and Supplementary Table 1) is on the left. Viral diversification, which precedes acquisition of breadth, is highlighted by vertical arrows to the right of each region. CD4 and CH103 contact residues, and amino acid position numbers based on HIV-1 HXB2, are shown along the base of each logo column.

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–CH106 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 (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

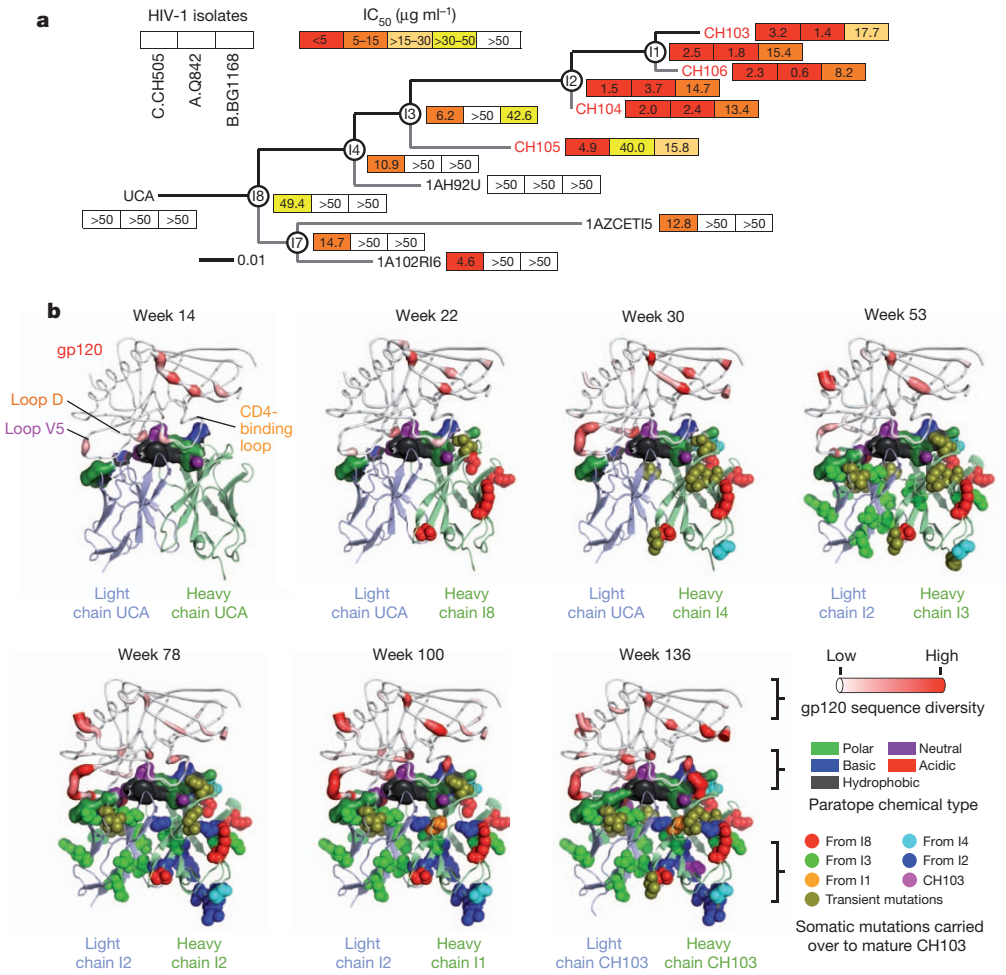

Figure 5 | Development of neutralization breadth in the CH103 clonal lineage. **a**, Phylogenetic CH103 clonal lineage tree showing the IC_{50} (μg ml^{-1}) of neutralization of the autologous transmitted/founder (C.CH505), heterologous tier clades A (A.Q842) and B (B.BG1168) viruses as indicated. **b**, 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 II 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.

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 were driven by sequential Env evolution beginning as early as 14 weeks after transmission, a time period compatible with induction of this type of BnAb lineage with a vaccine given the correct set of immunogens. Second, whereas heterologous Env proteins did not bind with UCAs or early intermediate antibodies of this lineage, the CH505 transmitted/founder Env bound remarkably well to the CH103 UCA, and subsequent Env proteins bound with increased affinity to later clonal lineage members. This suggests that immunizations with similar sequences of Env or Env subunits may drive similar lineages. Third, the CH103 lineage is less complicated than those of the VRC01 class of antibodies because antibodies in this lineage have fewer somatic mutations, and no indels, except CH103 V_L, which has a deletion of three amino acid residues in the LCDR1 region. It should also be noted that our study is in one patient. Nonetheless, in each BnAb patient, analysis of viral evolution should determine a similar pathway of evolved Env proteins that induce BnAb breadth. The observation that rhesus macaques infected with the CCR5-tropic simian/human immunodeficiency virus (SHIV)-AD8 virus frequently develop neutralization breadth[42] suggests that certain Env proteins may be more likely to induce breadth and potency than others.

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[21,43]. Alternatively, as CH103 affinity maturation involves adapting to the simultaneous presence of diverse co-circulating forms of the epitope[44], 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

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[11]. 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[11].

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[5–7,22,36]. V_HDJ_H and V_LJ_L 454 pyrosequencing was performed on samples from five time points after transmission[6]. 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://arxiv.org/abs/1303.0424). Additional V_HDJ_H and V_LJ_L genes were identified by 454 pyrosequencing[6,34,38] and select V_HDJ_H and V_LJ_L genes were used to produce recombinant antibodies as reported previously[34] and described in the Methods. Binding of patient plasma antibodies and CH103 clonal lineage antibody members to autologous and heterologous HIV-1 Env proteins was measured by ELISA and surface plasmon resonance[19,34,43,45], and neutralizing activity of patient plasma and CH103 antibody clonal lineage members was determined in a TZM-bl-based pseudovirus neutralization assay[5,37,46]. Crystallographic analysis of CH103 bound to the HIV-1 outer domain was performed as previously reported[7], 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.

1. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326,** 285–289 (2009).
2. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477,** 466–470 (2011).
3. Burton, D. R., Poignard, P., Stanfield, R. L. & Wilson, I. A. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337,** 183–186 (2012).
4. Kwong, P. D. & Mascola, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* **37,** 412–425 (2012).
5. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329,** 856–861 (2010).
6. Wu, X. *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333,** 1593–1602 (2011).
7. Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329,** 811–817 (2010).
8. Sattentau, Q. J. & McMichael, A. J. New templates for HIV-1 antibody-based vaccine design. *F1000 Biol. Rep.* **2,** 60 (2010).
9. Stamatatos, L. HIV vaccine design: the neutralizing antibody conundrum. *Curr. Opin. Immunol.* **24,** 316–323 (2012).
10. Haynes, B. F. *et al.* Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308,** 1906–1908 (2005).
11. Haynes, B. F., Kelsoe, G., Harrison, S. C. & Kepler, T. B. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnol.* **30,** 423–433 (2012).
12. Mouquet, H. & Nussenzweig, M. C. Polyreactive antibodies in adaptive immune responses to viruses. *Cell. Mol. Life Sci.* **69,** 1435–1445 (2012).
13. Scheid, J. F. *et al.* Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458,** 636–640 (2009).
14. Chen, W. *et al.* All known cross-reactive HIV-1 neutralizing antibodies are highly divergent from germline and their elicitation may require prolonged periods of time. *AIDS Res. Human Retrovirol.* **24,** Abstract DA03-03 (2008).
15. Dimitrov, D. S. Therapeutic antibodies, vaccines and antibodyomes. *MAbs* **2,** 347–356 (2010).
16. Ma, B. J. *et al.* Envelope deglycosylation enhances antigenicity of HIV-1 gp41 epitopes for both broad neutralizing antibodies and their unmutated ancestor antibodies. *PLoS Pathog.* **7,** e1002200 (2011).
17. Pancera, M. *et al.* Crystal structure of PG16 and chimeric dissection with somatically related PG9: structure-function analysis of two quaternary-specific antibodies that effectively neutralize HIV-1. *J. Virol.* **84,** 8098–8110 (2010).
18. Xiao, X. *et al.* Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochem. Biophys. Res. Commun.* **390,** 404–409 (2009).
19. Alam, S. M. *et al.* Differential reactivity of germ line allelic variants of a broadly neutralizing HIV-1 antibody to a gp41 fusion intermediate conformation. *J. Virol.* **85,** 11725–11731 (2011).
20. Bonsignori, M. *et al.* Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* **85,** 9998–10009 (2011).
21. Mouquet, H. *et al.* Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature* **467,** 591–595 (2010).
22. Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333,** 1633–1637 (2011).
23. Hoot, S. *et al.* Recombinant HIV envelope proteins fail to engage germline versions of anti-CD4-binding site bNAbs. *PLoS Pathog.* **9,** e1003106 (2013).
24. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **105,** 7552–7557 (2008).
25. Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100,** 4144–4149 (2003).
26. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422,** 307–312 (2003).
27. Corti, D. *et al.* Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS ONE* **5,** e8805 (2010).
28. Gray, E. S. *et al.* The neutralization breadth of HIV-1 develops incrementally over four years and is associated with $CD4^+$ T cell decline and high viral load during acute infection. *J. Virol.* **85,** 4828–4840 (2011).
29. Klein, F. *et al.* Broad neutralization by a combination of antibodies recognizing the CD4 binding site and a new conformational epitope on the HIV-1 envelope protein. *J. Exp. Med.* **209,** 1469–1479 (2012).
30. Lynch, R. M. *et al.* The development of CD4 binding site antibodies during HIV-1 infection. *J. Virol.* **86,** 7588–7595 (2012).
31. Moore, P. L., Gray, E. S. & Morris, L. Specificity of the autologous neutralizing antibody response. *Curr. Opin. HIV AIDS* **4,** 358–363 (2009).
32. Moore, P. L. *et al.* Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. *J. Virol.* **85,** 3128–3141 (2011).
33. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85,** 11502–11519 (2011).
34. Liao, H. X. *et al.* Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *J. Exp. Med.* **208,** 2237–2249 (2011).
35. Tomaras, G. D. *et al.* Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82,** 12449–12463 (2008).
36. Scheid, J. F. *et al.* A method for identification of HIV gp140 binding memory B cells in human blood. *J. Immunol. Methods* **343,** 65–67 (2009).
37. Seaman, M. S. *et al.* Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.* **84,** 1439–1452 (2010).
38. Boyd, S. D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1,** 12ra23 (2009).
39. Bonsignori, M. *et al.* Two distinct broadly neutralizing antibody specificities of different clonal lineages in a single HIV-1-infected donor: implications for vaccine design. *J. Virol.* **86,** 4688–4692 (2012).
40. Haynes, B. F., Moody, M. A., Verkoczy, L., Kelsoe, G. & Alam, S. M. Antibody polyspecificity and neutralization of HIV-1: a hypothesis. *Hum. Antibodies* **14,** 59–67 (2005).
41. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301,** 1374–1377 (2003).
42. Shingai, M. *et al.* Most rhesus macaques infected with the CCR5-tropic SHIV(AD8) generate cross-reactive antibodies that neutralize multiple HIV-1 strains. *Proc. Natl Acad. Sci. USA* **109,** 19769–19774 (2012).
43. Alam, S. M. *et al.* The role of antibody polyspecificity and lipid reactivity in binding of broadly neutralizing anti-HIV-1 envelope human monoclonal antibodies 2F5 and 4E10 to glycoprotein 41 membrane proximal envelope epitopes. *J. Immunol.* **178,** 4424–4435 (2007).
44. Malherbe, D. C. *et al.* Sequential immunization with a subtype B HIV-1 envelope quasispecies partially mimics the *in vivo* development of neutralizing antibodies. *J. Virol.* **85,** 5262–5274 (2011).
45. Alam, S. M. *et al.* Human immunodeficiency virus type 1 gp41 antibodies that mask membrane proximal region epitopes: antibody binding kinetics, induction, and potential for regulation in acute infection. *J. Virol.* **82,** 115–125 (2008).
46. Montefiori, D. C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.* **Chapter 12,** Unit 12.11 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the National Institutes of Allergy and Infectious Diseases (NIAID) and by intramural National Institutes of Health (NIH) support for the NIAID Vaccine Research Center, by grants from the NIH, NIAID, AI067854 (the Center for HIV/AIDS Vaccine Immunology) and AI100645 (the Center for Vaccine Immunology-Immunogen Discovery). The authors thank J. Pritchett, H. Chen, D. Pause, M. Cooper, E. Solomon, J. Blinn, K. Yarborough, E. Friberg, M. Smith, A. Hogan, C. Peckels, A. Foulger and T. Jeffries for technical assistance, and J. Kircherr

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 reasonance 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_HDJ_H and 174 V_LJ_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 Becker[1], Betty Benjamin[1], Robert Blakesley[1], Gerry Bouffard[1], Shelise Brooks[1], Holly Coleman[1], Mila Dekhtyar[1], Michael Gregory[1], Xiaobin Guan[1], Jyoti Gupta[1], Joel Han[1], April Hargrove[1], Shi-ling Ho[1], Taccara Johnson[1], Richelle Legaspi[1], Sean Lovett[1], Quino Maduro[1], Cathy Masiello[1], Baishali Maskeri[1], Jenny McDowell[1], Casandra Montemayor[1], James Mullikin[1], Morgan Park[1], Nancy Riebow[1], Karen Schandler[1], Brian Schmidt[1], Christina Sison[1], Mal Stantripop[1], James Thomas[1], Pam Thomas[1], Meg Vemulapalli[1] & Alice Young[1]

[1]NISC 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[45] and neutralization assays[46].

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 deoxynucleotide 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[47] 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 (*VHDJH* and *VLJL*) 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 sequences inferred to have been rearranged to the same VHDJH and JH, and to have the correct CDR3 length. For each sequence, we counted the number of mismatches between the sequence and the presumed VHDJH 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 two (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 VH UCA sequences resulting in four unique amino acid sequences that were all produced and assayed for reactivity with the transmitted/founder envelope gp140 (Supplementary Table 5).

Phylogenetic trees. Maximum-likelihood phylograms were generated using the dnaml program of the PHYLIP package (version 3.69) using the inferred ancestor as the outgroup root, 'speedy/rough' disabled, and default values for the remaining parameters. For the large antibody data sets, neighbour-joining phylogenetic trees were generated using the EBI bioinformatics server (http://www.ebi.ac.uk/Tools/phylogeny/) using default parameter values. All neighbour-joining trees were generated subsequent to the inference of the unmutated ancestors.

Isolation and expression of *VHDJH* and *VLJL* genes. The *VHDJH* and *VLJL* gene-segment pairs of the observed CH103, CH104 and CH106 antibodies, and the *VHDJH* gene segment of CH105 were amplified by reverse transcription followed by PCR (RT–PCR) of flow-sorted HIV-1 Env RSC3-specific memory B cells using the methods described previously[5–7,22,36]. To compare VH mutation frequency of CH103, CH104, CH105 and CH106 antibodies with that of previously published of CD4-binding site BnAbs VRC01, CH31 and NIH45-46, VH sequences of these antibodies were aligned to the closest *VH* gene segment from the IMGT reference sequence set, and differences between the target sequence and the *VH* gene segment up to and including the second invariant cysteine were counted. The comparison 3′ of Cys 2 is omitted because the unmutated form of the ancestral sequence is not as well known.

Additional *VHDJH* and *VLJL* genes were identified by 454 pyrosequencing. Clonally related *VHDJH* and *VLJL* sequences derived from either sorted single B cells or 454 pyrosequencing were combined and used to generate neighbour-joining phylogenetic trees (Fig. 2a, b). Antibodies that were recovered from single memory B cells are noted in the figure in red, and bold lines show the inferred evolutionary paths from the UCA to mature BnAbs. For clarity, related *VH* variants that grouped within monophyletic clades from the same time point were collapsed to single branches, condensing 457 *VHDJH* and 174 *VLJL* variants to 119 and 46 branches, respectively, via the 'nw_condense' function from the Newick Utilities package (v. 1.6)[49]. The frequencies of *VHDJH* variants in each B-cell sample are shown to the right of the VHDJH tree in Fig. 2a, and were computed from sample sizes of 188,793, 186,626 and 211,901 sequences from weeks 53, 92 and 144, respectively. Two *VHDJH* genes (IZ95W and 02IV4) were found at 14 weeks after transmission and paired with UCA *VLJL* for expression as IgG1 monoclonal antibodies. The IZ95W monoclonal antibody weakly bound the CH505 transmitted/founder Env gp140 with an end-point titre of 11 $\mu g\,ml^{-1}$. Among heavy-chain sequences in the tree, the mean distance of each to its nearest neighbour was calculated to be 8.1 nucleotides. The cumulative distribution function shows that, although there are pairs that are very close together (nearly 30% of sequences are 1 nucleotide from its neighbour), 45% of all sequences differ by 6 nucleotides or more from its nearest neighbour. The probability of generating a sequence that differs by 6 or more nucleotides from the starting sequence by PCR and sequencing is very small. The numbers of sequences obtained from a total of 100 million PBMCs were within the expected range of 50 to 500 antigen-specific B cells.

We have analysed the number of unique *VHDJH* and *VLJL* genes that we have isolated in several ways. First, we have clarified the calculations for the possible number of antigen-specific CD4-binding site memory B cells that could have been isolated from the samples studied. We studied five patient CH505 time points with pyrosequencing, with ~20 million PBMCs per time point for a total of 100 million PBMCs studied. In chronic HIV infection, there is a mean of 145 total B cells per microlitre of blood, and 60 memory B cells per microlitre of blood[50]. This high percentage of memory B cells of ~40% of the total B cells in chronic HIV infection is due to selective loss of naive B cells in HIV infection. Thus, in 100 ml of blood, there will be approximately 6 million memory B cells. If 0.1–1.0% are antigen specific, that would be 6,000–60,000 antigen-specific B cells sampled, and if, of these, 5% were CD4-binding site antibodies, then from 300 to 3000 CD4-binding site B cells would have been sampled in 100 million PBMCs studied. We studied 100 million PBMCs, therefore there should, by these calculations, be 1,000 CD4-binding site B cells sampled. This calculation therefore yields estimates that are completely compatible with the 474 *VHDJH* genes amplified.

To study the plausibility of sequences isolated further, the second method of analysis we used was as follows. Among heavy-chain sequences in the tree, one can compute the distance of each to its nearest neighbour. The mean distance to the nearest neighbour is 8.1 nucleotides. The cumulative distribution function shows that, although there are pairs that are very close together (nearly 30% of sequences are 1nt from its neighbour), 45% of all sequences differ by 6 nucleotides or more from its nearest neighbour. The probability of generating a sequence that differs by 6 or more nucleotides from the starting sequence by PCR and sequencing is very small. We believe the number of genes represented in our sample is closer to 200 than to 50, and most likely is larger than 200.

The third analysis we performed was to compute the distance of each heavy-chain sequences in the tree to its nearest neighbour. The mean distance to the nearest neighbour is 8.1 nucleotides. We used agglomerative clustering to prune the sequence alignment. At the stage where no pairs of sequences were 3 nucleotides apart or closer, there were 335 out of 452 sequences remaining; when no pairs are 6 nucleotides apart or closer, there are still 288 sequences remaining. Therefore, with this analysis, we believe the number of genes represented in our sample is closer to 300 than to 50, and may be larger. Thus, by the sum of these re-analyses, we believe that the number of genes in the trees in Fig. 2 is plausible.

The isolated Ig *VHDJH* and *VLJL* gene pairs, the inferred UCA and intermediate *VHDJH* and *VLJL* sequences, and select *VHDJH* gene sequences identified by pyrosequencing were studied experimentally (Supplementary Table 2), and used to generate a phylogenetic tree showing the percentage of mutated *VH* sites and time of appearance after transmission (Fig. 2c) and binding affinity (Fig. 2d). The isolated four mature antibodies are indicated in red, antibodies derived from 454 pyrosequencing are indicated in black, and inferred-intermediate antibodies (I1–I4, I7 and I8) are indicated by circles at ancestral nodes. The deep clades in this tree had modest bootstrap support, and the branching order and UCA inference were altered when more sequences were added to the phylogenetic analysis (compare the branching order of Fig. 2a and c). The tree depicted in Fig. 2c, d was used to derive the ancestral intermediates of the representative lineage early in our study, and marked an important step in our analysis of antibody affinity maturation. The *VHDJH* and *VLJL* genes were synthesized (GenScript) and cloned into a pcDNA3.1 plasmid (Invitrogen) for production of purified recombinant IgG1 antibodies as described previously[51,52]. The *VHDJH* genes of I1–I4, I7 and I8 as well as the *VHDJH* of CH105 were paired with either the *VL* gene of the inferred UCA or I2 depending on the genetic distance of the

VHDJH to either the UCA or mature antibodies for expressing as full-length IgG1 antibodies as described[51] (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 transmitted/founder Env were obtained from acutely infected HIV-1 subjects by single genome amplification[24], codon-optimized by using the codon usage of highly expressed human housekeeping genes[53], *de novo* synthesized (GeneScript) as gp140 or gp120 (AE.427299) and cloned into a mammalian expression plasmid pcDNA3.1/hygromycin (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-agarose (Vector Labs) column chromatography[16,52,54], and stored at −80 °C. Select Env proteins made as CH505 transmitted/founder Env were further purified by superose 6 column chromatography to trimeric forms, and used 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[34,52]. Plasma samples in serial threefold dilutions starting at 1:30 to 1:521,4470 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 K_d and rate constant (association rate (K_a)) measurements of monoclonal antibodies and all candidate UCAs to the autologous Env C. CH05 gp140 and/or the heterologous Env B.63521 gp120 were carried out on BIAcore 3000 instruments as described previously[19,43,45]. 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[55]. 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 herologous HIV-1 Env-pseudotyped viruses in TZM-bl-based neutralization assays using the methods as described[5,37,55]. Neutralization breadth of CH103 was determined using a previously described[5,37] panel of 196 of geographically and genetically diverse Env-pseudoviruses representing the major circulated genetic subtypes and circulating recombinant forms. The subtypes shown in Fig. 1c are consistent with previous publications[5,56], and the clades described in Los Alamos database (http://www.hiv.lanl.gov). HIV-1 subtype robustness is derived from the analysis of HIV-1 clades over time[57]. 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[7]. The extended gp120 core of HIV-1 clade C ZM176.66 was used to form a complex with Fab CH103 by using previously described methods[58]. In brief, deglycosylated ZM176.66, constructed as an extended gp120 core[59], that was produced using the method as described previously[7] and Fab CH103 were mixed at a 1:1.2 molar ratio at room temperature and purified by size-exclusion chromatography (Hiload 26/60 Superdex S200 prep grade, GE Healthcare) with buffer containing 0.35 M NaCl, 2.5 mM Tris, pH 7.0 and 0.02% NaN_3. Fractions of the Fab or gp120–CH103 complex were concentrated to ~10 mg ml^{-1}, flash frozen with liquid nitrogen before storing at −80 °C and used for crystallization screening experiments.

Commercially available screens, Hampton Crystal Screen (Hampton Research), Precipitant Synergy Screen (Emerald BioSystems), Wizard Screen (Emerald BioSystems), PACT Suite and JCSG+ (Qiagen) were used for initial crystallization screening of both Fab CH103 and its gp120 complex. Vapour-diffusion sitting drops were set up robotically by mixing 0.2 μl of protein with an equal volume of precipitant solutions (Honeybee 963, DigiLab). The screen plates were stored at 20 °C and imaged at scheduled times with RockImager (Formulatrix.). The Fab CH103 crystals appeared in a condition from the JCSG+ kit containing 170 mM ammonium sulphate, 15% glycerol and 25.5% PEG 4000. For the gp120–CH103 complex (Supplementary Table 8), crystals were obtained after 21 days of incubation in a fungi-contaminated[60,61] droplet of the PACT suite that contained 200 mM sodium formate, 20% PEG 3350 and 100 mM bistrispropane, pH 7.5.

X-ray data collection and structure determination for gp120–CH103. Diffraction data were collected under cryogenic conditions. Optimal cryoprotectant conditions were obtained by screening several commonly used cryoprotectants as described previously[7]. X-ray diffraction data were collected at beam-line ID-22 (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory, with 1.0000 Å radiation, processed and reduced with HKL2000 (ref. 62). For the Fab CH103 crystal, a data set at 1.65 Å resolution was collected with a cryo-solution containing 20% ethylene glycol, 300 mM ammonium sulphate, 15% glycerol and 25% PEG 4000 (Supplementary Table 8). For the gp120–CH103 crystals, a data set at 3.20 Å resolution was collected using a cryo-solution containing 30% glycerol, 200 mM sodium formate, 30% PEG 3350 and 100 mM bistrispropane, pH 7.5 (Supplementary Table 8).

The Fab CH103 crystal was in the $P2_1$ space group with cell dimensions at $a = 43.0$, $b = 146.4$, $c = 66.3$, $\alpha = 90.0$, $\beta = 97.7$ and $\gamma = 90.0$, and contained two Fab molecules per asymmetric unit (Supplementary Table 8). The crystal structures of Fab CH103 were solved by molecular replacement using Phaser[63] in the *CCP*4 program suite[64] with published antibody structures as searching models. The gp120–CH103 crystal also belonged to the $P2_1$ space group with cell dimensions at $a = 48.9$, $b = 208.7$, $c = 69.4$, $\alpha = 90$, $\beta = 107.2$ 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 gp120 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[60,61].

Structural refinements were carried out with PHENIX[65]. Starting with torsion-angle simulated annealing with slow cooling, iterative manual model building was carried out on COOT[66] with maps generated from combinations of standard positional, individual *B*-factor, TLS (translation/libration/screw) refinement algorithms and non-crystallographic symmetry (NCS) restraints. Ordered solvents were added during each macro cycle. Throughout the refinement processes, a cross validation (R_{free}) 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[67] and pdb-care[68]. X-ray crystallographic data and refinement statistics are summarized in Supplementary Table 8. The Kabat nomenclature[69] was used for numbering of amino acid residues in amino acid sequences in antibodies.

Protein structure analysis and graphical representations. PISA[70] was used to perform protein–protein interfaces analysis. *CCP*4 (ref. 66) was used for structural alignments. All graphical representation with protein crystal structures were made with Pymol[71].

Polyreactivity analysis of antibodies. All antibodies in CH103 clonal lineage were assayed at 50 μg ml^{-1} for autoreactivity to HEp-2 cells (Inverness Medical Professional Diagnostics) by indirect immunofluorescence staining and a panel of autogens by antinuclear antibody assays using the methods as reported previously[10]. 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 ProtoArray 5 microchip (Invitrogen) according to the instructions of the microchip manufacturer. In brief ProtoArray 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

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 CstF-77. *Acta Crystallogr. F* **63,** 135–138 (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. Otwinowski, 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 *CCP*4 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. Lütteke, T. & von der Lieth, C. W. pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics* **5,** 69 (2004).
69. Kabat, E. A., Wu, T. T., Gottesman, K. S. & Foeller, C. *Sequences of Proteins of Immunological Interest* 5th edn (DIANE publishing, 1991).
70. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
71. DeLano, W. L. The PyMOL Molecular Graphics System http://www.pymol.org (DeLano Scientific, 2002).