Comparison of Antibody Repertoires Produced by HIV-1 Infection, Other Chronic and Acute Infections, and Systemic Autoimmune Disease

Felix Breden1*, Christa Lepik2#, Nancy S. Longo3##, Marinieve Montero2##, Peter E. Lipsky3###, Jamie K. Scott2####

1 Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada, 2 Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada, 3 Repertoire Analysis Group, Autoimmunity Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, 4 Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

Abstract

Background: Antibodies (Abs) produced during HIV-1 infection rarely neutralize a broad range of viral isolates; only eight broadly-neutralizing (bNt) monoclonal (M)Abs have been isolated. Yet, to be effective, an HIV-1 vaccine may have to elicit the essential features of these MAbs. The V genes of all of these bNt MAbs are highly somatically mutated, and the VH genes of five of them encode a long (≥20 aa) third complementarity-determining region (CDR-H3). This led us to question whether long CDR-H3s and high levels of somatic mutation (SM) are a preferred feature of anti-HIV bNt MAbs, or if other adaptive immune responses elicit them in general.

Methodology and Principal Findings: We assembled a VH-gene sequence database from over 700 human MAbs of known antigen specificity isolated from chronic (viral) infections (ChI), acute (bacterial and viral) infections (AcI), and systemic autoimmune diseases (SAD), and compared their CDR-H3 length, number of SMs and germline VH-gene usage. We found that anti-HIV Abs, regardless of their neutralization breadth, tended to have long CDR-H3s and high numbers of SMs. However, these features were also common among Abs associated with other chronic viral infections. In contrast, Abs from acute viral infections (but not bacterial infections) tended to have relatively short CDR-H3s and a low number of SMs, whereas SAD Abs were generally intermediate in CDR-H3 length and number of SMs. Analysis of VH gene usage showed that ChI Abs also tended to favor distal germline VH-genes (particularly VH1-69), especially in Abs bearing long CDR-H3s.

Conclusions and Significance: The striking difference between the Abs produced during chronic vs. acute viral infection suggests that Abs bearing long CDR-H3s, high levels of SM and VH1-69 gene usage may be preferentially selected during persistent infection.
which bear CDR-H3s of 20 aa or more, based on the IMGT numbering system. Consistent with this, most HIV-1-infected individuals produce strong strain-specific Nt Ab responses against HIV-1 envelope (Env) soon after initial infection; yet rarely do they develop broad neutralization [10,11], and then only after a year or more [12].

While high levels of SM have been noted for all bNt MAbs, starting with Kunert et al. [13], a number of authors have proposed a connection between the length of the CDR-H3 region, and the broad neutralization of these MAbs [5,7,14, 15,16,17,18,19,20,21]. Mutagenesis experiments and/or X-ray crystal structures of Fab bound to protein or peptide Ag have implicated their CDR-H3s as being required for neutralization: b12 [15,22], 2F5 [16,23,24], 447-52D [17], and 4E10 [18,25]. This has been observed even in cases in which CDR-H3 appears to make minimal or no contact with envelope protein Ag [16,23]. It has been speculated that in these cases, CDR-H3 may contact other sites on HIV-1, such as the viral membrane [16,18,19,20,24,25,26]. Nevertheless, it is not clear whether long CDR-H3s are required, in general, for broad neutralization; certainly the exceptions, MAbs 2G12 and VRC01/02, disprove this as an absolute rule for broad Nt activity.

While it is generally acknowledged that high levels of SM are produced by T-cell driven processes in germinal centers, the conditions under which long CDR-H3 Abs appear in adaptive immune responses are less well understood, and this could help explain the origin of bNt Abs during HIV infection. Importantly, the long CDR-H3s have been associated with anti-protein Abs [27] and anti-viral Abs [28]. Long CDR-H3s found among polyreactive natural Abs [29,30] and autoreactive Abs produced by naive B cells in SLE patients [31] do not carry SMs; yet, those found among the memory B cells in healthy people do [32]. However, analyses to directly associate different types of adaptive immune response with CDR-H3 length have not been reported.

Our purpose was to identify the circumstances under which Abs bearing some of the features of bNt MAbs (i.e., long CDR-H3s and high levels of SM) appear during adaptive immune responses in humans.

As an initial approach, we examined heavy chain variable (VH) genes expressed by individuals undergoing Ag-specific immune responses, by compiling a database of expressed human VH genes of MAbs for which the Ag specificity of the MAb was known. The MAbs were taken from individuals with chronic infections (ChI) and non-protein Ag, and/or from different conditions under which Abs bearing some of the features of bNt MAbs (i.e., long CDR-H3s and high levels of SM) appear during adaptive immune responses in humans.

Table 1A summarizes the analysis of expressed VH genes for several MAbs categories with regard to CDR-H3 length, number of SMs relative to the predicted germline VH gene, and the distance in the IgH locus between an Ab’s germline VH gene and the VH isoform; the VH1 gene closest to the D region. As expected, the bNt HIV MAbs had the longest CDR-H3s with an average length of 20.9 aa. Fig. 1 compares the CDR-H3 length distributions for bNt HIV MAbs, the non-bNt HIV MAbs (i.e., excluding the bNt HIV MAbs), and the remaining ChI MAbs (excluding all HIV MAbs). It shows that both the bNt and non-bNt HIV MAbs have long CDR-H3s; the difference between these two MAbs groups was only marginally significant (Table 1A: unadjusted p = 0.0501), and not significant compared to the Bonferroni corrected value of 0.0018. Thus, CDR-H3 length does not appear to be restricted to broad neutralization. In addition, the CDR-H3s of all of the anti-protein HIV MAbs (mean 17.8 aa) were not significantly longer than those of non-HIV ChI MAbs (16.5 aa). Although our data set has only 34 non-HIV ChI MAbs, 12 of these have CDR-H3s of 19 aa or longer (Fig. 1), placing them in the upper quartile of the 427 Ag-specific MAbs. Thus, long CDR-H3s were associated with all types of ChI MAbs, including the anti-HIV MAbs.

Fig. 2 shows the distribution CDR-H3 length for anti-HIV MAbs, partitioned according the region of Env bound; CDR-H3 length was longest for MAbs against the CD4i site (19.6 aa), intermediate for those against the V3 loop (18.5 aa) and CD4bs (18.3 aa), and shortest for the anti-gp41 MAbs (15.9 aa). These distributions were significantly different (χ² test, p<0.05, PROC FREQ, SAS), indicating that, while the anti-HIV MAbs as a whole bear long CDR-H3s, epitope specificity also shapes CDR-H3 length.

The observation that long CDR-H3s are associated with chronic viral infections in general led us to compare MAbs from such infections to those from other immune responses, namely SAD and AcI. Categorization of MAbs by these types of immune response showed that the average length of the ChI MAbs (17.6 aa, Table 1B and Fig. 3A) was most different from the AcI MAbs (14.7 aa), with the SAD MAbs being intermediate (15.1 aa). This trend, of the AcI MAbs being the most different from the ChI MAbs, persisted when the categories were further divided into Abs against protein vs. non-protein Ags (Fig. 3B and Table 1B). For example, among anti-protein MAbs, those from the ChI group had significantly longer CDR-H3s than did those from the AcI or SAD groups (p<0.0001). As virtually all of the ChI Abs (except bNt 2G12) and all of the anti-protein AcI Abs are anti-viral, it is striking that a large difference in CDR-H3 length exists between the Abs elicited by the two types of viral infection. In addition, anti-protein MAbs had significantly longer CDR-H3s than MAbs against non-protein Ags (p<0.0001, Table 1C), even with the Bonferroni correction; while the difference between self and non-self Abs was not as great (p<0.005, not significant with the Bonferroni correction). Thus autoimmune status (self vs. non-self) had a much lower effect on CDR-H3 length than did protein vs. non-protein Ag.

SM of expressed VH genes in adaptive Ab responses

That long CDR-H3s were found mainly among anti-protein Abs suggests that Abs with long CDR-H3s are selected by protein Ag, and hence, by T-cell-driven processes; such responses typically occur in germinal centers and involve SMs introduced by activation-induced cytidine deaminase (AID) [36]. Predicting that

Comparison of Antibody Repertoires
SMs should increase among Abs bearing long CDR-H3s, we compared the patterns of SM for the same categories of MAb as for CDR-H3 length in Table 1. For many comparisons, the patterns observed for SMs paralleled those observed for CDR-H3 length. For example, the eight bNt HIV MAbs had both the highest average CDRH3 length and the highest average level of SM. For example, the eight bNt HIV MAbs had both the patterns observed for SMs paralleled those observed for CDR-H3 according to protein.

For this Tukey a posteriori test, let us denote conditions that differ at the 5% level. For example, ChI and AcI and SAD Abs were heterogeneous for CDR-H3 length at the p<0.0001 level according to the Kruskall-Wallis test, and the Tukey a posteriori test grouped AcI and SAD Abs together (letter B), but different from ChI Abs (letter A).

Distance between an Ab’s germline VH gene and the VH6-1 gene, the VH gene closest to the DH region.

Comparison of Antibody Repertoires

Table 1. Comparison of CDR-H3 length, SM and distance from D_{H} of V_{H} genes among MAbs of known Ag specificity associated with HIV infection (Table 1A), chronic infection (ChI), acute infection (AcI) or systemic autoimmune disease (SAD) (Table 1B), and self vs. non-self (Table 1C).

| MAb Type | # MAbs | CDR-H3 (aa) | # MAbs | V_{H} SM | p | # MAbs | V_{H} Distance | p |
|----------|--------|-------------|--------|----------|---|--------|----------------|---|
| All      | 427    | 16.3        | 426    | 21.0     | --| 413    | 445            | --|
| A. HIV bNt | 8     | 20.9        | 8      | 53.3     | 0.0024 | 8      | 303            | 0.0541|
| HIV Non-bNt | 185   | 17.7        | 184    | 27.3     | 167  | 508    |                |    |
| HIV/P    | 192    | 17.8        | 191    | 28.2     | 181  | 503    | 0.0952         |    |
| Non-HIV/P | 34    | 16.5        | 34      | 21.0     | 32   | 440    |                |    |
| B. ChI   | 227    | 17.6        | 226    | 27.3     | 0.0001 A | 214 | 492 | 0.0001 |   |
| AcI      | 113    | 14.7        | 113    | 10.9     | B   | 112    | 414            | B  |
| SAD      | 87     | 15.1        | 87      | 17.9     | B   | 86     | 369            | B  |
| ChI/NP   | 1      | 16.0        | 1      | 60.0     | 0.0836 | 1      | 286            | 0.6627|
| AcI/NP   | 43     | 14.0        | 43      | 15.5     | 43   | 357    |                |    |
| SAD/NP   | 60     | 15.0        | 60      | 13.2     | 60   | 379    |                |    |
| ChI/P    | 226    | 17.6        | 225    | 27.1     | 0.0001 A | 204 | 493 | 0.0018 A |   |
| AcI/P    | 70     | 15.1        | 70      | 7.9      | B   | 69     | 450            | A/B|
| SAD/P    | 27     | 15.4        | 27      | 28.5     | A   | 26     | 346            | B  |
| C. Self  | 87     | 15.1        | 87      | 17.9     | 0.0370 | 86     | 369            | 0.0003|
| Non-self | 340    | 16.6        | 339    | 21.8     | 327  | 466    |                |    |
| NP       | 104    | 14.6        | 104    | 14.6     | 0.0001 | 104   | 369            | 0.0001|
| P        | 323    | 16.9        | 322    | 23.1     | 309  | 471    |                |    |
| Self/NP  | 60     | 15.0        | 60      | 13.2     | 0.0001 | 60    | 379            | 0.6206|
| Self/P   | 27     | 15.4        | 27      | 28.5     | 26   | 346    |                |    |
| Non-self/NP | 44 | 14.0 | 44 | 16.6 | 0.0148 | 44 | 355 | 0.0001 |
| Non-self/P | 296    | 17.0        | 295    | 22.6     | 283  | 482    |                |    |

*MAbs are categorized by Ag type and condition: bNt, broadly neutralizing; NP, non-protein; P, protein; SAD, autoimmune disease; Chl, chronic infection; AcI, acute infection and vaccination; Self = SAD MAbs; non-Self = Chl plus AcI MAbs.

1Differences between means were analyzed by the Kruskall-Wallis Test (PROC NPAR1WAY, SAS), and the uncorrected p values are shown. This analysis involved testing 9 groupings of MAbs (e.g. HIV bNt vs. HIV Non-bNt) for 3 parameters (CDR-H3 length, SM, and distance of V_{H} gene from V_{H}6-1), or 27 tests of heterogeneity among sets of MAbs. Therefore the critical value adjusted for multiple tests (i.e., bonferroni correction, see Materials and Methods) was 0.05/27 = 0.0019; all p values satisfying this corrected level of significance are shown in bold). When the Kruskall-Walls test showed significant heterogeneity among 3 categories of MAb being compared (i.e., Chl, AcI and SAD), we used parametric, a posteriori Tukey pair-wise comparisons (PROC GLM, SAS) to determine which sets of MAbs were contributing to the heterogeneity.

2Distance between an Ab’s germline V_{H} gene and the V_{H}6-1 gene, the V_{H} gene closest to the DH region.

doi:10.1371/journal.pone.0016857.t001
shows little difference in the number of SMs between the anti-protein Abs for all disease categories and little difference in SM levels between the quartiles for medium-length and the longest CDR-H3s; however, the short CDR-H3 quartile tended to have lower levels of SM than did the longer two groups. Furthermore, the trend for MAbs against non-protein Ags was reversed: MAbs in the shortest CDR-H3 quartile tended to have the highest SM levels for both the SAD and AcI categories. Although long CDR-H3 MAbs against non-protein Ags are uncommon, their SM level was lower than that of their short CDR-H3 counterparts. This is consistent with the hypothesis that MAbs against non-protein Ags (even those with long CDR-H3s) may derive from different B-cell subsets and/or different immune processes than the MAbs against protein Ags.

Germline VH-gene usage in adaptive Ab responses

Early reports of VH gene family usage in anti-HIV Abs reported that VH3 was over-utilized and VH4 was under-utilized compared to the naïve repertoire [37,38]. Given that our dataset includes Abs of known Ag specificity for several disease conditions, we compared the use of V-gene families and of the specific gene VH1-69 among these conditions (Table 3 and Figure S1). Table 3 shows that there were significant differences when the proportions were tested across each gene family. The proportion of anti-HIV MAbs that use family VH3 genes (29%) was lower than that for SAD MAbs (52%) or AcI MAbs (53%). Concomitantly, use of family VH1 increased for HIV MAbs (38% for HIV compared to 23% for SAD MAbs and 22% for AcI MAbs), whereas the proportion of MAbs using family VH4 was not significantly different among the major categories, ranging from 18 to 25%. Thus, HIV infection was related to an increase in VH1 and decrease in VH3 gene usage that was not apparent for ChI Abs (perhaps because this sample size is small) or other conditions.

We were particularly interested in usage patterns for the VH1-69 germline gene, as it is not commonly used in the naïve repertoire (e.g., Wardemann et al. 2003 [39]), but is characteristic of Ab repertoires in several disease states (see Discussion). As shown in Table 3, analysis of the 193 HIV MAbs having verifiable germline VH genes revealed that 43 (22%) used VH1-69, whereas only five of 87 (6%) SAD MAbs and three of 113 (3%) AcI MAbs used this VH gene; VH1-69 usage by (non-HIV) ChI MAbs was intermediate, being four of 34 (12%). Thus, HIV MAbs, and, perhaps, ChI MAbs use VH1-69 at a higher frequency than do SAD or AcI MAbs.

Given this distinct difference, we directly compared the features of MAbs that use VH1-69 to those that do not (Table 4). Among ChI MAbs, those that used VH1-69 had significantly longer CDR-H3s (means of 20.1 vs. 17.0, P<0.001), whereas their SMs were not significantly different. Very few SAD or AcI MAbs used VH1-69, so no statistical comparisons could be made within those groups. Among MAbs that use VH genes other than VH1-69, the patterns among the disease categories of ChI, SAD and AcI were similar to those observed in the full data set; the CDR-H3 length of ChI MAbs remained significantly longer than those of SAD and AcI MAbs, and the SMs of all three groups were different. Thus, ChI Abs encoded by VH1-69 appear to have longer CDR-H3s but not more SMs than their counterparts that do not use this germline gene.

VH1-69 is a fairly distal gene, being approximately 764 Kb from VH6-1 [40]; only five of the approximately 40 functional VH1-genes are more distal. Thus we wondered if the use of VH1-69 among Abs bearing long CDR-H3s in HIV infection could be part of a
larger trend toward using distal genes in ChI. Table 1A shows that ChI MAbs use the most distal VH genes, consistent with this class having the longest CDR-H3s and the highest frequency of SMs. However, inconsistent with CDR-H3 length and SMs, anti-protein (anti-viral) AcI Abs had intermediate VH gene distances whereas SAD Abs had the most proximal ones. Importantly, the AcI MAbs, only three of which use VH1-69, used distal genes overall; the average distance for 69 anti-protein (antiviral) AcI MAbs is 450 kb, suggesting that distal genes besides VH1-69 may be selected in viral infections of all types.

Table 2 further analyzes the relationship between CDR-H3 length and VH gene distance, dividing the MAbs into short, medium and long. Table 2 shows that the pattern is also not consistent within disease condition. Analyzing CDR-H3 length by quartile, the trend between CDR-H3 length and VH gene distance held for ChI and SAD but not for AcI Abs. The medium and long quartiles of ChI and SAD MAbs used the most distal VH genes, whereas they were used by the short and medium quartiles of anti-protein AcI MAbs. Thus, within the viral infection groups, and following similar trends for SM, distal gene usage was related to CDR-H3 length for the ChI, but not the AcI, MAbs.

Discussion

The primary motivation for this analysis was to determine if other types of Ab share features with the bNt HIV MAbs, which might help explain the rarity of bNt MAbs in HIV-1 infection. We found that the bNt MAbs most closely resemble the other anti-HIV MAbs, and ChI MAbs as a group, in being enriched for long CDR-H3s and high numbers of SMs; this indicates that these features are not limited to broad neutralization, but appear to be common characteristics of the Abs involved in chronic viral infections. That all anti-HIV MAbs, including the bNt MAbs, share similar features indicates that unusual immunological processes, such as breaking of tolerance [19], are probably not responsible for the rarity of the bNt Abs during chronic HIV infection. Instead, processes involved in chronic viral infections in general may be at play in shaping the repertoire of Abs available for selection after viral persistence and/or multiple rounds of viral escape; such processes are probably linked to broadening of the Ab response beyond those involved in the response to initial infection [41]. Our results are consistent with the view that Abs having the features of the bNt MAbs are not rare, but arise as a result of chronic viral infection.

Strikingly, the Abs from acute viral infections (anti-protein AcI Abs) had significantly shorter CDR-H3s and lower numbers of SMs than did the ChI Abs. In addition, while both types of Ab tended to use distal VH genes, Abs from acute viral infections bearing long CDR-H3s tended not to use the most distal VH genes, nor did they use VH1-69 to the same extent as the ChI (and HIV) Abs. We speculate that, if the Abs involved in acute viral infections reflect those produced during the early phases of chronic viral infection, a shift in expressed VH gene composition (i.e., CDR-H3 length and VH gene usage) must occur over time, along with an increase in SMs.
Figure 3. Distribution of CDR-H3 length for different types of immune response and Ag type. A. Distribution of CDR-H3 length for ChI, SAD, and AcI mAbs. B. Distribution of CDR-H3 length for ChI anti-protein, SAD anti-protein, AcI anti-protein, SAD anti-non-protein, and AcI anti-non-protein mAbs. Arrows indicate the means for each category. Control MAb (blue line) are included for comparison [63].

doi:10.1371/journal.pone.0016857.g003
High SMs, but not long CDR-H3s nor use of distal VH genes, were also found among SAD-related anti-protein MAbs. This lack of shared features between the SAD and bNt MAbs (or ChI MAbs in general) suggests that the bNt MAbs against HIV are probably not drawn from an initial pool of autoimmune B cells bearing long CDR-H3s, as previously hypothesized [19,20,42]; were that the case, then the bNt MAbs would be expected to similar to the SAD MAbs but not the ChI ones. Clearly, Abs having the features of the bNt MAbs are not rare, and are routinely produced during other MAbs. This may be consistent with the suggestion that HIV-1 gp120 acts as a superAg that specifically deletes B cells bearing Abs encoded by genes from family VH3 [47,48]. In addition, and in contrast to some previous findings [45], we did not observe over-utilization of the VH4 family, which is consistent with the cryptic, flexible and/or transient nature of conserved epitopes on the neutralization-competent structure of HIV Env; such epitopes are not immunodominant on the virus, nor on the envelope “debris” shed by infected cells, and as such, multiple rounds of viral escape are likely required before the immune system can mount an effective Ab response against them. This study is to our knowledge the first to explicitly compare gene family usage in MAbs from HIV with those from other types of immune response. We observed a bias toward family VH1 and against family VH3 genes in the HIV and other ChI MAbs. The increased usage of family VH1 agrees with Scheid et al. [43]; and removal of the large number of Abs from Scheid et al. did not affect this conclusion (analysis not shown). A deficit in family VH3 usage is also consistent with the hypothesis that HIV-1 gp120 acts as a superAg that specifically deletes B cells bearing Abs encoded by genes from family VH3 [47,48]. In contrast, and in contrast to some previous findings [45], we did not observe over-utilization of the VH4 family, which is consistent with the cryptic, flexible and/or transient nature of conserved epitopes on the neutralization-competent structure of HIV Env; such epitopes are not immunodominant on the virus, nor on the envelope “debris” shed by infected cells, and as such, multiple rounds of viral escape are likely required before the immune system can mount an effective Ab response. This study is to our knowledge the first to explicitly compare gene family usage in MAbs from HIV with those from other types of immune response. We observed a bias toward family VH1 and against family VH3 genes in the HIV and other ChI MAbs. The increased usage of family VH1 agrees with Scheid et al. [43]; and removal of the large number of Abs from Scheid et al. did not affect this conclusion (analysis not shown). A deficit in family VH3 usage is also consistent with the hypothesis that HIV-1 gp120 acts as a superAg that specifically deletes B cells bearing Abs encoded by genes from family VH3 [47,48]. In contrast, and in contrast to some previous findings [45], we did not observe over-utilization of the VH4 family, which is consistent with the cryptic, flexible and/or transient nature of conserved epitopes on the neutralization-competent structure of HIV Env; such epitopes are not immunodominant on the virus, nor on the envelope “debris” shed by infected cells, and as such, multiple rounds of viral escape are likely required before the immune system can mount an effective Ab response. This study is to our knowledge the first to explicitly compare gene family usage in MAbs from HIV with those from other types of immune response. We observed a bias toward family VH1 and

### Table 2. Somatic mutations and VH-gene distance of MAbs grouped according to short (<= 13aa), medium (14-18 aa) or long (>= 19aa) CDR-H3 length.

| MAb Type | VH1 SM | VH1 Distance | P-value |
|----------|--------|--------------|---------|
|          | S  | M  | L  | S  | M  | L  |         |
| All      | 1.65| 21.8| 22.9| 353| 466| 488|         |
| A. HIV bNt | 76.5(21) | 45.6(6) | NS | 286(2) | 320(6) |         |
| B. ChI | 24.1 | 27.9 | 26.1 | 327 | 498 | 539 |         |
| AcI | 10.6 | 11.5 | 8.9(15) | 401 | 448 | 322(15) |         |
| SAD | 18.9 | 18.9 | 17.7(14) | 299 | 405 | 394(14) |         |
| Chi/NA | 60(1) | NA | 286(1) | NA | 286(1) | NA |         |
| Chi/NA | 24.1 | 28.2 | 26.9 | 333 | 504 | 534 |         |
| HIV/NA | 60(1) | NA | 286(1) | NA | 286(1) | NA |         |
| HIV/NA | 25.3 | 29.7 | 27.3 | 334 | 517 | 537 |         |
| Not HIV/NA | NA | NA | NA | NA | NA | NA |         |
| Not HIV/P | 19.0(6) | 19.6(16) | 24.0(12) | 328(6) | 424(14) | 516(12) |         |
| AcI/NA | 15.8(19) | 15.6(21) | 14.3(3) | 329(19) | 400(21) | 229(3) |         |
| AcI/P | 6.1(22) | 9.2 | 7.6(12) | 466(21) | 476 | 346(12) |         |
| SAD/NA | 15.0(22) | 11.9 | 12.8(10) | 327(22) | 402 | 429(10) |         |
| SAD/P | 22.0(6) | 30.4(17) | 30.0(4) | 197(6) | 411(16) | 307(4) |         |
| C. Self | 15.6 | 18.9 | 17.7(14) | 299 | 405 | 394(14) |         |
| Non-self | 16.2 | 22.6 | 24.3 | 375 | 483 | 502 |         |
| NP | 15.4 | 14.4 | 13.2(13) | 328 | 399 | 382(13) |         |
| P | 17.2 | 24.1 | 24.8 | 370 | 488 | 502 |         |
| Self/NA | 15.0(22) | 11.9 | 12.8(10) | 327(22) | 402 | 429(10) |         |
| Self/P | 22.0(6) | 30.4(17) | 30.0(4) | 197(6) | 411(16) | 307(4) |         |
| Non-self/NA | 15.8(22) | 17.6(22) | 14.3(3) | 329(19) | 395(22) | 229(3) |         |
| Non-self/P | 16.6 | 23.3 | 24.6 | 392 | 497 | 510 |         |

*MAbs are categorized by Ag type and condition: Naïve, broadly neutralizing; P, protein; SAD, autoimmune disease; ChI, chronic infection; AcI, acute infection and vaccination; Self = SAD MAbs; non-Self = ChI plus AcI MAbs.

**Distance between an Ab’s germline VH gene and the VH6-1 gene, the VH1 gene closest to the D region.

†Not available; i.e., there were no MAbs of this type in the database.

‡Sample sizes<25 are given in parentheses.

### Table 3. VH gene family and Vn1-69 gene usage among 4 types of MAbs.

| Gene Family | Condition | % HIV | % CHI, not HIV | % SAD | % AcI |
|-------------|-----------|-------|----------------|-------|-------|
| A. Vn1-1 | 39 (76)† | 21 (7) | 23 (20) | 22 (25) | ** |
| Vn1-2 | 2 (3) | 0 (0) | 1 (1) | 0 (0) | NS |
| Vn1-3 | 29 (56) | 44 (15) | 52 (45) | 53 (60) | ** |
| Vn1-4 | 24 (47) | 26 (9) | 18 (16) | 23 (26) | NS |
| Vn1-5 | 6 (11) | 9 (3) | 5 (4) | 1 (1) | NS |
| Vn1-6 | 0 (0) | 0 (0) | 1 (1) | 1 (1) | NS |
| Total | 100 (193) | 100 (34) | 100 (87) | 100 (113) |  |

### Table 4. Association between Vn1-69 gene usage, CDR-H3 length and number of somatic mutations.

| Condition | CDR-H3 length | Vn1 gene SM | P-value |
|-----------|---------------|-------------|---------|
|          |              |             |         |
| ChI | 20.1 (47)*** | 17.0 (180) | A | 27.6 (47) | NS | 27.2 (179) | A |
| AcI | 15.3 (5) | 14.7 (110) | B | 14.7 (5) | 10.7 (110) | B |
| SAD | 15.2 (5) | 15.1 (82) | B | 31.2 (5) | 17.1 (82) | C |

*Mean (% of MAbs). CDR-H3 length and SM were tested between ChI MAbs using Vn1-69 and those that do not; CDR-H3 length was significantly different (p<0.001), while the average number of SMs was not different (NS). Statistical comparisons were not made between SAD and AcI MAbs due to the low numbers of those MAbs using Vn1-69.

**Tukey test. Letters denote conditions (rows) that differ at the 5% level in a posterior comparisons among MAbs not using Vn1-69. Again, statistical comparisons were made only for those not using Vn1-69, due to the low numbers of SAD and AcI MAbs that used Vn1-69.

doi:10.1371/journal.pone.0016857.t003

doio:10.1371/journal.pone.0016857.t002
observations of Huang et al. [49], who noted that nine of twelve MAbs against the CD4i site of gp120, used V1H-69, and those of Gorny et al. [46], who showed that MAbs against all HIV Env epitopes, except the V3 loop, are enriched for V1H-69 usage. In addition, three studies have noted almost exclusive use of V1H-69 among cross-protective MAbs against influenza virus [50,51,52].

Both AcI and ChI MAbs tended to use distal V1H genes, but only in the latter group were long CDR-H3s present in Abs encoded by distal V1H genes, which tended to be V1H-69. Three mechanisms can produce long CDR-H3s: (i) longer V1H, D1H, or JH genes can be used preferentially, (ii) CDR-H3 can be lengthened by insertions induced by activation-induced cytidine deaminase [9,53], and (iii) secondary rearrangement (or receptor editing or revision [54,55]) can result in N and P additions at the N1 junction. Secondary V1H-gene rearrangement necessarily involves the use of distal V1H and JH genes, because once a V1H gene is somatically recombined with a D1H gene (i.e., after primary V-D-J rearrangement), only genes more distal to the D1H region are available for further joining. The features of the ChI Abs alone are consistent with secondary rearrangement model, in both using distal V1H genes and having long CDR-H3s for the same Ab population. Thus, viral infection appears to select for distal V1H genes, but if secondary rearrangement is playing a role in lengthening CDR-H3, it appears to be doing so only for the ChI Abs.

Many of our conclusions should be interpreted with caution, given that they are based on a limited dataset that may be biased in several ways. For example, there was significant bias related to Ag specificity for several categories of MAb; many of the anti-HIV MAbs were against the gp120 CD4bs, and were obtained via phage-displayed Ab libraries; most of the SAD MAbs were against the non-protein Ags DNA and phospholipid/cardioliopin; whereas most of the AcI MAbs were against streptococcal capsular polysaccharide and rotavirus. Another potential bias is related to the limited number of SADs we studied, with the preponderance being SLE and anti-phospholipid syndrome. Since we are studying MAbs, it is important to realize that particular antibodies are often selected for further study based on characteristics such as strength of binding, isotype or epitope, and thus the data set is not random with respect to these parameters. This is one reason why we adopted a conservative approach, and emphasize those results that satisfy a Bonferroni-adjusted p value based on the total number of tests conducted in Table 1. A larger dataset, indexed by disease and clinical condition, would overcome many of these potential biases. One roadblock to such a compilation is that many researchers do not routinely submit expressed sequences to public databases; this will become especially critical as high-throughput methods are employed to survey large sets of disease-specific MAbs.

This analysis of 427 Ag-specific MAbs should directly inform vaccine research. For example, the result that long CDR-H3s are associated with chronic and persistent Ag and anti-protein Abs motivates several questions. Are Abs bearing long CDR-H3s present at the beginning of an immune response, or do they “evolve” over time? If they accumulate over time, are they directly selected from a pre-existing minor compartment within the naive B-cell populations, do they comprise a specially recruited B-cell subset, and/or do they evolve by secondary processes (e.g., V1H-gene replacement, DNA insertion, or gene conversion)? All of the bNi MAbs against HIV are heavily mutated and five of the eight have long CDR-H3s. This line of reasoning raises the possibility that long CDR-H3s are not required to bind conserved epitopes on the HIV-1 envelope, but arise instead through processes that come into play during long-term persistence of protein Ag and viral escape. If so, an effective HIV vaccine may produce bNi Abs via “normal” immunization processes, by virtue of enhancing the immunogenicity of N1 sites on Env. Given this scenario, it remains unknown if Abs bearing the features of acute antiviral Abs (which we expect to be similar to the features of Abs elicited by a traditional vaccine) can act as bNi Abs. While bNi Abs have yet to be elicited by vaccines meant to mimic the epitopes on Env that mediate neutralization by the bNi MAbs, that should not be taken as evidence that they cannot be so produced. Our results indicate that HIV vaccine research should continue to follow “reverse vaccinology” approaches [56] that attempt to make the sites recognized by the bNi MAbs immunodominant [57,58]. Progress in this approach has recently been observed with an influenza vaccine that elicits broadly protective Abs [59]. Conversely, it is also possible, but not proven, that “chronic” type Abs bearing the features of the bNi MAbs will be required for broad neutralization. If this is the case, then research into the cellular and genetic origins of such Abs is required. Thus our second recommendation is for research efforts to be expanded in this area, with the goal of developing vaccination strategies that stimulate key features of these chronic processes, and in so doing, elicit bNi Abs.

Materials and Methods

Sequence database

Heavy chain sequences of expressed MAbs were retrieved from the IMGT/LIGM-DB on-line database (http://imgt.cines.fr/), from the literature, and from direct contacts with researchers (see Table S1). Our goal was to collect V1H sequences for all of the available human HIV MAbs. The Ag targets of the HIV MAbs included gp120, its CD4 binding site (CD4bs) and CD4 inducible site (CD4i), the gp120 V3 loop, gp41, Rev, Tat, p24, and p25. This MAb dataset was expanded to include human MAbs associated with other chronic infections (the ChI MAbs), including those against Epstein Barr virus, hepatitis B and C virus, herpes simplex virus and human cytomegalovirus. (Note that all of these MAbs are from viral infections.) For comparison, a similar group of MAbs from Systemic Autoimmune Disease (SAD) was assembled, including from systemic lupus erythematosus (SLE), anti-phospholipid syndrome, mixed connective tissue disease, rheumatoid arthritis, Sjogren’s disease, and cold agglutin disease, and against the Ags, cardioliopin (serum dependent and serum independent), phospholipids, DNA, beta-2-glycoprotein, Sm ribonucleoproteins, myelin basic protein, myelin-associated glycoprotein, acetylchyloline receptor, Ro/SSA and La/SSB. We concentrated on SAD MAbs based on the hypothesis that the bNi anti-HIV MAbs were derived from autoAb/auto-reactive precursors [19,20,42]. In addition, V1H sequences were collected for MAbs associated with acute infections (Pseudomonas aeruginosa, rotavirus, Pneumococcus pneumoniae, Ebola virus, Neisseria meningitidis, hepatitis A virus), and from vaccinated individuals (Haemophilus influenzae Type b conjugate vaccine, 23-valent pneumococcal polysaccharide vaccine, Streptococcus pneumoniae, tetanus toxoid, hepatitis B surface Ag), reported as the AcI MAbs. For each MAb, we attempted to obtain information on its isotype, Ag specificity, the methods used to obtain it (e.g., phage display, B-cell sorting, etc.), clinical data on the source-subject, and the bibliographic reference and GenBank accession number for the original MAb sequence. This information was entered into an Excel database by hand.

Sequence analysis

Nucleotide sequences were analyzed using a recent version of JoinSolver (http://joinsolver.niams.nih.gov/index.htm; [60]),
which provides the closest-matched VH, DH and JH genes, determines the limits of the CDR-H3 region, the length (in amino acids) of CDR-H3 region, the contributions of P and N nucleotides at both the V-D and D-J junctions, and the number of SMs in the MAbs relative to the predicted germline genes (this number is defined as of the number of base pair substitutions relative to a predicted germline gene). These results, including nucleotide sequence for CDR-H3 region, were also entered into the Excel database. For a few HIV MAbs, only CDR-H3 length, and not the expressed VH sequence, was available (e.g., Dittrich et al. [61]; see Table S1). Results from JoinSolver were compared to those produced by the V-QUEST and JunctionAnalysis algorithms of the IMGT system [62], which also analyzes VH sequences for gene usage and somatic mutations. In addition, assignments of each MAb to predicted germline VH, DH and JH genes were confirmed visually. Results from IMGT and JoinSolver differed systematically. For example, the size of the region of CDR-H3 contributed by the germline DH gene was consistently estimated to be greater using IMGT V-QUEST. This result can be explained by the fact that standard parameters for V-QUEST allow more mutations in the DH-gene core. However, differences between classes of MAbs were similar whether the comparisons were calculated using V-QUEST or JoinSolver, and these relative differences (e.g., the average CDR-H3 length of self vs. non-self MAbs) are the important parameters in our study.

JoinSolver results were used to screen for clonal expansions, which were identified as those Abs that used the same sets of VH, DH and JH regions with similar patterns of N and P nucleotides. A single, randomly chosen MAb was retained to represent each clonally-expanded set. Two recently reported bNt MAbs [7] expansions of the same B-cell lineage, so we randomly selected one, PG16, for analysis; taking the same approach we selected VRC01 from the set of two bNt MAbs reported by Wu et al. [9]. Thus, the final set that was statistically analyzed does not include all reported HIV MAbs, but only those representing independent clonal lineages (see below, and Tables S1 and S2).

In summary, the entire database consists of over 700 MAbs (Table S1), which underwent two screens to produce the final dataset (Table S2) for analysis of CDR-H3 length, SM and gene usage. In the first screen, each MAb had to have a specified Ag, and to be associated with a particular immune response. In the second screen, clones from the same clonal expansion were deleted from the dataset, resulting in a 427-MAb dataset comprising 227 Chl MAbs (including 193 HIV MAbs), 67 SAD MAbs and 113 AcI MAbs, which was exported to SAS (Rel. 8.2, 2001; SAS Institute Inc., Cary, NC) for statistical analysis. Of these 427 MAbs, 318 were identified as to IgM or IgG; 90% of these were IgG (Table S2).

Statistical analysis

For all MAb categories, PROC UNIVARIATE (SAS) was used to test the distributions of CDR-H3 length, total VH-gene mutations, distance of predicted VH gene used in the MAb relative to VH6-1, the VH-gene most proximal to the DH region (VH-distance), and distance of predicted JH gene from JH6, the JH gene most distal to the DH region (JH-distance), against the normal distribution. Most of the distributions were non-normal, even after log transformation, so a non-parametric Kruskall-Wallis Test was used to test for differences among sets of MAbs (PROC NPAR1WAY, SAS). To avoid zero values, the natural log of (3*CDR-H3 length in aa + 0.1) was used in tests for differences in CDR-H3 length. All results from the non-parametric tests were compared to one-way ANOVA (PROC GLM, SAS), and in all cases the results were similar in terms of levels of significance. When more than two categories were compared, (i.e., comparisons among Chl, AcI and SAD MAbs), Tukey a posteriori tests were used to determine what groups were statistically different, and these different groups were denoted by different letters (PROC GLM, SAS). In Table 1 we present the p values for the main statistical tests of this study. This Table reports 9 hypothesis tests for each of CDR-H3 length, number of SMs and VH-distance, for a total of 27 tests; therefore, to be conservative, all tests that passed a Bonferroni-corrected P value of 0.05/27 = 0.0018 were highlighted in bold. Given the many confounding factors in this data base, these probability values should be interpreted as indicators of strong differences among categories rather than strictly interpreted statistical tests (see Discussion). Distributions of CDR-H3 length for CD4bs, CD4i, V3 loop, and anti-gp41 MAbs presented in Figure 2 were tested for heterogeneity by χ². JH-distance did not vary among MAbs categories and is not reported. The difficulty of assigning germline DH genes to expressed Ab sequences, especially for highly mutated HIV MAbs, precluded a comprehensive analysis of DH gene usage or the number of P and N nucleotides.

Supporting Information

Figure S1 VH gene family usage in anti-protein and non-anti-protein MAbs for 3 disease conditions. See Table 3 for sample sizes; there is only 1 Chl Mab that is not anti-protein. MAbs utilizing VH1 family were separated into those using VH1-69 and others. (TIF)

Table S1 Database of antigen-specific expressed human MAbs. (XLS)

Table S2 Expressed MAbs analyzed for CDR-H3 length, somatic mutations, and gene usage. (XLS)

Acknowledgments

R. Kunert and H. Katunger (University of Natural Resources and Life Sciences, Vienna) kindly provided the sequences of MAbs 2F5, 2G12 and 4E10 for this analysis, and J. Scheid (Charite Universitaet medizin, Berlin) and M. Nussenzweig (Rockefeller University) provided a large number of anti-HIV-1 expressed Ab sequences [43]. We are grateful to Ralph Pantophlet (Simon Fraser University) for comments on the manuscript and Kevin Henry (Simon Fraser University) for help with graphics.

Author Contributions

Conceived and designed the experiments: FB JKS. Performed the experiments: FB CL MM NSL PEL JKS. Analyzed the data: FB CL MM NSL PEL JKS. Contributed reagents/materials/analysis tools: PEL NSL. Wrote the paper: FB JKS NSL PEL.

References

1. Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302: 575–581.  
2. Market E, Papavasiliou FN (2005) VDJ recombination and the evolution of the adaptive immune system. PLoS Biol 1: E10.

3. Wu TT, Johnson G, Kabat EA (1993) Length distribution of CDRH3 in antibodies. Proteins 16: 1–7.  
4. Schief WR, Ban YE, Stamatatos L (2009) Challenges for structure-based HIV vaccine design. Curr Opin HIV AIDS 4: 451–440.

3. Wu TT, Johnson G, Kabat EA (1993) Length distribution of CDRH3 in antibodies. Proteins 16: 1–7.
60. Souto-Carneiro MM, Longo NS, Russ DE, Sun HW, Lipsky PE. (2004) Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOIN-SOLVER. J Immunol 172: 6790–6802.

61. Ditzel HJ, Parren PW, Binley JM, Sodroski J, Moore JP, et al. (1997) Mapping the protein surface of human immunodeficiency virus type 1 gp120 using human monoclonal antibodies from phage display libraries. J Mol Biol 267: 684–695.

62. Lefranc MP. (2005) IMGT, the international ImMunoGeneTics information system: a standardized approach for immunogenetics and immunoinformatics. Immunome Res 1: 3.

63. Zemlin M, Klünger M, Link J, Zemlin C, Bauer K, et al. (2003) Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. J Mol Biol 334: 733–749.