HLA-A*7401–Mediated Control of HIV Viremia Is Independent of Its Linkage Disequilibrium with HLA-B*5703


*J Immunol* 2011; 186:5675-5686; Prepublished online 15 April 2011;
doi: 10.4049/jimmunol.1003711
http://www.jimmunol.org/content/186/10/5675

**Supplementary Material**
http://www.jimmunol.org/content/suppl/2011/04/15/jimmunol.1003711.DC1

**References**
This article cites 56 articles, 27 of which you can access for free at:
http://www.jimmunol.org/content/186/10/5675.full#ref-list-1

**Subscription**
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
HLA-A*7401–Mediated Control of HIV Viremia Is Independent of Its Linkage Disequilibrium with HLA-B*5703

Philippa C. Matthews,* Emily Adland,* Jennifer Listgarten,† Alasdair Leslie,‡ Nompumelelo Mkwanazi,§ Jonathan M. Carlson,† Mikkel Harndahl,‡ Anette Stryhn,* Rebecca P. Payne,* Anthony Owu,† Kuan-Hsiang Gary Huang,§ John Frater,# Paolo Paioni,* Henri Kloverpris,* Pieter Jooste,** Dominique Goedhals,†† Cloete van Vuuren,‡‡ Dewald Steyn,‡‡ Lynn Riddell,§§ Fabian Chen,¶¶ Graz Luzzi,¶¶¶ Thembiah Balachandran,## Thumbi Ndung’u,u,§ Søren Buus,玛丽 Carrington,***††† Roger Shapiro,‖‖ David Heckerman,† and Philip J. R. Goulder*§§

The potential contribution of HLA-A alleles to viremic control in chronic HIV type 1 (HIV-1) infection has been relatively under-studied compared with HLA-B. In these studies, we show that HLA-A*7401 is associated with favorable viremic control in extended southern African cohorts of >2100 C-clade–infected subjects. We present evidence that HLA-A*7401 operates an effect that is independent of HLA-B*5703, with which it is in linkage disequilibrium in some populations, to mediate lowered viremia. We describe a novel statistical approach to detecting additive effects between class I alleles in control of HIV-1 disease, highlighting improved viremic control in subjects with HLA-A*7401 combined with HLA-B*57. In common with HLA-B alleles that are associated with effective control of viremia, HLA-A*7401 presents highly targeted epitopes in several proteins, including Gag, Pol, Rev, and Nef, of which the Gag epitopes appear immunodominant. We identify eight novel putative HLA-A*7401–restricted epitopes, of which three have been defined to the optimal epitope. In common with HLA-B alleles linked with slow progression, viremic control through an HLA-A*7401–restricted response appears to be associated with the selection of escape mutants within Gag epitopes that reduce viral replicative capacity. These studies highlight the potentially important contribution of an HLA-A allele to immune control of HIV infection, which may have been concealed by a stronger effect mediated by an HLA-B allele with which it is in linkage disequilibrium. In addition, these studies identify a factor contributing to different HIV disease outcomes in individuals expressing HLA-B*5703. The Journal of Immunology, 2011, 186: 5675–5686.

Human leukocyte Ag class I genotype is a major determinant of HIV-1 viremic control and progression to AIDS (1–5). Much of this effect is mediated by the presentation of HIV-specific epitopes by different HLA class I molecules for recognition of virus-infected cells by CD8+ T cells (6–12). HLA-B is the class I locus with the strongest influence on median viral load in chronic infection (13), and HLA-B*5703 is the allele associated with most effective suppression of viremia in C-clade–infected subjects (6, 14–17). However, an effect on viremic control has also previously been identified in association with some HLA-A alleles (18–21), and with the HLA-Cw class I locus (3, 22, 23).

The starting point of these studies was the observation that HLA-A*7401 is associated with the lowest median HIV-1 viral load of...
all the 23 HLA-A alleles expressed at \( \geq 0.5\% \) phenotypic frequency in a large cohort of HIV-1–infected subjects from Durban, South Africa (3). Furthermore, in this C-clade–infected cohort, we noted that HLA-A*7401 is associated with the selection of escape mutations in Gag, Pol, and Nef (3, 5). These two strands of evidence both suggest that HLA-A*7401–restricted CD8\(^+\) T cell responses may contribute to disease control in chronic HIV-1 infection and are in keeping with recently published data documenting an association between HLA-A*7401 and disease control in Tanzania and Kenya (19, 24).

The role of HLA-A*7401 in viremic control in Durban is potentially confounded by its occurrence in lineage disequilibrium with HLA-B*5703 (23, 25). This raises the possibility that the apparent benefits to viremic control seen among subjects with HLA-A*7401 are, in fact, mediated by a linked HLA-B*5703–restricted response. Alternatively, it is possible that the impact of HLA-A*7401 is independent of HLA-B*5703 but obscured by the stronger effects of the HLA-B allele. Indeed, although HLA-B*5703 is the allele most strongly associated with successful control of viremia in sub-Saharan African populations, considerable variation in viral set point is seen among subjects possessing this allele (23). We therefore hypothesized that some of the stratification in disease outcome among subjects with HLA-B*5703 might be accounted for by the presence or absence of HLA-A*7401 in the haplotype.

HLA-A*7401 alleles belong to the HLA-A19 serotype, which comprises HLA-A*29, -A*30, -A*31, -A*32, and -A*74, and has been included within the HLA-A*03 supertype (26, 27). Although HLA-A*7401 is prevalent in sub-Saharan African populations, occurring at a phenotypic frequency of \( \sim 10\% \), http://www.allelefrequencies.net (28), the peptide binding motif for HLA-A*7401 has yet to be defined.

The aims of this study were, therefore, as follows: first, to investigate the contribution of HLA-A*7401 to control of HIV-1 infection, both alone and in combination with HLA-B*5703 in southern African populations; and second, to investigate the means by which HLA-A*7401 might contribute to viremic control via the presentation of HLA-A*7401–restricted CD8\(^+\) T cell epitopes.

**Materials and Methods**

**Study cohorts**

We studied a total of 2126 treatment-naive, adult subjects with chronic HIV-1 C-clade infection from southern Africa. We recruited subjects via five cohorts as follows: 1) Durban, South Africa (\( n = 1218 \), as previously described (3, 13, 16, 29); 2) Bloemfontein, South Africa (\( n = 261 \); in this cohort, HLA types were ascertained only for subjects with CD4\(^+\) T cell counts <100 or >500 cells/mm\(^3\), as previously described (30)); 3) Kimberley, South Africa (\( n = 31 \) postnatal mothers); 4) Gaborone, Botswana (\( n = 514 \) antenatal women) (10); and 5) southern African subjects attending outpatient HIV clinics in the Thames Valley area of the U.K. (\( n = 102 \) originating from Botswana, Malawi, South Africa, and Zimbabwe (\( n = 1, 11, 17, \) and 73, respectively)). Ethics approval was given by University of KwaZulu-Natal Research Board and the Massachusetts General Hospital Review Board (Durban cohort); the University of the Free State Ethics Committee (Kimberley and Bloemfontein cohorts); the Office of Human Research Administration, Harvard School of Public Health and the Health Research Development Committee, Botswana Ministry of Health (Gaborone cohort); and the Oxford Research Ethics Committee (Durban, Kimberley, and Thames Valley cohorts). Study subjects from all cohorts gave written informed consent for their participation. Viral load in chronic infection was measured using the Roche Amplicor version 1.5 assay; CD4\(^+\) T cell counts were measured by flow cytometry (data for each cohort are shown in Table I).

**HLA typing**

High-resolution HLA typing was performed from genomic DNA by single-stranded conformation polymorphism PCR. As in previous studies (3, 5, 23), for certain alleles in the Durban cohort, including HLA-A*74, the four-digit type was not resolved in a minority of study subjects. In the Durban cohort, 100/138 (72%) subjects with HLA-A*74 were typed to 4 digits, and 100/100 (100%) of these subjects typed to 4 digits had HLA-A*74. Of 81 HLA-A*74\(^+\) subjects from the Thames Valley, Botswana, Kimberley, and Bloemfontein cohorts typed to 4-digits, again 100% had HLA-A*74 confirmed by high-resolution typing. To test the generality of these findings, we also examined a random sample of 90 individuals of African descent with HLA-A*74 recently typed by the National Institutes of Health, and found that >97% had HLA-A*74 (M. Carrington, unpublished observations). In this study, we have therefore classified all HLA-A*74\(^+\) subjects together as HLA-A*74.

**Amplification and sequencing of proviral DNA**

Gag sequences were generated from genomic DNA extracted from PBMCs, amplified by nested PCR to obtain population sequences, as previously described (16, 31). Sequencing for the Durban and Bloemfontein cohorts was undertaken using the Big Dye Ready Reaction Terminator Mix (V3) (Applied Biosystems, U.K.) and manually aligned using Se_Al software, as previously described (31, 32). Sequencing of subjects from the Thames Valley, Botswana, and Kimberley cohorts was performed by Macrogen (South Korea). Sequences were analyzed using Sequencer v4.8 (Gene Codes Corporation). Online sequences are available at GenBank, FJ198407-FJ199088 (Durban) and FJ497801-FJ497950 (Botswana).

**HLA-A*7401 motif prediction and confirmation; HLA–peptide binding studies**

Amino acid sequences for HLA-A*7401 and related HLA-A alleles were downloaded from the International Immunogenetics Project (http://www.ebi.ac.uk/imgt/hla/). HLA-A*7401 optimal epitopes were initially predicted using motif inference based on the comparison of HLA-A*7401 with HLA-A alleles that share sequence homology for the B and F pockets of the peptide binding groove (33). These two pockets determine the preferred residues bound at the “anchor” positions of the epitope (corresponding to position two [2P] and the carboxyl-terminal position [P1]) (34). The validity of the predicted HLA-A*7401 peptide binding motif was tested by two methods. Optimal epitopes were first determined using IFN-\(\gamma\) ELISPOT assays to test recognition of peptides that corresponded to the predicted optimal epitope, and of peptides differing by one amino acid in length from each predicted optimal. We used Los Alamos HIV databases (epitopes “A-list”, http://www.hiv.lanl.gov) to search for previously published HLA-A*7401–restricted epitopes.

Binding of these predicted optimal peptides to HLA-A*7401 was tested in HLA–peptide-binding studies, undertaken using a luminoscent oxygen channeling immunoassay (LOCI), as previously described (35).

**IFN-\(\gamma\) ELISPOT assays**

We tested ex vivo PBMCs from 1010 study subjects from the Durban cohort against a panel of 410 overlapping peptides (OLPs) spanning the entire C-clade HIV-1 proteome to screen for IFN-\(\gamma\) ELISOTF responses, as previously described (29). A total of 119 of these subjects had HLA-A*7401 (91 of these 119 [76%] were confirmed to be HLA-A*7401 at 4-digit resolution; the remainder were typed to only 2 digits). Fisher’s exact test was performed to identify responses to OLPs that were significantly associated with expression of HLA-A*7401. Analysis was also repeated in the absence of the 28 subjects for whom HLA-A*74 was typed to only 2-digit resolution to remove any potential bias from subjects not confirmed to have HLA-A*7401 as the high-resolution type. We tested putative optimal epitopes by ELISPOT using ex vivo PBMCs from HLA-A*7401\(^+\) subjects recruited through the Thames Valley and Durban cohorts.

**Cell staining and flow cytometry**

Cell staining was undertaken from cryopreserved PBMCs using anti-CD3-Pacific orange (Invitrogen), anti–CD8-Alexa Fluor 700 (BD Biosciences), and HLA-A*7401-RR9 tetramer conjugated to PE. Dead cells were gated out using “live/dead” viability kit (Invitrogen).

**Statistical analysis**

We used Fisher’s exact test to screen for LD between HLA class I alleles, using the online tool available at http://www.hiv.lanl.gov/content/immunology/hla/hla_linkage.html, corrected for multiple comparisons using a Bonferroni approach. To identify MHC class I haplotypes, we used PyPop software [Python for Population Genomics, http://www.pypop.org (36)] (Supplemental Table I). Other statistical analysis was undertaken using Prism GraphPad software v. 5.0a.

To assess whether any possible pairs of HLA class I alleles might perform better than a single allele in mediating disease control in chronic infection,
we devised a novel statistical test for additive effects using pooled data from a total of 2126 C-clade–infected adult subjects from southern Africa. Each HLA combination was tested to see whether an additive model for two HLA alleles together, irrespective of linkage disequilibrium, performed better in predicting disease control than a model that considered only the most predictive of each single HLA allele. For the purposes of this test, each allele was collapsed to a two-digit resolution only, apart from exceptions in which it is known that the four-digit types cross supertype boundaries or have a strong bearing on outcome (HLA-A*68XX, HLA-B*15XX, and HLA-B*58XX) (13, 27).

For disease control based on the log of the absolute viral load and the log of the CD4+ T cell count, we used the log likelihood of linear-regression–based models to look for additive effects. In particular, we defined the null model for a given HLA pair as the one of the two models, \( p_{0X} \) or \( p_{0Y} \) (defined next), that had the maximum likelihood under a maximum likelihood parameter setting. Intuitively, it is a model where at most only one HLA at a time can have an effect within an individual patient; the HLA chosen is the one with the larger overall effect in such a model. Thus, the null model was either \( p_{0X}(\text{phen}|X,Y) = \text{normal}(w_0 + w_1X + w_2Y, \sigma_X) \) where \( X = 0 \) for people with \( Y = 1 \), and otherwise \( X = X \), or \( p_{0Y}(\text{phen}|X,Y) = \text{normal}(w_0 + w_1X + w_2Y, \sigma_Y) \) where \( Y = 0 \) for people with \( X = 1 \), and otherwise \( Y = Y \). The alternative model used is given by \( p_{1X}(\text{phen}|X,Y) = \text{normal}(w_0 + w_1X + w_2Y, \sigma_X) \).

The variable “phen” is, for example, the viral load; \( X \) and \( Y \) are binary variables representing two HLA alleles; \( w_1 \) and \( w_2 \) are parameters in the model for the effect of each allele; \( \sigma_X \) is the variance parameter of the normal distribution; and \( w_0 \) is an offset parameter. Parameters were fit independently for each model. The test statistic for an HLA pair was the difference in log likelihood between these null and alternative models after fitting each by maximum likelihood. The \( p \) values were obtained by 50,000 permutations of one HLA allele in the test. In addition, any pair of alleles for which one of the alleles did not have a univariate correlation in the direction of control (as defined by the phenotype of interest) was assigned a test statistic of zero. Note that two alleles in perfect linkage disequilibrium (LD) (or anti-LD) could not come up as having an additive effect with this test statistic. For binary outcomes (disease control based on criteria of viral load \( \leq 2000 \) RNA copies/ml plasma and CD4+ T cell count \( > 250 \) cells/mm³), we used an analogous test based on logistic regression.

To account for possible differences between cohorts that might cause artefact either by enhancing or obscuring additive effects between HLA alleles (i.e., confounding), we also corrected for geographical origin of the subject by adding a set of binary cohort covariates to each model (one for recruitment in each of Durban, Kimberley, Bloemfontein, or Gaborone; we removed Thames Valley subjects who originate from a wide variety of locations). An LRT test showed these cohort covariates to be highly significant (e.g., \( p = 2 \times 10^{-36} \)) in predicting the log CD4+ T cell count, so these covariates were used in all models. Note that all HLA pairs coming out as significant in these tests had parameter weights in the direction of control in the alternative model, indicating that better predictive power was obtained because both alleles were contributing to control rather than one negating the effect of the other.

Results

HLA-A*7401 is common in African populations and is found in linkage disequilibrium with HLA-B*5703 in South Africa

HLA-A*74 was found in 216 subjects from a total of 2126 southern African subjects (phenotypic frequency, 10.2%) (Table I). These were confirmed to be HLA-A*7401 in 181/216 cases (84%); the remaining cases were typed only to 2-digit resolution as HLA-A*74. Thus, no four-digit HLA-A*74 allele other than HLA-A*7401 was identified in any of our cohorts.

We confirmed LD between HLA-A*7401 and HLA-B*5703 among South African subjects from Durban, Kimberley, and Bloemfontein (\( n = 1510, p = 1.1 \times 10^{-5}, \) Fisher's exact test) (Fig. 1A). In the epidemiologically unlinked study population of adults in Botswana (\( n = 514 \)), we did not identify statistically significant LD between these two alleles (Fig. 1A, 1B). The apparent lack of LD between HLA-A*7401 and HLA-B*5703 in Botswana does not appear to be attributable to a smaller sample size than our South African cohorts, because in the Botswana cohort, only 1/46 (2.2%) HLA-A*7401* subjects also carried HLA-B*5703, whereas in the South African cohorts, 25/157 (15.9%) HLA-A*7401* subjects also carried HLA-B*5703. The HLA-A*74-B*57 haplotype has previously been identified in African Americans [http://www.allelefrequencies.net (28); both alleles resolved to two digits only in this source (3)], but to our knowledge has not been reported in other African populations to date.

To confirm the haplotypes in which HLA-A*7401 is transmitted, we analyzed HLA data for Durban, Bloemfontein, and Gaborone using PyPop software [http://www.pypop.org (36)] (Supplemental Table I). In keeping with our analysis of LD (Fig. 1), this demonstrates that HLA-A*7401 is transmitted on a haplotype with HLA-B*1503, HLA-B*3501, and HLA-B*5703 in South Africa.

The association of HLA-A*7401 with HLA-B*1503 and HLA-B*5703 in Botswana is also seen in Botswana, but in this population, there is no HLA-A*7401-B*5703 haplotype. In addition, these data show a weaker haplotype of HLA-A*7401-B*4201 in all three cohorts, but the LD between HLA-A*7401 and this allele was not statistically significant.

HLA-A*7401 is statistically associated with lowered viral load in chronic C-clade HIV-1 infection

To assess the relation between HLA-A*7401 and viral load in chronic infection, we compared subjects with and without HLA-A*7401 in four different cohorts: Durban, Gaborone, Bloemfontein, and the Thames Valley cohorts (excluding Kimberley because of small numbers). In each case, HLA-A*7401 was associated with a reduction in viral load in chronic infection \( p = 0.0007 \) (Fig. 2A), \( p = 0.02 \), \( p = 0.2 \), and \( p = 0.05 \), respectively, Mann–Whitney U test]. The trend toward lowered viremia in association with HLA-A*7401 in Bloemfontein may not have reached statistical significance at least in part because of the small number of HLA-A*7401 subjects in Bloemfontein (\( n = 8 \)). In all southern African subjects, HLA-A*7401 was strongly associated with lowered viremia (median viral set point, 14600 versus 32450 RNA copies/ml in HLA-A*7401+ versus HLA-A*7401−; \( n = 1965 \) subjects with viral load data available; \( p < 0.0001 \), Mann–Whitney U test) (Fig. 2B).

In subjects from Gaborone (where HLA-A*7401 is not in linkage disequilibrium with HLA-B*5703), CD4+ T cell counts were statistically higher in subjects with HLA-A*7401 (\( p = 0.0005 \), Mann–Whitney U test) (Fig. 2C). This illustrates that HLA-A*7401 may mediate a protective effect that is independent of linkage disequilibrium with HLA-B*5703. Associations between absolute CD4+ T cell count and expression of HLA-A*7401 did not reach statistical significance in the Durban, Bloemfontein, or Thames Valley cohorts (data not shown). However, in pooled data from all subjects, there was a significant association between the presence of HLA-A*7401 and higher CD4+ T cell count (median CD4+ cell count, 409 versus 350 in HLA-A*7401+ versus HLA-A*7401−; \( n = 1962 \) subjects with CD4+ cell count available; \( p = 0.001 \), Mann–Whitney U test) (Fig. 2D).
However, the combination of HLA-A*7401 and HLA-B*5703 resulted in significantly lower viral loads than in the presence of either allele alone, suggesting an additive effect of these alleles in mediating control of disease (Fig. 3A, 3B), as seen previously (23). Additional evidence for the independent role of HLA-A*7401 in lowering viremia, irrespective of the presence of HLA-B*5703, comes from the Botswana cohort in which there is no LD between these alleles (Fig. 1), but HLA-A*7401 is still associated with favorable outcomes (see earlier and Fig. 2C).

As HLA-A*7401 is also transmitted in strong LD with HLA-B*1503 and HLA-B*3501 in both Botswana and South Africa (Fig. 1A), we also sought to exclude either of these alleles as the driving force behind the apparent benefit of HLA-A*7401. There was no relation between HLA-B*1503 or HLA-B*3501 and lowered viremia in this cohort (data not shown); indeed, in Durban, HLA-B*1503 was actually statistically associated with higher viral loads (p = 0.03, Mann–Whitney U test; data not shown). The overall relation between HLA-A*7401 and lowered viral load is, therefore, not likely to be attributable to linkage with either of these alleles.

To further investigate the possible additive effect of HLA-A*7401 on disease control in subjects with HLA-B*5703, we ranked HLA-B*5703+ subjects from the whole pooled southern African cohort (n = 97) and from Durban (n = 55) according to viral load. We found significant enrichment of HLA-A*7401 in subjects with viral loads less than the median in each case (p = 0.015 in the whole cohort, p = 0.003 in Durban, Fisher’s exact test). Division of these HLA-B*5703+ subjects into four pools by viral load quartiles confirmed enrichment of HLA-A*7401 in subjects with the lowest viral loads (Fig. 3C, 3D). This phenomenon is not associated with the presence of other favorable HLA-B alleles, such as HLA-B*5702, -B*5801, or -B*8101, in subjects with the lowest viral loads (data not shown), suggesting that HLA-A*7401 may indeed be responsible for mediating the effect.

To investigate whether the beneficial effect of HLA-A*7401 occurs significantly in combination with HLA alleles other than HLA-B*5703, we undertook a computational analysis of all 2126 southern African subjects. Using this newly developed test, we sought evidence for additive effects of any combination of two HLA alleles (collapsed to two-digit resolution), irrespective of LD in improving disease control compared with the effect of any one allele alone. After correction for cohort location, we identified statistically significant additive effects between HLA-A*74 and two favorable HLA-B alleles, HLA-B*57 and HLA-B*81, in mediating control of viral load (Table II). No significant effect of HLA-A*74 was determined in conjunction with any other allele in

### Table I. Summary of 2126 C-clade HIV-1–infected study subjects from five southern African study cohorts

<table>
<thead>
<tr>
<th>Cohort Location</th>
<th>No. of Subjects</th>
<th>Number (%) with HLA-A*7401&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Viral Load (RNA Copies/ml Plasma)</th>
<th>CD4 T Cell Count (Cells/mm&lt;sup&gt;3&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number</td>
<td>Median</td>
</tr>
<tr>
<td>Durban, South Africa</td>
<td>1218</td>
<td>135 (11.1)</td>
<td>1218</td>
<td>38,200</td>
</tr>
<tr>
<td>Bloemfontein, South Africa&lt;sup&gt;b&lt;/sup&gt;</td>
<td>261</td>
<td>21 (8.0)</td>
<td>163</td>
<td>84,000</td>
</tr>
<tr>
<td>Kimberley, South Africa</td>
<td>31</td>
<td>1 (3.2)</td>
<td>14</td>
<td>47,000</td>
</tr>
<tr>
<td>Gaborone, Botswana</td>
<td>514</td>
<td>46 (8.9)</td>
<td>471</td>
<td>19,100</td>
</tr>
<tr>
<td>Thamess Valley, southern Africa</td>
<td>102</td>
<td>13 (12.7)</td>
<td>99</td>
<td>5090</td>
</tr>
<tr>
<td>All Southern Africa</td>
<td>2126</td>
<td>216 (10.2)</td>
<td>1965</td>
<td>29,500</td>
</tr>
</tbody>
</table>

<sup>a</sup>The 4-digit HLA type was determined in 181/216 subjects (84%); in all 181 of these (100%), the 4-digit type was HLA-A*7401. In the remaining 35 cases typed to 2 digits, the 4-digit type was undetermined.

<sup>b</sup>Bloemfontein data are for 261 HLA-typed subjects with CD4+ T cell count < 100 or > 500 cells/mm<sup>3</sup>. In an extended Bloemfontein cohort (including subjects without HLA types), median CD4+ cell count was 259 cells/mm<sup>3</sup>, with interquartile range (IQR) of 122–366 cells/mm<sup>3</sup> (n = 885), and median viral load was 76,000, with IQR of 20,000–21,000 cells/mm<sup>3</sup> (n = 500).

**FIGURE 1.** LD between HLA-A*7401 and other class I alleles in southern African populations. A, All HLA-B and HLA-Cw alleles that are significantly linked to HLA-A*7401 in a pooled cohort from South Africa (Durban, Bloemfontein, and Kimberley cohorts, n = 1510) and in Botswana (n = 514) are shown. The p values were calculated by Fisher’s exact test and corrected for multiple comparisons using the Bonferroni approach, using the online tool at: http://www.hiv.lanl.gov/content/immunology/hla/hla_linkage.html. No significant linkage associations were identified for HLA-A*7401 among Zimbabweans recruited via the Thames Valley cohort (n = 73, data not shown). All subjects with HLA-B*35 are pooled together as HLA-B*35(01). B, Percentage of HLA-A*7401+ subjects with HLA-B*5703 in each of five southern African cohorts. Statistically significant LD between these alleles is present only in the Durban cohort.
having predicted a peptide binding motif for HLA-A*7401, we identified regions of the HIV-1 proteome-containing potential HLA-A*7401–restricted epitopes using two approaches. First, we screened IFN-γ ELISPOT responses from 1010 Durban subjects. We identified two statistically significant IFN-γ responses made by 119 subjects with HLA-A*7401: these were to OLP-19 in p24 Gag (p = 1.7 × 10⁻⁵) and OLP-100 in Rev (p = 1.8 × 10⁻⁵; Fisher’s exact test) (Fig. 4, Table IV). Even after exclusion of the minority of subjects with HLA-A*74 to only two digits (n = 28), both of these associations remained statistically significant (p = 2.5 × 10⁻⁴ and p = 1.8 × 10⁻⁵, respectively). To define the optimal epitope within each of these OLPs, we identified a putative HLA-A*7401–restricted epitope based on the earlier predicted binding motif for HLA-A*7401 (L/I/M/Q at P2 and R/K at PC) (Table III). Within OLP-19 (IVNLQQGMVHQAISPR), the peptides GR11 (GQMVHQAISPR, p24-Gag 8–18) and QR10 (QMVHQAISPR, p24-Gag 9–18) fitted this motif. Within the Rev 17mer OLP-100 (RWRARQRQHISIERSLR), two putative overlapping HLA-A*7401–binding peptides were identified: Rev-RR9 (RQHISIERSLR, Rev 50–58) and Rev-RR11 (RQRQHISIERSLR, Rev 48–58).

The second approach we used to identify HIV-specific HLA-A*7401–restricted epitopes involved reference to previously determined HLA-A*7401–associated HIV amino acid polymorphisms within Gag, Pol, and Nef population sequences from the Durban cohort (3). Six sites of HIV polymorphism associated with HLA-A*7401 were identified (3, 40) (Table V). Based on the predicted HLA-A*7401 binding motif, we used these sites of point mutation to identify further putative HLA-A*7401–restricted epitopes, p17-KR9 (p17 12–20, KLDKWEKIR), and Nef-SR9 (QMVHQAISPR, p24-Gag 9–18) fitted this motif. Within the Rev 17mer OLP-100 (RWRARQRQHISIERSLR), two putative overlapping HLA-A*7401–binding peptides were identified: Rev-RR9 (RQHISIERSLR, Rev 50–58) and Rev-RR11 (RQRQHISIERSLR, Rev 48–58).

To confirm the optimal epitope, we used IFN-γ ELISPOT assays to quantify CD8⁺ T cell responses to the optimal epitope and variant truncations of the peptide in individual subjects selected on the basis that they made a response to the relevant OLP. Using this method, we acquired data to support the optimal epitope in three of our predicted HLA-A*7401–restricted epitopes: p24-Gag-GR11 (p24 Gag 8–18, GQMVHQAISPR), RT-QR9 (RT 269–277, QIYPGKVR), and Rev-RR9 (Rev 50–58, RQRQHISIERSLR) (Fig. 5A–C, respectively). In the case of Rev-RR9, we also synthesized the peptide–MHC class I tetramer to confirm the HLA restriction of the response (Fig. 5D).

We also used LOCI (35) to confirm binding of HLA-A*7401 to optimal peptides Gag-GR11, RT-QR9, and Rev-RR9 (Fig. 6). These binding studies also add to the evidence for two other putative HLA-A*7401–restricted epitopes, p17-KR9 and Nef-SR9.
There was no substantial difference in the strength of binding of overlapping variants of our three confirmed epitopes (p24-Gag epitopes p24-GR11/QR10, RT-SR10/QR9, and Rev-RR11/RR9) (Fig. 6), suggesting that the optimal peptide for presentation is not necessarily determined by differences in HLA class I binding.

Together, these ELISPOT, tetramer, and binding data strongly support three new HLA-A*7401–restricted epitopes (Gag-GR11, RT-QR9, and Rev-RR9) and point to two further likely epitopes (Gag-KR9 and Nef-SR9). These data also substantiate our proposed motif for HLA-A*7401 as L/I/M/Q at P2. At the PC of the epitope, R appears to be the preference (our two putative epitopes bearing K at this position, p17-SK10 and RT-SK9, were both low binders to HLA-A*7401). We also tested binding of the only HLA-A*7401 HIV-1 epitope that has previously been reported by Los Alamos HIV databases (http://www.hiv.lanl.gov), Prot-IV9 (Pol 58–67, ITLWQRPLV) (41); this does not match our motif prediction at the PC, and was also a low binder.

In the presence of HLA-A*7401, the HLA-B*5703 Gag footprint is diminished and HLA-B*5703–restricted IFN-γ ELISPOT responses are preserved.

To investigate the way in which HLA-A*7401 may ameliorate disease progression in subjects with HLA-B*5703, we investigated the impact of HLA-A*7401 on the selection of polymorphisms and Table II. Additive effects between HLA-A*74 and other HLA class I alleles in mediating good HIV-1 disease control

<table>
<thead>
<tr>
<th>Criterion for Disease Control</th>
<th>HLA-1</th>
<th>HLA-2</th>
<th>p Value</th>
<th>q Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute viral load</td>
<td>A*74</td>
<td>B*57</td>
<td>2.00E-05</td>
<td>0.023</td>
</tr>
<tr>
<td>Viral load &lt; 2000 copies/ml</td>
<td>A*74</td>
<td>B*81</td>
<td>2.00E-05</td>
<td>0.023</td>
</tr>
<tr>
<td>Absolute CD4⁺ T cell count</td>
<td>A*74</td>
<td>NS</td>
<td>2.00E-05</td>
<td>0.023</td>
</tr>
<tr>
<td>CD4 T cell count &gt; 250 cells/mm³</td>
<td>A*74</td>
<td>NS</td>
<td>2.00E-05</td>
<td>0.023</td>
</tr>
</tbody>
</table>

Analysis of HLA type, viral load (n = 1972), and CD4 count (n = 1871) in adults with chronic HIV-1 infection from pooled subjects from southern Africa (South Africa, Botswana, Zimbabwe, and Malawi). All associations between HLA-A*74 and any other allele (q < 0.2) are shown.

NS, not significant.
likely than HLA-A*7401 and B*4801 (10). Subjects with HLA-A*7401 were also less position (including HLA-B*3910, HLA-B*1510, HLA-B*5801, number making response/total number of subjects.

ELISPOT data from 46 HLA-B*5703+ southern African subjects impact on HLA-B*5703 responses to Gag epitopes, we studied on IFN-γ ELISPOT responses associated with HLA-B*5703. HLA-B*5703 is strongly associated with the selection of five mutations in p24 Gag, at positions A146, I147, A163, S165, and T242, within and flanking the epitopes ISW9 (p24 Gag 147–155, ISPRTLNAW), KF11 (p24 Gag 162–172, KAFSPEVIPMF), and TW10 (p24 Gag 240–249, TSTLQEIQAW) (3, 6, 16, 17).

In 64 subjects with HLA-B*5703, there was a trend toward preservation of the wild-type amino acid in the presence of HLA-A*7401 at 4 of the 5 sites of polymorphism (I147, A163, S165, and T242), although this only reached statistical significance at S165 (p = 0.004, Fisher’s exact test; Fig. 7A). The exception was A146, flanking the ISW9 epitope. However, selection at this position may occur as a consequence of many different alleles because of the extensive overlap of CD8+ T cell epitopes at this position (including HLA-B*3910, HLA-B*1510, HLA-B*5801, and B*4801) (10). Subjects with HLA-A*7401 were also less likely than HLA-A*7401+ subjects to have mutations at all five sites (p = 0.03, Fisher’s exact test) (Fig. 7A).

To determine whether the presence of HLA-A*7401 has an impact on HLA-B*5703 responses to Gag epitopes, we studied ELISPOT data from 46 HLA-B*5703+ southern African subjects from the Durban and Thames Valley cohorts. Although responses to individual OLPs containing HLA-B*5703 epitopes were not statistically associated with the presence or absence of HLA-A*7401, we found overall preservation of an IFN-γ response to these OLPs in the presence of HLA-A*7401 (p = 0.006 for response to any OLP containing an HLA-B*5703 epitope, Fisher’s exact test) (Fig. 7B). The trend toward maintenance of these IFN-γ ELISPOT responses is in keeping with the sequence data showing preservation of wild-type epitopes in the presence of HLA-A*7401.

Discussion
These studies add to recently published literature (19, 23) substantiating the beneficial influence of HLA-A*7401 on disease control in adult subjects with HIV-1 infection, both alone and in tandem with HLA-B*5703. HLA-A*7401 is common in black African populations (occurring at a 10.2% phenotypic frequency in our extended cohort of >2100 southern African subjects), but has been relatively understudied compared with many HLA-B alleles.

The transmission of HLA-A*7401 in LD with HLA-B*5703 in the Durban cohort highlights the difficulties of appreciating the influence of a single allele. Although our calculations to verify LD do not, of course, demonstrate, in an individual subject, on which haplotype the HLA-A*7401 and HLA-B*5703 alleles are located, they show that statistically these alleles are likely to arise on the same haplotype when coexpressed in a particular individual, and this is verified by haplotype analysis using PyPop (36) (Supplemental Table I). Given the well-established benefits of HLA-B*5703, it would be easy to dismiss the apparent association between HLA-A*7401 and low viremia as simply caused by linkage with HLA-B*5703. However, in addition to our data demonstrating a significant HLA-A*7401–restricted response through CD8+ T cell epitopes, we also present two further strands of evidence for an independent role for HLA-A*7401. First, the favorable effect of this allele on viremic control observed in a Botswana cohort (in which these two alleles are not transmitted in LD) suggests a true independent benefit of HLA-A*7401. Second, we have identified a consistent additive effect of HLA-A*74 and HLA-B*57 through a novel computational approach to the analysis of all possible pairs of two alleles at the A, B, and C loci in >2100 subjects.

Our results demonstrate that HLA-A*7401 is associated with lower viral loads and higher CD4+ T cell counts in several
southern African cohorts. Although these associations do not reach statistical significance in every cohort, the trends are broadly consistent in different settings. There are several possible explanations for the lack of strong correlation with CD4+ T cell counts in cohorts other than Gaborone. First, viral load and CD4+ T cell count are correlated, but not always strongly (typical correlation coefficients reported are of $r^2 = 0.28$) (42), so an association between HLA-A*7401 and lowered viremia does not necessarily translate into a statistically significant association with higher CD4+ T cell count, especially where cohort size is small. Indeed, this same observation has been made in a previous study of HLA associations with viral set point or absolute CD4+ T cell count (23), in which some alleles were significantly associated with low viral set point, for example, but not with high absolute CD4+ T cell count. An alternative possible explanation of this observation might be that certain HLA alleles have a stronger impact on CD4 count than on viral load, or vice versa, for reasons unknown.

Using motif prediction undertaken according to previously validated methodology (33), in combination with ELISPOT and sequence data, we identified eight potential sites of HLA-A*7401–restricted CD8+ T cell epitopes in HIV-1 Gag, Pol, Rev, and Nef (Tables IV, V). Putative epitopes at five of these sites are intermediate or good binders to HLA-A*7401. In particular, the presence of epitopes in Gag may contribute to viremic control in two specific ways. First, the presence of Gag epitopes suggests the potential for an HLA-A*7401–restricted response to emerge early in the course of acute infection because of the abundance of Gag protein at the initiation of infection (43). Second, of the six sequence polymorphisms we identified in association with HLA-A*7401, three have previously been predicted to revert to wild-type after transmission to an HLA-mismatched host (Table V), suggesting a fitness cost imposed by the mutation (3). In these instances, we predict that selection of the mutation itself may contribute to suppression of viremia in subjects with HLA-A*7401 (3) through an effect on viral infectivity and/or replicative capacity. This effect has previously been shown in mutations in HLA-B*57 epitopes that contribute to a reduced fitness virus (6, 17) and are therefore associated with clinical control of disease (44–46). In particular, the presence of two reverting sites in Gag highlights specifically that HLA-A*7401 may contribute to viremic control through the selection of costly mutations in the highly structurally constrained Gag protein (17, 47–49).

To date, only one putative HIV-1 epitope restricted by HLA-A*7401 is listed by Los Alamos HIV databases “A-list” (http://www.hiv.lanl.gov), Prot-IV9 (Pol 58–67 ITLWQRPLV) (41). However, this epitope sequence was determined through pre-
diction, and the HLA restriction has not been demonstrated. The binding studies shown in this article suggest this peptide is a very low binder of HLA-A*7401 (Fig. 6), and together with two of the putative epitopes presented in this article (Prot-NR10 and RT-SK9), it is therefore less likely to be a commonly targeted HLA-A*7401–restricted epitope. A further two putative overlapping HLA-A*7401–restricted epitopes have previously been proposed on the basis of an HLA-A*7401–associated polymorphism at position Gag-441 (24). One of these did fit the HLA-A*7401 binding motif proposed in this article, with Lys at the PC; however, the other lacked a positively charged residue at PC. No HLA-A*7401 binding data were provided for these proposed epitopes, and these peptide sequences did not meet the more rigorous criteria for epitopes set out in the Los Alamos Immunology Database “A-list.” These criteria include titration curves showing recognition of the optimal epitope at serial peptide dilutions compared with four additional peptides, each differing from the optimal by one amino acid, either longer or shorter by one amino acid at the N or C termini, respectively (see Fig. 5A, 5B), or unequivocal staining with a peptide–MHC tetramer (see Fig. 5D). Peptides have been described that do not precisely fit the established peptide binding motif for the respective HLA molecule (50), but these are the exception rather than the rule. In contrast, more typically, a peptide that does not fit the peptide binding motif (such as Prot-IV9; Fig. 6) proves not to bind once the peptide–MHC binding assays are undertaken.

Notably, the approach we adopted using HLA-associated polymorphisms to identify location of an epitope successfully enabled us to identify optimal HLA-A*7401–restricted epitopes, but this approach is challenging because, by definition, few HLA-A*7401+ individuals make responses to epitopes where escape mutations are so readily selected. Despite screening cryopreserved PBMCs from 15 HLA-A*7401+ subjects from the Thames Valley cohort against OLPs and optimal peptides for putative epitopes restricted by this allele, we identified only two responses: one each in subjects H005 and R070 (Fig. 5). On a follow-up visit (after an interval of 23 mo), H005 had lost the IFN-γ response to RT-QR9, and we were therefore unable to proceed with further restriction of this epitope. These findings suggest that HLA-A*7401 CD8+ T cell responses may be short-lived (arising early in the course of infection and being lost as a consequence of immune escape), or may be subdominant compared with greater magnitude responses restricted by HLA-B alleles, and therefore difficult to detect using this IFN-γ ELISPOT screening method. However, despite the challenges of confirming HLA-A*7401–restricted responses, in this article, we present in vitro data (binding studies and tetramer data) that support three new epitopes: Gag-GR11 (GQMVHQAISPR), RT-QR9 (QIYPGIKVR), and Rev-RR9 (RQIHISER).

In these in vitro experiments, HLA-A*7401 constructs were used, confirming the binding of this allele (resolved to high resolution) to predicted peptides. Although ELISPOT responses to a particular peptide cannot necessarily be attributed to one specific allele (in this case, HLA-A*7401), the combined approach of analysis of ELISPOT and sequence data, motif inference, in vitro binding studies, and tetramer binding is a robust method of highlighting putative novel epitopes.

**FIGURE 6.** Binding of HLA-A*7401 to putative optimal epitopes, including overlapping variants of epitopes in p17, RT, and Rev and sequence variants of the epitope in Nef. Binding affinity KD (nM) of HLA-A*7401 to optimal peptides assessed by LOCI (35). Strength of binding is classified on a log scale in accordance with previous methods (35).

**FIGURE 7.** Relation between HLA-A*7401 status and frequency of HLA-B*5703–restricted CD8+ T cell mutations/IFN-γ ELISPOT responses. A. Frequency of HLA-B*5703–selected Gag polymorphisms [defined by previous studies (3, 6, 16, 17)], in 64 subjects with HLA-B*5703 according to the presence/absence of HLA-A*7401. p24-Gag sequences from Bloemfontein (n = 8), Durban (n = 38), Gaborone (n = 11), and Thames Valley cohorts (n = 7), with sequence data available for TW10 (n = 63), KF11 (n = 64), ISW9 (n = 63), and all three epitopes (n = 62). B. Frequency of IFN-γ ELISPOT response to three HLA-B*5703 epitopes in p24-Gag in 46 subjects with HLA-B*5703 according to the presence or absence of HLA-A*7401. The p values were by Fisher’s exact test. Data from Durban (n = 31) and Thames Valley cohorts (n = 15).
The mechanism by which the additive effect of HLA-A*7401 and HLA-B*5703 arises is uncertain. The preservation of wild-type HLA-B*5703 epitopes and the related maintenance of an HLA-B*5703-restricted CD8+ T cell response in the presence of HLA-A*7401 may be the cause or effect of lowered viremia. Several potential mechanisms may explain the favorable effect of this haplotype. First, there may simply be an additive effect of multiple CD8+ T cell responses consequent on the dual presentation of epitopes restricted by both alleles on the surface of infected cells, particularly given the Gag-directed responses restricted by both alleles. The preservation of wild-type sequence in HLA-B*5703-restricted Gag epitopes in the presence of HLA-A*7401 suggests the possibility that subdominant responses from the HLA-A allele delay HLA-B–selected escape mutations, allowing preservation of responses to epitopes restricted by HLA-B*5703. Further protection might be afforded by HLA-A*7401–peptide complexes as killer Ig-like receptor ligands (51), activating a favorable NK response. Finally, we cannot exclude the presence of another confounding variable, either within or outside the MHC locus, that is transmitted in linkage with HLA-A*7401, which is responsible for enhanced viremic control.

It is well-known that HLA-B*5703 and other HLA-B*57 alleles have better immune control than others. One possibility is that HLA-B*5703–restricted CD8+ T cell response in the presence of HLA-A*7401 are shown (Durban and Thames Valley cohorts, as shown in Fig. 6; p values by Fisher’s exact test). Arrows indicate the anchor positions at which the optimal binds the class I allele (position 2 and C terminus); these residues are shown in bold and underlined.

The putative HLA-A*7401–restricted epitopes predicted using motif inference (32) combined with analysis of IFN-γ ELISPOT responses in subjects expressing HLA-A*7401 are shown (Durban and Thames Valley cohorts, as shown in Fig. 4; p values by Fisher’s exact test). Arrows indicate the anchor positions at which the optimal binds the class I allele (position 2 and C terminus); these residues are shown in bold and underlined.

### Table IV: Putative HLA-A*7401–restricted epitopes predicted using motif inference (32) combined with analysis of IFN-γ ELISPOT assays

<table>
<thead>
<tr>
<th>Protein</th>
<th>OLP Number</th>
<th>C-clade OLP Sequence</th>
<th>p Value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p24-Gag</td>
<td>19</td>
<td>I V Q N L Q G Q M V H Q A I S P R</td>
<td>1.7 x 10^-4</td>
</tr>
<tr>
<td>Rev</td>
<td>100(^b)</td>
<td>R W R A R Q R Q I H S I R E L</td>
<td>1.8 x 10^-5</td>
</tr>
</tbody>
</table>

\(^a\)The p values both remain statistically significant if analysis repeated after removal of minority of subjects with HLA-A*74 restricted to only two-digit resolution: OLP-19, p = 2.5 x 10^-5, OLP-100, p = 1.8 x 10^-5.

\(^b\)Responses to OLP-100 are corrected by removal of subjects with HLA-B*1510 from analysis, because of presence of HLA-B*1510 epitope IL9 (HSISERIL, Rev 52–60; http://www.fandl.gov).

### Table V: Putative HLA-A*7401–restricted epitopes predicted using motif inference (32) combined with analysis of HLA-A*7401 associations with sequence polymorphisms

<table>
<thead>
<tr>
<th>Protein</th>
<th>HXB2 Position of Polymorphism</th>
<th>A*7401 Polymorphism</th>
<th>Reversion</th>
<th>p Value</th>
<th>q Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>p17 Gag</td>
<td>12</td>
<td>G A R A S T L R G K L D K W E K IR L</td>
<td>R</td>
<td>2.71E-14</td>
<td>0.00</td>
</tr>
<tr>
<td>p17 Gag</td>
<td>20</td>
<td>G E R L D K W E K R L P G K K H Y M</td>
<td>R</td>
<td>6.47E-10</td>
<td>0.00</td>
</tr>
<tr>
<td>Protease</td>
<td>9</td>
<td>F E R N L A F F Q G A R E K P S E</td>
<td>R</td>
<td>4.83E-05</td>
<td>0.02</td>
</tr>
<tr>
<td>RT</td>
<td>277</td>
<td>A S Q I Y G P I K V R Q L C K L L G A K</td>
<td>R</td>
<td>3.13E-04</td>
<td>0.13</td>
</tr>
<tr>
<td>RT</td>
<td>476</td>
<td>I V S L T E T T N Q K T E L Q A I Q L A L</td>
<td>R</td>
<td>6.03E-04</td>
<td>0.15</td>
</tr>
<tr>
<td>Nef</td>
<td>192</td>
<td>K W F D S L A R H L A R L H P E Y</td>
<td>R</td>
<td>4.95E-05</td>
<td>0.00</td>
</tr>
</tbody>
</table>

\(^a\)Six sites of HLA-A*7401–associated polymorphism identified from lineage-corrected analysis of sequences from a total of 710 Durban subjects (q < 0.2), as previously published (3). The position of the polymorphism is marked \(\uparrow\), with consensus sequence 10 aa upstream and downstream of the site of polymorphism. Arrows indicate anchor positions (shown in bold and underlined) as in Table IV. R indicates polymorphisms that are predicted to revert to wild-type after transmission to an HLA-mismatched recipient [methods as previously described (17, 57)].

\(^b\)Sequences based on Durban consensus sequence from Gag (n = 446) and Nef (n = 436) subjects.
tial for HIV class I alleles to work in tandem to influence HIV-1 disease control, and demonstrate the importance of studying HIV class I alleles that arise in different populations; HIV-A*7401 is not found in white subjects, but is common, and appears to exert considerable influence, in black African cohorts. Furthermore, our data highlight the importance of a broad CD8+ T cell response, especially targeting Gag peptides, and suggest the utility of combining multiple CD8+ T cell responses in a potential vaccine.

Disclosures

The authors have no financial conflicts of interest.

References


