Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design


Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design

Understanding human immunodeficiency virus type 1 (HIV-1) transmission is central to developing effective prevention strategies, including a vaccine. We compared phenotypic and genetic variation in HIV-1 env genes from subjects in acute/early infection and subjects with chronic infections in the context of subtype C heterosexual transmission. We found that the transmitted viruses all used CCR5 and required high levels of CD4 to infect target cells, suggesting selection for replication in T cells and not macrophages after transmission. In addition, the transmitted viruses were more likely to use a maraviroc-sensitive conformation of CCR5, perhaps identifying a feature of the target T cell. We confirmed an earlier observation that the transmitted viruses were, on average, modestly under-glycosylated relative to the viruses from chronically infected subjects. This difference was most pronounced in comparing the viruses in acutely infected men to those in chronically infected women. These features of the transmitted virus point to selective pressures during the transmission event. We did not observe a consistent difference either in heterologous neutralization sensitivity or in sensitivity to soluble CD4 between the two groups, suggesting similar conformations between viruses from acute and chronic infection. However, the presence or absence of glycosylation sites had differential effects on neutralization sensitivity for different antibodies. We suggest that the occasional absence of glycosylation sites encoded in the conserved regions of env, further reduced in transmitted viruses, could expose specific surface structures on the protein as antibody targets.
in the blood of the donor. Such differences could arise if the virus population in the donor’s genital tract is a distinct subset of the virus in the donor (i.e., compartmentalization). In addition, phenotypic differences may arise due to selection at any of several steps in the transmission process. First, selection may favor viruses that are better able to cross the mucosal surface to reach an initial target cell. Second, selection at the site of transmission may favor viruses capable of infecting available target cells, such as T cells or macrophages, or interacting with dendritic cells (DCs). Third, selection may favor viruses capable of rapid replication in the initial target cell. Fourth, selection may favor viruses whose phenotypes direct them to specific sites, such as the gut-associated lymphoid tissue, where the amount of virus is greatly amplified early after infection. One or more of these features of transmission may impact the biological properties of the virus that initiates infection.

The nature of the transmitted virus has been examined in several settings. Initially, these variants were thought to be macrophage-tropic (9, 10). While the transmitted virus is predominantly an R5 virus (8, 10–15), more recent evidence indicates that macrophage tropism is not required for transmission, and in several studies, there were no examples of transmission of a virus that could efficiently replicate in macrophages (12, 16–18) or infect cells with low levels of CD4 (a surrogate for macrophage tropism) (11, 13).

Other features of the transmitted variant have also been examined. In several studies, there has been consistent observation of reduced length of the variable regions in the Env protein, with roughly two-thirds encoded in the main part of the Env protein, with approximately 30 glycosylation sites encoded in the extracellular portion of the Env protein, it is important that concepts concerning HIV-1 transmission be formulated based on large sample sizes. Here, we compared the sequences of a large number of viral Env proteins from acute/early infections (n = 68) to Env proteins present in contemporaneous chronic infections (n = 62) in the setting of heterosexual transmission of subtype C HIV-1. We found that the Env protein of the transmitted virus was 5% underglycosylated on average compared to Env proteins in chronically infected subjects, with the virus found in acutely infected men being 7% underglycosylated relative to the virus found in chronically infected women. The difference between acutely infected women and chronically infected men was much less pronounced, suggesting that underglycosylation is principally a feature of female-to-male transmission and not a general feature of all transmission types. A subset of the sequences analyzed were cloned into an expression vector to assess the phenotypic characteristics of the Env protein in pseudotyped virus assays. The transmitted viruses were not differentially sensitive to heterologous neutralizing antibodies (with one exception), consistent with the transmitted virus having a conformation similar to that of the virus in chronically infected subjects with respect to antibody sensitivity. In addition, we found, using a more quantitative assay for CD4 dependence in entry, that the transmitted viruses required high levels of CD4 to infect cells, consistent with an activated CD4+ T cell and not a macrophage being the initial target cell for replication after transmission. Both types of viruses could use a form of CCR5 that was inhibited by maraviroc, a CCR5 antagonist. However, only a fraction of the transmitted viruses were able to utilize an alternative conformation of CCR5 that was insensitive to maraviroc, while a majority of viruses from chronic infections were able to use this conformation. Collectively, these differences suggest selective pressures at work during transmission. In addition, the underglycosylation of the transmitted/founder virus highlights the fact that the presence of many of the glycosylation sites is variable. We propose that these glycosylation sites can be grouped by their proximity to specific surface structures of the Env protein and that the absence of glycosylation sites may make the virus vulnerable to antibodies targeting these protein surface structures.

(These results were presented in part at the March 2012 CROI Meeting.)

MATERIALS AND METHODS

Cohort description. Plasma samples from 68 subjects identified as having acute/early subtype C HIV-1 infection were previously described (3) (the previous study included one acute-to-acute transmission pair, one of which has been omitted from this study to avoid overrepresentation of this sequence in the analysis). This group included 27 subjects in Fiebig stages (51) I to III, 32 in stages IV and V, and 8 in stage VI. One participant was indicated as stage V or VI. All subjects were identified and enrolled in the CHAVI 001 or CAPRISA 002 acute infection study. Subjects who did not have acute HIV-1 infection were identified as chronics if a fully positive HIV-1 Western blot result was obtained (p31+). A total of 62 subjects were identified with chronic HIV-1 infection from cross-sectional cohorts.
in Malawi and South Africa. Subjects with advanced HIV-1 disease (CD4+ T cell counts <200 cells/ml) were excluded. All samples were collected with written informed consent, and all protocols were approved by institutional review boards at the collection sites.

**Isolation of viral RNA and SGA.** Plasma was collected in the presence of either acid-citrate-dextrose (ACD) or EDTA. HIV-1 viral loads from blood plasma were determined by using Abbott RealTime assays. An aliquot of between 140 and 200 µl was extracted using the Qiagen QIAamp viral RNA minikit, with the final eluted volume being 50 µl. For each sample, approximately 10,000 viral RNA copies were reverse transcribed in a 100-µl reaction mixture that included SuperScript III reverse transcriptase (Invitrogen) and either the OFM-19 primer or oligo(dT) as previously described (3, 8, 52). The cDNA product was then serially diluted and used in a nested PCR amplification strategy to amplify the viral env gene using primers as described previously (3, 8, 52). Between 20 and 30 amplicons were generated at a dilution where approximately 30% of the PCR amplifications gave an appropriate product. The amplicons were sequenced bidirectionally and further screened to ensure that a single template gave rise to each amplicon by eliminating sequences that showed evidence of mixtures. Single-genome amplification (SGA)-derived env amplicons with frameshift mutations that resulted in premature stop codons were excluded.

**Analysis of viral sequences.** DNA sequence alignments were performed using ClustalW (53) and hand edited. Phylogenetic trees were generated using neighbor-joining methods in MEGA 4.0 (54). Phylogenetic trees and highlighter plots (www.hiv.lanl.gov) were used to determine the number of transmitted variants in subjects with acute infection. The consensus sequence was determined for all subjects with acute infection using Consensus Maker (www.hiv.lanl.gov), and the consensus sequences in subjects with multiple transmitted variants were determined within each clade. The number of encoded N-linked glycosylation sites was determined using N-Glycosite (www.hiv.lanl.gov) for the ectodomain of gp160. To determine the frequency of glycosylation site conservation, the variable-length regions of V1/V2 and V4/V5 were deleted from the sequences, which were then realigned to allow a tally at each site. Glycosylation sites are numbered according to HXB2.

**Cloning of env genes and generation of pseudotyped virus.** Full-length rev-env-env cassettes were cloned into pcDNA 3.1 from the directionTopo expression kit (Invitrogen) as previously described (8, 52, 55). The entire insert of the molecular clones was sequenced to ensure a match with the transmitted variant in subjects with acute infection. In subjects with chronic infection, a representative SGA-derived env amplicon was selected from the predominant clade for cloning. All inserts were confirmed to match the amplicon sequence. A subset of env genes were mutagenized by plasmid resynthesis using overlapping PCR primers containing the mutation of interest.

Viral stocks for neutralization assays were generated by transfecting 293T cells with a subtype C backbone vector (ZM247Fv1deltaenv) and the HIV-1 env clones as previously described (56). Viruses were harvested 48 h after transfection, filtered through a 0.45-µm filter, and frozen in aliquots. Viral titers were determined by measuring luciferase expression after infecting TZM-bl cells (catalog no. 8129; NIH AIDS Research and Reference Reagent program). Viruses used in Affinofile entry assays were produced identically except that the 293T cells were cotransfected with env clones and the pNL4-3.lucR“E” plasmid.

**Virus neutralization assay.** Single-round neutralizing antibody assays were performed as previously described (56). Neutralization of virus was quantified by the reduction in luciferase reporter gene expression after infection of the TZM-bl reporter cell line. Neutralization titers were reported as the antibody concentrations for monoclonal antibodies and the polyclonal sera which were used as purified IgG fractions, with the value that gave a 50% reduction in relative luminescence units (RLU) being reported. These values were interpolated using 5 parameter curve-fitting.

**Affinofile cell entry assay.** In order to quantify virus entry at high versus low concentrations of CD4 on the cell surface, Affinofile cells (57) were seeded into 96-well plates. Twenty-four hours later, ponasterone A and doxycycline were added at concentrations that induced the desired CD4 (low, ~2,000 antibody binding sites [ABS]/cell; high, ~70,000 ABS/cell) and CCR5 levels (high, ~30,000 ABS/cell). Twenty hours later, we removed the drugs, quantified receptor densities with flow cytometry (anti-CD4 monoclonal antibody RPA-T4 and anti-CCR5 monoclonal antibody 2D7), and infected the cells with pseudotyped viruses using spinoculation (58). Forty-eight hours after spinoculation, the cells were lysed and luciferase activity in the cell lysate was measured.

**Analyses.** Standard statistical analyses were performed using R statistical software (version 2.14.1). Model fitting of infection data was performed using the R statistical package (drc) designed to analyze dose response models (drm). These analyses fit a four-parameter, log- logistic model to the data.

Direct statistical comparison of genetic data can be biased by phylogenetic relatedness among the data. To avoid biases caused by closely related sequences in comparing the glycosylation site density between sequences from acutely and chronically infected subjects, we used the difference between glycosylation site density of an extant sequence and its most recent ancestor. Comparison of extant sequences to the most recent common ancestors to correct for phylogenetic bias in statistical comparisons of genetic variables has been adopted successfully in the past, e.g., by Bhattacharya et al. (59). In cases where more than one sequence was present from the same subject, in addition to the direct most-recent common ancestor, we used the most closely related sequence from another subject to account for phylogenetic signals in the data and used a statistical modeling approach to account for similarities (due both to shared ancestry and shared intra-subject environment) between sequences from the same subject. In all cases, the counts of changes were divided by the total number of residues in the sequence to calculate the change in glycosylation site density since the most recent ancestor. This was used as the variable for nonparametric comparisons between sequences from acutely and chronically infected groups.

The glycosylation site count of the most-recent ancestor was estimated in two main ways. First, the most recent ancestor amino acid sequence was reconstructed using GASP (gapped ancestral sequence prediction for proteins) (60), which considers the possibility of indels occurring between an ancestor and its descendants and is thus appropriate for identifying gains and losses of glycosylation sites. GASP determines ancestral states based on the most-probable residue (including gaps) obtained by considering residues at the descendant nodes, branch lengths, and employment of the PAM1 matrix of amino acid substitutions. We inferred and counted mutations that had occurred at the glycosylation sites of each extant sequence since evolution from the most-recent ancestor. Second, continuous values for the glycosylation site density of the most-recent ancestor of each sequence were directly estimated using the ancestral character estimation (ACE) function under the Analyses of Phylogenetics and Evolution package in R (61). The ancestry estimates in ACE were done using (i) phylogenetically independent contrasts (PIC) of ancestral values and (ii) general least-squares (GLS), assuming Brownian motion of evolution along the tree (62, 63). We used linear mixed-effects models to compare changes in glycosylation site density between groups of subjects, considering the subject as a random effect and including the country of origin, infection stage, or both as covariates.

**Nucleotide sequence accession numbers.** The sequences obtained from subjects with acute infection have been described previously (3). The sequences obtained from subjects with chronic infection have been submitted to GenBank under accession numbers KC862610–KC863648, HM638963–HM638965, HM638988–HM638989, HM638993–HM638994, HM638997–HM639008, HM639101–HM639103, HM639105–HM639116, HM639136–HM639139, HM639143–HM639147, HM639152–HM639157, HM639159–HM639162, HM639162–HM639184, HM639199–HM639200.
TABLE 1 Cohort description

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Acutely infected</th>
<th>Chronically infected/controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>27.0 yr (30/38)</td>
<td>29.4 yr (17/38)</td>
</tr>
<tr>
<td>Mean CD4^+ T cell count/µl</td>
<td>508 (21/20)</td>
<td>370 (20/40)</td>
</tr>
<tr>
<td>Mean plasma viral RNA load (log_{10})</td>
<td>6.9 (30/38)</td>
<td>5.4 (19/32)</td>
</tr>
</tbody>
</table>

Dependent: Age, CD4^+ T cell count, and viral load not available for all subjects.

RESULTS

Cohort characteristics and sequence generation. In this study, we examined a large set of viral env genes from people in acute/early infection with subtype C HIV-1 who were infected via a heterosexual route. The HIV-1 env sequences from 68 acutely infected subtype C subjects have been described previously (3). Briefly, subjects were identified as being in acute infection by the presence of viral-specific antibodies and viral RNA and were assigned to Fiebig stages I to VI (51). As a control group, we chose contemporaneous seroreplicable cases (Table 1). This control/chronic group is a surrogate for the real population of linked donors, which were not available for this large cohort. Since the time of infection was not known for the control group, we used the inclusion criterion of a CD4^+ T cell count between 200 and 500 cells/µl, with a total sample size of 62 subjects. Molecular clones were made of the env consensus sequence from 34 acutely infected subjects who were infected with a single variant. For comparison, one env clone was made from each of 31 chronically infected subjects. The env genes were cloned into a plasmid expression vector, and these clones were used to complement an env^− viral genome encoding a reporter gene to create pseudotyped viruses. We also used a set of 33 samples obtained from the South African Blood Bank as a validation group for subjects in acute infection (viral RNA positive/antibody negative). Viral RNA was extracted from blood plasma for each subject, the RNA converted to cDNA, and the cDNA used as the template in the template endpoint dilution strategy, single-genome amplification (SGA), to generate amplions spanning the viral env gene. An average of 20 amplions were sequenced per sample.

The transmitted virus in heterosexual transmission requires high levels of CD4 to enter cells. We used the pseudotyped viruses to test infectivity on cells expressing a wide range of CD4 densities. The ability to infect cells with low levels of surface CD4 is a feature of viruses that are able to infect macrophages (64–67) and of the Ba-L isolate, which was originally derived by culturing on macrophages (68). This phenotype of CD4 density dependence is conveniently measured using Affinofile cells, which have both CD4 and CCR5 under the control of inducible promoters (57). We have recently shown that viruses produced from long-lived cells in the central nervous system (CNS) have the capability to infect Affinofile cells at the lowest CD4 levels, a property shared with the Ba-L isolate, while viruses found in the blood generally do not, suggesting that viruses with the ability to infect cells expressing low levels of CD4 (e.g., macrophage/microglia) are not readily found in the blood (69). Here, we used the titration of CD4 levels on Affinofile cells as a much more reliable and quantitative assay for the determination of CD4 dependence, rather than using macrophages, which can vary significantly in the level of CD4 expressed in different cell preparations (unpublished data).

Entry assays were performed using pseudoviruses to infect Affinofile cells with CD4 levels titrated over a range from 1,780 antibody binding sites (ABS)/cell (uninduced) to 63,498 ABS/cell (fully induced) (Fig. 1A). Pseudoviruses were generated using env genes cloned from subjects with acute infection and with chronic infection and with two control clones: Ba-L and JR-CSF. Ba-L is the prototypic macrophage-tropic clone, while JR-CSF is a prototypic R5 T cell-tropic clone (requiring high levels of CD4 to enter a cell). All of the viruses from both the acute and chronic infections, as well as the R5 T cell-tropic clone JR-CSF, had dose-response curves that went to very low levels of infectivity at the lowest levels of CD4 tested and had concave curves that often did not saturate at the highest levels of CD4. This was in contrast to the macrophage-tropic Ba-L isolate, which retained approximately 20% infectivity at the lowest level of CD4 (the same as viruses in the CSF produced from long-lived cells [69]) and had a convex dose-response curve with respect to the CD4 concentration. This indicated that all of the viruses, with the exception of Ba-L, required high levels of CD4 in order to infect. Finally, the four parameters (50% inhibitory concentration [IC50], slope, upper limit, and lower limit) used to fit each dose-response curve (4-parameter log-logistic) did not differ between the transmitted viruses and viruses from chronic infections, indicating that there were no consistent differences in how transmitted viruses and chronically infected controls used CD4. Based on the requirement for high levels of CD4 to mediate entry, we conclude that the transmitted viruses were selected for replication in activated T cells and not macrophages. Similarly, in this set of chronically infected subjects, we did not find evidence for any macrophage-tropic viruses in the blood plasma.

The transmitted viruses preferentially use a maraviroc-resistant conformation of CCR5 to enter cells. We next examined the coreceptor tropism of these pseudotyped viruses. First, all of the pseudotyped viruses were sensitive to the CCR5 antagonist TAK-779 and, with the exception of one clone from a chronic infection (3011), all of the viruses were insensitive to the CXCR4 antagonist AMD3100. Thus, consistent with previous reports (8, 10–15), the transmitted viruses in this cohort all used CCR5 for entry. In our assays, we have seen that reducing Affinofile expression of CCR5 to its lowest level only reduced HIV-1 infectivity by approximately 50% (data not shown). To add more information about the utilization of CCR5 by these two groups of viruses, we carried out titration with maraviroc (another CCR5 antagonist) at both high and low expression levels of CCR5 on the Affinofile cells to assess the efficiency of CCR5 utilization. In the first experiment, we induced maximum levels of CCR5, treated the Affinofile cells with a wide range of maraviroc concentrations, and infected them with the transmitted viruses and the viruses from chronic infection (Fig. 1B). We found no difference in the mean IC50 for the transmitted viruses and the viruses from chronic infections. How-
A

CD4 antibody binding sites / cell (X1000)

Percent infection

0089  0393  0626  0665  0682  0985  1086  1176  1394  2010  2052

2103  703010010  703010131  703010200  703010217  704010017  704010042  704010069  704809221  704810053

706010016  706010164  CAP177  CAP188  CAP200  CAP210  CAP221  CAP45  Ba-L  JRCSF

3003  3004  3011  3012  3022  3025  3026  3027  3029  3031  3036

3037  3038  3040  3042  3045  3048  3050  704010034  704010075  704010124  705010154

705010212  705010212  705010223  C030  C034  C047  C061  C083  C109  C111

Percent infection

0089  0393  0626  0665  0985  1086  1172  1176  1394  2010  2052

2060  2103  703010010  703010054  703010131  703010200  703010217  703010228  704010017  704010042  704010069

704010083  704809221  704810003  70601018  706010164  CAP177  CAP188  CAP200  CAP210  CAP221  CAP45

3003  3004  3012  3022  3026  3027  3029  3031  3036  3037  3038

3040  3042  3045  3048  3050  704010034  704010075  704010124  705010154  706010223  C030

C034  C047  C061  C083  C109  C111

B

Maraviroc concentration μM
ever, we noted that some of the viruses were not fully inhibited at the highest concentrations of maraviroc and, in fact, had a plateau of infectivity resistant to maraviroc (i.e., the lower limit of the curve was greater than zero). When we compared the frequency of viruses in each group with and without a plateau in infectivity of at least 10% in the presence of maraviroc, we found a difference between the transmitted viruses and the viruses from chronic infection (Fig. 2A), i.e., in the presence of high levels of CCR5, a greater fraction of the viruses from chronic infection had a plateau of infectivity of at least 10%. We performed an additional experiment using 34 of the viruses (13 transmitted and 21 chronic) that represented a range of maraviroc sensitivities when CCR5 levels were high. When this subset of viruses was used to infect Affinofile cells expressing the lower uninduced levels of CCR5 in the presence of maraviroc (10 μM) we found that none of these viruses displayed any plateau and all were fully inhibited by high levels of maraviroc (Fig. 2B). We interpret these results to indicate that CCR5 can exist in at least two conformations when expressed at high levels, a maraviroc-sensitive conformation and a maraviroc-resistant conformation, and that the transmitted viruses disproportionately use the maraviroc-sensitive conformation.

Our results are consistent with two potential mechanisms of maraviroc resistance. Either CCR5 could exist in an alternative conformation that maraviroc is unable to bind to but that a subset of viruses can use for entry, or CCR5 could exist in an alternative conformation that maraviroc binds and a subset of viruses are able to use this maraviroc-bound alternative conformation.

**Comparison of encoded glycosylation sites in HIV-1 env gene sequences from acutely infected and chronically infected subjects.** A previous study comparing 8 linked donor-recipient pairs infected with subtype C HIV-1 showed differences in variable loop length and N-linked glycosylation site count (Asn-X-Ser/Thr) in the transmitted virus (4), an observation that was confirmed in a cohort infected with subtype A HIV-1 (19) but was less apparent in two cohorts infected with subtype B HIV-1 (21, 22). We examined the encoded N-linked glycosylation count in this large sampling of unlinked acutely infected and chronically infected subjects. In this analysis, we assume that all encoded glycosylation sites in the Env protein ectodomain are utilized, as previously reported for recombinant gp120 (29).

We used the env consensus sequence as the representative of the transmitted viruses in the 68 acutely infected subjects (n = 87) but included all of the sequences from the 62 chronically infected subjects (n = 1,111), since each person has significant intrasample variation in glycosylation site count and the level of sampling for each chronically infected subject was similar. Equivalent results were obtained when single sequences were selected from each chronically infected subject in an iterative way to create multiple comparator groups (data not shown). Figure 3A shows the distribution of glycosylation site counts in env sequences from the acutely infected subjects compared to that in env sequences from the chronically infected subjects. There was a statistically different distribution, with Env proteins from the acutely infected subjects having a mean glycosylation site count of 28.3, compared to a mean of 29.8 for the chronically infected subjects (a 5% difference in total glycosylation count number; Wilcoxon W = 33,681, P = 0.01). Thus, consistent with previous reports for heterosexual transmission (4, 19), we observed that transmission favors Env proteins that are modestly underglycosylated.

**Reduced glycosylation count encoded in the viral env gene is a feature of acutely infected men.** We considered the possibility that underglycosylation of the virus in acutely infected subjects was specifically associated with transmission from male to female or female to male. To address this question, we compared the glycosylation site counts encoded in the env genes of acutely infected men and women to the glycosylation site counts for the subsets of men and women in the chronically infected group, with the chronically infected group serving as a surrogate for the actual donors. As shown by the results in Figure 3B, the viral env genes from acutely infected men encoded fewer glycosylation sites than the env genes from chronically infected women. Similarly, when we compared env genes from acutely infected women to those from chronically infected men, we found that env genes from the acute population encoded fewer glycosylation sites, but this dif-
The difference was much smaller and did not reach statistical significance (Fig. 3C). The mean glycosylation count value for each of these groups is summarized in Table 2.

We considered several sources of error in this analysis, especially given that the samples were collected from two different countries (South Africa and Malawi) and there was a different proportion of males (higher in Malawi) and females (higher in South Africa) from each country. When we corrected for phylogenetic relatedness among the sequences (59, 60), the difference in
glycosylation levels between the transmitted viruses and the viruses in chronically infected subjects persisted. In addition, the biggest part of this difference remained in the comparison of viruses from acutely infected males versus viruses from chronically infected females. We were able to detect a difference in glycosylation count between viruses from chronically infected females in South Africa versus Malawi, but the intracountry comparison of viruses from acutely infected males and viruses from chronically infected females in Malawi remained significant, and the same comparison in South Africa approached significance in spite of the small number of males in that portion of the cohort. Thus, overall, the data showed a significant reduction in glycosylation count of the viruses in acutely infected people and, specifically, in acutely infected men even after considering confounding factors of phylogenetic relatedness and geographic isolation.

Given that the glycosylation count difference between these groups was relatively small (with the greatest difference being a mean of 27.9 versus 30.2, or about 7%, for the comparison of acutely infected men with chronically infected women), we used a second set of samples for independent validation that selection for underglycosylated viruses was most pronounced in female-to-male transmission. Sequences were generated from 35 subjects in acute infection based on their identification as being viral-RNA positive but antibody negative at the time of blood donation. The subjects were black South African blood donors who were infected with subtype C HIV-1 and presumed to have been infected via a heterosexual mode. For this analysis, we compared acutely infected men \((n = 15)\) to acutely infected women \((n = 20)\) and saw a similar trend, with the men having, on average, underglycosylated Env proteins (mean of 28.6) compared to the glycosylation of Env for the women (mean of 29.8) (Table 2). The high variance in glycosylation count precludes this comparison from reaching statistical significance, but the direction of the trend is the same as in our larger analysis.

**Distribution of glycosylation sites across Env.** We next asked if the underglycosylated sites (i.e., the missing sites) were randomly distributed across the Env protein or if specific sites were responsible for the differences in the transmitted viruses. We first carried out an alignment of the \(env\) genes to identify all conserved glycosylation sites. Because of the difficulty of aligning the length-variable regions V1/V2 and V4/V5, these were excised from the alignment and simply represented by the count of the number of glycosylation sites encoded within the variable regions for each sequence. In a few instances, adjacent glycosylation sites within the conserved regions were pooled if they occurred in a mutually exclusive (i.e., overlapping) way among the sequences. Using this alignment, we identified 22 sites in the ectodomain of Env (outside the V1/V2 and V4/V5 variable-length regions), where a glycosylation site occurred in at least 50% of the sequences. Some of the glycosylation sites were present in nearly 100% of the sequences (highly conserved), while the rest were present at frequencies ranging down to 60 to 70% of sequences (moderately conserved). On average, there were 8 to 10 glycosylation sites encoded in the variable regions, or about one-third of the total glycosylation sites in Env.

The frequencies of each N-linked glycosylation site in the sequences generated from the acutely and chronically infected subjects were compared (Fig. 3D). When the counts of glycosylation sites within the variable regions were analyzed, we found that approximately 60% of the reduced glycosylation count signal of the acutely infected subjects could be accounted for within the variable regions, with just over half of the effect in V1/V2 and just under half in V4/V5 (data not shown). Within the conserved regions of the Env ectodomain, we identified 8 glycosylation sites that were more common in the transmitted viruses and 14 sites that were more common in viruses from chronically infected subjects (Fig. 3D). Of these, only one position was statistically less common in the viruses from the chronically infected subjects (position 356/358), and four positions (234, 442, 611, and 624/625, with the latter two being in the gp41 ectodomain of Env) were statistically less common in the acutely infected subjects. These data indicate that Env proteins from acutely infected subjects are underglycosylated in both the highly variable regions and the relatively conserved regions.

The transmitted virus in heterosexual transmission is not globally neutralization sensitive but can show selective sensitiveness. We next tested the pseudotyped viruses for neutralization sensitivities to a group of monoclonal antibodies (Fig. 4A) and to a group of broadly neutralizing polyclonal sera (Fig. 4B). The viruses, pseudotyped with these subtype C Env proteins, were largely resistant to b12, 2G12, and 2F5 (at the concentrations tested), as previously noted for this subtype (70). Those viruses with some sensitivity to 2F5 had substitutions at A662, S665, and K667 within the known 2F5 epitope in the subtype C consensus sequence. There was no significant difference in sensitivity to 4E10 between the viruses pseudotyped with Env proteins derived from acute versus chronic infection. For subtype B HIV-1, the acute/transmitted virus has been reported to be either more sensitive (71) or more resistant to 4E10 (8). In addition, there was no difference in the sensitivity to soluble CD4 (sCD4), suggesting similar conformations of the Env proteins with respect to their interaction with CD4 and an absence of an open Env conformation associated with culture-adapted viruses (72). Similarly, the CD4 binding site monoclonal antibodies VRC01 and CH31 showed no difference in their ability to neutralize acute and chronic viruses (Fig. 4A). Finally, the monoclonal antibodies PG9, PG16, and CH01, which are directed at the \(\beta\)-sheet scaffold of the V1 and V2 loops, also did not distinguish the viruses from acute infection and those from chronic infection. It is worth noting that the IC\(_{50}\) titers for each of the more-active antibodies ranged from 100- to 1,000-fold, indicating profound differences in sensitivity to neutralization even for these broadly neutralizing antibodies.

The pseudotyped viruses with the different Env proteins displayed different but measurable neutralization sensitivities to a set of seven sera previously characterized as broadly neutralizing, as well as to pooled sera from subjects infected with subtype B HIV-1 (HIV Ig) or subtype C HIV-1 (HIV Ig-C) (Fig. 4B). For most of the sera, including the HIV Ig pools, there were no detectable differences in the sensitivities of the viruses from acute infection and the viruses from chronic infection. Only one serum, SA-C26, neutralized Env proteins from acutely and chronically infected individuals differently, with viruses pseudotyped with chronically derived Env proteins being significantly more sensitive to SA-C26 than those derived from acute infections \((P = 0.04)\) (Fig. 4B). Thus, the determinants of neutralization sensitivity to heterologous sera (i.e., protein sequence and conformation) are largely shared and do not appear to be differentially selected during transmission, at least within the limits of detection given the size of this study.
The presence or absence of glycosylation sites can differentially affect the potency of monoclonal antibodies and polyclonal sera. Given the large data set of Env protein sequences and neutralization sensitivities, we explored the contribution of the presence or absence of specific glycosylation sites in determining the neutralization potencies for the different monoclonal and polyclonal antibodies. We used a natural log transformation of the neutralizing titers in a linear regression model to explore the potential contribution of each of the 22 glycosylation sites within the conserved regions of the Env ectodomain. Because viruses from acute and chronic infections typically do not differ in their neutralization sensitivities, we pooled all of the data from both sets of viruses to enhance our statistical power to detect a correlation between specific glycosylation sites and neutralization sensitivities. The results are shown in Figure 4C, with filled bars indicating that the presence of a glycosylation site reduced sensitivity to the antibody (escape) and open bars indicating enhanced sensitivity to neutralization if the glycosylation site was present (either as part of the epitope or through a conformational change). Several correlations supported the validity of this approach. First, two of the CD4 binding site monoclonal antibodies were negatively impacted by the presence of the glycosylation sites near the CD4 binding site, at positions 276 and 386 for VRC01 and 386 for b12, a glycosylation site known to affect b12 potency (36, 73); however, we did not detect a signal for the other CD4 binding site antibody, CH31. Second, the potency of PG9 was enhanced when a glycosylation site was present at position 160, which is known to be part of the PG9 epitope in the V1/V2 sheet scaffold (74, 75). PG16, which has specificity similar to that of PG9, also had enhanced potency when a glycosylation site was present at position 160; in addition, PG16 showed enhanced potency when position 197 was glycosylated (also part of the V1/V2 sheet scaffold), although this association lost statistical significance after correction for multiple comparisons (data not shown).

Given our ability to observe correlations for interactions that are known to exist, we analyzed the data for new interactions. We note that SA-C26, which showed more potency toward viruses from chronic infection than from acute infection (Fig. 4B), was more potent when there was a glycosylation site at position 616 (Fig. 4C), a site that is underrepresented in the acutely infected viruses (Fig. 3D). Thus, the differential glycosylation at this site between viruses from acute infection and chronic infection may account for the differential activity of this polyclonal serum with these two groups of viruses. The monoclonal antibody CH01, which is thought to target the sheet scaffold of the V1 and V2 loops (76), was more potent in the absence of a glycosylation site at position 130. Sensitivity to soluble CD4 was complex, with glycosylation at positions 160 and 197 around the V1/V2 sheet scaffold conferring increased resistance and the juxtaposed glycosylation sites at positions 611 and 616 in gp41 differentially affecting sensitivity to sCD4. These differences may reflect the impact on either protein conformation or ease of conformational change; a similar pattern for these two glycosylation sites in gp41 was also seen for the polyclonal antibodies SA-C26 and SA-C62. Finally, glycosylation at position 230 was associated with reduced sensitivity to the monoclonal antibodies PG16 and 4E10 and the polyclonal antibodies HIV Ig and SA-C26 by an unknown mechanism, since the epitopes for PG16 and 4E10 are dispersed between gp120 and gp41 (77).

As one test of the validity of these correlations (beyond the known relationships) we examined the role of glycosylation at position 130 in neutralization by CH01. The initial report of epitope mapping of monoclonal antibody CH01 suggested that CH01 and PG9 bind overlapping epitopes (76). However, in this large panel of viruses, we saw that the presence of an N-linked glycan at site 130 made viruses more resistant to CH01 but did not alter sensitivity to PG9. To further explore the relationship between glycosylation at position 130 and sensitivity to CH01, we mutated a panel of 10 of these subtype C env clones to encode Env proteins fully glycosylated at the moderately conserved sites. In addition, we mutated each of the otherwise fully glycosylated clones to remove the glycan motif at 130. We saw an increase in neutralization in three viruses when the clone lacked only the 130 glycan, consistent with the pattern of neutralization observed in the larger panel. Two of the fully glycosylated viruses became sensitive to CH01 when a glycosylation site at 160 was added to the
parent, consistent with the earlier mapping that highlighted the importance of glycan 160 in neutralization by CH01 (76). The remaining 5 viruses were resistant to neutralization by CH01 in all cases. These results also emphasize the complexity of the interaction between glycosylation sites in individual Env proteins and neutralization by glycan-binding antibodies.

**DISCUSSION**

Understanding the selective pressures acting on the Env protein during transmission informs our knowledge of the mechanism of transmission and is essential to the development of prevention strategies. Several previous studies have observed that the transmitted virus is underglycosylated in an HIV-1 subtype C cohort (4) and in an HIV-1 subtype A cohort (19), although this effect was less apparent overall in a subtype B cohort (19, 21, 22), with the exception of underrepresentation of a glycosylation site at position 413 (24). We have been able to confirm this observation of reduced glycosylation of the transmitted virus in this large subtype C cohort (Fig. 3A). Consistent with the underglycosylation of the Env protein in the transmitted virus, the virus accumulates glycosylation sites during the course of infection (70, 78, 79). Thus, there appears to be a cycling of glycosylation sites, with a reduction at the time of transmission and an accumulation of sites over time.

Based on our and other studies (4, 19), it is clear that viruses with fewer glycosylation sites can have a transmission advantage in certain settings, but the magnitude and mechanism of this advantage is unknown. We have estimated what the magnitude of this advantage would have to be in order to transform the distribution of glycosylation counts among viruses from chronically infected subjects into the distribution seen in the transmitted viruses. We estimate that an increase in the probability of transmission of 20% for each lost glycosylation site could account for the differences in these distributions (Fig. 3A). A persistent question has been whether this transmission advantage is due to selection at specific glycosylation sites or selection for an overall reduction in glycan number. Our observation that two-thirds of the reduction in glycosylation sites was associated with the variable regions of Env indicates that the transmission advantage is likely not due to a specific glycan but, rather, that transmission of subtype C HIV-1 in the context of heterosexual transmission favors viruses whose Env proteins have overall fewer glycans.

The fact that subtypes differ in the extent to which they favor transmission of underglycosylated viruses may provide insight into the mechanism of this transmission advantage. We hypothesize that this difference between subtypes B and C is due to differences in their predominant modes of transmission. This suggestion is supported by two observations. First, the subtype B HIV-1 epidemic is dominated by male-to-male transmission, and the magnitude of the transmission advantage of reduced glycosylation is much less apparent. In contrast, subtype C HIV-1 is predominantly heterosexual transmitted, and transmitted HIV-1 subtype C variants have a more pronounced reduction in glycan number. Second, we show that female-to-male transmission is associated with a larger reduction in glycosylation count than male to female (Fig. 3). Together, this suggests that transmission from females selects for underglycosylated viruses.

Transmission from females requires that the virus maintain infectivity after being secreted into the cervicovaginal mucus. Humans make a protein that binds mannose, mannose binding lectin (MBL), and this protein has been shown to be capable of neutralizing HIV-1 (80). It is possible that MBL in vaginal secretions traps or neutralizes virus in a way that is enhanced by higher glycosylation density. This model is consistent with the recent observation of a reduced glycosylation count in viruses transmitted intrapartum but not intrauterine (81). Alternatively, differences in Langhans cells (LCs) in the male epithelium and the female epithelium (such as the differences observed between skin and vaginal LCs [82]) could result in either enhancement of trans-infection of T cells by underglycosylated viruses or inhibition of trans-infection of T cells by viruses that bind surface lectins too tightly due to higher levels of glycosylation.

It is now clear that transmission involves the infection of a cell with high levels of CD4, which identifies CD4+ T cells as the target. Studies of tissue at the site of infection in recently infected macaques found that T cells were the predominant cell type infected (83). Several studies of human cohorts have failed to find examples of transmitted viruses that can efficiently enter macrophages (12, 16–18) or infect cells with low levels of CD4 (a surrogate for macrophage tropism) (11, 13). We have used a large sample size and a quantitative assay for CD4 dependence to show that the viruses transmitted in heterosexual transmission require high levels of CD4 for entry, i.e., the high levels found on activated T cells and not on macrophages (Fig. 1A). Given the low levels of CD4 on dendritic cells from the peripheral blood (84) or the gut mucosa (85), it is unlikely that infection of DCs plays a role in the transmission process. Finally, our failure to find viruses isolated from the blood of chronically infected individuals that have the ability to enter cells with low levels of CD4 indicates that the evolution of macrophage-tropic virus is likely restricted to very specific circumstances, such as the CNS, where we have been able to show a link between the infection of a long-lived cell and the presence of virus able to enter cells with low levels of CD4 (69, 86).

Thus, the transmitted virus, and most examples of HIV-1 found in the blood, are appropriately called R5 T cell-tropic.

Transmitted viruses and viruses from chronic infections do not differ in how their Env proteins interact with CD4. This is based on two assays: inhibition by soluble CD4 (Fig. 4) and entry efficiency as a function of CD4 density (Fig. 1A). Previous work suggested that a property of transmitted viruses is high α4β7 integrin binding and low binding of gp120 to a monomer of soluble CD4, with the typical HIV-1 Env protein having the reverse property (25). Similar to our results with subtype C HIV-1, no difference in sensitivities to soluble CD4 has been seen for subtype B HIV-1 in comparing acute/transmitted viruses and viruses from chronic infection (8, 71), although a difference has been reported for neutralization sensitivities to CD4 binding site antibodies for subtype B (15) but not subtype C HIV-1 (13) (Fig. 4A). Thus, if there are differences in the interaction with CD4, they are small and largely assay specific. Similarly, a recent analysis of a panel of transmitted viruses and viruses from chronic infection failed to confirm a role for α4β7 binding as a specific feature of the transmitted virus (13). Overall, we found no difference in neutralization sensitivities between viruses from acute versus chronic infection, with the exception of one polyclonal serum (Fig. 4B). Thus, we conclude that there are not significant differences in either conformation or heterologous neutralization sensitivity that distinguish the transmitted virus. The relationship between underglycosylation of the transmitted virus and neutralization sensitivity is less clear. In 13 of the 18 examples where glycosylation alters neutralization sensitivity, underglycosylation increases sensitivity to neutralization...
(Fig. 4C), but connecting this pattern to transmission is difficult given that underglycosylation of the transmitted virus is likely a small signal in the background of high variability of both neutralization sensitivity and overall glycosylation diversity.

Another feature of the transmitted virus phenotype is an increased ability to use a maraviroc-sensitive conformation of CCR5. G protein-coupled receptors (GPCRs) have extreme levels of structural flexibility that allow them to bind many different ligands and achieve ligand-specific conformations (reviewed in reference 87). Epitope mapping studies have found that some anti-CCR5 monoclonal antibodies recognize a greater fraction of the total CCR5 molecules than other antibodies, thus suggesting that, like other GPCRs, CCR5 exists in a number of conformations (88, 89). While the nature of CCR5 conformational variation has not been extensively studied, it appears to be substantial on both cultured (88, 89) and primary T cells (88) and could be generated by a number of mechanisms, including posttranslational modification (90, 91), the lipid environment (92), and ligand or G protein binding (87). Furthermore, some CCR5 antagonist-resistant viruses have been shown to differ in their sensitivities to anti-CCR5 monoclonal antibodies, raising the possibility that HIV-1 can evolve the ability to bind these alternative CCR5 conformations as a resistance pathway (88). Similarly, resistance to maraviroc has been shown to vary by cell line and by donor (93), consistent with the idea that different cells can display different forms of CCR5.

Here, we show that viruses isolated from chronically infected subjects are more often partially resistant to maraviroc than viruses isolated from acutely infected subjects. While 75% of the viruses from chronic infections had the ability to use this alternative conformation, only about 40% of the transmitted/founder viruses could use this conformation (Fig. 1B and 2). Because the subjects used in this study were treatment naive, these innate resistance levels reveal variation in how viruses interact with CCR5. This interpretation implies that viruses from chronic infection are more variable in how they interact with CCR5 and that they are able either to infect using maraviroc-bound CCR5 or using a CCR5 conformation that maraviroc is unable to bind. It is possible that this diversity in the ability to use different CCR5 conformations is analogous to the evolution to use CXCR4 as a coreceptor late in infection, possibly allowing the virus to grow in a different subset of T cells. Several studies have shown that R5 viruses taken from late in disease are more difficult to inhibit with agents that have an expanded CCR5 coreceptor usage capacity, it is difficult to interpret the selective pressure that would restrict this capacity in the transmitted virus. Perhaps the expanded coreceptor usage capacity allows the virus to infect a cell type that is on average less productive for successful transmission. Alternatively, this expanded coreceptor capacity may be linked to some other feature of Env that is important for transmission or to the level of glycosylation. We do not know which form of CCR5 predominates on CD4+ T cells, although the therapeutic efficacy of maraviroc suggests that it is largely the sensitive form (95). However, a recent failure of maraviroc to block rectal transmission of the simian-human immunodeficiency virus (SHIV) isolate 162p3 in macaques in the face of high drug exposure (96) may suggest a role for this alternative conformation in at least some settings. Parrish et al. (13) did not observe a resistance plateau using a largely nonoverlapping set of subtype C HIV-1 isolates, although this analysis was done using a cell line (NP2/CD4/CCR5) that expresses a single, lower level of CCR5. These investigators have recently repeated this experiment based on the results reported in our manuscript and have confirmed the difference in transmitted viruses versus chronic viruses (97).

The modest underglycosylation of transmitted viruses (Fig. 3) emphasizes the fact that glycosylation is a dynamic state for HIV-1, with some glycosylation sites in the highly variable loops being poorly conserved and other sites in the conserved regions of Env being moderately or highly conserved. There are a number of examples, from this and previous studies, indicating that there is an intimate relationship between glycosylation at specific sites in Env and sensitivity or resistance to antibody neutralization. We propose that a vaccine that targets structural epitopes underlying multiple, variable glycosylation sites could neutralize a large fraction of viruses.

There are several examples where inclusion of part of the carbohydrate structure into the epitope has resulted in an antibody that has broad neutralizing capacity, such as the monoclonal antibody 2G12 (98, 99), the monoclonal antibody PGT 128, which shows a specificity for the presence of a glycan at position 332 (100), and the broadly neutralizing antibodies PG9 and PG16, which include specific glycans in the V1/V2 β-sheet scaffold as part of the epitope (35, 74, 101). However, epitopes that include carbohydrate must be the exception to what the virus experiences, since selection maintains such a high glycosylation density as a general strategy to avoid neutralization.

There is also evidence that glycosylation sites can confer resistance to neutralization, as opposed to being the target of neutralization. Changes in glycosylation within the variable loops represent an important path of escape from autologous neutralizing antibodies (77). There are also examples where moderately conserved glycans have been implicated in protecting specific features of the Env protein surface. The glycan at position 386 is a major determinant of sensitivity to the CD4 binding site antibody b12 (36, 73), and we were able to see this effect across our data set, as well as a potency effect on the CD4 binding site antibody VRC01 by glycosylation at 386 and the spatially adjacent site 276 (Fig. 4C). Similarly, the α2 helix can be an important target for neutralization (42, 102), with one example where escape was mediated by the addition of a glycan at position 339 (103). More recently, Moore et al. (104) have observed escape from an autologous neutralization response by the addition of a glycan at position 352, which then conferred sensitivity to the glycan-dependent monoclonal antibody PGT 128. Given that glycans are typically present in the conserved regions of Env (Fig. 3D), the antibodies that react with the surface protein structures of Env otherwise covered by these glycans would appear as largely autologous, i.e., of restricted range in their neutralization properties. A more systematic approach is needed to determine the breadth of the neutralization properties of antibodies to these surface structures among isolates where the same glycan is missing.

If these moderately conserved glycans are viewed as masks for the Env protein surface, then these carbohydrate side chains can be grouped as protecting specific structures (Fig. 5). The glycans at positions 262, 332, 442, and 446/448 are all placed such that they could occlude the surface of the three-stranded β-sheet β12/13/22. The glycans at positions 276 and 386 are positioned to protect the protein surface around the CD4 binding site. Similarly, a group of glycans (positions 88, 230, 234, and 241) in the region where gp120 and gp41 are thought to interact (105, 106) could...
represent a distinct domain where glycans are occluding putative neutralization epitopes. The glycan at position 337/339 sits on the α2 helix. The single glycans on loop C (position 289) and loop E (position 356/358) could similarly occlude specific epitopes in these loops. Finally, the newly described four-stranded β-sheet that provides a scaffold for the V1 and V2 surface loops is covered with four glycans (positions 130, 156, 160, and 197) (74).

Given that these glycosylation sites are variably present, we asked what the mean occupancy of these sites was collectively among the transmitted viruses. As can be seen by the results in Figure 5, 93% of the transmitted viruses are missing a glycan associated with at least one of these structures, and 74% have at least one glycan missing from two or more separate structures. If these moderately conserved glycosylation sites protect epitopes that can be recognized by the host, then the absence of some of these glycosylation sites in most isolates should make the virus sensitive to antibodies targeted to these surface structures of Env. In much the same way that the virus toggles between cytotoxic T lymphocyte epitope escape mutations and reversion to wild-type sequence (107–109), the virus may also toggle between exposing an otherwise carbohydrate-occluded epitope in the absence of the selective pressure of neutralizing antibodies to that site in a specific host and retaining (or evolving anew) the glycosylation site when there is a host response to that protein surface epitope. A vaccine approach that was able to elicit a response to these surface structures could represent a combinatorial approach to neutralizing a large majority of transmitted viruses, targeting the differing subsets of structures exposed on different viruses, an approach that is already being explored for the CD4 binding site (110). This strategy would take advantage of the variable presence of these glycosylation sites in the conserved domains of Env among the entire viral population, an advantage that is enhanced by the further reduction in glycosylation of the transmitted virus.

ACKNOWLEDGMENTS

We thank all of those who contributed to the collection of these samples and all of the subjects who provided the samples.

This work was funded by awards from the National Institutes of Health (U01 AI067854 [CHAVI], R01 AI10273 and R37 AI44667 to R.S., and R01 AI092218 to B.L. We also received support from the UNC Center for AIDS Research (P30 AI50410) and the UCLA Center for AIDS Research (P30 AI028697).

REFERENCES

Ping et al.


tropism phenotypes and identifies envelopes in the brain that confer an enhanced tropism and fusogenicity for macrophages. J. Virol. 78:6915–6926.


