

Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120

Anthony P. West, Jr.^a, Ron Diskin^a, Michel C. Nussenzweig^b, and Pamela J. Bjorkman^{a,c,1}

^aDivision of Biology and ^cHoward Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125; and ^bLaboratory of Molecular Immunology and Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065

Contributed by Pamela J. Bjorkman, May 30, 2012 (sent for review May 2, 2012)

A large number of anti-HIV-1 antibodies targeting the CD4-binding site (CD4bs) on the envelope glycoprotein gp120 have recently been reported. These antibodies, typified by VRC01, are remarkable for both their breadth and their potency. Crystal structures have revealed a common mode of binding for several of these antibodies; however, the precise relationship among CD4bs antibodies remains to be defined. Here we analyze existing structural and sequence data, propose a set of signature features for potent VRC01-like (PVL) antibodies, and verify the importance of these features by mutagenesis. The signature features explain why PVL antibodies derive from a single germ-line human V_H gene segment and why certain gp120 sequences are associated with antibody resistance. Our results bear on vaccine development and structure-based design to improve the potency and breadth of anti-CD4bs antibodies.

analysis | selection

Understanding the humoral response against HIV-1 is critical for developing new approaches to preventing HIV-1 infection. Much effort has been devoted to understanding why HIV effectively evades most antibodies (Abs). Accepted explanations include rapid mutation of the two glycoproteins that comprise the envelope spike, gp120 and gp41, and structural features that enable the spike to hide conserved epitopes from Abs. These structural features include a shield of host-derived carbohydrates (1), conformational masking (2), steric occlusion (3), protecting conserved regions at interfaces by oligomerization (4–6), and the presence of highly variable flexible loops that shield conserved epitopes (4, 7). The small number and low density of envelope spikes on HIV virions may also contribute to HIV's ability to evade Abs by preventing most IgGs from binding simultaneously with both Fabs (8, 9). Combined with HIV's ability to rapidly mutate, these features of the HIV spike make it difficult for the host to develop antibodies with high levels of breadth and potency.

Although strain-specific Abs are more common, a subset of HIV-infected individuals develops broadly neutralizing Abs (bNAbs), i.e., Abs that neutralize many viral strains, and this happens only several years after infection (10, 11). Until recently, very few (e.g., 4E10, 2F5, 2G12, and b12) monoclonal human Abs with broadly neutralizing activity had been isolated (12–15). However, with the advent of new single-cell cloning techniques (16, 17), the number of promising bNAbs has been expanded greatly to include quaternary-specific Abs, whose epitopes involve gp120 glycosylation and the gp120 variable loops (18, 19), and a series of Abs that recognize the CD4-binding site (CD4bs) on gp120 [e.g., VRC Abs (20, 21), HJ16 (22), and highly active agonistic anti-CD4bs (HAAD) Abs (23)]. In addition, sequences of isolated heavy and light chains related to these Abs were obtained by deep sequencing methods (21). Some of the new CD4bs Abs, e.g., VRC01, NIH45–46, 3BNC117, VRC-PG04, and VRC-CH31, are remarkable for their breadth (neutralizing ~90% of strains) and the relative inability of HIV to escape these Abs by altering its glycan shield (20, 21, 23).

Interestingly, despite being isolated from different donors, the HAAD and VRC01-like CD4bs Abs arose from two closely related germ-line *IgV_H* genes (VH1-2 and VH1-46) (23), which underwent extensive somatic hypermutation (65–91 somatic mutations within 288 nucleotides) (21) to produce Abs with divergent sequences, including some related by <50% amino acid identity. Structures of the Fabs of VRC01-like Abs have been solved as complexes with HIV gp120 (21, 24, 25), revealing that these Abs all bind to gp120 by mimicking CD4; specifically, V_H chain residue Arg71 (Arg71_{VH}) forms a favorable ionic interaction with Asp368_{gp120} to mimic Arg59_{CD4}, and backbone atoms in the V_H domain C' strand form direct and water-mediated hydrogen bonds with the CD4-binding loop in gp120. Here we present analyses of the available structural and sequence data for the CD4bs Abs and propose a classification system that can be used to predict their binding and neutralization potencies and that rationalizes their origin from specific germ-line precursors. Site-directed mutagenesis is used to verify these predictions. This information should assist in vaccine development as well as in efforts to improve these antibodies by structure-based design.

Results

Sequence Signatures of Potent CD4bs Abs. The starting point of our analyses is the correlation between neutralization potency and the length of two of the light-chain CDR loops. The relatively small CDRL1 of VRC01, which has a two-residue deletion relative to its germ-line precursor, was previously correlated with increased neutralization potency (25). We noticed that sequences of VRC01, NIH45–46, and VRC-PG04 revealed a more striking correlation for the length of CDRL3, which is only 5 residues in these Abs (Fig. 1A). Examination of the large Abysis database for human Ab sequences (<http://www.bioinf.org.uk/abs/>) showed that only ~1% of V_L domains have a CDRL3 length of 5 aa, compared with more typical 9- to 11-residue lengths (Fig. 1B). The structure of the NIH45–46–gp120 complex (24) suggests that the 5-residue length may be because there is not enough space for a longer CDRL3: The end of the NIH45–46 CDRL3 (residues 91_{LC} and 96_{LC}) directly contacts the side chain of Trp100_{BHC} to stabilize the Ab V_H–V_L interface and the CDRL3 tip is near gp120 loop D residues 279_{gp120} and 280_{gp120} (Cα 91_{LC}–Cα 279_{gp120} distance of 6.9 Å). Larger CDRL3 loops would place critical side chains at the tip of CDRL3 in different locations and thus would not be able to interact with gp120 in the same manner. In Abs with longer CDRL3s, the tip of CDRL3

Author contributions: A.P.W. designed and performed research; A.P.W., R.D., and P.J.B. analyzed data; and A.P.W., M.C.N., and P.J.B. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: bjorkman@caltech.edu.

See Author Summary on page 11916 (volume 109, number 30).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1208984109/-DCSupplemental.

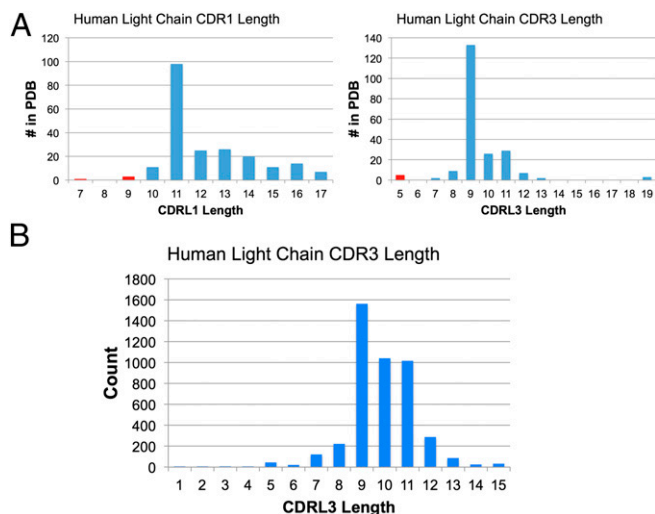


Fig. 1. Length distributions of human CDRL1 and CDRL3. (A) Length distributions of human CDRL1 and CDRL3 in the PDB. The red bars all represent CD4bs Ab structures. (B) Distribution of CDRL3 lengths of human Abs. The Abysis database (<http://www.bioinf.org.uk/abysis/index.html>) was queried for CDRL3 lengths. CDRL3 was defined according to the Chothia definition in the Abysis server.

interacts with Trp47_{HC}, a highly conserved residue (found in 63 of 69 germ-line V_H gene segments) (26) that plays a similar role to that of Trp100B_{HC} in the Abs with 5-residue CDRL3s to stabilize the V_H-V_L interface. This analysis of CDRL3 length suggests that potent anti-CD4bs Abs such as 8ANC131, 8ANC134, 1B2530, and 1NC9 (23), which have 9- or 11-residue CDRL3s (Table 1), are unlikely to bind gp120 in the same manner as the VRC01-like antibodies. In addition, this group of anti-CD4bs Abs lacks other sequence features found in VRC01, NIH45–46, and VRC-PG04: specifically, Trp50_{HC}, Asn58_{HC}, and Trp100B_{HC} (23) (Fig. S1 A and B).

V-domain alignments (Fig. 2 A and B) reveal the following sequence characteristics of the most potent of the VRC01-like Abs: complete conservation of heavy-chain Arg71_{HC}, Trp50_{HC}, Asn58_{HC}, and Trp100B_{HC} and light-chain Glu96_{LC} and Trp67_{LC}/Phe67_{LC} and a CDRL3 length of exactly 5 aa. [Kabat numbering is used herein for antibody sequences. Residues that have different Kabat numbering in different potent VRC01-like (PVL) Abs are named by their VRC01 numbering; in particular, Trp100B_{HC}, although at a fixed location relative to the C-terminal end of CDRH3, has a different numbering in each PVL Ab: 100B in VRC01, 100D in VRC-PG04, and 100F in NIH45–46. In the NIH45–46/gp120 structure, Protein Data Bank (PDB) code 3U7Y, this residue is numbered 102.] The heavy-chain

sequence characteristics are found in CD4bs Abs that descended from the VH1-2 germ line, but not from the VH1-46 germ line (23). Analysis of the per residue variability of VH1-2*02-derived Abs indicates that the conservation of Trp50_{HC} and Asn58_{HC} is unlikely to be coincidental (Table 2 and Fig. S2).

The roles that conserved residues play in the V_H domain structure and in binding to the CD4bs on gp120 are shown schematically in Fig. 3 and in stereo in Fig. S3. Fig. 3 and Fig. S3 are based on interactions present in the gp120 complexes of VRC01, NIH45–46, and VRC-PG04 (21, 24, 25). The side chains of Trp50_{HC}, Trp100B_{HC}, and Trp47_{HC} form an unusual propeller-like arrangement on the surface of the V_H domain. (Although Trp47_{HC} participates in the “propeller,” we do not consider it to be a signature residue of potent CD4bs Abs because it is commonly found in V_H domains) (26). The main interactions of the characteristic V_H domain residues with gp120 are as follows: Trp50_{HC}, indole N-H hydrogen bonds with the side-chain oxygen of Asn280_{gp120}; Asn58_{HC}, side chain N-H hydrogen bonds with the backbone carbonyl of Arg456_{gp120}; Arg71_{HC}, side chain hydrogen bonds/salt bridges with the side chain of Asp368_{gp120}; and Trp100B_{HC}, indole N-H hydrogen bonds with the side-chain oxygen of Asn/Asp279_{gp120}. Trp100B_{HC} also buries 85 Å² of surface area at the V_H/V_L interface—contacting residues Tyr91_{LC} and Glu96_{LC}.

In the light chains, the side chain of Glu96_{LC} forms a hydrogen bond with the backbone nitrogen of Gly459_{gp120} and/or the side chain of Asn280_{gp120}. The conservation of Trp67_{LC}/Phe67_{LC} is surprising as this position is distant from gp120 in the available crystal structures. However, this hydrophobic residue may interact with a portion of the glycan attached to Asn276_{gp120} that was disordered in these structures; this potential N-linked glycosylation site is present in ~90% of HIV-1 strains.

For those interactions that depend on specific gp120 side chains, the degree of conservation of the relevant gp120 residues is 96.4% for Asn/Asp279_{gp120}, 96.4% for Asn280_{gp120}, and 99.7% for Asp368_{gp120} (based on the 2010 filtered web alignment of 2,869 HIV-1 sequences in the Los Alamos HIV database: <http://www.hiv.lanl.gov>). Arg456_{gp120}, which is involved in a main-chain hydrogen bond with the side chain of Asn58_{HC}, is also highly conserved (95.0%).

PVL Class of CD4bs Abs. Individual CD4bs Abs bear a variety of names that do not generally describe their properties (20, 21, 23). For this reason, we propose a nomenclature to describe the class of highly potent CD4bs Abs that include the following common features: conservation of Arg71_{HC}, Trp50_{HC}, Asn58_{HC}, Trp100B_{HC}, Glu96_{LC}, and Trp67_{LC}/Phe67_{LC}; exactly 5 aa in CDRL1; and neutralization of >75% of HIV strains with IC₅₀ (the concentrations at which half-maximal neutralization is achieved in an in vitro neutralization assay) <50 μg/mL (it should be noted that some Abs cannot be classified with certainty

Table 1. High potency, broadly neutralizing CD4bs Abs

	Origin	VH	CDRH3 length	VL κ/λ	CDRL1 length	CDRL3 length
VRC01, NIH45–46	Donor 45	1-2	12, 16	K3-11	9	5
VRC-PG04, VRC-PG04b	Donor 74	1-2	14	K3-20	9	5
VRC-CH30–VRC-CH34	Donor 0219	1-2	13	K1-33	11	5
3BNC117, 3BNC60, 3BNC62, 3BNC95, 3BNC176	Patient 3	1-2	10	K1-33	7	5
12A12	Patient 12	1-2	13	K1-33	11	5
8ANC131, 8ANC134	Patient 8	1-46	16	K3-11	10	9
1B2530	Patient 1	1-46	16	L1-47	13	11
1NC9	Patient 1	1-46	19	L1-47	13	11
8ANC195	Patient 8	1-69	20	K1-5	12	9

Antibodies are described in refs. 20, 21, and 23. Germ-line V segment genes and selected CDR lengths (Chothia definition) are given.

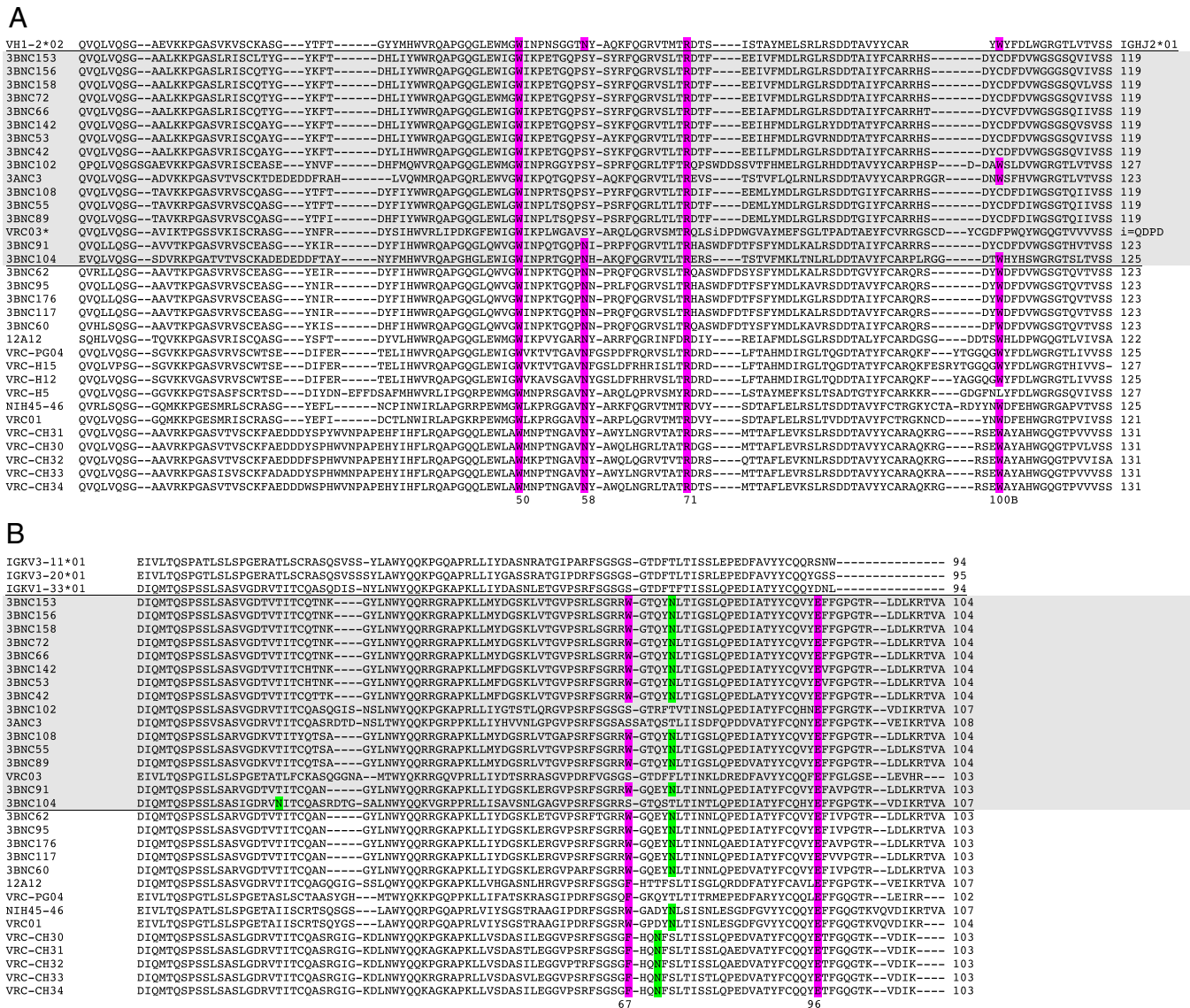


Fig. 2. Sequence alignments of PVL Abs. (A) Alignment of the heavy chains of PVL Abs, their less potent relatives, and their germ-line precursor. The less potent Abs are shaded in gray. PVL characteristic residues are highlighted in purple. [Note that W100B was not consistently aligned in previous lists of CD4bs Ab sequences (21, 23)]. Antibodies VRC-H5, VRC-H12, and VRC-H15 are heavy chains isolated by deep sequencing of donor 74 paired with the VRC-PG04 light chain. *VRC03 has several notable differences from other VRC01-like antibodies: Trp100_{HC} is replaced by Phe, there is a 7-aa insertion in framework region 3, there is a disulfide bond within CDRH3 (in contrast to the CDRH1-CDRH3 disulfide found in VRC01 and NIH45-46), it is the only isolated VRC01-like antibody with Trp54_{HC}, and the VRC03 light chain lacks a hydrophobic residue at position 67, instead retaining the germ-line Ser residue at position 67. (B) Alignment of the light chains of PVL Abs, their less potent relatives, and their germ-line precursors. The less potent Abs are shaded in gray. PVL characteristic residues are highlighted in purple; glycosylation sites are highlighted in green.

because of the large margin of error for the breadth of Abs evaluated only on small viral panels). We refer to Abs with these characteristics as PVL Abs, for “Potent VRC01-Like” Abs, reflecting the first of this class to be isolated (20). PVL Abs would include subsequently described CD4bs Abs such as NIH45-46, 3BNC117, 12A12, VRC-PG04, VRC-CH31, and a number of related variants (21, 23) (Table 1 and Fig. 2).

We refer to CD4bs Abs that include some, but not all, of the signature residues of PVL Abs and neutralize 10–75% of HIV strains with IC₅₀s < 50 μg/mL as “almost PVL” Abs. This category includes VRC03, 3BNC55, 3BNC91, 3BNC104, 3BNC89, and 12A21 (Table S1). VRC03 is unusual with respect to its neutralization breadth because when the VRC03 heavy chain is paired with its cognate light chain, it does not neutralize well enough to belong to the PVL class of Abs (nor does it contain

all of the signature sequence characteristics); however, when the VRC03 heavy chain is paired with the VRC01 light chain, the chimeric Ab exhibits broader neutralization (21). Although the heavy and light chains of members of the PVL Ab class are often interchangeable (21), the VRC03 example illustrates the difficulty of using potency to categorize Abs identified by deep sequencing (21) because the recombinant Abs used in neutralization assays may not have been paired with their physiological light chains.

3BNC55 represents an interesting example of an almost PVL Ab that has less breadth than the PVL Abs, but nonetheless neutralizes a viral strain, T250-4, that is resistant to PVL Abs (23). The T250-4 gp120 sequence includes Pro459_{gp120} (a substitution from Gly459_{gp120}), which lacks a backbone NH group to hydrogen bond with Glu96_{LC} in the Ab light chain (Fig. 3). The replacement of Trp100_{HC} with Cys100_{HC} in 3BNC55 and other defective PVL

Table 2. PVL residues conserved from the germ-line VH1-2*02 gene segment listed in order of improbability of conservation based on variability in a set of 253 (non-PVL) VH1-2*02-derived Abs

Residue	P value
Trp50 _{HC}	0.00005*
Asn58 _{HC}	0.001*
Gln6 _{HC} [†]	0.07
Arg71 _{HC}	0.27

*Likelihood < 0.05.

[†]Residue does not contact gp120. The most common VH1-2*02 mutation of this residue is to Glu; nonoccurrence of this mutation may be due to an unfavorable effect on gp120 binding of adding a negative charge to the V_H domain.

Abs may represent an adaptation of the Ab to allow Glu96_{LC}, which normally hydrogen bonds with the Gly459_{gp120} backbone, to move into the space otherwise occupied by the bulky Trp100B_{HC} side chain and perhaps hydrogen bond with the side chain of Asn280_{gp120}. Thus, production of less effective almost PVL Abs such as 3BNC55 may be selected in HIV-infected patients in response to evolution of viruses such as T250-4 that are resistant to PVL Abs. A similar in vivo selection scenario has been proposed for VRC03 and VRC06 (27).

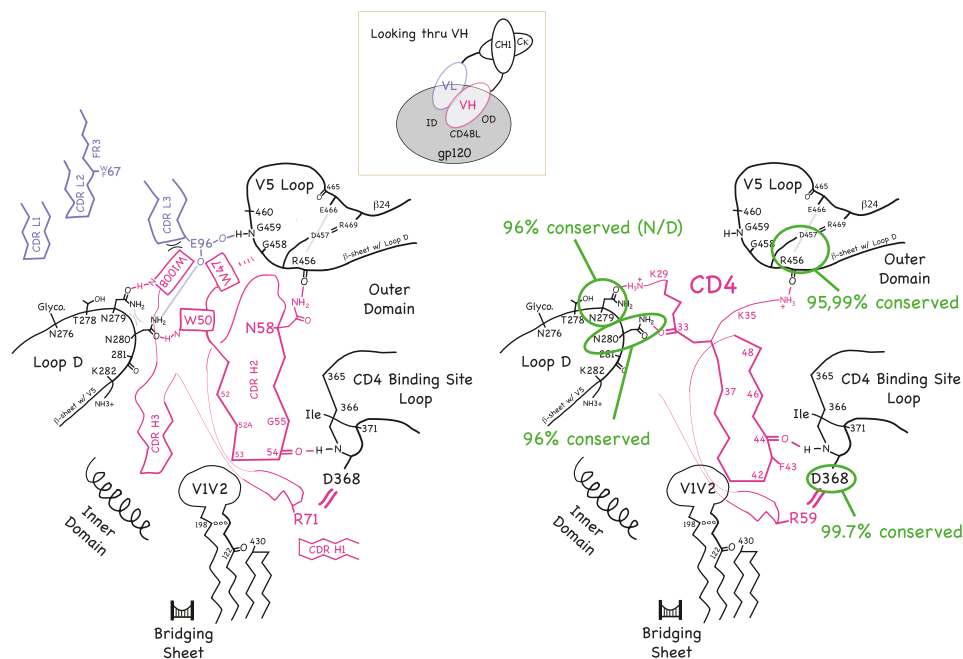
Some of the CD4bs Abs with 3BNCx names (where x is a number) (23) fall into a category we refer to as “defective

PVL” Abs, defined as Abs that lack some PVL signature residues and that neutralize <10% of HIV strains with IC₅₀s < 50 μg/mL. These Abs show some common sequence patterns: In 12 of 15 defective PVL 3BNCx Abs, Trp100B_{HC} is replaced by Cys, and in 13 of the 15, Asn58_{HC} is replaced by Ser (Fig. 24). Of the defective 3BNCx Abs, the only one to include both Trp100B_{HC} and Asn58_{HC} is 3BNC104, which comes closest to neutralizing as well as the PVL Abs (23).

Germ-Line PVL Binding to HIV—Effects of Mutating Critical Residues.

All of the PVL Abs are derived from a single germ-line V_H gene segment, IGHV1-2, and from the 02 allele of this gene segment (IGHV1-2*02) (20, 21, 23). One explanation for this finding is that the signature PVL residues identified above need to be present in the initial rearranged germ-line B-cell receptor Ab. Testing this hypothesis is difficult because germ-line versions of PVLs and other anti-gp120 bNAbs have been reported to show little or no binding to purified HIV envelope proteins (23, 25, 28). However, these binding assays are often conducted with low-micromolar protein concentrations. In addition, because the exact sequence of the HIV envelope protein that originally stimulated the B cell expressing the germ-line B-cell receptor cannot be determined, a lack of detectable binding to one or even several gp120s does not rule out the possibility of germ-line Ab binding to the original virus.

Half-germ-line versions of VRC01 were reported to retain some binding and neutralization activities (25), providing a potential method to evaluate germ-line Ab interactions with gp120.



PVL Characteristic feature	Role
Trp50 _{HC}	H bond with Asn280 _{gp120}
Asn58 _{HC}	H bond with Arg456 _{gp120}
Arg71 _{HC}	H bond/salt bridge with Asp368 _{gp120}
Trp100B _{HC}	H bond with Asn/Asp279 _{gp120}
Glu96 _{LC}	H bonds with Gly459 _{gp120} and Asn280 _{gp120}
Trp67 _{LC} /Phe67 _{LC}	interaction with Asn276 _{gp120} glycan
5-residue CDRL3	prevent steric clashes, position 91 _{LC} & 96 _{LC} side chains

Fig. 3. (Left) Schematic illustration of the common interactions of PVL Abs (magenta and blue-gray) with gp120 (black, left) and CD4 (magenta) with gp120 (black, right). Inset shows the approximate viewpoint of the diagram. (Right) The degree of conservation of critical gp120 residues in HIV-1 strains is highlighted. A stereo version of the Left is shown in Fig. S3.

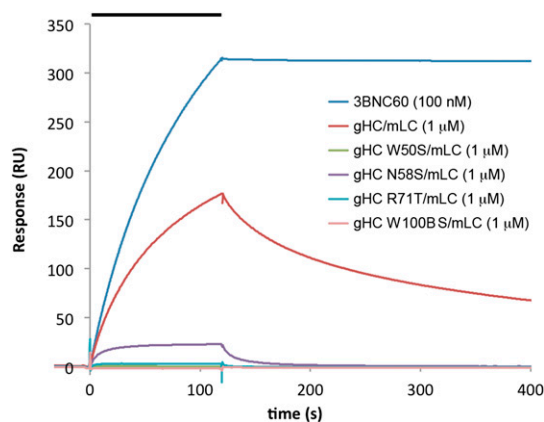


Fig. 4. Binding to immobilized YU2 gp140 of 3BNC60 IgG, 3BNC60 germ-line heavy chain/mature light chain (gHC/mLC), and germ-line heavy-chain mutants paired with mLC. The concentration of injected Ab is indicated in parentheses after each Ab. The injection period is indicated as a bar at the top.

For our experiments, we paired a germ-line 3BNC60 heavy chain with the mature 3BNC60 light chain to provide sufficient binding strength for comparisons with mutated germ-line heavy chains. An SPR-based binding assay demonstrated detectable binding of the germ-line heavy-chain/mature light-chain IgG to immobilized gp140 trimers (Fig. 4). We then compared the binding of germ-line heavy-chain IgGs with substitutions in the four signature heavy-chain residues (W50S, N58S, R71T, and W100BS) (again paired with the mature 3BNC60 light chain) (Fig. 4 and Fig. S4). The W50S, R71T, and W100BS mutants showed little or no gp140 binding, and the N58S mutation diminished binding by ~20-fold, consistent with the corresponding PVL characteristic residues playing key roles in recognition of the HIV-1 envelope spike by the germ-line PVL B-cell receptor (Table 3).

Sequence Patterns That Select the Germ-Line Parents of PVL Abs.

A requirement for the PVL signature residues to be present in the germ-line heavy chain greatly restricts the possible parent V_H gene segments of PVL Abs to those that include Trp50_{HC}, Arg71_{HC}, and Asn58_{HC} (Trp100B_{HC} is encoded outside the V_H gene segment). A requirement for Trp50_{HC} would narrow possible germ-line V_H gene segments to the VH1 and VH7 families (29), excluding common VH1 gene segments that lack Trp50_{HC} such as VH1-69, VH1-24, VH1-46, and VH1-f. A requirement for Arg71_{HC} would eliminate VH7 germ-line segments, as well as the common VH1 segments VH1-18, VH1-69, VH1-f, and VH1-24. A requirement for Asn58_{HC} would exclude other VH1 segments: VH1-c, VH1-3, VH1-46, and VH1-8. The only germ-line V_H gene segments that include all of the VRC01 signature residues are VH1-2, VH1-45, VH1-58, and VH1-68, of which VH1-68 is a pseudogene (Fig. S5). The rank order of the frequency with which VH1 gene segments are used is VH1-3 (6.2%), VH1-18 (6%), VH1-2 (2.2%), VH1-46 (2%), VH1-69 (1.8%), VH1-8 (0.8%), VH1-f (0.25%), VH1-58 (0.2%), VH1-24 (0.1%), VH1-45 (0.1%), and VH1-c (0.1%) (segments containing the signature heavy-chain residues of PVL Abs are underlined) (30). VH1-2 is >10-fold more common than the next most frequent germ-line segment containing all PVL signature residues (VH1-58), thus rationalizing the common germ-line V_H gene segment origin of known PVLs. It may be possible that PVL Abs can be derived from the V_H gene segments VH1-45 and/or VH1-58; if so, such Abs have yet to be discovered.

The VH1-2 locus has several common alleles. Two alleles, VH1-2*01 and VH1-2*05, lack Trp50_{HC} and would be unlikely to generate PVL Abs. The VH1-02*02 allele, however, encodes Trp50_{HC}; thus the presence of this residue together with Arg71_{HC}

Table 3. Binding of 3BNC60 Abs with mature, germ-line, or mutant variable domains to gp140

3BNC60 HC	3BNC60 LC	Apparent gp140 affinity (K_d values)
Mature	Mature	<0.1 nM
Germ line	Mature	0.5 μ M
Germ-line W50S	Mature	No binding
Germ-line N58S	Mature	~10 μ M
Germ-line R71T	Mature	~100 μ M
Germ-line W100BS	Mature	No binding
Germ line	Germ line	No binding

A surface plasmon resonance-based binding assay was used to evaluate the interaction of injected Ab with immobilized YU2 gp140. For samples with no observed binding, Abs were injected at a minimum concentration of 30 μ M. Derived K_d values are apparent affinities that implicitly include avidity effects due to bivalent binding. Low-affinity interactions did not fit well to a 1:1 binding model and apparent affinities listed are approximate.

and Asn58_{HC} may be sufficient to explain the use of the VH1-02*02 gene segment for PVL Abs. Individuals with no alleles encoding VH1-2 gene segments with Trp50_{HC} may have a reduced likelihood of successful immunization to produce PVL Abs. However, if polymorphism at residue 50_{HC} is distributed randomly in the population, only 5.3% of individuals would lack any VH1-2 allele with Trp50_{HC} (*SI Materials and Methods*).

Of relevance to efforts to raise PVL Abs in experimental animals, we asked whether mouse, rat, rabbit, and macaque genomes include V_H gene segments containing these residues. An examination of mouse, rat, and rabbit V_H gene sequences (29) [ImmunoGeneTics (IMGT): <http://www.imgt.org/IMGTrepertoire/Proteins/index.php>] indicates that none have V_H segments that include the combination of Trp50_{HC}, Arg71_{HC}, and Asn58_{HC} (Table S2). However, there is a VH1-2*02-like segment in the macaque genome that possesses all of the PVL characteristic residues except Asn58_{HC} (31). In addition, the guinea pig genome contains two germ-line V_H genes (GenBank accession nos. XM_003464985 and XM_003464972) with the PVL Ab characteristic residues; however, it is not clear whether these germ-line genes are functional because there are no corresponding sequences of mature guinea pig Abs in GenBank.

The final signature PVL residue, Trp100B_{HC}, is encoded either at the 5' end of the J_H gene segment or at the junction between the D and J_H segments. Only one J_H gene segment, IGHJ2*01, encodes Trp100B_{HC}. IMGT/V-QUEST (version 3.2.24) (32) assigns the most probable J gene segment used by VRC01, VRC-PG04, 3BNC117, and 12A12 as IGHJ2*01 (see *SI Materials and Methods* for additional analysis). Notably, IGHJ2*01 is the least frequently used J_H segment (<5%) (30), suggesting a selection for the use of IGHJ2*01 by PVL Abs. Considering that Trp100B_{HC} is a signature residue of PVL Abs and is critical for gp140 binding by the half-germ-line Ab (Table 3), we suggest that a number of these Abs originated from a IGHJ2*01-derived recombined ancestor that encodes Trp100B_{HC}. Other Abs, such as VRC03 (which lacks Trp100B_{HC}), do not use IGHJ2*01.

The light chains of PVL Abs derive from a variety of V_L gene segments, suggesting that PVL Abs do not show much restriction in V_L gene use. Unlike the heavy chain, the V_L gene segment region in mature PVL light-chain sequences does not contain any characteristic residues shared with the germ-line precursors; the PVL signature light-chain hydrophobic residue (Trp/Phe67_{LC}) is somatically changed from the germ-line residue Ser (Fig. 2B). However, in addition to being unusually short, the CDRL3 of PVL Abs (encoded by the V_L - J_L join and J_L gene segment) includes a partially conserved consensus sequence: QQYEF (underlined residues are more conserved; Fig. 2B). The beginning of this sequence (QQ or QQY) derives from the 3' end of the germ-line

V_L gene segment, whereas the rest is the result of hypermutation and/or nucleotide addition between the V_L and J_L gene segments. Acquisition of Tyr91_{LC} and Glu96_{LC} (the Y and E in the QOYEF sequence) may represent key steps in the maturation of PVL Abs.

gp120 Sequence Features Correlated with Complete Resistance to PVLs. To examine the importance of the signature PVL Ab residues to their activity, we analyzed the gp120 sequences of HIV-1 strains resistant to neutralization by NIH45–46. The gp120 residue variants associated with resistance were identified by three criteria: First, they are contact residues with NIH45–46 (24); second, they are present in NIH45–46-resistant viruses; third, they do not appear in NIH45–46-sensitive viruses. The critical positions identified were 279_{gp120}, 280_{gp120}, 456_{gp120}, 458_{gp120}, and 459_{gp120}; the common (i.e., sensitive) residues at these positions are Asx, Asn, Arg, Gly, and Gly, respectively. These sites make significant contacts with the characteristic PVL residues (Fig. 3). Viral strains with variations at these sites are generally neutralized poorly by all PVL Abs, as expected if substitutions at these positions interfere with common interactions (Table S3).

To verify the significance of gp120 variations at these positions, we engineered point mutants within the gp160 gene of HIV-1 strain YU2, created pseudoviruses carrying the mutant gp160s, and determined the neutralization potencies of the PVL NIH45–46^{G54W} (24) (as characterized by IC₅₀ values). Mutations at 279_{gp120} and 280_{gp120} rendered the virus resistant to neutralization by NIH45–46^{G54W}, and substitution of 458_{gp120} diminished the neutralization potency by >1,500-fold (Fig. 5).

We also examined sequences for HIV envelopes isolated from donor 45 (from whom VRC01, VRC03, and NIH45–46 were isolated) (20, 23), including 3 early proviral clones sensitive to VRC01 and 26 later clones that were insensitive to VRC01, plus additional Env clones from other individuals with strongly HIV-1–neutralizing sera (27). Variation at the critical gp120 sites (279_{gp120}, 280_{gp120}, 456_{gp120}, 458_{gp120}, and 459_{gp120}) correctly predicted sensitivity to PVL Abs for all 29 gp120 sequences from donor 45 (Fig. S6). Specifically, variation at position 279 was sufficient to explain the sensitivity/resistance characteristics: The 3 early (proviral) clones sensitive to VRC01 encoded Asp279_{gp120}, whereas none of the 26 VRC01-resistant clones that arose later encoded either Asp or Asn at position 279 (Glu279_{gp120} was the most common residue in this set) (Fig. S6).

For the full set of 100 clones with reported neutralization data (27), we find that VRC01 sensitivity is strongly correlated with the residue at position 279. In this group of clones, 34 had Asx at position 279; other common residues were Glu (33 clones)

and Lys (14 clones). Of the 31 well-neutralized viruses (VRC01 IC₅₀ < 10 μg/mL), 26 included the PVL-sensitive pattern (Asx279_{gp120}, Asn280_{gp120}, Arg456_{gp120}, Gly458_{gp120}, and Gly459_{gp120}) (average IC₅₀ = 0.83 μg/mL), whereas 5 had the resistant pattern (other residues at positions 279_{gp120}, 280_{gp120}, 456_{gp120}, 458_{gp120}, and 459_{gp120}). Of the 69 poorly neutralized viruses (IC₅₀ > 10 μg/mL), 63 had the PVL-resistant pattern, and 6 had the sensitive pattern. Together, the PVL sensitivity pattern correctly predicted the activity of this class of antibodies against 89 of 100 viruses. In particular, it appears that Asx at 279_{gp120} has been strongly selected against in the viral population in these individuals. Although position 279_{gp120} is usually predictive, two Env clones had Lys279_{gp120}, yet were well neutralized by VRC01. For YU2, we found that the N279K mutation increased the IC₅₀ for NIH45–46^{G54W} by >5,000-fold (Fig. 5). Thus, resistance to viruses encoding Lys279_{gp120} can be context (i.e., strain) dependent.

The PVL Ab 12A12 neutralizes HIV-1 strains 6540_v4_c1 and 6545_v4_c1, which other PVLs do not. These strains encode potential PVL-resistant variations Ser280_{gp120}, Ser456_{gp120}, and Tyr458_{gp120}. The explanation for the ability of 12A12 to neutralize PVL-resistant strains may involve an unusual CDRH3 residue, Leu100D_{HC}, in place of Phe or Tyr at this position. This substitution may indirectly permit a larger residue to be accommodated at position 458_{gp120}; e.g., Tyr458_{gp120} could induce Trp47_{HC} of the antibody to flip into a cavity that exists only with a smaller side chain such as Leu100D_{HC} in place of Phe or Tyr.

Discussion

Designing an HIV-1 vaccine that provides robust protection from HIV-1 infection remains a challenge despite many years of effort. The nature of the problem is most clearly illustrated by the essentially uniform inability of the immune system to clear the infection. Despite these difficulties, the finding of broadly neutralizing sera in a subset of infected individuals (33) provides a promising starting point for vaccine design. Neutralizing activity targets a variety of sites on the envelope spike—the CD4-binding site, the coreceptor-binding site, quaternary sites involving variable loops and glycosylation sites, and the V3 loop (16). Cloning, structural, and mechanistic studies have provided a base of knowledge to design antigens capable of generating these specificities. Nevertheless, immunization attempts to date have not been successful in eliciting strongly neutralizing Abs. For example, antigens designed to give 2G12-like (34), 4E10-like (35), or 2F5-like Abs (36) have had limited success, and although immunogens designed for greater CD4bs recognition showed an improved ability to elicit CD4bs neutralizing Abs, the potencies of these Abs were generally limited to the most neutralization-sensitive (Tier 1) viruses (37). In addition, a gp120 designed specifically to focus the immune response to the initial CD4-binding site on gp120 failed to elicit broadly potent CD4bs Abs (38), although this designed gp120 was the bait used to isolate VRC01 (20), a remarkably potent Ab that is the prototype member of the PVL Abs analyzed in our study.

The requirement or near requirement for PVL Abs to be derived from the VH1-2 gene segment has several implications. For testing antigens designed to elicit PVL-like Abs, it is likely that animals lacking a germ-line gene containing the signature residues will fail to generate PVL Abs. In particular, the absence of any germ-line V_H segments with the PVL motif in mice or rabbits indicates that an alternative testing platform would be preferable. For example, transgenic rodent strains incorporating some or all of the human Ig locus (39) would be a superior platform for evaluating the potential of gp120 antigens for generation of this unique class of Abs.

Understanding germ-line PVL Ab binding to gp120 is critical for immunogen design. Factors working against elicitation of PVL Abs include the restriction of possible parental germ-line V_H segments, the infrequency of rearranged light-chain genes

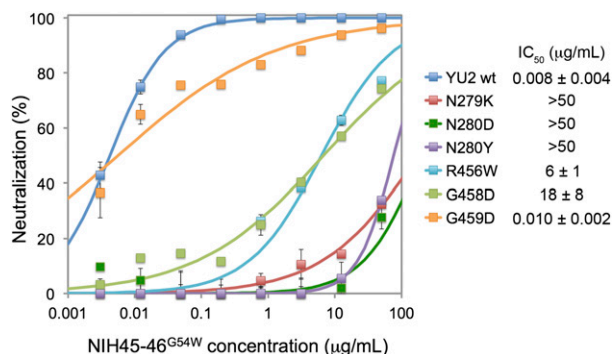


Fig. 5. Effects of mutations at critical residues in YU2 gp120 on neutralization by PVL antibody NIH45–46^{G54W}. IC₅₀ values are the mean of several independent experiments, one of which is shown on the *Left*. Although the IC₅₀s for the G459D YU2 mutant and wild-type YU2 are similar, the G459D curve is more shallow, resulting from incomplete neutralization at higher concentrations.

encoding five-residue CDRL3 regions, and the high degree of somatic hypermutation in the mature PVL Abs. With respect to the CDRL3 length, it is not clear whether PVL Abs originate as germ-line Abs with five-residue CDRL3s or whether the short length results from deletion during somatic mutation, as occurs for achieving the short length of CDRL1 (25). For choosing the strain of HIV-1 for a gp120 immunogen, a critical question is whether the germ-line PVLs bind weakly to gp120s from a wide variety of strains or whether rare forms of gp120 bind to the germ-line Abs initiating the maturation process. In light of this question, one general approach is to modify gp120 antigens to enhance binding to germ-line or half-germ-line versions of PVL antibodies, while maintaining contacts with the PVL signature residues.

The characteristic PVL residues we have identified suggest a common pathway for development of potent CD4bs Abs—beginning with heavy chains derived from VH1-2*02 and (usually) IGHJ2*01. Although some CD4bs Abs with variations at position 100B_{HC} and/or different germ-line origins are capable of being strongly neutralizing, we suggest that the PVL class represents the mostly probable form of broadly neutralizing, highly potent CD4bs Ab that can develop in humans, as evidenced by their independent isolation from five different individuals (20, 21, 23).

Materials and Methods

Sequence Analysis. Protein sequences of antibodies were aligned using the ClustalW2 multiple-sequence alignment tool at <http://www.ebi.ac.uk/Tools/msa/clustalw2/>, followed by manual adjustments based on published structures of the Abs alone or in complex with gp120. HIV-1 envelope sequences were aligned on the basis of DNA sequences using the Gene Cutter tool at http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html. Multiple-sequence alignments (along with IC₅₀s from neutralization assays when analyzing envelope sequences) were then read and analyzed to identify patterns, using a custom Objective-C program.

Analysis of Likelihood of Conservation of Germ-Line Residues. To estimate the probability of conservation of germ-line residues, we combined a position-dependent relative rate of mutation (at the protein level) with an estimate of the amino acid frequency distribution per position expected in a population of “fully” hypermutated V_H domains. By fully hypermutated, we mean that additional mutation no longer changes the per residue amino acid frequencies within the population of Abs. An approximate rate of mutation of VH1-2*02 gene segment residues was estimated by calculating the frequency of conservation, $\text{freqCons}(i)$, for each position i by aligning and examining a set of 253 VH1-2*02-derived Abs (identified by IMG_T; none of these are PVL Abs) (29). Fig. S2, which shows variability by residue [i.e., $1 - \text{freqCons}(i)$], displays the approximate rate of mutation, which results from a combination of factors: the probability of nucleotide mutation, the tolerance of V_H positions to mutation, and selection for improved antigen binding.

A distribution of V_H amino acid frequencies for fully somatically mutated V_H domains was calculated from a set of 1,461 heavy chains from the IMG_T/3Dstructure database. These sequences include both V_H gene diversity and somatic hypermutation, constrained to maintain V_H domain function. This distribution is denoted as $\text{relaxedFreq}(i, r)$, for the frequency of amino acid r at position i .

We estimate the probability of conservation of germ-line amino acid r at position i after a degree of hypermutation parameterized by x as

$$\text{prob}(i, r) = (1 - \text{relaxedFreq}(i, r)) \cdot \text{freqCons}(i)^x + \text{relaxedFreq}(i, r),$$

where one overall value for x ($x = 17.2$) was chosen so that the average number of mutations expected in a mature V_H domain matched the average number of mutations (37.8) seen in the five PVL Abs considered. For the 253 VH1-2*02 Ab V_H regions used to calculate $\text{freqCons}(i)$, there were an average of 9.3

residue mutations compared with the germ line. The scaling by x was done to adjust for the much higher level of somatic hypermutation in the PVL Abs.

For our set of five independent PVL Abs (each isolated from a different individual), we chose 3BNC117, 12A12, VRC01, VRC-CH31, and VRC-PG04. The likelihood estimate for complete conservation of germ-line residue r at position i in our five PVL Abs is $28 \cdot \text{prob}(r, i)^5$. This estimate includes a Bonferroni correction for multiple tests using the number of V_H gene segment residues ($n = 28$) completely conserved in the set of five PVL Abs compared with the germ line. A variety of alternative models were also used to provide likelihood estimates; although the calculated values differed, a strong signal for significance was found for PVL signature residues Trp50_{HC} and Asn58_{HC} (Table 2). The conservation of the third signature residue encoded within the V_H domain, Arg71_{HC}, was not as improbable, because Arg is often found in this position.

Materials. The 3BNC60 germ-line heavy-chain gene and purified YU2 gp140 trimers were provided by Johannes Scheid (The Rockefeller University).

Mutagenesis. QuikChange mutagenesis was used to generate Ab and gp160 mutants. All gene constructs were verified by complete sequencing.

Protein Expression and Purification. Abs were expressed transiently in a suspension of HEK 293–6E cells (NRC Biotechnology Research Institute, Montréal, QC, Canada), using 25 kDa linear polyethylenimine (PEI) (Polysciences) for transfection as described in refs. 24 and 40. Supernatants were passed over MabSelect SuRe protein A resin (GE Healthcare), eluted using pH 3.0 citrate buffer, and then immediately neutralized. Antibodies tested in neutralization or binding assays were further purified by size exclusion chromatography, using a Superdex 200 or 75 10/300 GL column.

In Vitro Neutralization Assays. A TZM-bl/pseudovirus neutralization assay was used to evaluate the neutralization potencies of the Abs (41, 42) as described in ref. 24. Pseudoviruses were generated by cotransfection of HEK 293T cells with an Env expression plasmid and a replication-defective backbone plasmid. Neutralization was determined by measuring the reduction in luciferase reporter gene expression in the presence of Ab NIH45–46^{G54W} following a single round of pseudovirus infection in TZM-bl cells. Nonlinear regression analysis was used to calculate the concentrations at which half-maximal inhibition was observed (IC₅₀ values) by fitting the observed neutralization to the expression $1/(1 + (\text{IC}_{50}/x)^n)$, where x is the antibody concentration and n is the Hill coefficient.

Surface Plasmon Resonance. A Biacore T100 (GE Healthcare) with the T200 sensitivity enhancement was used to evaluate the interactions of antibodies with gp140. Approximately 200, 750, and 2,500 response units (RUs) of YU2 gp140 were covalently immobilized on three flow cells of a CM5 biosensor chip, using standard primary amine coupling chemistry (Biacore manual). Concentration series of Abs were injected at room temperature in 10 mM HEPES with 150 mM NaCl, 3 mM EDTA, and 0.05% surfactant P20 at pH 7.4. Equilibrium dissociation constants (K_{D} s) were determined from kinetic constants derived from sensorgram data, using simultaneous fitting to the association and dissociation phases of the interaction to a 1:1 binding model. Because the injected proteins (the Abs) are capable of bivalent binding to the immobilized gp140, this model produces apparent K_{D} s that implicitly include the avidity effects.

ACKNOWLEDGMENTS. We thank the Caltech Protein Expression Center, Terri Lee, Paola Marcovecchio, Tim Feliciano, Gloria Tran, and Han Gao for DNA preparation, protein expression, and purification; Jost Vielmetter for assistance with Biacore experiments; Johannes Scheid, Florian Klein, and Ari Halper-Stromberg for helpful discussions and reagents; Grace Aldrovandi and Kyle Nakamura for helpful discussions; and Thiago Olivera for help with the data for Fig. S2. This project was supported by Award DP10D006961 (to P.J.B.) from the Office of The Director, National Institutes of Health. This work was also supported by Collaboration for AIDS Vaccine Discovery grants with support from the Bill and Melinda Gates Foundation [Grant 38660 (to P.J.B.) and Grant 38619s (to M.C.N.)]

- Wei X, et al. (2003) Antibody neutralization and escape by HIV-1. *Nature* 422: 307–312.
- Kwong PD, et al. (2002) HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* 420:678–682.
- Labrijn AF, et al. (2003) Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1. *J Virol* 77:10557–10565.

- Kwong PD, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393:648–659.
- Wyatt R, et al. (1997) Analysis of the interaction of the human immunodeficiency virus type 1 gp120 envelope glycoprotein with the gp41 transmembrane glycoprotein. *J Virol* 71:9722–9731.
- Moore JP, Sodroski J (1996) Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J Virol* 70:1863–1872.

7. Starcich BR, et al. (1986) Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 45:637–648.
8. Klein JS, Bjorkman PJ (2010) Few and far between: How HIV may be evading antibody avidity. *PLoS Pathog* 6:e1000908.
9. Mouquet H, et al. (2010) Polyreactivity increases the apparent affinity of anti-HIV antibodies by heterologation. *Nature* 467:591–595.
10. Gray ES, et al.; and the CAPRISA002 Study Team (2011) 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.
11. Mikell I, et al. (2011) Characteristics of the earliest cross-neutralizing antibody response to HIV-1. *PLoS Pathog* 7:e1001251.
12. Muster T, et al. (1993) A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J Virol* 67:6642–6647.
13. Buchacher A, et al. (1994) Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Res Hum Retroviruses* 10:359–369.
14. Trkola A, et al. (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* 70:1100–1108.
15. Burton DR, et al. (1994) Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* 266:1024–1027.
16. Scheid JF, et al. (2009) Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* 458:636–640.
17. Scheid JF, et al. (2009) A method for identification of HIV gp140 binding memory B cells in human blood. *J Immunol Methods* 343:65–67.
18. Walker LM, et al.; Protocol G Principal Investigators (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326:285–289.
19. Walker LM, et al.; Protocol G Principal Investigators (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477:466–470.
20. Wu X, et al. (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 329:856–861.
21. Wu X, et al.; NISC Comparative Sequencing Program (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 333:1593–1602.
22. Corti D, et al. (2010) Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS ONE* 5:e8805.
23. Scheid JF, et al. (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333:1633–1637.
24. Diskin R, et al. (2011) Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* 334:1289–1293.
25. Zhou T, et al. (2010) Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* 329:811–817.
26. Lefranc MP, Lefranc G (2001) *The Immunoglobulin FactsBook* (Academic, London).
27. Wu X, et al. (2012) Selection pressure on HIV-1 envelope by broadly neutralizing antibodies to the conserved CD4-binding site. *J Virol* 86:5844–5856.
28. Xiao X, et al. (2009) 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.
29. Lefranc MP, et al. (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 37(Database issue):D1006–D1012.
30. Arnaout R, et al. (2011) High-resolution description of antibody heavy-chain repertoires in humans. *PLoS ONE* 6:e22365.
31. Thullier P, Chahboun S, Pelat T (2010) A comparison of human and macaque (*Macaca mulatta*) immunoglobulin germline V regions and its implications for antibody engineering. *MAbs* 2:528–538.
32. Brochet X, Lefranc MP, Giudicelli V (2008) IMGT/QUEST: The highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* 36(Web Server issue):W503–W508.
33. Mascola JR, Montefiori DC (2010) The role of antibodies in HIV vaccines. *Annu Rev Immunol* 28:413–444.
34. Dunlop DC, et al. (2010) Polysaccharide mimicry of the epitope of the broadly neutralizing anti-HIV antibody, 2G12, induces enhanced antibody responses to self oligomannose glycans. *Glycobiology* 20:812–823.
35. Correia BE, et al. (2010) Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* 18:1116–1126.
36. Guenaga J, et al. (2011) Heterologous epitope-scaffold prime:boosting immunofocuses B cell responses to the HIV-1 gp41 2F5 neutralization determinant. *PLoS ONE* 6:e16074.
37. Feng Y, et al. (2012) Biochemically defined HIV-1 envelope glycoprotein variant immunogens display differential binding and neutralizing specificities to the CD4-binding site. *J Biol Chem* 287:5673–5686.
38. Nabel G, Kwong P, Mascola J (2011) Progress in the rational design of an AIDS vaccine. *Philos Trans R Soc Biol Sci* 366(1579):2759–2765.
39. Lonberg N (2005) Human antibodies from transgenic animals. *Nat Biotechnol* 23:1117–1125.
40. Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res* 30:E9.
41. Li M, et al. (2005) Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 79:10108–10125.
42. Montefiori DC (2005) Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Current Protocols in Immunology*, eds Coligan JE, et al. (John Wiley & Sons, New York), Chap 12, Unit 12.11.