DIFFERENTIAL ESCAPE PATTERNS WITHIN THE DOMINANT HLA-B*57:03-
RESTRICTED HIV GAG EPITOPE REFLECT DISTINCT CLADE-SPECIFIC
FUNCTIONAL CONSTRAINTS

Payne RP1, Branch S2, Kløverpris H1,3, Matthews PC1, Koofhethile CK1, Strong T1, Adland E1,
Leitman E1, Frater J4,5, Ndung’u T6, Hunter E7,8, Haubrich R9, Mothe B10, Edwards A11, Riddell
L12, Chen F13, Harrigan PR14, Brumme ZL14,15, Mallal S16,17, John M16,17, Jooste JP18, Shapiro R19,

1Department of Paediatrics, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road,
Oxford OX1 3SY, UK; 2Ladymeade Reference Unit, Queen Elizabeth Hospital, Bridgetown, Barbados; 3 K-RITH,
University of Kwa-Zulu Natal, Duran, South Africa; 4 Nuffield Department of Clinical Medicine, University of Oxford,
Oxford, UK; 5 Oxford NIHR Biomedical Research Centre, Oxford, UK; 6 HIV Pathogenesis Program, Doris Duke
Medical Research Institute, University of KwaZulu-Natal, Durban, 4013, South Africa; 7 Emory Vaccine Center,
Yerkes National Primate Research Centre, Emory University, Atlanta, Georgia, USA; 8 Department of Global Health,
Rollins School of Public Health, Emory University, Atlanta, Georgia, USA; 9 Antiviral Research Center, University of
California San Diego, San Diego, California, USA; 10 IrsiCaixa AIDS Research Institute - HIVCAT, Hospital Germans
Trias i Pujol, Badalona, Spain; 11 Oxford Department of Genitourinary Medicine, The Churchill Hospital, Oxford, UK;
12 Department of Genitourinary Medicine, Northamptonshire Healthcare National Health Service Trust, Northampton
General Hospital, Cliftonville, Northampton, UK; 13 Department of Sexual Health, Royal Berkshire Hospital, Reading,
UK; 14 British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada; 15 Faculty of Health Sciences,
Simon Fraser University, Burnaby, British Columbia, Canada; 16 Centre for Clinical Immunology and Biomedical
Statistics, Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, Western Australia,
Australia; 17 Department of Clinical Immunology, Royal Perth Hospital, Perth, Western Australia, Australia;
18 Paediatric Department, Kimberley Hospital, Northern Cape, South Africa; 19 Department of Immunology and
Infectious Diseases, Harvard School of Public Health, 110 Francis Street, Boston, MA 02215, USA; 20 Dept of
Medicine, University of California, San Francisco, CA, USA; 21 Ragon Institute of MGH, MIT, and Harvard, AIDS
Research Center, Rm. 5212D, 149 13th Street, Charlestown, MA 02129, USA; 22 Institució Catalana de Recerca i
Estudis Avançats (ICREA), Barcelona, Spain; 23 Microsoft Research, eScience Group, Los Angeles CA, USA

1 Corresponding author - rebecca.payne@hotmail.co.uk
HLA-B*57:01 and HLA-B*57:03, the most prevalent HLA-B*57 subtypes in Caucasian and African populations, respectively, are the HLA alleles most protective against HIV disease progression. Understanding the mechanisms underlying this immune control is of critical importance and yet remains unclear. Unexplained differences are observed in the impact of the dominant CTL response restricted by HLA-B*57:01 and HLA-B*57:03 in chronic infection towards the Gag epitope KAFSPEVIPMF (‘KF11’, Gag162-172). We previously showed that the HLA-B*57:03-KF11 response is associated with a >1 log lower viral setpoint in C-clade infection and that this response selects escape mutants within the epitope. We first examined the relationship of KF11 responses in B-clade infected subjects with HLA-B*57:01 to immune control and observed that a detectable KF11 response was associated with a >1 log higher viral load (p=0.02). No evidence of HLA-B*57:01-KF11 associated selection pressure was identified in previous comprehensive analyses of >1800 B-clade infected subjects infected. We then studied a B-clade infected cohort in Barbados where HLA-B*57:03 is highly prevalent. In contrast to B-clade infected subjects expressing HLA-B*57:01, we observed strong selection pressure driven by the HLA-B*57:03-KF11 response for the escape mutation S173T. This mutation reduces recognition of virus-infected cells by HLA-B*57:03-KF11 CTL, and is associated with a >1 log increase in viral load in HLA-B*57:03-positive subjects (p=0.009). We demonstrate functional constraints imposed by HIV clade relating to the residue at Gag-173 that explain the differential clade-specific escape patterns in HLA-B*57:03 subjects. Further studies are needed to evaluate the role of the KF11 response in HLA-B*57:01-associated HIV disease protection.
IMPORTANCE SECTION

HLA-B*57 is the HLA class I molecule that affords the greatest protection against disease progression in HIV infection. Understanding the key mechanism(s) underlying immune suppression of HIV is of importance in guiding therapeutic and vaccine-related approaches to improve the levels of HIV control occurring in nature. Numerous mechanisms have been proposed to explain the HLA associations with differential HIV disease outcome but no consensus exists. These studies focus on two subtypes of HLA-B*57, prevalent in Caucasian and African populations, HLA-B*57:01 and HLA-B*57:03, respectively. These alleles appear equally protective against HIV disease progression. The CTL epitopes presented are in many cases identical, and the dominant response in chronic infection in each case is to the Gag epitope KF11. However, there the similarity ends. This paper seeks to better understand the reasons for these differences and what this teaches us about which immune responses are contributing to immune control of HIV infection.
INTRODUCTION

HLA polymorphism has a substantial impact on HIV disease outcome (1-5) and yet the principal mechanisms underlying these effects remain unresolved (5). The most protective HLA class I molecule is HLA-B*57. HLA-B*57:01 is the most prevalent subtype in Caucasian populations, apparently conferring a similar level of protection against HIV disease progression as HLA-B*57:03, the most prevalent subtype in African populations (6). One proposal is that the HLA-B*57-mediated protection is at least in part due to the breadth of the Gag-specific CD8+ T-cell response, and that HLA-associated immune control of HIV is related to the ability of the CTL response to drive selection pressure on the virus, such that escape can only be achieved at significant cost to viral replicative capacity (5, 7-10). However, although HLA-B*57:01 and HLA-B*57:03 appear to present the identical Gag epitopes, previous studies suggest that significant differences exist in the impact of these responses on immune control.

The dominant HIV-specific CD8+ T-cell response in each case is directed towards the Gag epitope KAFSPEVIPMF (‘KF11’, Gag 162-172). Published studies in HLA-B*57:01-positive subjects have been almost exclusively conducted in B clade infection and these have suggested that a response to KF11 is not associated with immune control (2, 11) and that the magnitude of HLA-B*57:01-KF11 responses may even be higher in progressors. Evaluation of full-length viral sequences in 1888 B clade infected subjects failed to identify any sequence polymorphisms within the KF11 epitope or flanking it that were directly associated with HLA-B*57:01 (12) and that would have suggested strong selection pressure imposed on the virus by this response. By comparison, 20 HIV amino acid polymorphisms were identified elsewhere in the HIV proteome that were directly associated with HLA-B*57:01. In contrast, in studies undertaken in C clade infected African subjects expressing
HLA-B*57:03, a KF11 response is associated with a >1 log lower viral load (13), and there is strong evidence of selection pressure within this epitope, with approximately 70% of subjects carrying mutants at Ala-163 and/or Ser-165, positions 2 and 4 in the epitope (P2 and P4) (9, 10, 14).

The escape mutants that are selected within KF11 in C clade-infected HLA-B*57:03-positive subjects are typically A163G and S165N. The A163G mutation is selected first, reducing CTL recognition, but also significantly lowering viral replicative capacity (14). The S165N mutant is then selected, substantially restoring viral replicative capacity at the same time as entirely abrogating recognition: an ideal result for the virus (14) that has been associated with higher viral loads (15). Thus the impact of the HLA-B*57:03-KF11 response on the virus is consistent with the mechanism of HLA-mediated immune control described above, with CTL activity forcing the selection of viral escape mutants that reduce viral replicative capacity, whereas that of the HLA-B*57:01-KF11 response is not consistent with this mechanistic model.

Initial studies proposed that the HLA-B*57:01-KF11 TCRs are highly conserved, with a dominant or exclusively expressed V5/V19 TCR in 60-100% of subjects (15, 16). These HLA-B*57:01-KF11 TCRs were more likely than the HLA-B*57:03-KF11 TCRs to recognise epitopes containing A163G and S165N; hence these mutations would not be selected in individuals expressing HLA-B*57:01 (15, 17). However subsequent studies of HLA-B*57:01-KF11 TCR usage have been contradictory, with the V5/V19 TCRs identified as the dominant receptor in 0/6 and 2/10 HLA-B*57:01-positive subjects, respectively (11, 18). Thus it appears unlikely that HLA B*57 subtype-specific effects, namely a public TCR clonotype with high functional avidity for HLA-B*57:01-
KF11 (19) fully explain the observed differences in selection of KF11 escape mutants in HLA-B*57:01 and HLA-B*57:03.

An alternative hypothesis is that escape mutations within KF11 are tolerated in the context of the C-clade, but not in the B-clade Gag sequence, and that this may contribute to the high frequency of escape mutations within the KF11 epitope in HLA-B*57:03-positive individuals infected with C-clade virus. To address this possibility, we here compared a cohort of B-clade infected subjects in Barbados, in which the HLA-B*57:03-subtype predominates, with a cohort of C-clade infected subjects in Botswana, where the HLA-B*57:03-subtype also predominates. We identified clade-specific sequence differences that influence the dynamics of viral escape within the HLA-B*57:03-restricted KF11 epitope. These differences were confirmed in a large multi-cohort dataset featuring 3298 subjects (including the Barbados and Botswana cohorts), including 1732 clade C infected Africans and 1566 clade B infected North Americans.
METHODS

Study cohorts and subjects
HIV-1 B clade infected subjects expressing HLA-B*57:01 were studied from ART-naïve cohorts in Oxford, UK (the Thames Valley Cohort, as previously described (20) and in Barcelona, Spain (21). Additional study cohorts for evaluation of HLA-B*57:03 in the context of B clade and C clade infection in Barbados and Botswana, respectively, were: (i) Bridgetown, Barbados (B-clade, n=246 Median age: 38 years IQR 31-47, female:male ratio 60:40, samples collected between 2008-2010), where study subjects were attendees at the Ladymeade Reference Clinic; and (ii) Gaborone, Botswana (C-clade, n=514, Median age; 27 years IQR 23-32, female:male 100:0, samples collected between 2007-2008), where study subjects were antenatal women from the Mma Bana Study, as previously described (8, 13, 22, 23).

Ethics approval was given by the Health Research Development Committee, Botswana Ministry of Health, by the Barbados Ministry of Health, the Hospital Germans Trias i Pujol' Ethics Committee, and by the Oxford Research Ethics Committee. Subjects received voluntary testing and counselling and written informed consent was obtained from all individuals. Viral load in chronic infection was measured using the Roche Amplicor version 1.5 assay; CD4+ T cell counts were measured by flow cytometry. Viral load and absolute CD4 count measurements were obtained at study entry (baseline) for all individuals. All study subjects were ART-naïve.

Four-digit HLA typing of the Class I locus was performed from genomic DNA as previously described (24) by sequence-based typing at the ASHI® accredited HLA typing laboratory, University of Oklahoma Health Sciences Centre, USA. Exons 2 and 3 of HLA Class I were
amplified by locus-specific PCR and then sequenced. Resolution of ambiguities was undertaken according to the ASHI committee recommendations.

Additional viral sequence analyses were performed on two previously described, multi-center cohorts: (1) the International HIV Adaptation Collaborative (IHAC) consisting of 1443 clade B Gag sequences (12); and (2) 1470 African clade C Gag sequences from cohorts based in Durban (8), Bloemfontein (25) and Kimberley (20) South Africa, Zambia and the Thames Valley area of the United Kingdom (20). Where high-resolution HLA typing was unavailable, we employed a published machine learning algorithm trained on a dataset of high resolution HLA class I types from >13,000 individuals with known ethnicity to complete these data to high resolution. (26).

IFN-γ ELISPOT assays

IFN-γ enzyme-linked immunospot (Elispot) assays were performed as previously described (13, 27), using optimally defined epitopes and 18mer overlapping peptides (OLP) with input cells/well ranging from 30,000 to 100,000. The number of specific spot-forming cells (SFC) was calculated by subtracting the mean number of spots in the negative control wells from the number of spots counted in each well. The magnitude of epitope-specific responses was calculated as SFC per million cells.

Site-directed mutagenesis of NL43

The mutation S173T (Serine to Threonine at Gag HXB2 position 173) was introduced by site directed mutagenesis (Quikchange I, Stratagene, UK) into wild-type NL43 plasmid DNA, as well as NL43 containing the mutations A163G and/or S165N (14). Whole plasmid DNA p83-2 (the 5’-half
of HIV-1NL43 strain) was PCR-amplified in a mutagenesis reaction with two overlapping primers, containing the target mutation. Primers used for the mutagenesis reaction were: F 5'-
CCCAGAAGTAATACCCATGTTTACGGCATTATCAGAAGGAGC-3' and R 5´-
GCTCCTCTGATAATGCCGTAAACATGGGTATTACTTCTGGG-3 (the mutagenesis site is underlined). The presence of mutations were verified by DNA Gag sequencing in newly generated plasmid clones. The DNA fragment ranging from SapI to Apa I restriction sites was then subcloned into a new p83-2 vector to avoid potential carry over of additional mutations during the mutagenesis, and the coding region sequence was verified again as previously described (28)

Virus production and Replication kinetics

Viral stocks were produced by cotransfection of the different site-directed mutant plasmids (5’ half of HIV-1NL43 strain) with p83-10cGFP (3’ half of HIV-1NL43 strain) into MT4 cells (29). Viral stocks were harvested and viral RNA extracted (Qiagen, UK). The gag p24/p17 coding region was PCR amplified and sequenced to confirm the presence of the mutations in the viral RNA and the absence of any other potential polymorphisms. The 50% tissue culture-infectious dose (TCID₅₀) for each viral stock was determined in MT4 using the Reed and Muench method (30). For replication experiments, Jurkat, MT4 and H9 T cells were infected in triplicate with a multiplicity of infection (MOI) of 0.005 in a total volume of 3ml with wild-type or mutant HIV-1 NL43 virus and incubated at 37°C for 2 hours. Pellets were washed twice with PBS and cultured at 37°C and 5% of CO₂. After infection around 50,000 cells were harvested daily in order to measure infectivity by percentage of eGFP-positive cells by Fluorescent Activated Cell Sorting (FACS). Replication kinetics were determined by calculating the mean viral slope using the LOGEST function (Microsoft Excel) and
converted to natural logs. Variation in replication slopes was assessed using Student’s T test. All statistical calculations were performed in Prism 5.0 (Graphpad).

Amplification and sequencing of proviral DNA

Gag p17/p24 sequences (Cohort; Barbados n=125, Botswana n=322) were generated from genomic DNA extracted from peripheral blood mononuclear cells (PBMC) where available, amplified by nested PCR using previously published primers to obtain population sequences, as previously described (31). Sequencing was undertaken using the Big Dye Ready Reaction Terminator Mix (V3.1) (Applied Biosystems, UK). Sequences were analysed using Sequencher v4.8 (Gene Codes Corporation) and aligned by SeAl to HXB2 B-clade reference strain. Sequences were submitted to Genbank and accession numbers are as follows FJ497801-FJ497875, FJ497885-FJ497899, FJ497901-FJ497905, FJ497907-FJ497916, FJ497918-FJ497950.

Identification of HLA-associated viral polymorphisms from proviral DNA

HLA-associated viral polymorphisms were identified from proviral DNA using a previously described method that corrects for phylogeny, HLA linkage disequilibrium and codon-covariation (8, 32). A q-value statistic, representing the p-value analogue of the false discovery rate (FDR), was computed for each association. The FDR is the expected proportion of false positives among the associations identified at a given p-value threshold; for example, among associations $q \leq 0.2$, we expect 20% to be false positives. The phylogenetically corrected methods rely on an inferred phylogeny. We constructed two phylogenies for this study: (i) a phylogeny consisting of clade B and C sequences from Barbados and Botswana was constructed using Phym v2.4.5, under the general time reversible (GTR) model (33) (ii) a phylogeny consisting of $N=3298$ p17/p24
sequences from all cohorts described in this study. This phylogeny was too large for Phyml, so we employed a 3 stage process to infer the phylogeny. (1) A combined alignment was created, then sites with >10% missing data were removed, after which sequences with missing data in >10% of remaining sites were removed (resulting in the above noted N’s); (2) a phylogeny was inferred separately for clade B and C alignments, using Phyml v2.4.5 under the GTR model); (3) the resulting phylogenies were joined by adding single common ancestor to the two clade trees, and the branch lengths were optimized using hyphy, under the GTR model (34).

PhyloDOR ratio

Identification of HLA-associated polymorphisms and assessment of differential escape between viral clades and/or closely-related HLA alleles were performed as previously described (12, 32, 35). Briefly, a maximum-likelihood phylogenetic tree was constructed for each gene, and a model of conditional adaptation was inferred for each observed amino acid at each codon (32). In this model, the amino acid is assumed to evolve independently along the phylogeny until it reaches the observed hosts (tree tips). In each host, the HLA-mediated selection pressure is modeled using a weighted logistic regression, in which the individual’s HLA repertoire is used as predictors and the bias is determined by the transmitted sequence (35). Because the transmitted sequence is not observed, we average over the possible transmitted sequences, and all possible phylogenetic histories, as inferred from the phylogeny. Similarly, where high resolution HLA types are not available, we perform a weighted average over possible completions (12).

To test for differential escape between HLA-B*57:01 and B*57:03, or to test for clade-specific effects on selection, interaction variables were added to the phylogenetically-corrected logistic
regression model and significance was determined via a likelihood ratio test, as previously described (35).

Effect of S173T mutation on epitope recognition by KF11-specific CD8+ T cells

CD4+ T cells were enriched from PBMCs from healthy donors expressing HLA-B*57:03 using negative selection (Dynabeads) and activated for 3-6 days using IL-2 (50U/ml Roche) and PHA (3ug/ml). KF11-specific CD8+ T cells (<98% specificity) were enriched from PBMCs from HIV-infected donors using tetramers as previously described (36). B*57:03-positive CD4+ T cells were infected with NL43GFP or NL43GFP containing the S173T mutation as described above. To test for epitope recognition, epitope-specific CD8+ T cells (<98% specificity) were cocultured with the HIV-infected CD4+ T cells in the presence of CD107a antibodies (PE-Cy5), 10ug/ml Brefeldin A, Golgi stop (BD), CD49d and CD28 for 6.5 hours at 37°C in a 5% CO2 incubator. Cells were stained for surface and intracellular antibodies against CD4 (APC), CD8 (Alexa Fluor 700), MIP1B (FITC), p24 (PE), IFN-y (PE-Cy7), Live/dead marker (Pacific Blue), and then immediately acquired by FACS (BD LSRII).
RESULTS

B clade HLA-B*57:01-KF11 responders have higher viral setpoints than non-responders

Previous studies of B clade infected subjects using peptide-MHC class I tetramers have suggested that a detectable HLA-B*57:01 response is more frequently observed in progressors (including those with viral loads of >90,000) than in elite controllers/long-term non-progressors (2, 11). These studies however were not sufficiently powered to demonstrate a statistically significant result. We therefore started by comparing responses to KF11 in B clade infected, ART-naïve individuals expressing HLA-B*57:01 whose viral setpoints ranged from undetectable to 500,000 copies/ml (Fig 1). Here the association between KF11 responders and high viral setpoint reaches statistical significance (p=0.02, Mann Whitney test). These findings are consistent with the earlier studies cited of B clade infected subjects expressing HLA-B*57:01, and provide the opposite result to that obtained in HLA-B*57:03-positive individuals infected with C clade virus (13), using the identical approach of measuring IFN-γ elispot responses to KF11, where a response was associated with a >10-fold lower viral setpoint. Equivalent studies of KF11 responses in 17 HLA-B*57:03-positive subjects infected with B clade virus similarly showed substantially lower median viral loads in KF11 responders compared to non-responders (median viral load 1,629 versus 6,127 c/ml, respectively), although here this difference did not reach statistical significance (p=0.28, data not shown).

Differential escape in the B*57:03-KF11 epitope in B-clade versus C-clade infection

In order to evaluate further the potential differences between HLA-B*57:01 and HLA-B*57:03, we investigated a B clade infected, ART-naïve study cohort in Barbados where HLA-B*57:03 is highly prevalent. It has been noted in several other studies that HLA-B*57:03 is associated with immune
control of HIV in B clade and C clade infection (1-5). Consistent with these studies, HLA-B*57:03-positive subjects in Barbados exhibited significantly lower median viral loads than HLA-B*57:03-negative subjects (median 3,450 versus 13,350, p=0.015, Mann Whitney test) and significantly higher CD4⁺ counts (median 565 versus 398, p=0.003 Mann Whitney test) (Fig. 2).

To determine the nature of any selection pressure imposed in the B clade virus through the HLA-B*57:03 KF11 response, we analysed viral sequences in \textit{gag} in the Barbados cohort in order to identify associations between HLA-B*57:03 and viral polymorphisms in the region of the KF11 epitope. This revealed that HLA-B*57:03 expression was associated with the previously described escape mutations T242N, in the epitope TW10 (TSTLQEIQGW; Gag HXB2 240-249) (7, 39), and I147X, in the epitope ISW9 (ISPRTLNAW; Gag HXB2 147-155) (Table I) (40, 41). However, the intra-epitope escape mutations within KF11 (KAPEVPEIPMF; Gag HXB2 162-172), namely A163G and S165N, selected in approximately 70% of C-clade infected HLA-B*57:03-positive subjects (3, 14), were not associated with HLA-B*57:03 in this Barbadian study cohort (Table I-II).

However, in this same Barbados cohort, we identified an HLA-B*57:03-associated viral polymorphism located at Gag HXB2 position 173, which immediately flanks the C-terminus of the KF11 epitope. This mutation has not been observed in association with HLA-B*57:03 in studies of C-clade infected cohorts, which in any case have Threonine as the consensus residue at position 173 (14, 42, 43). The high frequency of selection of S173T by HLA-B*57:03-positive subjects (61% versus 24% in HLA-B*57:03-positive versus HLA-B*57:03-negative subjects) together with the lack of any selection of intra-epitope KF11 mutations led to the hypothesis that selection of S173T...
in B-clade virus may mitigate against further selection of KF11 escape mutations A163G and/or S165N (Table II) (see below).

We performed a further analysis using a phylogenetically-corrected method (12) to compare the impact of HLA-B*57:03 on the selection of Gag escape mutants in B-clade versus C-clade HIV, using data from the study cohorts in Barbados (B-clade) and in Gaborone, Botswana (C-clade). We found no statistical difference between odds of HLA-B*57:03-mediated escape in the two cohorts for T242N (p=0.82) or I147L (p=0.29). In contrast, we observed substantial clade differences for all three KF11 escape mutations: the strength of selection for A163G and S165N was significantly greater in the C clade cohort (p=0.006 and p=0.08, respectively), whereas 173T was only selected in the B clade cohort (p=0.0006). In fact, Gag-173T, the consensus in C-clade, arises at significantly lower frequency in HLA-B*57:03-positive subjects in Botswana (p=0.0062; discussed further below) (Table III). These data demonstrate clade-specific differences in the impact of HLA-B*57:03 on Gag escape mutant selection, with differential effects at Gag-163, 165 and 173, within or immediately flanking the dominant KF11 epitope.

Impact of S173T on recognition of virus-infected target cells and on viral setpoint

The location of the HLA-B*57:03-associated mutation immediately downstream of the KF11 epitope suggests that the S173T mutant reduces processing of the epitope. To test whether the HLA-B*57:03-associated S173T polymorphism reduces recognition of virus-infected target cells, CD4+ T cells from HLA-B*57:03+ healthy subjects were infected with NL43 HIV that was either wildtype, expressing Ser-173, or engineered to express the S173T viral polymorphism. Infected cells were incubated with HLA-B*57:03-KF11-specific CD8+ T-cells (>98% specific) and the level
of CD8+ T-cell activation monitored by CD107a and MIP1β expression. We observed that the
S173T mutant indeed significantly reduced recognition by the KF11-specific CD8+ T cells (Fig 3, p=0.0038, Student’s t test). In the same assay, using CD8+ T-cells specific for the HLA-B*57:03-restricted Pol-specific epitope IATESIVIW (‘IAW9’), no difference were observed in the level of stimulation by the two viruses on the HLA-B*57:03-restricted IAW9-specific CD8+ T cells (Fig 3). These data support the hypothesis that S173T specifically reduces presentation of the KF11 epitope by HLA-B*57:03. Furthermore, mismatched CD4+ T cells induced consistently low levels of stimulation confirming that activation of the KF11- and IAW9-specific CD8+ T cells was HLA-B*57:03-dependent.

We next examined the viral setpoints and CD4 counts in HLA-B*57:03-positive subjects with and without the S173T mutation. Viral loads in HLA-B*57:03-positive subjects with the B-clade wildtype, Serine at Gag-173, were more than 10-fold lower than in B*57:03-positive subjects with the S173T polymorphism (median viral load 520 versus 6,905 respectively; p=0.009 Mann Whitney test). Furthermore, 173S was associated with a substantially higher CD4 count in HLA-B*57:03-positive subjects than 173T (median CD4 count 787 versus 375 respectively; p=0.036 Mann Whitney test) (Fig. 4). However no differences in median viral load or CD4 counts were observed in B*57:03-negative subjects with Serine versus Threonine at Gag-173 (median viral load 14,450 versus 10,600 respectively; p=0.949 and median CD4 count 358 versus 374 respectively; p=0.522 Mann Whitney test). These data together support the conclusion that HLA-B*57:03-KF11 responses drive the selection of the S173T mutation in B clade infected individuals expressing HLA-B*57:03; and that this is an escape mutation in that it reduces recognition of virally infected targets. These findings are consistent with the hypothesis that this response contributes to HLA-
B*57:03-associated control of HIV, since viral loads are significantly higher in those with the S173T escape mutation.

**S173T with A163G and S165N significantly reduces viral replicative capacity**

The observations above prompt the question: if A163G and S165N are escape mutations frequently selected in HLA-B*57:03-positive subjects infected with C clade virus, why are they not selected in HLA-B*57:03-positive subjects with B clade virus infection? To assess the functional significance of the HLA-B*57:03-associated S173T mutation and the possible impact of this polymorphism on the selection of A163G and S165N, the viral polymorphisms S173T, A163G, and S165N were introduced by site-directed mutagenesis into the B-clade backbone of NL43GFP. Infectious viral stocks were generated by transfecting MT4 T-cells with the relevant DNA constructs. H9, MT4 or Jurkat T cells were then infected and the rate of viral growth was determined by monitoring the percentage of GFP-infected cells over 14 days.

Analysis of the rate of viral growth in MT4, H9 and Jurkat T-cells showed, first, that the S173T polymorphism had no significant effect on viral fitness in this *in vitro* system in any of these three cells lines used (Fig.5 and data not shown). We previously showed that the introduction of A163G or A163G/S165N into the NL43 backbone significantly reduced viral replicative capacity, with S165N acting as a partial compensatory mutant for A163G that also completely abrogated recognition of KF11 (14, 15). Here we observe that the introduction of either A163G or S165N into the NL43 backbone in combination with S173T also significantly reduces viral spread, but substantially more so than in the absence of S173T. Furthermore, the combination of S173T and both of the KF11 mutations, A163G and S165N, dramatically reduced viral spread even further,
indicating a significant cost to viral fitness of this combination of viral mutations in a B clade virus (Fig. 5). These data together suggest that the KF11 escape mutant S173T is more commonly selected in B clade infected subjects expressing HLA-B*57:03 because the cost to replicative capacity is negligible, and less than that resulting from A163G or S165N. Subsequent mutations in addition to S173T result in such a substantial reduction in replicative capacity, without any apparent amelioration from S165N to reduce these fitness costs, that these arise very rarely (Table II).

As mentioned above, Gag-173T, the consensus in C-clade, arises at a significantly lower frequency in HLA-B*57:03-positive compared to HLA-B*57:03-negative subjects in Botswana (p=0.0062). A larger analysis of the KF11 epitope region of 1899 C-clade sequences confirmed that the presence of A163G, S165N or both in combination, was significantly associated with Serine at position 173 and that this was the case both for HLA-B*57:03-positive and HLA-B*57:03-negative individuals (Fig 6). Thus, although 173T is consensus in C-clade, it appears unfavourable in the context of the KF11 intra-epitope escape mutations, supporting the findings in B-clade suggesting that this combination of mutations has a detrimental impact on viral fitness.
HLA-B*57:01 and HLA-B*57:03 are the two most protective HLA molecules against HIV disease progression in both B and C clade infection (5). These molecules differ by only two amino acids (D114N and S116Y respectively), and the peptide binding motifs are almost indistinguishable (44, 45). In chronic infection, the dominant HIV-specific CD8+ T cell response in subjects expressing HLA-B*57:01 or HLA-B*57:03 is to the Gag epitope KAFSPEVIPMF (‘KF11’, Gag HXB2 162-172) (2, 10, 13). Studies of HLA-B*57:03-positive subjects infected with C clade virus indicate that this KF11 response makes an important contribution to immune control (3, 15), and contributes to the superiority of HLA-B*57:03 as a protective HLA class I molecule over the closely-related HLA-B*57:02 and HLA-B*58:01. (10).

This study sets out to investigate the observations, first, that whilst the HLA-B*57:03-KF11 response is associated with significantly lower viral setpoints in C clade infection (13), studies of HLA-B*57:01-KF11 responses had suggested the opposite in B clade infected individuals (2, 11); and, second, whereas the HLA-B*57:03-KF11 response frequently drives escape mutations within KF11 in C clade infection (A163G and S165N) (8), these are not selected in response to HLA-B*57:01 responses in B clade infection (12).

We first confirmed a statistically significant association between response to the KF11 epitope in B clade infected subjects expressing HLA-B*57:01 and a >1 log higher viral load. This result arose from the identical assays that were used in the studies that showed a KF11 response was a associated with a >1 log lower viral load in C clade infected subjects expressing HLA-B*57:03 (13). We next showed that HLA-B*57:03 has a similar impact in B and C clade infection in terms...
of the escape mutations selected in the Gag epitopes ISW9 and TW10, but a differential impact in
the KF11 epitope. The S173T mutation selected in B clade infection reduces recognition of virus-
infected targets and is associated with a >1 log increase in viral setpoint. This is consistent with
studies in C clade infection (15) suggesting that this HLA-B*57:03 KF11 response contributes to
HLA-B*57:03-associated immune control of HIV infection. The clade specific differences in the
selection of KF11-driven escape mutants observed in Barbados (B clade infected cohort) and
Botswana (C clade infected cohort) were corroborated in analyses of larger datasets.

The position of S173T one residue downstream (P1’) of the KF11 epitope suggests that it may be a
processing mutation since this residue would be involved in the cleavage site of the proteasome
(46). Previous studies of peptide cleavage motifs have suggested that the constitutive- and immuno-
proteasome have a strong preference for Alanine at P1’ but prefer Serine over Threonine; thus the
mutation S173T could affect efficient cleavage of the C-terminal end of the KF11 epitope by the
proteasome (46).

We show that the HLA-B*57:03-associated S173T mutation effectively precludes further selection
of the KF11 intra-epitope viral mutations, A163G and S165N, since the combination of these three
mutations in a clade B backbone results in a virus with severely reduced replicative capacity.
Indeed, the close proximity of the amino acid positions 173, 163 and 165 between helix 1 and helix
2 of the Gag p24 structure, suggests that structural constraints prevent selection of A163G and
S165N if S173T has already been selected. Previous work has shown that, using a B-clade
backbone, and in the presence of S173, the mutation A163G reduces replicative capacity, but that
the further addition of S165N, as observed in vivo, partially restores replicative capacity (14, 47).
This fits with the order of selection of A163G and S165N, with S165N apparently always arising subsequent to A163G (14). However in B clade infection it appears that the selection of S173T prevents the selection of further mutants within the epitope because the fitness cost is too high. S173T appears to be the preferred choice of viral escape from the KF11-specific response since it has minimal effect on viral fitness. Our inference from the data described above and from previous studies (23) indicating that the HLA-B*57:03-KF11 response contributes to immune control of B and C clade HIV infection. The reduced recognition of S173T-virus-infected cells by KF11-specific CTL together with the lack of cost to viral replicative capacity resulting from S173T is consistent with the observation that viral loads are higher and CD4 counts lower in B clade infected subjects expressing HLA-B*57:03. In view of the substantial reduction in viral replicative capacity resulting from the A163G/S165N/S173T combination in B clade infection, it is perhaps surprising to observe the selection of A163G/S165N at high frequency in HLA-B*57:03-positive subjects infected with C clade virus in which the vast majority of sequences carry Thr at Gag-173. It may be inferred from this that the presence of consensus 173-Thr in the context of C clade Gag does not have the same prohibitive effect on viral fitness, as it does not prevent the selection of A163G and S165N. Nevertheless, in C clade infection, both in HLA-B*57:03-positive and HLA-B*57:03-negative individuals, A163G/S165N are significantly associated with Ser at Gag-173 (Fig. 6), as opposed to the consensus Thr at this position, suggesting that the combination of A163G/S165N/S173T is not favoured in either B or C clade infection.
Gag-173 has been well-studied in relation to HLA-B*27, another protective HLA molecule, because of the S173A mutation that accompanies the R264K escape variant within the dominant HLA-B*27-restricted epitope KRWIILGLNK (KK10) (48-50). It is noteworthy that in C clade infection, R264K escape in HLA-B*27-positive subjects is typically accompanied by compensatory mutations not at Gag-173, but at S165N (Brener et al, in preparation). These data underline the tight constraints on amino acid substitutions in the capsid protein, the interdependence of residues at certain key positions in the structure, including Gag-163, Gag-165, and Gag-173, and therefore the impact that clade can have on the escape options for the virus.

These data help to explain why HLA-B*57:03 is not associated with the ‘usual’ KF11 intra-epitope mutations A163G/S165N in clade B, but they do not explain why HLA-B*57:01 is not associated with either the S173T flanking mutation, nor with any KF11 intra-epitope mutations. Previous studies have suggested that TCR usage for the HLA-B*57:01-KF11 response allows recognition of the KF11 variants (15), but these initial TCR studies indicating conservation of a ‘public’ HLA-B*57:01-KF11 TCR have not been borne out by subsequent studies (11, 18). One possible explanation is that the potency of the HLA-B*57:01-KF11 response is so great that a moderate reduction in processed epitope would not affect killing sufficiently to be selected, however preliminary data suggests that HLA-B*57:03 response is, if anything, the more potent. Further studies with a large number of KF11-specific clones would be needed to establish whether clear-cut differences between the responses restricted by HLA-B*57:03 and HLA-B*57:01 exist in terms of potency, and the relevance of this to viral escape patterns. A recent study comparing the impact of individual HLA class I molecules on immune control (viral load <2000 copies/ml) versus non-control (VL>10,000 copies/ml) of B clade infection showed the identical odds ratio for protection.
via HLA-B*57:01 in a European American cohort as via HLA-B*57:03 in an African American cohort (6).

These studies therefore provide an explanation for the distinct clade-specific selection of escape mutants by the HLA-B*57:03-KF11 response but do not resolve the question of why the HLA-B*57:01-KF11 response does not select escape mutants. Insufficient studies have been undertaken in C clade infected subjects who express HLA-B*57:01 to be certain of whether this response selects no escape mutants in C clade as well as in B clade infection. The absence of the KF11 response in elite controllers with HLA-B*57:01 does not necessarily mean that these responses have not contributed to immune control in these subjects, since it is possible that the period of detectability may be transient. It is clear that many responses that are undetectable in elite controllers can become detectable after peptide stimulation (51). However, if the KF11-specific CTL response contributes to immune control of HIV in HLA-B*57:01-positive subjects in B clade infection, it would be unique in failing to select escape mutants in the process and the mechanism would be invaluable for directing successful vaccine targets.

Funding acknowledgements

PJRG is supported by the Wellcome Trust and the NIH (RO1 AI046995). ZLB is the recipient of a New Investigator Award from the Canadian Institutes of Health Research and a Scholar Award from the Michael Smith Foundation for Health Research.
REFERENCES


expressing green (EGFP) or red (DsRed2) fluorescent proteins to examine the contribution of pol and env genes to overall HIV-1 replicative fitness. Journal of virological methods 136:102-117.


FIGURE LEGENDS

Figure 1. Median Viral of KF11 responders versus nonresponders in B clade-infected individuals expressing HLA-B*57:01

PBMCs from B clade-infected, ART-naïve individuals expressing HLA-B*57:01 were analysed by IFN-γ ELISPOT assay for responses to the KF11 epitope. Viral loads of responders and nonresponders were compared. Mann-Whitney U tests were performed.

Figure 2. Median Viral Load and CD4 count of B*57:03-positive versus B*57:03-negative subjects in Barbados cohort (B-clade).

B*57:03-positive subjects were compared to B*57:03-negative subjects for (A) Viral Load and (B) CD4 count. Median and 5-95 percentiles are shown. Mann-Whitney U tests were performed.

Figure 3. Effect of viral mutation S173T on epitope recognition of HIV-infected cells by KF11-specific CD8+ T-cells

Ex vivo CD4+ T-cells from B*57:03+ and B*57:03- donors were infected with wildtype NL43 virus or NL43 virus harbouring the S173T viral mutation. Infected CD4+ T-cells were then cultured with KF11-specific CD8+ T-cells (A) or IAW9-specific CD8+ T-cells (B) and the level of CD8+ T-cell activation monitored by expression of CD107 and Mip1β. Data from both experiments were standardized relative to % recognition by wildtype virus (C). Experiments were performed in triplicate, mean and SD are shown. Student’s t test were performed. p value summary is as follows; p<0.01 *, p<0.001 **, p<0.0001 ***.
Figure 4. Median Viral Load and CD4 count of B*57:03-positive HIV-infected subjects with viral polymorphisms S173 and T173.

Proviral DNA sequences from B*57:03-positive subjects from the Barbados cohort (B-clade) were analysed for the presence of the viral polymorphisms S173 and T173. Viral Loads (A) and CD4 counts (B) were compared. Median and 5-95 percentiles are shown. Mann-Whitney t tests were performed.

Figure 5. Viral replication capacity of NL43GFP virus with multiple B*57:03-associated viral mutations

NL43 GFP virus was engineered to contain combinations of the viral mutations 173T, 163G and 165N. MT4 cells were infected and monitored for GFP-positive cells over 14 days (A). The slope of the curve was calculated from the exponential growth phase using the LOGEST function and converted to natural logs (B). Experiments were performed in triplicate and mean and SD are shown. Dunnett’s multiple comparison tests were performed. p value summary is as follows; p<0.01 *, p<0.001 **, p<0.0001 ***

Figure 6. Frequency of KF11 mutations (A163G and S165N) and S173T in C-clade gag sequences (n=1899)

HIV-1 p24 gag sequences (n=1899) were analysed for the presence of the KF11 mutations, A163G and S165N, in the presence of 173T and 173S. B*57:03-positive subjects (A) and B*57-negative subjects (B) were analysed. Fisher’s exact tests were performed.
Table I. Summary of HLA-B*57:03-associated viral mutations in Barbados cohort (n=125).

A phylogenetically-corrected method was used to determine the location of HLA-B*57:03-associated viral mutations in proviral gag p17 and p24. Polymorphism location and HXB2 number are shown. A q value cut-off of q<0.2 was used.

Table II. Summary of amino acid polymorphisms in the KF11 epitope region (HXB2 162-173) in Barbados cohort (n=125).

Proviral sequences were grouped into those from HLA-B*57:03-positive and B*57:03-negative donors. A summary of all HIV-1 polymorphisms is shown for HXB2 region 162-173.

Table III. Summary of PhyloDOR of B*57:03-associated mutations located in HIV-1 p24 gag between Barbados [BB] and Botswana [BW] cohorts.

Proviral sequences from Barbados (n=125) and Botswana (n=322) were analysed from HLA-B*57:03-positive and -B*57:03-negative individuals for the presence of B*57:03-associated viral mutations located in known viral epitopes in p24 gag. HXB2 position of mutations is shown. p values were calculated from phylogenetically-corrected Odds ratios (PhyloDOR) using an online tool (35).
<table>
<thead>
<tr>
<th>Protein</th>
<th>HLA</th>
<th>Polymorphism location</th>
<th>HXB2</th>
<th>p value</th>
<th>q value</th>
</tr>
</thead>
<tbody>
<tr>
<td>p17</td>
<td>B*5703</td>
<td>ASVLSGCELDYWKEIRLRPGG</td>
<td>15</td>
<td>0.00211</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>B*5703</td>
<td>QPSLQTGSEERKSLYNVTATL</td>
<td>75</td>
<td>0.00388</td>
<td>0.14</td>
</tr>
<tr>
<td>p24</td>
<td>B*5703</td>
<td>NLOGMNYHQAISPERLNAWVK</td>
<td>147</td>
<td>0.00544</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>B*5703</td>
<td>AFSPDVPMFALSEGATFQD</td>
<td>173</td>
<td>0.000301</td>
<td>0.0218</td>
</tr>
<tr>
<td></td>
<td>B*5703</td>
<td>RGSDIAGTSLQEQICGMN</td>
<td>242</td>
<td>3.02E-11</td>
<td>6.56E-09</td>
</tr>
<tr>
<td>HXB2 position 162-173</td>
<td>B*5703 positive (n=15)</td>
<td>B*5703 negative (n=110)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>K A F S P E V I P M F S</td>
<td>1</td>
<td>7</td>
<td>77</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>- G - N - - - - - - - - -</td>
<td>7</td>
<td>47</td>
<td>25</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>- S - - - I - - - - - - -</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>- - - K - - - - - - - -</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- - - N - - - - - - - -</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>- - - - - - - - - - -</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- N - - - - - - - - - T</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- S - - - - - - - - - T</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>- - - - - - - - - I -</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>- - - - - - - - - I</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>- - - - - - - - - A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>- G - N - - - - - - - A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>X (HXB2)</td>
<td>Country</td>
<td><strong>Log(PhyloDOR)</strong></td>
<td>*Difference?</td>
<td>p-value</td>
<td>Y/N</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>-------------------</td>
<td>--------------</td>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>242N</td>
<td>Botswana</td>
<td>4.9</td>
<td>3.26E-07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Barbados</td>
<td>11.67</td>
<td>6.15E-14</td>
<td>N</td>
<td>0.815</td>
</tr>
<tr>
<td>147L</td>
<td>Botswana</td>
<td>10.5</td>
<td>4.99E-05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Barbados</td>
<td>13.26</td>
<td>4.00E-03</td>
<td>N</td>
<td>0.2937</td>
</tr>
<tr>
<td>163G</td>
<td>Botswana</td>
<td>13.22</td>
<td>0.0003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Barbados</td>
<td>9.343</td>
<td>0.2006</td>
<td>Y</td>
<td>0.0067</td>
</tr>
<tr>
<td>165N</td>
<td>Botswana</td>
<td>4.213</td>
<td>0.0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Barbados</td>
<td>11.88</td>
<td>0.0887</td>
<td>N</td>
<td>0.0818</td>
</tr>
<tr>
<td>173T</td>
<td>Botswana</td>
<td>-2.798</td>
<td>0.0062</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Barbados</td>
<td>13.52</td>
<td>0.0033</td>
<td>Y</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

*Is B*5703's effect on X different in Botswana versus Barbados?

**Phylogenetically-corrected Odds ratio
Viral load (RNA copies/ml plasma)

KF11 Responders: median = 994, n=28
KF11 Non-Responders: median = 76, n=8

p=0.02
A.

Viral load (RNA copies/ml plasma)

- Non-B*5703: median = 13,350, n=210
- B*5703: median = 3,450, n=29

p=0.015

B.

CD4+ count

- Non-B*5703: median = 398, n=189
- B*5703: median = 565, n=30

p=0.003