Published in final edited form as:

Exp Mol Pathol. 2012 December ; 93(3): 399–407. doi:10.1016/j.yexmp.2012.09.022.

Characterization of human IgG repertoires in an acute HIV-1 infection

Weizao Chen^{a,*}, Ponraj Prabakaran^{a,b,*}, Zhongyu Zhu^a, Yang Feng^a, Emily D. Streaker^{a,b}, and Dimitar S. Dimitrov^a

^aProtein Interactions Group, National Cancer Institute, National Institutes of Health (NIH), Frederick, Maryland 21702, USA

^bBasic Research Program, Science Applications International Corporation-Frederick, Inc., Frederick National Laboratory, Frederick, Maryland 21702, USA

Abstract

All known broadly neutralizing antibodies (bnAbs) are highly somatically mutated and therefore significantly differ from their germline predecessors. Thus although the mature bnAbs bind to conserved epitopes of the HIV-1 envelope glycoprotein (Env) with high affinity their germline predecessors do not or weakly bind Envs failing to initiate an effective immune response. The identification of less somatically mutated bnAbs and/or antibody maturation intermediates that are clonally related to bnAbs may be useful to circumvent the major problem of initiating immune responses leading to elicitation of bnAbs. Here, we describe the identification of IgG antibodies from an acutely HIV-1-infected patient using a combination of phage display and high-throughput sequencing. We found two antibodies with only a single point mutation in the V region of their heavy chain variable domains compared to their putative germline predecessors which bound with high affinity to several Envs. They targeted the Env gp41 and did not neutralize HIV-1. Using high-throughput sequencing, we identified several highly abundant CDR3s, germline-like as well as somatically mutated V genes in the VH/VL repertoires of the patient which may provide antibody intermediates corresponding to known bnAbs as templates for design of novel HIV-1 vaccine immunogens.

Keywords

HIV-1; Human monoclonal antibody; IgG; gp140; Envelope glycoprotein; Immunogen; High-throughput sequencing; Vaccine

Introduction

Advancement in high-throughput screening technologies has led to the recent discoveries of several potent broadly neutralizing antibodies (bnAbs) against HIV-1 recovered from peripheral blood mononuclear cells (PBMCs) of HIV-1-seropositive donors (Pietzsch et al., 2010; Scheid et al., 2009; Walker et al., 2011; Wu et al., 2010, 2011). Elicitation of such bnAbs remains a major challenge attributing to virus evasion strategies and host immune regulatory mechanisms. We found that germline-like predecessors of bnAbs bind weakly or undetectably to all tested HIV-1 envelope glycoproteins (Envs) (Xiao et al., 2009). This

*Corresponding author at: Protein Interactions Group, NCI, NIH, Bldg 469, Rm 150B, Frederick, MD 21702, USA. Fax: +1 301 846 5598. chenw3@mail.nih.gov (W. Chen), prabakaran.ponraj@nih.gov (P. Ponraj).

Conflict of Interest statement The authors declare that there are no conflicts of interest.

finding suggests that HIV-1 could have evolved a strategy to reduce or eliminate the immunogenicity of the highly conserved epitopes of bnAbs by using holes in the human germline B cell receptor repertoires — absence of or reduced binding of germline antibodies to the conserved epitopes that is not sufficient to initiate and/or maintain an effective immune response. To overcome this fundamental issue, we resorted to explore large naïve IgM repertoires for identifying antibody maturation intermediates that are clonally related to HIV-1 bnAbs (Prabakaran et al., 2012b). We also found that several human monoclonal antibodies (mAbs) selected from another large naïve IgM phage-displayed library were able to bind with high affinity to recombinant Envs of HIV-1 isolates from different clades (Chen et al., 2010). Although those antibodies enhanced or did not neutralize infection by some of the HIV-1 primary isolates, they could have implications for the B-cell-lineage vaccine design strategy (Dimitrov, 2010; Haynes et al., 2012).

For this study, we constructed two IgG antigen-binding fragment (Fab) phage display libraries from an acutely HIV-1-infected patient, which were panned against the Env to identify HIV-1 specific binders. We analyzed the genetic origin, diversity and level of maturation of the selected antibody binders; in parallel, we used high-throughput sequencing to analyze the extent of germline diversity, complementarity determining region 3 (CDR3) lengths, somatic mutation and most frequently expressed clones in the two libraries. We found germline-lineaged antibodies exhibiting cross-reactivity against the Envs, in contrast to bnAbs and other antibodies capable of binding to Env which respective germline-versions are unable to bind (Chen et al., 2010; Xiao et al., 2009). This may have significant bearing on the development of predecessor antibodies useful in the new vaccine design approach. Further, combined phage display and high-throughput sequencing methods helped identify several long, highly abundant CDR3s and highly mutated V-genes in V_H/V_L chains of the HIV-1-infected patient which may be potential candidates useful in the development of effective HIV-1 vaccines.

Materials and methods

Library construction and selection of antibodies against HIV-1 Envs

We isolated mRNA from frozen PBMCs that had been derived from an HIV-1 patient at two different time points in the period of about 40 days and 8 months post infection, converted them into cDNA, and constructed two separate Fab libraries encoding IgG heavy and light chains as briefly described previously (Chen et al., 2008). The two libraries were panned against Envs, a homologous gp140 and a consensus gp140, respectively. Procedures followed in this study were in accordance with the ethical standards of concerned institutional policies and the Research Donor Program of Frederick National Laboratory.

Expression, purification and binding of antibodies

Soluble Fabs of antibodies were expressed in *E. coli* and purified by affinity chromatography methods. To analyze the genetic origin and properties of the selected antibodies, heavy and light chain variable domains of these binders were sequenced. Further, binding and competition ELISAs were performed as described previously (Chen et al., 2010).

Analysis of antibody sequence diversity from the libraries and statistical calculations

The heavy (V_H) and light (V_L) chain variable domains of antibodies from the two libraries were sequenced using the high-throughput 454 sequencing pipeline method. The complete set of primers used in the PCR amplification of heavy and light chains, which also included adaptor sequences along with target amplification sequence, were described in detail elsewhere (Prabakaran et al., 2012a). For quality control of antibody sequences, we trimmed

the sequence data and retained only sequences of length more than 300 nucleotides (nt), covering the entire antibody variable domains consisting of the three CDRs along with framework regions (FRs). We used the IMGT/HighV-QUEST analysis tool for genetic diversity analysis of antibody sequences (Alamyar et al., 2010). The output results from the IMGT/HighV-QUEST analysis in CSV files were stored at PostgreSQL database, and Structured Query Language (SQL) was effectively used to retrieve the data for the sequence analysis. Statistical calculations involving germline usage, distribution of antibody CDR3 lengths and somatic mutations were carried out using SAS JMP10® statistical software (SAS Institute, Cary, NC).

Results

Construction of antibody libraries from an HIV-1 patient and selection of antibodies against HIV-1 Envs

HIV-1 patient blood obtained from two time points at approximately 40 days and 8 months post infection was used to construct two Fab phage display libraries. To identify human IgG-derived antibodies specific for HIV-1 Envs, we panned the libraries against the homologous gp140 (CH12.0544.2 gp140) isolated from the patient and a consensus gp140 (Cons gp140) designed by aligning >1,000 sequences of group M. Upon panning of the library generated using serum collected at the first time point, we were able to select six unique specific binders (Fig. 1). Selection of antibodies were performed and screened by ELISA for specific binding to the targets and not to unrelated bovine serum albumin (BSA) (Fig. 2). From the other library generated using blood collected at the second time point with CH12.0544.2 gp140, we identified one binder by panning against Cons gp140 (Fig. 3).

High level of sequence diversity and low level of somatic diversification of selected antibodies

To analyze genetic origin, diversity and level of maturation of the selected antibodies, we sequenced all the antibodies and performed the IMGT V-QUEST analysis (Lefranc et al., 2008). Among the six antibodies isolated from the first library made of the sample at ~40 days post infection, there were two groups of antibodies – group 1 with ma9, ma7, ma12 and ma4 antibodies sharing the V_H germline IGHV1-46*01 and V_L germline IGLV3-25*03 (Fig. 1A) while group 2 with ma5 and ma11 antibodies sharing the V_H germline IGHV5-51*01 and V_L germline IGKV2-28*01 (Fig. 1B). Alignments of the antibodies from the two groups along with their germline genes, and annotated FRs and CDRs are shown in the Fig. 1. The only antibody, mc5, selected from the second library made of the sample at ~8 months post infection was found to have V_H germline IGHV3-48*02 and V_L germline IGKV3-20*01 (Fig. 3A). The numbers of somatic mutations (Figs. 1 and 3A with highlighted in gray) observed in the selected antibodies were found to be relatively very low as compared to the HIV-1 bnAbs. Remarkably, the lengths of V_H CDR3s of the four antibodies that shared GHV1-46*01 germline were only five amino acids (AA) long. The selected binders originated from three different IGHV and IGKV/IGLV germline genes indicating sequence diversity albeit low level of somatic diversification.

Characterization of HIV-1 Env binding, competition and epitopes of selected antibodies

To approximately map antibody epitopes and understand the underlying mechanisms of antibody binding and functional activities, we measured binding activities of selected antibodies with Envs from different isolates and performed antibody competition assays with well-characterized antibodies. Fig. 2 shows the binding of six antibodies from two groups; four antibodies in group 1 showed specific binding to three different Envs, CH12.0544.2 gp140, JRFL gp140 and Cons gp140, to different extents (Figs. 2A–C); and two antibodies in group 2 also showed similar specific binding to the Envs except for JRFL

gp140 (Figs. 2D–F). Among these antibodies, ma9 from group 1 and ma5 from group 2 were found to have higher binding activity than the others although closer to the corresponding germlines, just one AA point mutation compared to their closest IGHV germlines (Figs. 1A and B). We measured the binding of the six antibodies to a gp41-Fc fusion protein, 89.6 gp41Fc (Fig. 4A). ma9 showed significant binding, the binding of other antibodies was very weak but measurable. Although cross-reactive binding activities of the mAbs were observed, they did not show any neutralizing activities against the primary HIV-1 isolates (data not shown). To find out the approximate binding sites of two antibodies (m9 and m5), we measured competition of ma9 (Figs. 4B) and ma5 (Figs. 4C) with well-characterized mAbs, including b12Fc, m9Fc, 4A18Fc, IgG1 m43, IgG1m44, IgG1 2F5 and IgG1 4E10, in binding to Envs. We found that ma9 did not compete with the known mAbs, but ma5 competed with IgG1 m44 in binding to the Envs. Further, we measured the binding activity of mc5 antibody selected from the second time point library with different Envs. It showed specific binding to Cons gp140 and CH12.0544.2 gp140 but not to JRFL gp140 and Bal gp120 (Fig. 3B). These results suggest that the antibodies target gp41, and that ma5 epitope could be shared by m44, which is an anti-gp41 antibody (Zhang et al., 2008).

High-throughput sequence analysis reveals germline usages, CDR3 length diversity, somatic mutations and clonally related antibodies

We employed high-throughput 454 sequencing method to characterize distinct antibodies produced in the two libraries, and selected hundreds of unique V_H and V_L (κ and λ) antibody sequences that were found as good quality productive sequences without any stop codons and/or frameshifts (Table 1). Fig. 5 shows the frequencies of the IGHV subgroups (Fig. 5A) and IGKV and IGLV subgroups (Fig. 5B). As previously observed in high-throughput sequencing studies of antibody repertoires (Boyd et al., 2010; Glanville et al., 2009; Prabakaran et al., 2012a), we noted biased germline usages in the library sequences but representatives from most of the subgroups were present. We observed that, in the IGHV subgroups, HV1-46, HV3-33 and HV5-51 germlines were most frequently used (10% for each subgroup) in the sequences of first library corresponding to ~40 days post infection serum. KV3-20 (15%), KV2-28 (8%) and KV3-11 (8%) dominated in the V_L germlines of that library. In the second library corresponding to ~8 months post infection serum, HV1-69 (26%), HV5-51 (12%) and HV1-2 (10%) were most frequently used IGHV germline families. LV3-21 (15%), LV1-44(10%), LV1-40 (8%) and KV3-20 (6%) were dominating the corresponding V_L germline families.

The productive V_H CDR3 (positions 105–117 in IMGT numbering) is the result of the V-D-J rearrangement and largely determines the diversity of expressed antibody V_H repertoire. We calculated the distribution of V_H CDR3 lengths in the two libraries, which ranged from 4 to 27 AA lengths (Fig. 6A). We also calculated the total numbers of AA changes in the V regions of V_H and V_L (κ and λ) for the two libraries and compared in box-plot diagram (Fig. 6B). The mean values of numbers of somatic mutation of IGHVs were 10.4 and 13.7 for ~40 days and ~8 months post infection, respectively. The mean values of numbers of somatic mutations for IGKV/IGLV ranged from 7 to 9. However, there were antibodies with minimum mutations, germline-like or intermediates antibodies, as well as hypermutated ones (Fig. 6B).

To find out most frequently expressed clones in the two antibody libraries we calculated the frequencies of CDR3s and associated IGHV, IGKV and IGLV germline genes. We found several highly abundant CDR3s that occurred more than 5 times, which were listed in Tables 2 and 3, respectively, for the first and second libraries. We found that, at ~40 days post infection, a particular V_H CDR3 with the IGHV3-33 germline lineage occurred 71 times (7% of the total V_H sequences obtained). The same V_H CDR3 was also observed in more than one V_H germline but less abundant — 4 times in IGHV3-30, 2 times in IGHV3-11,

IGHV3-13, IGHV3-21, IGHV3-15 and 1 time in IGHV4-61, IGHV3-23, IGHV3-20, and IGHV3-48. The same V_H CDR3 sequence to occur in more than one V_H germline sequence was previously noted in antibodies selected against the BLyS antigen (Edwards et al., 2003). In the V_L sequences of that library, the most abundant V_L CDR3 with the IGKV3-11 germline lineage occurred 30 times (2.5% of the total V_κ sequences obtained). In the second time point library, at ~8 months post infection, the most frequent CDR3s were from IGHV1-69, which occurred 8 times (0.9% of the total V_H sequences obtained) and IGLV3-21, which occurred 16 times (1.1% of the total V_L sequences obtained). We noted that the four group-1 antibodies from the first time point library had a 5 AA long V_H CDR3 with AA sequence ARFDY, one of the highly abundant CDR3s. Although that V_H CDR3 occurred 12 times, a point mutation at the third residue “F” into other hydrophobic residues (Y, I, L and M) resulted in a total of 34 similar V_H CDR3s, making it second most abundant V_H CDR3 (3.4% of the total V_H sequences obtained). Interestingly, that 5 residue long V_H CDR3 was always found to be associated with the IGHV1-46 germline, except one instance with IGHV1-2 and IGHV1-58 germline each. Therefore, some of the highly abundant V_H and V_L CDR3s revealed the clonally related antibodies or B cell clonal expansion in the acute HIV-1 patient.

Discussion

Phage display enables the rapid selection of antibodies from large libraries by nature of their binding to almost any target antigen (Smith, 1985). High-throughput sequencing can be applied to explore antibody variable region repertoires and to bypass *in vitro* screening or immunization steps (Cheung et al., 2012; Reddy et al., 2010). Importantly, massive amounts of sequence data encoding large antibody repertoires are useful for germline-lineage and maturation analyses of binders selected by display methods or immunizations. In this study, we used a combined approach of phage display and high-throughput sequencing technologies to identify cross-reactive anti-HIV-1 antibodies from an acutely HIV-1-infected patient. PBMCs were obtained at approximately 40 days and 8 months post infection and were used to construct two Fab format phage display libraries from which selection of antibodies was performed against HIV-1 gp140. We isolated six unique antibodies from the first library forming two groups based on different V gene germline origins and analyzed their variable region sequences (Fig. 1). These antibodies were able to bind specifically with different Envs, CH12.0544.2 gp140, JRFL gp140 and Cons gp140, showing the potential cross-reactivity (Fig. 2), except the group 2 antibodies, ma5 and ma11, that did not show any binding to JRFL gp140. We also sequenced the binder selected using the second library, which showed binding to Cons gp140 and CH12.0544.2 gp140 but not to JRFL gp140 and Bal gp120 (Fig. 3). All six antibodies from the first time point library were tested for gp41 binding, and then ma9 and ma5 were further characterized by using competition assays with known mAbs in binding to Env targets (Fig. 4), which suggested that the two antibodies targeted gp41. We found that these two antibodies used almost IGHV germplines (with a single point mutation at the FR2 for ma9 and at the CDR1 for ma5), and exhibited cross-reactivity against Envs. This finding could be significant as previously identified HIV-1 bnAbs as well as other antibodies in their respective germline versions do not bind to Envs (Chen et al., 2010; Xiao et al., 2009).

To find the extent of diversity, frequently expressed clones or highly abundant CDR3s and analyze clonally-related sequences of the selected binders in acute HIV-1 patient libraries, we performed high-throughput sequencing for the two libraries, and processed hundreds of unique V_Hs and V_Ls (Table 1). Despite limited sequencing depth, we observed most of the V genes representing different subgroups of IGHV and IGKV/IGLV (Fig. 5), and a wide range of V_H CDR3 lengths (Fig. 6A). We noted that some of the most dominant V_H and V_L

chains identified by high-throughput sequencing, for example HV1-46, HV5-51 and KV2-28, corresponded to the germlines of phage display selected antibodies.

The V_H CDR3 length distribution in the two libraries ranged from 4 to 27 AA lengths (Fig. 6A). Notably, at ~40 days post infection, a particular V_H CDR3 with IGHV3-33 gene was found to be the most dominant and accounted for 7% of the total V_H repertoire. The second most dominant V_H CDR3 with IGHV1-46 gene accounted for 3.4% of the total V_H sequences. Actually, the second most dominant clone was found the same as the one in group 1 antibodies selected by phage display. These and other highly abundant CDR3s of V_H and V_L in the libraries (Tables 2 and 3) revealed the clonally related antibodies and provided evidence for B cell clonal expansion in the HIV patient (Chong et al., 2001). We found that antibody diversity of the libraries was further increased by somatic mutation (Fig. 6B), and there were antibodies with low numbers of mutations or germline-like sequences, intermediates and hypermutations.

In conclusion, by panning of large phage-displayed antibody libraries derived from an HIV patient we identified germline-like antibodies that specifically bound to different Envs but were non-neutralizing. Further, high-throughput sequence analysis revealed germline V gene usages, V_H CDR3 length diversity, somatic mutations and B cell clonal expansion in the HIV patient. Thus, combined analysis of display and next-generation sequencing methods using different B-cell sources can facilitate rescuing of missed binders due to wrong baits or host tolerance mechanisms. It could also help in identifying intermediate antibodies binding to Envs and understanding maturation pathways of bnAbs that could lead to novel approaches for vaccine design.

Acknowledgments

We thank Drs. B. Haynes, H. Liao, C. Broder, and T. Fouts for reagents. We thank the Laboratory of Molecular Technology and Advanced Biomedical Computing Center of SAIC-Frederick Inc. for sequencing service. We thank Ms. Maria G. Singarayan for constructing the PostgreSQL database and developing JAVA applications. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, and the Gates Foundation to DSD, and by Federal funds from the NIH, National Cancer Institute, under Contract No. NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U. S. Government.

References

- Alamyar, E.; Giudicelli, V.; Duroux, P.; Lefranc, MP. IMGT/HighV-QUEST: A High-Throughput System and Web Portal for the Analysis of Rearranged Nucleotide Sequences of Antigen Receptors. Paper presented at: Journées Ouvertes en Biologie, Informatique et Mathématiques (Montpellier, France); 2010.
- Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *Journal of Immunology*. 2010; 184:6986–6992.
- Chen W, Zhu Z, Feng Y, Dimitrov DS. Human domain antibodies to conserved sterically restricted regions on gp120 as exceptionally potent cross-reactive HIV-1 neutralizers. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:17121–17126. [PubMed: 18957538]
- Chen W, Zhu Z, Liao H, Quinnan GV Jr. Broder CC, Haynes BF, Dimitrov DS. Cross-reactive human IgM-derived monoclonal antibodies that bind to HIV-1 envelope glycoproteins. *Viruses*. 2010; 2:547–565. [PubMed: 21755021]
- Cheung WC, Beausoleil SA, Zhang X, Sato S, Schieferl SM, Wieler JS, Beaudet JG, Ramenani RK, Popova L, Comb MJ, et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nature Biotechnology*. 2012; 30:447–452.

- Chong Y, Ikematsu H, Ariyama I, Chijiwa K, Li W, Yamaji K, Kashiwagi S, Hayashi J. Evidence of B cell clonal expansion in HIV type 1-infected patients. *AIDS Research and Human Retroviruses*. 2001; 17:1507–1515. [PubMed: 11709095]
- Dimitrov DS. Therapeutic antibodies, vaccines and antibodyomes. *MAbs*. 2010; 2:347–356. [PubMed: 20400863]
- Edwards BM, Barash SC, Main SH, Choi GH, Minter R, Ullrich S, Williams E, Du Fou L, Wilton J, Albert VR, et al. The remarkable flexibility of the human antibody repertoire; isolation of over one thousand different antibodies to a single protein, BLyS. *Journal of Molecular Biology*. 2003; 334:103–118. [PubMed: 14596803]
- Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L, Sundar PD, Day GM, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:20216–20221. [PubMed: 19875695]
- Haynes BF, Kelsoe G, Harrison SC, Kepler TB. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnology*. 2012; 30:423–433.
- Lefranc MP, Brochet X, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research*. 2008; 36:W503–W508. [PubMed: 18503082]
- Pietzsch J, Scheid JF, Mouquet H, Klein F, Seaman MS, Jankovic M, Corti D, Lanzavecchia A, Nussenzweig MC. Human anti-HIV-neutralizing antibodies frequently target a conserved epitope essential for viral fitness. *The Journal of Experimental Medicine*. 2010; 207:1995–2002. [PubMed: 20679402]
- Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streaker E, Feng Y, Dimitrov DS. Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics*. 2012a; 64:337–350. [PubMed: 22200891]
- Prabakaran P, Zhu Z, Chen W, Gong R, Feng Y, Streaker E, Dimitrov DS. Origin, diversity, and maturation of human antiviral antibodies analyzed by high-throughput sequencing. *Frontiers in Microbiology*. 2012b; 3:277. [PubMed: 22876240]
- Reddy ST, Ge X, Miklos AE, Hughes RA, Kang SH, Hoi KH, Chrysostomou C, Hunnicke-Smith SP, Iverson BL, Tucker PW, et al. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nature Biotechnology*. 2010; 28:965–969.
- Scheid JF, Mouquet H, Feldhahn N, Seaman MS, Velinzon K, Pietzsch J, Ott RG, Anthony RM, Zebroski H, Hurley A, et al. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature*. 2009; 458:636–640. [PubMed: 19287373]
- Smith GP. Filamentous fusion phage — novel expression vectors that display cloned antigens on the virion surface. *Science*. 1985; 228:1315–1317. [PubMed: 4001944]
- Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*. 2011; 477:466–470. [PubMed: 21849977]
- Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*. 2010; 329:856–861. [PubMed: 20616233]
- Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*. 2011; 333:1593–1602. [PubMed: 21835983]
- Xiao X, Chen W, Feng Y, Zhu Z, Prabakaran P, Wang Y, Zhang MY, Longo NS, Dimitrov DS. 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. *Biochemical and Biophysical Research Communications*. 2009; 390:404–409. [PubMed: 19748484]
- Zhang MY, Vu BK, Choudhary A, Lu H, Humbert M, Ong H, Alam M, Ruprecht RM, Quinnan G, Jiang S, et al. Cross-reactive human immunodeficiency virus type 1-neutralizing human

monoclonal antibody that recognizes a novel conformational epitope on gp41 and lacks reactivity against self-antigens. *Journal of Virology*. 2008; 82:6869–6879. [PubMed: 18480433]

Fig. 1. The amino acid sequences of heavy and light chains of the selected group 1 (A) and group 2 (B) mAbs in alignment with the corresponding germlines of human antibody V genes. The CDRs and FRs are indicated according to the ImMunoGeneTics annotation (<http://imgt.cines.fr/>). The somatic mutations in the V genes of the selected antibodies are highlighted with gray color.

Fig. 2. Binding of four antibodies (group 1) to Envs CH12.0544.2 gp140, JRFL gp140 and Cons gp140 (A–C), and binding of two antibodies (group 2) to Envs CH12.0544.2 gp140, JRFL gp140 and Cons gp140 (D–F).

Fig. 3. The amino acid sequences of heavy and light chains of mc5 in alignment with the corresponding germlines of human antibody V genes (A) and (B) binding of mc5 to Envs. The CDRs and FRs are indicated according to the ImMunoGeneTics annotation (<http://imgt.cines.fr/>). The somatic mutations in the V genes of the selected antibodies are highlighted with gray color.

Fig. 4. Binding of the six antibodies to 89.6 gp41Fc (A), and competition of ma9 and ma5 with well-characterized mAbs 12Fc, 4A18Fc, IgG1 m43, IgG1 m44, IgG1 2F5 and IgG1 4E10 in binding to Envs (B).

Fig. 5. Gene usage frequencies observed for the IGHV subgroups (A), and IGKV and IGLV subgroups (B) in the two libraries.

Fig. 6. Distribution of V_H CDR3 lengths (A), and V_H and V_L (κ and λ) mutations (B) for the two libraries.

Table 1

Numbers of antibody sequences (AA) from IgG libraries generated from an acute HIV-1 patient serum at two time points using high-throughput sequencing method.

Data sets	VDJ/VJ			CDR3s		
	V _H	V _κ	V _λ	V _H	V _κ	V _λ
First time point (~40 days post infection)	1009	1199	687	611	785	529
Second time point (~8 months post infection)	878	466	1448	774	425	1215

Table 2

High-frequency CDR3s observed at ~40 days post infection. CDR3 sequences are listed according to the frequency (N). Shown are the CDR3s with N ≥ 5.

V _H CDR3	Length	IGHV	N
ARDLRVTYLDY	11	HV3-33	71
ARDGSGPVEFDF	12	HV4-61	30
ARDNLWSNGVDV	12	HV1-46	16
ARDARHYGDNDY	12	HV1-46	13
TRGLNY	6	HV4-34	13
ARFDY	5	HV1-46	12
ASSRRMVRGSWGFDY	15	HV1-2	12
ARRADHNSKFDY	12	HV6-1	10
ARDGWGRVELDY	12	HV4-61	9
ARRIGTRESTISFYGMDI	18	HV5-51	8
TTPGHPPALFDH	12	HV3-15	7
ARVDY	5	HV1-46	7
VKDMGWGTPATSPRADF	17	HV3-23	7
AKDYSNYVY	9	HV3-23	7
AREGERGGSYPGY	13	HV1-18	7
ARRGEINEWFDP	12	HV5-51	6
ARDGLGTVELDY	12	HV4-61	5
ARYDY	5	HV1-46	5
ARCDSSWYGRGRYHYGMDV	20	HV4-34	5
ARRIGTRESTITFYGMDI	18	HV5-51	5

V _L CDR3	Length	IGKV	N
QQRFNWPLT	9	KV3-11	30
QQRVNWPLFT	10	KV3-11	14
QQRVAWPLFT	10	KV3-11	12
HQSSSLPPWT	10	KV6-21	11
MQDSSYPLT	9	KV1-6	11
QQYGSLIT	8	KV3-20	10
QQSHSAPRT	9	KV1-39	10
MQSIYLPIT	9	KV2D-29	9
QQYGSSPRT	9	KV3-20	9
QQYGSSLFT	9	KV3-20	8
QQLDIYPLT	9	KV1-9	8
QHRIRGPPLLT	11	KV3-11	8
QQYNNWPPGLT	11	KV3-15	7
MQALQKLT	8	KV2-28	7
MQGLRPPHS	9	KV2-28	7
QQYGSSPLT	9	KV3-20	6
MQGLQTPIA	9	KV2-28	6

MQALQTPWT	9	KV2-28	6
QQYDRSRT	8	KV3-20	6
QQYGSSPYT	9	KV3-20	6
QQYSTYPLT	9	KV1-5	5
QQYNDWLS	8	KV3-15	5
QQYNNWPPWT	10	KV3-15	5
QQTYSKWA	8	KV1-39	5
QQYNNWPLT	9	KV3-15	5
QQYGDSPLT	9	KV3-20	5
QQYKSYRT	8	KV1-5	5

V _L CDR3	Length	IGLV	N
QVWDASSDHPVL	12	LV3-21	14
AAWDDSLNGVV	11	LV1-44	11
AAWDDSLSGVV	11	LV1-44	6
QAWDSSTVV	9	LV3-1	6
QVWDASRDHSWV	12	LV3-21	5
QAWDSSHVV	9	LV3-1	5
SSYAGSNNLV	10	LV2-8	5
TSYTNSNTGV	10	LV2-14	5

Table 3

High-frequency CDR3s observed at ~8 months post infection. CDR3 sequences are listed according to the frequency (N). Shown are the CDR3s with N ≥ 5.

V _H CDR3	Length	IGHV	N
ARVRSGYDLQGDY	13	HV1-69	8
ARGKQDCSSVTCWLEY	16	HV5-51	6
ARGYYETGYYYFDY	14	HV1-8	6
ARDWTADSGTFPGSY	15	HV1-46	5
ARGGRDAYGSGAHFDN	16	HV1-2	5

V _L CDR3	Length	IGKV	N
MQSTHWPPWT	10	KV2-30	6
LQHDNFPLT	9	KV5-2	5

V _L CDR3	Length	IGLV	N
QVWDSDS DHV	11	LV3-21	16
QAWDSSTVV	9	LV3-1	9
QSYDSSLGSGV	11	LV1-40	8
QVWDSSSDHV	11	LV3-21	6
QSYDSSLSGYV	11	LV1-40	6
AAWDDSLNGPwV	11	LV1-44	5
GTWDSLSAVV	11	LV1-51	5
QAWDSSTAV	9	LV3-1	5
AAWDDSLNAYV	11	LV1-44	5
QVWHSSSDHYV	11	LV3-21	5
QSYDSSLSGRGV	12	LV1-40	5
AAWDDSLNGYV	11	LV1-44	5