Differential Clade-Specific HLA-B*3501 Association with HIV-1 Disease Outcome Is Linked to Immunogenicity of a Single Gag Epitope


Differential Clade-Specific HLA-B*3501 Association with HIV-1 Disease Outcome Is Linked to Immunogenicity of a Single Gag Epitope


Department of Paediatrics, University of Oxford, Oxford, United Kingdom; Center for AIDS Research, Kumamoto University, Kumamoto, Japan; AIDS Clinical Center, National Center for Global Health and Medicine, Tokyo, Japan; Department of International Health, Immunology and Microbiology, University of Copenhagen, Copenhagen, Denmark; Centre for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico; Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, United Kingdom; Microsoft Research, eScience Group, Los Angeles, California, USA; Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana; HIV Pathogenesis Programme, The Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, South Africa; Department of Sexual Health, Royal Berkshire Hospital, Reading, United Kingdom; Department of Genitourinary Medicine, Northamptonshire Healthcare NHS Trust, Northampton General Hospital, Cliftonville, Northampton, United Kingdom; Department of Sexual Health, Wycombe Hospital, High Wycombe, Bucks, United Kingdom; Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA; Fundación irsiCaixa-HIVACAT, Hospital Universitari Germans Trias i Pujol, Autonomous University of Barcelona, Barcelona, Spain; Instituto Catalana de Recerca i Estudis Avancats, Barcelona, Spain; The Ragon Institute of Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; and Emory Vaccine Center, Atlanta, Georgia, USA

The strongest genetic influence on immune control in HIV-1 infection is the HLA class I genotype. Rapid disease progression in B-clade infected has been linked to HLA-B*35 expression, in particular to the less common HLA-B*3502 and HLA-B*3503 subtypes but also to the most prevalent subtype, HLA-B*3501. In these studies we first demonstrated that whereas HLA-B*3501 is associated with a high viral set point in two further B-clade-infected cohorts, in Japan and Mexico, this association does not hold in two large C-clade-infected African cohorts. We tested the hypothesis that clade-specific differences in HLA associations with disease outcomes may be related to distinct targeting of critical CD8+ T-cell epitopes. We observed that only one epitope was significantly targeted differentially, namely, the Gag-specific epitope NPPIPVGDIY (NY10, Gag positions 253 to 262) (P = 2 × 10−7). In common with two other HLA-B*3501-restricted epitopes, in Gag and Nef, that were not targeted differentially, a response toward NY10 was associated with a significantly lower viral set point. Nonimmunogenicity of NY10 in B-clade-infected subjects derives from the Gag-D260E polymorphism present in ~90% of B-clade sequences, which critically reduces recognition of the Gag NY10 epitope. These data suggest that in spite of any inherent HLA-linked T-cell receptor repertoire differences that may exist, maximizing the breadth of the Gag-specific CD8+ T-cell response, by the addition of even a single epitope, may be of overriding importance in achieving immune control of HIV infection. This distinction is of direct relevance to development of vaccines designed to optimize the anti-HIV CD8+ T-cell response in all individuals, irrespective of HLA type.

Several genome-wide association studies now indicate that the host HLA class I genotype is the major genetic determinant of HIV-1 disease progression (19, 20, 61). Previously it had been established that differences in HLA allele expression have a substantial impact on HIV disease outcome, in both B-clade (10, 19, 20, 59) and C-clade (38, 44, 54, 63) infection. Variation at the HLA-B locus has the greatest impact on viral set point (20, 38). This may result from the increased diversity of HLA-B compared to non-HLA-B alleles (28), affecting the repertoire, protein specificity, and peptide-binding characteristics of epitopes presented by HLA-B alleles (38, 39, 41). In addition, HLA-Bw4 alleles can act as KIR ligands and modulate an NK response, with certain HLA-KIR combinations resulting in selection pressure on HIV and/or significantly influencing viral set point (2, 3, 52, 68).

The mechanisms by which certain HLA alleles are consistently linked with particular HIV disease outcomes remain unresolved. Several possible mechanisms have been proposed. First, HLA-associated immune control has been linked to the specificity of the CD8+ T-cell response (39, 54). In this way, HLA alleles such as HLA-B*57 or HLA-B*27, associated with immune control (4, 46, 59), restrict dominant Gag-specific responses, escape from which results in a substantial reduction in viral replicative capacity (13, 15, 46, 53, 65). In contrast, HLA alleles such as HLA-B*35, associated with rapid disease progression (12), restrict dominant epitopes in Nef, Env, and other non-Gag proteins (7, 39, 58, 67, 69, 72, 73).

A second mechanism proposed for the association of particular HLA types with characteristic HIV disease outcomes is through an impact on antiviral NK activity, since certain HLA alleles have the...
potential to act as KIR ligands. The HLA alleles associated with lowest viral set point tend to be HLA-Bw4 alleles (22). HLA-Bw4 alleles expressing Ile at HLA residue 80 significantly reduce the viral set point in combination with either KIR3DS1 or KIR3DL1 (51, 52). However, the impact of HLA-KIR combinations only partially explains the effect of protective alleles such as HLA-B*27 and HLA-B*57 or of disease susceptibility alleles such as HLA-B*35 (5, 52).

A third mechanism, more recently proposed (41), suggests that disease susceptibility (61) HLA alleles such as HLA-B*0702 and HLA-B*3501 have peptide-binding motifs such that large numbers of self peptides can bind, and hence a relatively large proportion of the T-cell receptor (TcR) repertoire would be lost through negative selection of autoreactive T cells in the thymus. In contrast, protective alleles such as HLA-B*2705 and HLA-B*5701 have more restrictive peptide-binding motifs, with a requirement for Arg at P2 in the case of HLA-B*2705 and a strong preference for Trp at the C-terminal position in the case of HLA-B*5701 (50). This would result in fewer autoreactive T cells being deleted in the thymus via negative selection and therefore a relatively large TcR repertoire remaining to accommodate the challenge of epitope variation inevitably presented by viruses such as HIV.

An additional mechanism proposed to explain the status of HLA-B*3503 as linked with more rapid disease progression than HLA-B*3501 (25), from which it differs by only one amino acid, derives from the observation that HLA-B*3503 binds with significantly greater affinity than HLA-B*3501 to immunoglobulin-like transcript 4 (ILT-4), an inhibitory major histocompatibility complex (MHC) class I receptor expressed on dendritic cells (34). These data suggest the possibility that dendritic cell function may be significantly affected by a variety of HLA molecules, thereby explaining a range of differential HLA associations with HIV disease outcome.

We here describe an observation that allows us to test the first of these hypothetical mechanisms. While HLA-B*3503 is associated with less rapid progression to HIV disease than the less common subtypes of HLA-B*35 in Caucasians, B*3502 and B*3503 (25), HLA-B*3501 itself has also been associated with higher-than-average viremia in B-clade HIV-1 infection (42). For example, in a recent study of 3,622 B-clade-infected study subjects, HLA-B*3501 was strongly associated with HIV disease progression (61). However, in a cohort of C-clade-infected study subjects (n = 1,210) in Durban, South Africa, we noted that HLA-B*3501 is somewhat protective: viral set points tend to be somewhat lower in HLA-B*3501-positive subjects. Indeed, having removed the effect of HLA-B*57, HLA-B*5801, HLA-B*1801, and HLA-B*5802, the alleles having the strongest impact on viral set point and absolute CD4 count (38), HLA-B*3501 was the HLA-B allele associated with the highest absolute CD4 counts (44, 54) in this C-clade-infected cohort.

We show here, first, that this observation of clade specificity of the HLA-B*3501 effect on viral set point could be replicated in two additional B-clade-infected cohorts, namely, in Japan and in Mexico, and in an additional C-clade-infected cohort in Botswana. We then tested the hypothesis that the clade-specific difference in HLA-B*3501-associated HIV disease outcome could be related to altered specificity of the CD8+ T-cell response. Based on the “Gag hypothesis” as described above, HLA-B*3501-restricted responses in C-clade infection would tend to be more Gag directed and less Nef/Env directed than in B-clade infection.

### MATERIALS AND METHODS

**Ethics statement.** Ethics approval was given by the following: the University of KwaZulu-Natal Review Board and the Massachusetts General Hospital Review Board (Durban cohort); the Office of Human Research Administration, Harvard School of Public Health, and the Health Research Development Committee, Botswana Ministry of Health (Gaborone cohort); the Oxford Research Ethics Committee (Thames Valley and other cohorts); and the Ethics Committees of Kumamoto University and National Centre for Global Health and Medicine (Kumamoto cohort). Study subjects from all cohorts gave written informed consent for their participation.

**Study cohorts.** We studied a total of 3,132 adults with chronic, antiretroviral therapy (ART)-naïve HIV-1 infection, recruited from six cohorts as follows: (i) Durban, South Africa (C clade; n = 1,218), as previously described (38, 39, 46, 54); (ii) Gaborone, Botswana (C clade; n = 514) via the Mma Bana study, as previously described (66); (iii) Kumamoto, Japan (B clade; n = 242), as previously described (37); and (iv) Mexico City, Mexico (B clade; n = 771), as previously described (6) (see Table S1 in the supplemental material); (v) the Thames Valley cohort, United Kingdom (mixed clades; n = 237), as previously described (60, 62); and (vi) a B-clade-infected cohort of 150 subjects drawn from multiple ethnicities, also as previously described (24). Viral loads were determined using Roche Amplicor version 1.5 assay; CD4+ T-cell counts were determined by flow cytometry.

**HLA typing and classification.** HLA typing from genomic DNA was undertaken by sequence-based typing as previously described (38), locus-specific PCR products of exons 2 and 3 were amplified and sequenced. In the Kumamoto cohort, 32/37 subjects with HLA-B*35 were typed to 4 digits, and all 32 of these were HLA-B*3501 positive; because of this, and because of a previous analysis of 1,018 Japanese subjects (36) which showed that 158/159 subjects with HLA-B*35 had HLA-B*3501, the remaining 5 Japanese subjects were designated HLA-B*3501 positive. Likewise, in the southern African cohorts, 96/102 HLA-B*35-positive subjects typed to 4 digits were HLA-B*3501 positive. For 23 Durban subjects in whom HLA-B*35 typing had been undertaken only to 2-digit resolution, we used an HLA completion tool (http://atom.research.microsoft.com /HLACompletion) (47) to predict the most likely 4-digit HLA-B*35 allele. In all cases HLA-B*3501 was predicted as the 4-digit type with a high level of statistical certainty (probability of B*3501, 0.86 to 0.98; median, 0.97). For this reason, we designated all 23 Durban subjects with HLA-B*35 typed to 2-digit resolution as HLA-B*3501.

**Definition of HLA-B*3501-restricted epitopes.** To define a comprehensive list of HLA-B*3501-restricted epitopes, we identified previously characterized epitopes from studies of predominantly B-clade-infected subjects (Los Alamos “A list”; www.lanl.gov) (48) and also identified five novel HLA-B*3501-restricted epitopes by testing recognition of 410 overlapping 18-mer peptides in a cohort of C-clade-infected subjects (see Table S2 in the supplemental material). One of these (HA9) has, since the start of this study, now been confirmed by another group (74). From this dual approach, 13 HLA-B*3501-restricted epitopes were identified for further analysis (Table 1).

**IFN-γ ELISpot assays.** Gamma interferon (IFN-γ)–enzyme-linked immunosorbent spot (ELISpot) assays were undertaken using fresh or cryopreserved peripheral blood mononuclear cells (PBMCs). We screened for HIV-1-specific responses statistically associated (q < 0.05) with the expression of HLA-B*3501 by testing a total of 1,010 chronically infected subjects (n = 795 from Durban; n = 215 from the Thames Valley) against a panel of 410 overlapping peptides (OLPs) spanning the entire HIV proteome, as previously described (38, 39, 54). Significant associations were determined using Fisher’s exact test and corrected for multiple comparisons using a q value (false-detection rate [FDR]) approach as previously described (11, 40, 54).

In order to screen subjects with HLA-B*3501 for specific responses to HLA-B*3501 epitopes, B-clade-infected subjects were tested for IFN-γ responses to optimal peptides (Japan, n = 30) or against overlapping
TABLE 1 Thirteen HLA-B*3501-restricted epitopes in HIV-1 from Gag, Pol, Rev, Env, and Nef proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Clade</th>
<th>Epitope sequence</th>
<th>HXB2 position</th>
<th>Epitope designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p24 Gag</td>
<td>B</td>
<td>HPVHAGPI</td>
<td>Gag 216-224</td>
<td>Gag HA9*</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>NPIPPIPVEIY</td>
<td>Gag 253-262</td>
<td>Gag NY10</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>B</td>
<td>TVLDRVGA</td>
<td>Pol 262-270</td>
<td>RT TY9</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>VPLDKDFRK</td>
<td>Pol 273-282</td>
<td>RT VY10</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>NPDPVIVIQY</td>
<td>Pol 330-338</td>
<td>RT NY9</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>HPIVVAETFY</td>
<td>Pol 587-596</td>
<td>RT EY10*</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int</td>
<td>B</td>
<td>TPAETGQETAY</td>
<td>Pol 804-814</td>
<td>Int IY11*</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rev</td>
<td>B</td>
<td>KTRVLIRKFL</td>
<td>Rev 14-23</td>
<td>Rev KY10, Rev QY10*</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>QA--I--I--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gp120</td>
<td>B</td>
<td>VFWKEATTTL</td>
<td>Env 42-52</td>
<td>Env VL11</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>DNPQEVVL</td>
<td>Env 78-86</td>
<td>Env DL9</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gp41</td>
<td>B</td>
<td>TAVPNNAS</td>
<td>Env 606-614</td>
<td>Env TW9</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nef</td>
<td>B</td>
<td>VRPLRMKT</td>
<td>Nef 73-81</td>
<td>Nef YY8, Nef VF8</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>YPLTFGWC</td>
<td>Nef 135-143</td>
<td>Nef YY9, Nef YF9*</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 13 epitopes include 8 from the Los Alamos database “A list” (www.lanl.gov) and 5 new HLA-B*3501-restricted optimal epitopes (indicated by asterisks).

The B- and C-clade consensus sequences of each epitope are listed; a dash indicates no difference between clades. Residues at position 2 and at the C terminus are in bold.
ing HLA-B*3501 pulsed with NY10-260E or NY10-260D peptide for 2 h at 37°C. Brefeldin A (10 μg/ml) was then added, and the cocultures were continued for additional 4 h. Cells were stained with phycoerythrin (PE)-labeled anti-CD8 monoclonal antibody (Dako Corporation, Glostrup, Denmark) and subsequently fixed (4% paraformaldehyde), permeabilized (0.1% saponin and 20% NCS in phosphate-buffered saline), and intracellularly stained with fluorescein isothiocyanate (FITC)-labeled anti-IFN-γ monoclonal antibody (MAb) (PharMingen, San Diego, CA). Samples were acquired on a FACSCalibur instrument within 24 h of staining and data analyzed using FlowJo version 8.8.6.

Generation of mutant virus NL43-2 E260D and epitope processing assay. The NL43-2-Gag260D mutant virus was generated by introducing the Gag260D mutation into the NL43-2 Gag260E backbone using site-directed mutagenesis (Invitrogen). After virus generation, 721.221-CD4- B*3501 and 721.221 target cells were infected with NL43-2 B-clade WT(Gag260E) or NL4-32-Gag260D mutant virus. The infection rates were determined by the level of intracellular p24-positive cells stained with FITC-conjugated anti-p24 MAb (KC57-FITC; BD Biosciences) and followed over 6 days. When the level of p24-positive target cells reached 80%, the Gag NY10-specific CD8+ T-cell line and the control Pol-EY10 CD8+ T-cell clone were cocultured with the target cells for 5 h in the presence of brefeldin A and subsequently stained for intracellular IFN-γ as described above. The level of IFN-γ-positive CD8+ T cells after coculture was used as a measure of the level of specific epitope presentation and controlled by uninfected HLA-matched and infected HLA-negative 721.221 cells. Peptide-pulsed HLA-matched target cells were used as positive controls for optimal epitope presentation. Samples were acquired on a FACSCalibur instrument within 24 h of staining and data analyzed using FlowJo version 8.8.6.

Peptide-MHC binding studies. HLA-peptide binding studies were undertaken using a luminescent oxygen channeling immunoassay (LOCI) as previously described (29). We tested binding for 12 HLA-B*3501 epitopes as shown in Table S3 in the supplemental material. Binding assays were performed in quadruplicate; the reported result is the mean of the four values obtained.

Stability of binding (binding half-life) was determined as described previously (30). Briefly, biotinylated HLA-1 heavy chain, 125I-labeled beta-2-microglobulin (B2m), and peptide were allowed to fold into peptide–HLA-I complexes in streptavidin-coated scintillation microplates (Flashplate Plus; Perkin-Elmer, Boston, MA) for 24 h at 18°C. Excess unlabelled B2m was added, and dissociation was initiated by placing the microplate in a scintillation reader (TopCount NXT; Perkin-Elmer, Boston, MA) operating at 37°C. The scintillation signal was monitored by continuous reading of the microplate for 24 h. Half-lives were calculated from dissociation curves using the exponential decay equation in Prism v.5.0a (GraphPad, San Diego, CA). Assays were performed in duplicate; the mean value from two experiments is reported.

Statistical analysis. Statistical analysis was undertaken using GraphPad Prism v.5.0a (GraphPad, San Diego, CA). To define the sites of new putative HLA-B*3501 epitopes, relationships between HIV-1 sequence polymorphisms and HLA class I expression and between ELISpot responses and HLA class I expression were determined using Fisher’s exact test (corrected for viral lineage in the case of sequence analysis) and corrected for multiple comparisons using a q value (false-detection rate), as previously described (11, 54).

RESULTS
Consistent differential HLA-B*3501-association with viral set point in B- and C-clade infection. We first sought to test the consistency of our initial observation that, in contrast to its impact in B-clade infection (7, 21, 59), HLA-B*3501 is not associated with high viral set point in C-clade infection (38, 44). In B-clade-infected cohorts in Mexico and in Japan, HLA-B*3501 is associated with a high viral set point (P = 0.06 and P = 0.0005, respectively) (Fig. 1 and 2). In contrast, in a C-clade-infected Botswanan cohort, HLA-B*3501 is somewhat protective, although this did not reach statistical significance (Fig. 1 and 2).

HLA-B*3501 is also associated with higher absolute CD4+ T-cell counts in subjects with C-clade infection (Durban, P = 0.06;
In contrast, HLA-B*3501 is associated with lower absolute CD4 counts in subjects with B-clade infection (Mexico, \( P = 0.01 \); Japan, \( P = 0.3 \); \( P = 0.01 \) when data were pooled; median absolute CD4 counts, 249 versus 370 cells/mm\(^3\) in HLA-B*3501-positive versus HLA-B*3501-negative subjects) (data not shown).

Thus, in two large C-clade-infected cohorts, HLA-B*3501 is associated with lower viral loads and higher CD4\(^+\) T-cell counts in chronic HIV infection, whereas in B-clade-infected cohorts, such as those studied in Japan and in Mexico, HLA-B*3501 tends to be associated with a higher viral set point and lower absolute CD4 count.

HLA-B*3501-restricted CD8\(^+\) T-cell responses in B- and C-clade infection. In order to investigate whether the observed difference in HLA-B*3501-associated HIV disease outcome in B- and C-clade-infected cohorts is related to clade-specific differences in the CD8\(^+\) T-cell activity, we measured responses in HLA-B*3501-positive subjects infected with B- or C-clade virus to a comprehensive panel of HLA-B*3501-restricted epitopes (Table 1). This panel comprised epitopes previously defined from studies of B-clade-infected subjects with HLA-B*3501 and published in the Los Alamos Immunology database “A list” (www.hiv.lanl.gov) (48), together with 5 additional novel epitopes that were identified by analysis of CD8\(^+\) T-cell responses in a cohort of 1,010 study subjects (40) on a panel of 410 overlapping 18-mer peptides (OLPs) spanning the C-clade proteome (see Table S2 in the supplemental material). An illustration of the approach that was used to identify these HLA-B*3501-restricted epitopes is shown for HPVHAGPIA (Gag positions 216 to 224) (HA9) (see Fig. S1 in the supplemental material), which was recently also described by another group (74) as a p24 Gag epitope restricted by HLA-B*3501.

For all the epitopes identified that were not listed in the Los Alamos Immunology database (www.hiv.lanl.gov) (48), in each case strong binding avidities to HLA-B*3501 (with the \( K_d \) [dissociation constant] ranging between 1 and 55 nM) were demonstrated (data not shown), and a CD8\(^+\) T-cell response to each was detected in \( \geq 2 \) study subjects tested (see below). In the process of validating the novel and previously published HLA-B*3501-restricted epitopes using HLA-class I tetramers (40), we noted one epitope that had been previously identified via an epitope prediction approach as PPIPVGDIY (PY9) (Gag positions 254 to 262) (64). We demonstrated that the true optimal epitope is the 10-mer NPPIPVGDIY (NY10) (Gag positions 253 to 262), which is consistently recognized at \( \leq 1/1,000 \) of the concentration of PY9 (Fig. 3A). HLA-B*3501 tetramer staining of antigen-specific cells was readily observed using the 10-mer NY10 (Fig. 3B) but was never achieved using the 9-mer PY9. This process of distinguishing the correct epitope, NY10, from the incorrect epitope, PY9, was of crucial significance in understanding the differential impact of HLA-B*3501 in B- and C-clade HIV infection (see below).

FIG 2 Median viral load in subjects with and without HLA-B*3501 in B- and C-clade-infected study cohorts. Boxes show median and 25th and 75th percentiles; whiskers show 10 to 90% confidence intervals. (A) Kumamoto, Japan (B clade); (B) Mexico City, Mexico (B clade); (C) Gaborone, Botswana (C clade); (D) Durban, South Africa (C clade). \( P \) values are by the Mann-Whitney test.

FIG 3 Optimization of the NY10 Gag epitope (NPPIPVGDIY). (A) IFN-\( \gamma \) ELSpot responses to titrated amounts of the 9-mer PPIPVGDIY versus the 10-mer NPPIPVGDIY peptides made by an HLA-B*3501-positive adult subject with chronic B-clade HIV-1 infection (Thames Valley subject H033, HLA-A*3601, -A*7401, -B*3501, -B*5301, -Cw*0401, -Cw*0401). (B) Unequivocal definition of the correct HLA-B*3501-restricted optimal epitope NY10 using an HLA-B*3501-NY10 tetramer to stain the NY10 responder PBMCs from the same subject (H033) as used for panel A. Results from one representative of two independent experiments are shown.
Gag NY10 is the single epitope differentially targeted by HLA-B*3501 subjects with B- and C-clade infection. Reactivity to the panel of HLA-B*3501-restricted epitopes was determined in HLA-B*3501-positive subjects with B-clade infection (n = 74) and in subjects with C-clade infection (n = 42) using ELISPOT assays (Fig. 4). Overall, p24 Gag-specific epitopes were targeted significantly more frequently by the C-clade-infected B*3501-positive study subjects (55% versus 31%; P = 0.02 by Fisher’s exact test), whereas Env-specific epitopes were targeted more frequently by B-clade-infected B*3501-positive study subjects (10% versus 26%; P = 0.05 by Fisher’s exact test) (Fig. 4A). At the individual epitope level, the single statistically significant clade-specific difference was in the response to the Gag NY10 epitope (Gag positions 253 to 262; P = 2 × 10⁻⁵). A response to this epitope was seen in only 5% of B-clade-infected subjects, versus 38% of C-clade-infected subjects. Although the Rev epitope KY10 (Rev positions 14 to 23) was also predominantly targeted in C-clade infection, this difference in recognition in B- and C-clade-infected HLA-B*3501-positive subjects did not reach statistical significance after correction for multiple comparisons.

Both p24 Gag responses and one Nef response are consistently associated with lower viral load in subjects with HLA-B*3501. Having determined which HLA-B*3501-restricted epitopes are targeted in B- and C-clade-infected subjects with HLA-B*3501, we next investigated which of these responses appear to be most effective in bringing about a low viral set point. Two responses were consistent in being associated with a lower set point in the responders compared to the nonresponders in both B- and C-clade cohorts, Gag HA9 and Nef VY8 (Fig. 4D). These two epitopes are targeted equally well in B- and C-clade infection, and therefore these responses do not help to explain why HLA-B*3501 is associated with lower viral set points in C-clade infection. In the case of Gag NY10, however, in B-clade infection there was only 1 responder among 31 B-clade subjects for whom viral loads were available. However, in the C-clade-infected cohort, a response toward Gag NY10 was also associated with a lowered viremia (P = 0.03 by Mann-Whitney test) (Fig. 4D), Thus, the only HLA-B*3501-restricted response associated with a lower viral set point for which there was a significant difference in epitope targeting comparing the B- and C-clade cohorts was the Gag NY10 response.
Lack of immunogenicity of NY10-260E indicated by strong selection of the Gag-D260E polymorphism in B- and C-clade infection and lack of NY10-260E-specific CD8$^+$ T-cell responses. We next addressed the question of why the B-clade version of Gag NY10, which differs from the C-clade version only at position 8 in the epitope, in the replacement of Asp by Glu (Gag-D260E), appears to be nonimmunogenic, whereas the C-clade version is highly immunogenic. Although 38% of HLA-B*3501-positive subjects with chronic C-clade infection show detectable responses to NY10-260D, analysis of gag sequences in the cohort indicates that exactly twice that figure, 76%, of HLA-B*3501-positive subjects carry the Gag-D260E mutation, compared to 28% of the HLA-B*3501-negative study subjects (Fig. 5A) ($P = 9 \times 10^{-11}$). We confirmed that, in every case tested, the NY10-D260E variant is substantially less well recognized than the C-clade wild-type NY10-260D and is therefore an escape mutant. Strikingly, NY10-260E is also selected in HLA-B*3501-positive subjects with B-clade infection (Fig. 5A), in spite of the fact that close to 90% of B-clade sequences carry Gag-260E (37). These data suggest that NY10-260E is nonimmunogenic and that only the small fraction of B-clade-infected HLA-B*3501-positive subjects presented with virus expressing the Gag-260D vari-
ant can make an NY10-260D-specific response, from which the viral escape mutant D260E is selected.

To test this hypothesis, i.e., that NY10 responses in B-clade-infected subjects are either cross-reactive between the two NY10-260D and NY10-260E variants or specific to the NY10-260D form but are never specific for the NY10-260E variant, we generated HLA-B*3501-NY10-260D and HLA-B*3501-NY10-260E tetramers with which to stain NY10-specific CD8⁺ T cells. Staining of PBMCs and antigen-specific cell lines with these two HLA-B*3501-NY10 tetramers was consistent with the hypothesis (Fig. 5C and D).

In vitro expansion of NY10-specific CD8⁺ T cells in the rare B-clade-infected persons showing a response to this epitope showed, irrespective of which variant had been used to stimulate PBMCs, preferential recognition of the NY10-260D (C-clade version) of the epitope (Fig. 5B). Where there is apparent cross-reaction of NY10-260D-specific CD8⁺ T cells to the NY10-260E variant (Fig. 5C), following in vitro expansion of these cells using either the NY10-260D or the NY10-260E peptide, preferential recognition of the NY10-260D epitope consistently emerges. Dual NY10-260D and NY10-260E tetramer staining confirms that only cross-reactive or NY10-260D-specific CD8⁺ T cells exist, with no detection of NY10-260E-specific CD8⁺ T cells (Fig. 5D).

It is noteworthy that had PY9 as opposed to NY10 been the optimal epitope in this case, it would not have been able to explain lack of immunogenicity of the B-clade variant in this way. Both B- and C-clade versions of PY9 had low peptide-binding affinities to HLA-B*3501, in particular the Gag-260D (C-clade) version (Kd = 76 and 407 nM for PY9-260E and PY9-260D, respectively), and very low peptide-B*3501 binding stabilities, again lower for the

FIG 6 Binding of NY10 Gag and PY9-Gag to the HLA-B*3501 molecule. Strength of binding affinity (Kd, nM) of HLA-B*3501 was determined using the luminescent oxygen channeling immunoassay, as previously described (29) (top panels), and stability half-life (t1/2) of binding (h) was determined using scintillation proximity assay, as previously described (30) (bottom panels), for NY10 (A) and PY9 (B). Results from one representative of four independent experiments are shown.
Gag-260D C-clade version of PY9 (half-life of 0.62 h and 0.34 h for PY9-260E and PY9-260D, respectively).

Together these data suggest that the observed differential HLA-B*3501 association with HIV disease progression in B- and C-clade infection may hinge on a single Gag epitope, NY10, and that the lack of immunogenicity of this epitope in B-clade infection rests on the presence of Glu at Gag-260 in the consensus B-clade sequence, in contrast to Asp at Gag-260 in the consensus C-clade sequence.

DISCUSSION

The data presented here demonstrate that subjects with HLA-B*3501 control HIV-1 more effectively in C-clade than in B-clade infection. This difference was associated with greater targeting of p24 Gag epitopes and less frequent targeting of Env epitopes overall. However, the single epitope significantly targeted differentially was the Gag NY10 epitope, targeted by 38% of HLA-B*3501-positive subjects with chronic C-clade infection and only 5% of HLA-B*3501-positive subjects with chronic B-clade infection. The reason for this difference is the replacement of Asp by Glu at Gag-260, position 8 within the NY10 epitope: in C-clade infection, ~75% of sequences carry Asp at Gag-260, whereas in B-clade infection, ~90% of sequences carry Glu at Gag-260. NY10-260E is nonimmunogenic and insufficiently recognized from infected cells (<25% C8α+ T-cell activation) because this variant fails to bind sufficiently stably to HLA-B*3501. In contrast, the NY10-260D version is recognized more efficiently (>60% C8α+ T-cell activation) and binds relatively stably to HLA-B*3501 (off-rate half-life of 1.6 h, compared to 0.5 h for NY10-260E). The binding affinity of HLA-B*3501 for NY10-260D was also substantially higher than that for NY10-260E (Kd of 10 nM versus 113 nM, respectively), consistent with the difference in antigen processing of this epitope. These findings provide a plausible explanation for why NY10-260E is an escape variant in B- and C-clade infection and why only the NY10-260D variant is immunogenic.

Several hypotheses have previously been proposed to explain the rapid disease progression of HLA-B*3501-positive subjects infected with B-clade HIV, including a paucity of HLA-B*3501-restricted Gag-specific C8α+ T-cell epitopes (39), failure to optimize antiviral NK activity (51, 52) and narrowness of the TcR repertoire available to counter epitope sequence variability (41). The data presented here support the "Gag hypothesis" (39), in that even the addition of a single extra Gag response appears to significantly alter the impact of HLA-B*3501 in HIV infection. This is consistent with previous findings that increasing Gag-specific C8α+ T-cell breadth is correlated with increasing viral suppression (39) and that the p24 Gag protein is infrequently targeted by HLA-B*3501-restricted C8α+ T-cell responses in B-clade infection (67). These data also support previous studies that have suggested that even one effective CTL response can mediate long-term immune control of immunodeficiency virus infection, such as the KK10 (Gag positions 263 to 272) response in HIV-infected subjects with HLA-B*27 (27) or the SW9 (Gag positions 241 to 249) response in simian immunodeficiency virus (SIV)-infected Burmese macaques expressing the MHC 90-120-Ia haplotype (35, 70).

These data show that inadequate HLA-B*3501 binding of the peptide, as opposed to TcR paucity, as has been proposed as a mechanism for HLA-B*3501-associated rapid progression (41), may provide the explanation for the lack of a response to NY10-260E in B-clade infection. The two hypotheses are not mutually exclusive, and it remains possible that HLA-B*3501 is associated with some degree of protection against C-clade progression in spite of TcR paucity. However, the distinction is of direct relevance to vaccine design, since we show here that HLA-associated disease outcome is dependent on the epitopes being targeted, irrespective of any deficiencies attributed to the respective HLA molecule. Furthermore, it is striking that HLA-B*0702 and HLA-B*3501, the two alleles proposed to predispose to rapid HIV progression as a result of TcR paucity (41), both have a more successful impact on the viral set point in C-clade infection (Fig. 1), as do many other alleles within the HLA-B7 supertype whose peptide-binding motifs are very similar, namely, HLA-B*8101, B*4201, B*0705, and B*3910 (45).

These studies also draw attention to caveats associated with epitope prediction approaches using peptide-binding motifs or even those using the most sophisticated software that takes account of the possible contribution to MHC binding of every amino acid of every peptide known to bind to a particular MHC class I molecule. Although PPIPVGDIY (PY9) has appeared in the "A" list of HIV-specific C8α+ T-cell epitopes since 1995 (48) and epitope prediction programs predict that PY9 would bind better than NY10 to HLA-B*3501 (17), nonetheless PY9 is not the epitope. It is significant that 0/377 peptides eluted from HLA class I molecules and sequenced have Pro at P1 (44). Bearing in mind the specificity of ERAP-1, which cleaves neither at X-P nor at P-X bonds (32), it seems that epitopes carrying Pro at P1, if they exist at all, are rare. The importance of defining the precise optimal epitope correctly is underlined by this study, in the demonstration that the 10-mer NY10 could only be immunogenic with Asp at P8 (Fig. 5). In contrast, although PY9-260E bound with stronger avidity than PY9-260D to HLA-B*3501, neither version of the 9-mer PY9 appeared to bind HLA-B*3501 with adequate stability to be immunogenic. It may be helpful in the future to confirm the identification of novel epitopes using peptide-MHC I tetramers, as now can be done readily (40, 43).

The critical contribution to MHC binding of the residue at P8 in an HLA-B*3501-restricted epitope was unexpected, given the peptide-binding motif of HLA-B*3501, which describes proline at P2 and Tyr at PC as the primary anchor residues, with various residues less strongly preferred at P2, P3, P4, and PC (18, 33). Explanation of this awaits the solution of the crystal structure of the HLA-B*3501–NY10-260D complex. However, an HLA-B*3501–EBV epitope structure has been solved (71), and modeling the HLA-B*3501-NY10 structure based on these data suggests that Asp at P8 in the NY10 epitope indeed points into the groove (Fig. 7). The model suggests that replacement of Asp by Glu at P8 would lead to steric hindrance between the longer side chain of Glu and the side chain of Ala-150 in the MHC α2 helix. The resulting altered conformation of the peptide would explain the observed reduction in stability of NY10-260E (Fig. 6B). This is consistent with the reduced but detectable processing of the NY10-260E peptide (data not shown) and is directly explained by the reduced affinity to the MHC molecule and thereby suggests that the limiting step in processing of the NY10-260E peptide occurs when the fully trimmed epitope is loaded onto the HLA-B*3501 molecule by the peptide-loading complex. This reduction in processing of the NY10-260E epitope may be critical to distinguish immunogenicity, especially at low infection levels of pri-
infection suggests a highly functional Gag NY10-specific CD8

A complex (71). (A) HLA-B*3501 (shown in gray cartoon)-NPPIPVGDIY

(shown in blue sticks), looking down at the MHC-binding groove. Position

P8D in the peptide is circled. (B) HLA-B*3501 (shown in gray cartoon)-NPP

IPVGEIY (shown in red sticks), looking down at the MHC-binding groove.
P8E is circled. (C) Modeled interaction with NY10 residue P8D (blue stick)

and MHC residues A150 and V152 (green sticks). (D) Modeled interaction

with NY10 residue P8E (red stick) and MHC residues A150 and V152 (green

sticks). The longer side chain of E in the escape mutant NY10 compared to D

in the wild-type NY10 could generate steric hindrance with MHC residue

A150. This could destabilize, and change the conformation of, the NY10 escape

mutant peptide.

FIG 7 Modeled B3501-NY10 structure using the B3501-EPLPQGQLTAY

complex (71). (A) HLA-B*3501 (shown in gray cartoon)-NPPIPVGDIY

(shown in blue sticks), looking down at the MHC-binding groove. Position

P8D in the peptide is circled. (B) HLA-B*3501 (shown in gray cartoon)-NPP

IPVGEIY (shown in red sticks), looking down at the MHC-binding groove.
P8E is circled. (C) Modeled interaction with NY10 residue P8D (blue stick)

and MHC residues A150 and V152 (green sticks). (D) Modeled interaction

with NY10 residue P8E (red stick) and MHC residues A150 and V152 (green

sticks). The longer side chain of E in the escape mutant NY10 compared to D

in the wild-type NY10 could generate steric hindrance with MHC residue

A150. This could destabilize, and change the conformation of, the NY10 escape

mutant peptide.

mary CD4+ T cells in vivo, in contrast to the higher multiplicity of

infection used in vitro in this assay.

The high frequency (~75%) of the D260E selection in C-clade

infection suggests a highly functional Gag NY10-specific CD8+

T-cell response in vivo. However, when we undertook sequencing

of the Gag NY10 region from position 253 to 262 of 17 HLA-

B*3501-positive recipients with known Gag NY10 sequences of

their linked donor viruses, we did not find any selection of D260E

escape mutation at very early viral load set points (CD4 count

nadir) in 10 HLA-B*3501 individuals infected with the 260D virus

(0/10) (data not shown). This suggests that the D260E selection

occurs after the CD4 nadir during chronic infection and that the

Gag NY10 response therefore may operate during chronic infec-

tion rather than during acute infection. This is consistent with a

previous study showing that the HLA-B*3501-D260E mutation is

selected outside acute infection (54). Moreover, we did not ob-

serve any change in the viral load set point for individuals carrying

260D versus 260E within linked recipients (37,720 versus 39,740

RNA copies/ml plasma; \( P = 0.6 \)) early after infection or during

chronic infection (17,550 versus 26,600 RNA copies/ml plasma;

\( P = 0.58 \)) (data not shown). However, the small numbers in com-

bination with the potential compensatory mutations identified,

which may restore viral fitness, may mask differences in viral load

set point.

Although the residue at Gag-260 appears to play an important

part in immunogenicity of the HLA-B*3501-NY10 epitope, it is

also important to note that, as with many amino acid substitutions

in p24 Gag, this single-amino-acid substitution at Gag-260 is of-

f ten observed in association with a number of other variations

elsewhere in p24 Gag. In a covariation analysis (14), we identified

9 statistically significant associations (\( q < 0.05 \)) between Gag-

260D and variation at other positions (see Table S3 in the supple-

mental material), which may indicate that the D260E escape in

C-clade virus may require compensatory mutations to minimize

the impact on viral replicative capacity.

One further observation with respect to epitope definition

highlighted by this study is the value of using a panel of overlap-

ping peptides to comprehensively map responses made by HIV-

infected subjects, as opposed to using epitope prediction. The

other p24 Gag epitope defined here in HLA-B*3501-positive sub-

jects, HPVHAGPIA (HA9), may have gone unnoticed previously

because HLA-B*3501 typically shows a binding preference for Tyr

or a larger hydrophobic residue than Ala at the C terminus. Be-

tween 40 and 60% of subjects studied here with HLA-B*3501

made a response to HA9, and, like for the NY10 Gag response,

responders had significantly lower viral loads than non-

responders. Thus, a critical epitope within p24 Gag would have re-

mained undetected had we used an approach based on predicted

epitopes only.

Of note, we unexpectedly showed that a response to one of the

Nef epitopes, NY8, was also associated with a lowered viremia in

both B- and C-clade infection. CD8+ T-cell responses to Nef have

not typically been associated with disease control (39), but the

data presented here suggest that specific responses within Nef may

also mediate viremic suppression. In a previous study, it was ob-

served that a substantial number of the Nef escape mutations re-

vert following transmission to an HLA-mismatched host (54),

suggesting a cost to viral fitness; the escape polymorphism itself

may therefore contribute to disease control via an effect on viral

reproductive capacity. This finding is also consistent with data de-

scribing effective control of SIV in Mamu-B*08- and Mamu-

B*17-positive rhesus macaques, which tend to target dominant

epitopes that exists between B- and C-clade-infected subjects. Thus, although a broad Gag-specific CD8+

T-cell response may be more likely to be effective against HIV, it remains possible that

CD8+ T-cell responses targeting epitopes in non-Gag proteins may also be effective in containment of immunodeficiency virus

infection.

It is important also to consider the limitations of this study. In

particular, attention should be drawn to the fact that optimal

HLA-B*3501-restricted epitopes 8 to 11 amino acids in length

were tested for recognition in the B-clade-infected Japanese study

subjects, whereas the C-clade-infected subjects were tested for rec-

ognition of the 18-mer overlapping peptides containing those op-

timal epitopes. Although responses to the 18-mer and to the op-

timal epitope have been strongly correlated (16) \( (r = 0.85; P <

0.0001 \) [H. N. Kloverpris et al., unpublished data]), the magnitude of

response to the 18-mer tends to be somewhat lower than that to the

optimal epitope, particularly if the location of the optimal

epitope is in the central part of the 18-mer peptide (16, 55). How-

ever, this likely underestimation of the responses in the C-clade-

infected study subjects, where response frequencies were deter-

mined using the 18-mer overlapping peptides, would likely have

reduced the estimates of the frequency of Gag NY10 responses and

of Gag HA9 responses, detected in 38% and 52% of subjects, re-

spectively. Therefore, the difference in targeting of p24 Gag

epitopes that exists between B- and C-clade-infected subjects is

likely, if anything, to be even greater than shown in Fig. 4.

In summary, the impact of HLA alleles such as HLA-B*3501 on

HIV disease outcome differs according to clade of infection. These

data suggest that the critical difference in C-clade infection is the ability of HLA-B*3501-positive subjects to make two p24 Gag-

specific responses restricted by this allele, NY10 and HA9, com-

pared to only one (HA9) in B-clade-infected subjects. This result

provides the clearest data yet that HLA-associated disease out-

come is dependent on the epitopes being targeted, irrespective of
the nature of the restricting HLA molecule (55), and this provides hope that a vaccine which can induce effective CD8+ T-cell responses can successfully bring about immune control even in people who carry HLA alleles traditionally regarded as associated with rapid disease progression.

ACKNOWLEDGMENTS

We declare that no competing interests exist.

This work was supported by the Welcome Trust (P.J.R.G.), National Institutes of Health grant R01 AI46995, and the Global COE program “Global Education and Research Center Aiming at the Control of AIDS,” launched as a project commissioned by the Ministry of Education, Science, Sports, and Culture, Japan.

REFERENCES

1. Reference deleted.
8. Reference deleted.
9. Reference deleted.
26. Reference deleted.