HLA-B molecules target more conserved regions of the HIV-1 proteome

Ana I. Fontaine Costa[a,*], Xiangyu Rao[b,*], Emmanuelle LeChenadec[b], Debbie van Baarle[a,d] and Can Keşmir[b,c]

Background: HLA-B alleles of HIV-infected individuals have been shown to have a major impact on their rate of progression toward AIDS, and the T-cell responses they restrict are immunodominant.

Objective: We sought to identify whether the association of HLA-B alleles with rate of progression toward AIDS is due to targeting of more restricted and thus more conserved regions of the HIV-1 proteome.

Methods: Each residue of the HIV-1 consensus subtype B sequence was coded according to the presence/absence of an epitope, using the compiled epitope data available in the HIV-*LANL* immunology database. The Shannon entropy for each HXB2 position was calculated using pre-aligned HIV-1 clade B sequences as a measure of its degree of conservation. We then compared the entropy of empty versus epitope-containing positions and HLA-B-restricted versus HLA-A-restricted positions.

Results: Positions containing $CD8^+$ epitopes were significantly more conserved than corresponding empty positions. Moreover, residues targeted by HLA-B alleles in the HIV-1 proteome were significantly more conserved than the ones targeted by HLA-A alleles. Analysing a recent dataset, we found that B epitope regions contain significantly more escape mutations and reversions, which might be the reason why we find them to be more conserved.

Conclusion: Our results suggest that epitopes in HIV-1 targeted by HLA-B alleles lie in more constrained regions of its proteins, in which mutations might have a higher fitness cost and tend to revert. Consequently, HLA-B-restricted cytotoxic T-lymphocyte (CTL) responses may persist longer. This may be one of the factors contributing to the immunodominance and impact of HLA-B-restricted CTL responses on disease progression.

AIDS 2010, **24**:211–215

Keywords: conservation, CTL epitopes, database, entropy, genome-wide analysis, HLA, immunodominance

[a]Department of Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, [b]Department of Theoretical Biology/Bioinformatics, Utrecht University, [c]Academic Biomedical Centre, Utrecht University, Utrecht, and [d]Department of Internal Medicine and Infectious Diseases, University Medical Center Utrecht, Heidelberglaan, The Netherlands.

Correspondence to Dr Can Keşmir, Department of Theoretical Biology/Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Tel: +31 30 2534212; fax: +31 30 2513655; e-mail: C.Kesmir@uu.nl

[*] A.I.F.C. and X.R. contributed equally to this work.

Received: 17 July 2009; revised: 7 October 2009; accepted: 15 October 2009.

DOI:10.1097/QAD.0b013e328334442e

Introduction

Cytotoxic T lymphocytes (CTLs) are believed to have a central role in controlling HIV-1 infection (reviewed in [1]). T cells responding to HLA-B-restricted epitopes seem to be immunodominant [2], and, through not yet understood reasons, have a major impact on progression toward AIDS (reviewed in [1]). It seems reasonable to assume that eliciting T-cell responses that are efficient, preserved and not evaded by HIV would be beneficial. These T cells would, thus, target regions of proteins with less mutational flexibility. As such, the mutability of the presented peptides might play a role: if HLA-B alleles present more constrained regions of HIV-1, the corresponding CTL responses may be better maintained and thus standing out as immunodominant. In addition, if escape mutations occur in these constrained epitopes, the fitness costs might be high and set strict limits to growth of the escape mutants, which would have a large impact on the rate of disease progression. In this study, we measured the degree of conservation of HIV residues targeted by HLA-B versus HLA-A alleles, and found that those targeted by HLA-B alleles are indeed more conserved.

Materials and methods

Pre-aligned clade B HIV-1 protein sequences (Gag, Pol, Env, Vif, Tat, Rev, Vpu, Vpr, Nef sequences dated 2007 or older) were downloaded from the *LANL* database (www.hiv.lanl.gov, July 2008). Only one sequence per patient is present in this selection, and recombinant sequences were excluded. Gag, Vpu, and Env were further manually curated. There was no clear bias in the number of sequences per protein that would hamper the calculation of the entropy per position (ranging from 194 sequences for Tat to 824 sequences for Nef). Similarly, the sampling year of the database sequences used for this analysis (ranging from 125 sequences from 1981 to 1985 to 866 sequences from 2001 to 2005) was not different than expected from the progressive increasing number of studies from the beginning of the epidemic. The Shannon entropy [3] at each position i of HIV-1 protein alignments was calculated to measure the conservation in terms of a score S, which is defined as $S = 1 - H$, where H represents the normalized Shannon entropy. Statistical analysis was performed in R package (www.rproject.org).

Results and discussion

Analysis of HIV-1 clade B major histocompatibility complex class I epitopes

The HIV-1 clade B sequence HXB2 has been widely used as B-consensus sequence, and epitopes have been annotated in relation to their relative amino acid position within it. We have downloaded HXB2 protein sequences and defined each residue relative to the epitope it has been reported to contain: unique HLA-A, HLA-B, or HLA-C epitope (positions A, B, or C), or both an A and B epitope (X), or empty (E), using publicly available CTL epitope lists at the *LANL* immunology database (http://www.hiv.lanl.gov/content/immunology, details of this coding schema are explained in the legend of Table 1).

Table 1 summarizes the fraction of residues comprised in HLA-A, HLA-B, and HLA-C epitopes across the total proteome and within each encoded HIV-1 protein. Approximately 41% of the total protein residues do not contain any described epitopes so far. Although p17 (matrix), p24 (capsid), and protease have 5–13% empty positions, the remainder proteins have large 'epitope empty' regions (28–78%).

There is a great variation in epitope density among proteins, which could be due to lower immunogenicity and/or just not being so thoroughly studied. For example, the low fraction of empty positions in p17 (~13%) and p24 (~6%) concurs with the fact that their precursor polyprotein Gag has been intensively studied, both because it is a main target for CTL responses associated with significant reduction in viral load [4–11] and is also highly immunogenic, even in different ethnicities [12]. An additional likely contribution for an under-representation of epitopes in more variable proteins is the general use of peptides derived from consensus sequences to measure responses in in-vitro settings [13,14].

In the total proteome, the fraction of unique A and B positions is equivalent (Table 1, 23.5 versus 23.0%, respectively). Still, Gag-p24 and Nef seem to be preferentially targeted by HLA-B alleles (57.6 and 35.4% of the total protein residues, respectively), the B-fraction being over three-fold higher than the A-residues. These proteins, the former being highly conserved when compared with Nef, have been previously shown to dominate the total HIV-specific response, in both breadth and magnitude [15]. Tat is also preferentially targeted by HLA-B alleles (16.8 versus 5.0% targeted by HLA-A), although this preference may be biased given the large proportion of epitope-free amino acids (57.4%) in this protein. In contrast, more A positions are described for the structural gp160 and the regulatory Rev proteins (25.8 and 28.4% of the total protein residues; two-fold and three-fold higher than B positions, respectively), which are among the most variable proteins in the proteome [16].

Degree of conservation of amino acid positions in clade B HIV-1

In order to assess the degree of conservation in all HIV-1 proteins, pre-aligned sequences from clade

Table 1. Fraction of epitope-free and epitope-coded amino acid positions per protein in the HIV-1 clade B proteome.

Protein	Length (aa)	Percentage A	Percentage B	Percentage C	Percentage X (A and B)	Percentage E	B/A
Gag-p17	132	34.1 (6)	30.3 (6)	0.0	22.7	12.9	0.9
Gag-p24	231	13.9 (4)	57.6 (19)	0.4	22.1	6.1	**4.2**
Gag-p2p7p1p6	137	34.3 (6)	27.7 (5)	0.0	4.4	33.6	0.8
Pol-protease	99	43.4 (6)	33.3 (5)	3.0	15.2	5.1	0.8
Pol-RT + RNaseH	560	28.6 (22)	23.8 (19)	0.2	14.1	33.4	0.8
Pol-integrase	288	13.2 (5)	20.5 (8)	2.8	3.8	59.7	1.6
Vif	192	22.4 (6)	26.0 (7)	0.0	3.1	48.4	1.2
Vpr	96	34.4 (5)	21.9 (3)	0.0	2.1	41.7	0.6
Tat	101	5.0 (1)	16.8 (2)	7.9	12.9	57.4	**3.4**
Rev	116	28.4 (5)	8.6 (1)	7.8	0.0	55.2	*0.3*
Vpu	82	11.0 (1)	0.0	11.0	0.0	78.0	0.0
gp160	856	25.8 (30)	12.1 (15)	5.0	5.0	52.0	*0.5*
Nef	206	9.2 (3)	35.4 (10)	0.5	26.7	28.2	**3.8**
Total proteome	3096	23.5 (100)	23.0 (100)	2.7	10.0	40.8	1.0

The protein sequences of the clade B consensus sequence HXB2 (accession number K03455) were downloaded from the *LANL* database (http://www.hiv.lanl.gov/content/sequence) and each residue was defined according to the epitope(s) it has been reported to contain: unique HLA-A, HLA-B, or HLA-C epitope (positions A, B, or C), or both an A and B epitope (X), or empty (E). Epitopes were retrieved from publicly available CTL epitope lists at the *LANL* immunology database (http://www.hiv.lanl.gov/content/immunology, July 2008). Two CTL epitope lists are available: the best defined CTL epitope list (epitopes with defined optimal length and HLA restriction, verified by several independent research groups) and the CTL epitope summary list (epitopes thus far described in the literature). Every epitope stated in the best defined CTL epitope list with an HLA restriction was used to assign residues of HXB2 proteins to one of the categories above. Epitopes that, although not having been assigned to subtype B or any other, are nonetheless found in the HXB2 protein sequences, were also included (a total of 169 curated epitopes were thus included, as compared to the initial 86). To avoid overestimation of empty positions, every *E* position in the best defined epitope coding was redefined as A, B, C, or X if, in the summary list, it contained an epitope of the corresponding category (following the above-mentioned criteria, yielding a total of 630 curated epitopes versus the initial 520). The under-representation of C epitopes (2.7%, although some are in fact included in the category *x*) most likely reflects the fact that they have been seldomly studied thus far. Highlighted in bold are the proteins in which the fraction of B positions is over three-fold higher than that of A residues (p24, Nef, and Tat), whereas italic highlights the ones with a B/A fraction of 0.5 or less (Rev, gp160). Numbers in parentheses indicate the contribution of each HIV-1 protein to the total fraction of residues that are exclusively targeted by HLA-A or HLA-B alleles. The premature stop codons in Tat and Nef (codons 87 and 124, respectively) were coded in HXB2 as a gap (−), because epitopes have been described beyond them, and thus the remainder of the HXB2 sequence was also epitope-coded. The same criteria were applied to the frame-shift mutation in Vpr (position 5772).

B HIV-1-infected patients available in the *LANL* database were used to calculate the entropy per residue, as described in Materials and methods section. The entropy analysis showed that CTL epitope-free positions in general are significantly more variable than epitope-containing regions (at the whole proteome: $P < 0.001$; at the single protein level for gp160, Nef, p2p7p1p6, and Tat: $P < 0.05$; Mann–Whitney tests). This is in agreement with the findings of Yusim *et al.* [16] that an inverse correlation exists between protein sequence variability and the presence of HIV-specific CTL epitopes. In Rev and Vif, epitope-free positions are more conserved than the rest of the protein ($P < 0.025$).

Focusing on the epitope-containing regions, we found that HLA-B-targeted residues in the HIV-1 proteome are significantly more conserved than residues targeted by HLA-A ($P < 0.01$). The large contribution (see Table 1) of conserved p24 to HLA-B-targeted positions (19% of HLA-B-targeted residues) and of variable gp160 to HLA-A-targeted positions (30% of HLA-A-targeted residues) may partially explain this observation: excluding either p24 or gp160, HLA-B-targeted residues remain more conserved than HLA-A counterparts; however, the difference is no longer significant. Within each HIV-1 protein, the conservation of HLA-A-targeted and HLA-B-targeted regions is not significantly different.

Conservation: lack of selection pressure or being constrained?

The results above do not directly show that HLA-B-targeted regions are more functionally and/or structurally constrained. In fact, one might argue that this lower entropy could reflect that these positions are not under enough selection pressure by CTLs to mutate. Alternatively, the higher degree of conservation of HLA-B-targeted positions can be the local net effect of escape mutations and subsequent reversion. The best way of exploring which of the two scenarios is more likely would be to analyse large-scale transmission data. However, we are not aware of such data being publicly available to date. As an alternative, we analysed data published recently by Wang *et al.* [17]. Briefly, they have analysed near full-length viral genomes from 98 chronically infected individuals and reported 76 HLA class I-associated mutations (within and flanking regions of described and predicted epitopes). These were classified as mutations in the presence (escape) or absence (reversion) of the restricting HLA allele. We analysed the data of Wang *et al.* and found that HLA-B-associated reversions and escapes are significantly

enriched when compared with HLA-A counterparts (reversions: HLA-A = 5; HLA-B = 22; escapes: HLA-A = 6, HLA-B = 26) ($P < 0.01$, χ^2-test; expected values were determined using total A + B positions identified according to our coding). Some of the reported HLA-associated polymorphisms in the study by Wang *et al.* overlap with a verified epitope from another loci and thus cannot be used as HLA-A-specific or HLA-B-specific positions. After correcting for this effect, the number of escapes and reversions associated with HLA-B alleles was still significantly different than expected ($P = 0.002$). These data, together with our finding that HLA-B-targeted positions are more conserved, suggest that HLA-B alleles target more constrained regions of HIV-1 than HLA-A alleles. In line with this, Li *et al.* [18] have illustrated that mutations at conserved sites revert more rapidly, suggesting they might be structurally or functionally constrained and thus impact viral fitness. Escape mutations in epitopes restricted by low-risk hazard HLA-B alleles (B51, B27, and B57) become fixed in the population (Schellens *et al.*, manuscript submitted) and correlate with the prevalence of the corresponding HLA [19]. Taken together, HLA-B-targeted positions thus seem to be under strong selection pressure. However, as they are in constrained regions of the HIV-1 proteome, either HLA-B escape mutations are rapidly converting or becoming fixed in the population (when accompanied with compensatory mutations) and, as a net result, HLA-B epitopes remain more conserved.

To our knowledge, this is the first formal demonstration of a preferential targeting of conserved regions in the HIV-1 proteome by HLA-B alleles. The reason behind why HLA-B molecules target conserved regions is largely unknown. Still, we believe it is not accidental and is partially due to the known binding motifs of HLA-B molecules. For example, less easily mutable amino acids, tryptophan (W) and proline (P), are overexpressed in the HLA-B positions (data not shown). These two amino acids occur almost exclusively in the binding motifs of HLA-B molecules (e.g., B7 and B58 supertypes) [20].

We acknowledge that our analysis is limited to the current epitopes described in the database and we cannot exclude that more epitopes, unidentified to date, may be targeted in the thus far empty regions, as previously illustrated by Schellens *et al.* [21], or that HLA specificities of A and B alleles may overlap in the thus far 'exclusive' A or B positions. In addition, we used database-curated sequences for each protein to determine the entropy at each amino acid position, irrespective of the time after seroconversion. Notwithstanding, the indication that HLA-B alleles target residues that are more constrained to mutate may allow preservation of responses targeting more conserved epitopes and, thus, be one of the factors contributing to the immunodominance of HLA-B-restricted CTL responses and their stronger/greater impact on disease progression.

Acknowledgements

The work was supported by High Potential grant (2006) from Utrecht University.

There are no conflicts of interest.

References

1. Goulder PJ, Watkins DI. **Impact of MHC class I diversity on immune control of immunodeficiency virus replication.** *Nat Rev Immunol* 2008; **8**:619–630.
2. Bihl F, Frahm N, Di GL, Sidney J, John M, Yusim K, *et al*. **Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses.** *J Immunol* 2006; **176**:4094–4101.
3. Shannon CE. **A mathematical theory of communication.** *Bell System Tech J* 1948; **27**:379–423.
4. Borghans JA, Molgaard A, de Boer RJ, Kesmir C. **HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24.** *PLoS ONE* 2007; **2**:e920.
5. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, *et al*. **Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage.** *PLoS Pathog* 2008; **4**:e1000033.
6. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. **Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma.** *J Virol* 2002; **76**:2298–2305.
7. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, *et al*. **Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients.** *J Exp Med* 2008; **205**:1009–1017.
8. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, *et al*. **CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.** *Nat Med* 2007; **13**:46–53.
9. Novitsky V, Gilbert P, Peter T, McLane MF, Gaolekwe S, Rybak N, *et al*. **Association between virus-specific T-cell responses and plasma viral load in human immunodeficiency virus type 1 subtype C infection.** *J Virol* 2003; **77**:882–890.
10. Rolland M, Heckerman D, Deng W, Rousseau CM, Coovadia H, Bishop K, *et al*. **Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads.** *PLoS ONE* 2008; **3**:e1424.
11. Zuniga R, Lucchetti A, Galvan P, Sanchez S, Sanchez C, Hernandez A, *et al*. **Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control.** *J Virol* 2006; **80**:3122–3125.
12. Frahm N, Korber BT, Adams CM, Szinger JJ, Draenert R, Addo MM, *et al*. **Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities.** *J Virol* 2004; **78**:2187–2200.
13. Altfeld M, Addo MM, Shankarappa R, Lee PK, Allen TM, Yu XG, *et al*. **Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences.** *J Virol* 2003; **77**:7330–7340.
14. Betts MR, Ambrozak DR, Douek DC, Bonhoeffer S, Brenchley JM, Casazza JP, *et al*. **Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection.** *J Virol* 2001; **75**:11983–11991.

15. Addo MM, Yu XG, Rathod A, Cohen D, Eldridge RL, Strick D, *et al*. **Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load.** *J Virol* 2003; **77**:2081–2092.
16. Yusim K, Kesmir C, Gaschen B, Addo MM, Altfeld M, Brunak S, *et al*. **Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation.** *J Virol* 2002; **76**:8757–8768.
17. Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, *et al*. **Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1.** *J Virol* 2009; **83**:1845–1855.
18. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, *et al*. **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution.** *J Virol* 2007; **81**:193–201.
19. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, *et al*. **Adaptation of HIV-1 to human leukocyte antigen class I.** *Nature* 2009; **458**:641–645.
20. Rapin N, Hoof I, Lund O, Nielsen M. **MHC motif viewer.** *Immunogenetics* 2008; **60**:759–765.
21. Schellens IM, Kesmir C, Miedema F, van Baarle D, Borghans JA. **An unanticipated lack of consensus cytotoxic T lymphocyte epitopes in HIV-1 databases: the contribution of prediction programs.** *AIDS* 2008; **22**:33–37.