Humans with chimpanzee-like major histocompatibility complex-specificities control HIV-1 infection

Ilka Hoof[a], Can Keşmir[a,b], Ole Lund[a] and Morten Nielsen[a]

Background: Major histocompatibility complex (MHC) class I molecules allow immune surveillance by presenting a snapshot of the intracellular state of a cell to circulating cytotoxic T lymphocytes. The MHC class I alleles of an HIV-1 infected individual strongly influence the level of viremia and the progression rate to AIDS. Chimpanzees control HIV-1 viral replication and develop a chronic infection without progressing to AIDS. A similar course of disease is observed in human long-term non-progressors.

Objective: To investigate if long-term non-progressors and chimpanzees have functional similarities in their MHC class I repertoire.

Methods: We compared the specificity of groups of human MHC molecules associated with different levels of viremia in HIV-1 infected individuals with those of chimpanzee.

Results and conclusion: We demonstrate that human MHC with control of HIV-1 viral load share binding motifs with chimpanzee MHC. Moreover, we find that chimpanzee and human MHC associated with low viral load are predicted to elicit broader Gag-specific immune responses than human MHC associated with high viral load, thus supporting earlier findings that Gag-specific immune responses are essential for HIV-1 control.

AIDS 2008, **22**:1299–1303

Keywords: HIV-1, human leukocyte antigen, major histocompatibility complex, *Pan troglodytes*, viral load

Introduction

Chimpanzees (*Pan troglodytes*) are, unlike the majority of the human population, able to fight the closely related lentiviruses HIV-1 and SIVcpz. They keep viral replication under control and develop a chronic infection without progressing to AIDS [1], a characteristic they share with human long-term non-progressors (LTNP). A key role in controlling viral replication has been assigned to HIV-1-specific cytotoxic T lymphocytes (CTL) [2]. More specifically, the major histocompatibility complex (MHC) molecules, in humans known as the human leukocyte antigen (HLA) molecules, have been shown to influence the progression rate to AIDS of HIV-1 infected individuals [3–5]. In line with this, the relative resistance of chimpanzees to AIDS has been proposed to be due to a

From the [a]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark, and the [b]Department of Theoretical Biology/Bioinformatics and the Academic Biomedical Centre, Utrecht University, The Netherlands.

Correspondence to Ilka Hoof, Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Building 208, Lyngby 2800, Denmark.

Tel: +45 45 25 61 27; fax: +45 45 93 15 85; e-mail: ilka@cbs.dtu.dk
Received: 31 January 2008; revised: 26 March 2008; accepted: 1 April 2008.

selective sweep in the MHC class I gene repertoire caused by an HIV-like pandemic in the distant past [6]. Various studies that address the mechanism behind the successful control of HIV have reported targeting of Gag to be associated with viral control, both in HIV-1 clade B and clade C infection [7–12]. In particular, increasing the breadth of Gag-specific responses has been associated with lowering of viremia [7,10–12].

Interestingly, HIV-1 infected chimpanzees target two conserved HIV-1 epitopes that are also recognized by human MHC class I molecules associated with LTNP [13]. Here, we propose that human MHC class I molecules that have been associated with control of HIV-1 viral load share binding specificities with chimpanzee MHC class I molecules. We support this by systematically comparing the binding motifs of a wide range of human and chimpanzee MHC class I molecules. We use data of Kiepiela *et al.* [14] who performed a thorough analysis of HLA alleles present in an HIV clade C infected southern African population. They determined sets of HLA alleles that were significantly associated with high or low viral load. The plasma viral load is a commonly accepted predictor for the rate of HIV-1 disease progression, in which a low viral load infers a delay in the onset of AIDS [15]. To address the potential mechanism behind the difference in viral control of different MHC class I molecules, we compared their preference for presenting Gag-epitopes.

Methods

Major histocompatibility complex class I protein sequences

The protein sequences of 1001 HLA molecules (355 HLA-A and 646 HLA-B) were obtained from the Anthony Nolan HLA sequence database (http://www.anthonyno lan.org.uk/HIG/data.html), 66 chimpanzee (Patr) MHC class I molecules were downloaded from the IPD-MHC database (http://www.ebi.ac.uk/ipd), and 68 rhesus macaque (Mamu) MHC class I molecules were retrieved from the Entrez Protein database (http://www.ncbi.nlm. nih.gov/sites/entrez?db=protein).

Combining human leukocyte antigen–HIV-association studies

On the basis of the Kiepiela study and additional studies, which involved HIV clade C infected southern African populations [11,16–18], we defined three groups of HLA alleles: those associated with low viral load (lowVL), those associated with high viral load (highVL) and a group including all the remaining HLA alleles (other): lowVL comprises B*5703 [10,14], HLA-B*0702 [11], B*4201, B*5801, B*8101 [14], B*1516, B*1517 [16], B*1302 [17], B*1301, B*3901, and B*3902 [18]; highVL consists of B*1801 [10,14], B*1503, B*5301 [11], B*4501, and B*5802 [14].

NetMHCpan method

The NetMHCpan method employs artificial neural networks to generate quantitative predictions of peptide-MHC class I binding affinity [19]. The networks not only consider the peptide sequence but also include the primary sequence information of the MHC molecule. By this, NetMHCpan is able to extrapolate with considerable accuracy the binding specificity of MHC class I molecules that have not been included in the training process. The performance of the method has been thoroughly tested for an extensive number of HLA-A and HLA-B alleles [19]. To establish its predictive performance for chimpanzee (Patr) and rhesus macaque (Mamu) MHC class I alleles, we applied the method to eight Patr and six Mamu alleles described in the IEDB database (www.immuneepitope.org) [20]. For 11 out of 14 alleles, the performance was significant with P less than 0.05 (data not shown). These results justify the application of NetMHCpan to predict binding affinities for chimpanzee and rhesus macaque MHC class I alleles.

Correlation distance and nearest neighbor

The distance d between the binding motifs of two alleles is defined as $d = 1 - P_{\text{corr}}$ where P_{corr} is the Pearson's correlation between the predicted binding affinities of a set of 10 000 random natural peptides for the two alleles. The correlation is close to one if the two alleles share common binding motifs and close to zero if the motifs are dissimilar. The nearest Patr neighbor of a given HLA allele is the Patr allele with the smallest correlation distance.

Results

Protective human leukocyte antigen molecules share peptide binding specificities with chimpanzee MHC class I molecules

Using the NetMHCpan method [19], we investigated if any of the HLA allele groups associated with different levels of viral load defined in Methods would share binding motif similarities with chimpanzee MHC molecules. In this context, we chose MHC molecules of rhesus macaque (*Macaca mulatta*) as a negative control, as rhesus macaques are not naturally infected with an SIV strain and, thus, have not been under selective pressure induced by an immunodeficiency virus. For each HLA allele, we determined the closest chimpanzee (Patr) and macaque (Mamu) MHC neighbors in terms of binding specificity. Comparing the distances of all HLA alleles with their closest Patr neighbor, the lowVL alleles were found to have a significantly smaller distance to Patr alleles than highVL alleles ($P = 0.035$) (see Fig. 1a). In addition, the group of other HLA alleles shows a significantly larger

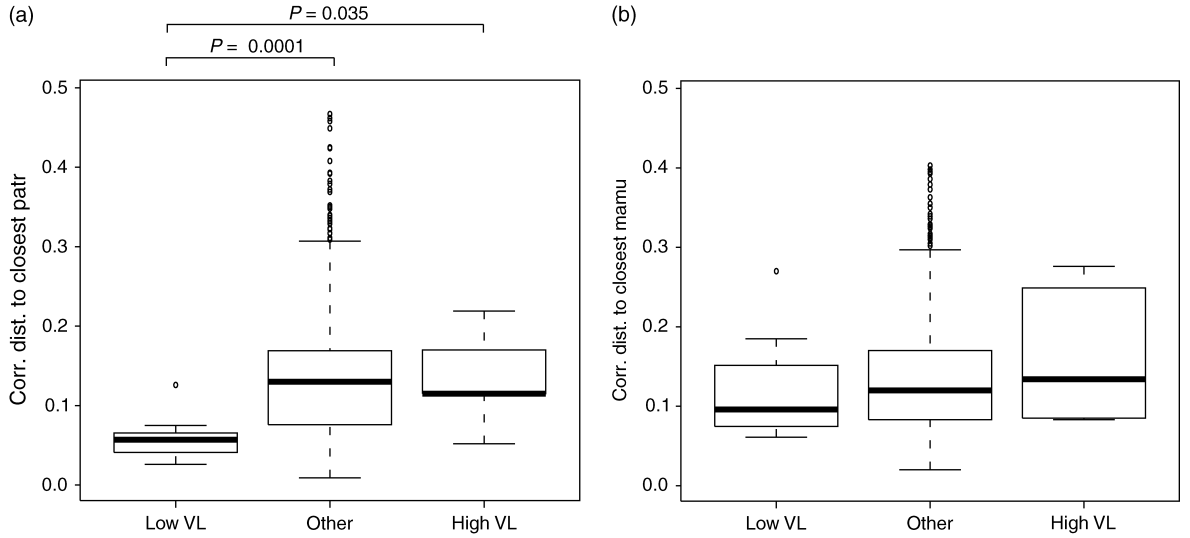

Fig. 1. Comparison of the distances between HLA alleles and (a) chimpanzee MHC class I (Patr) and (b) rhesus macaque MHC class I (Mamu) neighbors. The distances were calculated as described in Methods. The boxplot shows the distribution of the correlation distances of the HLA alleles associated with low viral load (lowVL, $n = 11$), high viral load (highVL, $n = 5$) and all other HLA alleles (other, $n = 985$) to the closest among (a) 66 Patr and (b) 68 Mamu alleles. *P*-values were derived from two-tailed Mann–Whitney tests. Only significant *P*-values ($\alpha = 0.05$) are displayed.

distance to Patr than the alleles in the lowVL group ($P = 0.0001$). Limiting the distance analysis to nine common Patr alleles [21] gave similar results. Conducting the same analysis for Mamu alleles did not result in a significant difference between either two of the three groups (see Fig. 1b). Comparison of the MHC protein sequences and phylogenetic reconstruction (data not shown) revealed that the strong similarity in binding specificity for lowVL human and chimpanzee MHC molecule pairs did not result from close phylogenetic relationships. Instead, the similarity is limited to the functionality of the MHC molecules in that they present similar sets of peptides to their respective immune systems.

Protective human leukocyte antigen molecules show broader targeting of the HIV-1 Gag protein than human leukocyte antigen molecules associated with high viral load

To investigate whether the defined groups of MHC class I alleles show differences in their preference to target the HIV-1 Gag protein, we applied NetMHCpan for each HLA and Patr allele in our dataset to all peptides in the HIV clade C proteome. We obtained the HIV clade C consensus proteome from the Los Alamos HIV sequence database (http://www.hiv.lanl.gov/content/hiv-db, August 2007). In general, we expect only 0.5–3% of a set of peptides to bind a certain MHC class I molecule [22,23]. For each HLA and Patr allele, we ranked the 3059 nonamers in the HIV clade C consensus proteome with respect to their predicted affinity scores and counted the number of Gag-derived peptides among the top-ranking 1% of all peptides. Patr and lowVL alleles were predicted to target a significantly higher number of Gag peptides than highVL alleles ($P < 0.02$, see Fig. 2). To study SIVcpz-Gag targeting by Patr, we repeated this analysis for the proteomes of three SIVcpz strains (accession numbers U42720, AF382828, AY169968). Also here, the number of SIVcpz-Gag peptides targeted by Patr is significantly higher than the number of top-ranking HIV-Gag peptides predicted for highVL ($P = 0.01$).

Discussion

Contemporary chimpanzees possess a reduced MHC class I repertoire [6], but the hypothesis that an HIV-like pandemic might have caused this repertoire reduction has been subject to debate [24]. HIV-specific CTL responses have been reported to play a crucial role in the immune response to HIV/SIV. This and the fact that chimpanzees are relatively resistant to the development of AIDS following SIV/HIV infection suggest that the chimpanzee MHC class I repertoire protects contemporary chimpanzees from progression to AIDS. This is strongly supported by our finding that protective HLA molecules show a significantly higher similarity in binding specificity to chimpanzee MHC class I molecules than all other HLA molecules. Our work extends the study of Balla-Jhagjhoorsingh *et al.* [13] who showed that chimpanzee alleles and HLA alleles associated with long-term non-progression to AIDS share conserved Gag epitopes. Their study is based on two chimpanzee alleles, two HLA alleles and two epitopes, whereas we here

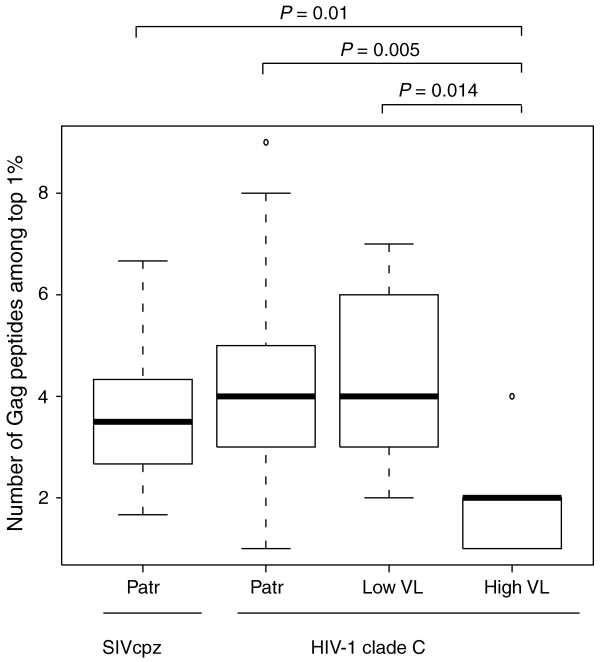

Fig. 2. Targeting of SIVcpz Gag and HIV-1 Gag. We predicted the binding affinity of 66 Patr alleles to all 9-mer peptides in the proteomes of three SIVcpz strains and counted the number of Gag-peptides that fall within the 1% top-scoring fraction. The mean over the three Gag-counts for each Patr allele was calculated and used to generate the left-most boxplot. The three boxplots on the right show the number of Gag peptides targeted in the HIV-1 clade C consensus proteome by chimpanzee MHC (Patr: $n = 66$), HLA alleles associated with low viral load (lowVL, $n = 11$), and high viral load (highVL, $n = 5$). *P*-values were derived from two-tailed Mann–Whitney tests.

confirm their finding in an exhaustive set of HLA-A, HLA-B and chimpanzee alleles. A possible limitation of our study might be that the Patr sequences were taken from chimpanzees in captivity, and may therefore not represent the MHC diversity in wild animals.

The predicted similarity in binding specificity between chimpanzee and protective HLA molecules clearly supports the hypothesis that the MHC class I repertoire of chimpanzees has been formed under selective pressure induced by an HIV-like virus. Along these lines, the results of our work agree with the assumption that pathogens influence the evolution of the host towards the selection of the most protective genotypes. If the MHC class I repertoire of contemporary chimpanzees has been formed under selective pressure induced by an HIV-like virus, we could expect a similar evolution in human populations with a high incidence of HIV-1 infection and when patients do not have access to antiretroviral drug therapy. As a matter of fact, the first signs of a selection are appearing in human populations in whom protective alleles are more prevalent in infected pregnant women than in infected infants, suggesting that protective alleles will become more prevalent in future generations [14].

What is the mechanism behind the different impact of certain MHC class I alleles on HIV-1 viral load? The HIV-1 Gag protein has repeatedly been reported to play an important role as an immunogen to induce an effective immune response against HIV-1. In several studies [7,10,11,25], the number of Gag-epitopes targeted was found to be inversely correlated with plasma viral load. A potential reason for this is the high fitness cost of CTL escape mutations [26]. In this study, we determined for each HLA and chimpanzee allele in our data set its preference to bind Gag-derived peptides. In agreement with previous studies on the effect of targeting Gag, we find that chimpanzee and protective HLA molecules are predicted to elicit broader Gag-specific immune responses than HLA molecules that have been associated with high plasma viral load.

The NetMHCpan method makes it possible to analyze the binding specificity of hitherto uncharacterized HLA class I molecules as well as the specificity of chimpanzee and rhesus macaque MHC class I molecules. The method is especially useful to study the binding motifs of those HLA molecules that are prevalent among southern African populations, in particular those HLA molecules that have been shown to be associated with extreme (high or low) levels of plasma viral load during HIV-1 infection. Up until now this has not been possible, as the set of experimentally characterized alleles is strongly biased towards alleles that are prevalent in Caucasian populations. Application of NetMHCpan to predict peptide binding affinities to non-human primate MHC class I alleles enabled us to compare the binding motifs of HLA alleles with those of chimpanzee and rhesus macaque alleles.

To conclude, our results indicate that the binding motifs of protective HLA and chimpanzee molecules are directed towards peptides able to elicit effective CD8 T-cell responses against HIV-1 and that similarity to chimpanzee MHC specificity plays a central role in controlling HIV-1 viral load in humans.

Acknowledgements

We thank Rob de Boer for valuable discussions and Bianca Mothé for the collection of rhesus macaque MHC class I sequences.

The present work was supported by NIAID Contract no. HHSN266200400083C.

I.H. performed the data analysis and wrote the manuscript; C.K. formulated the biological idea for the study;

I.H., C.K. and M.N. designed the study; all authors provided critical input to the data analysis and reviewed the manuscript.

The authors have no conflict of interest.

References

1. Bontrop RE, Watkins DI. **MHC polymorphism: AIDS susceptibility in nonhuman primates.** *Trends Immunol* 2005; **26**:227–233.
2. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, *et al*. **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nat Med* 1997; **3**:205–211.
3. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, *et al*. **Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection.** *Nat Med* 1996; **2**:405–411.
4. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, *et al*. **HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage.** *Science* 1999; **283**:1748–1752.
5. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, *et al*. **Advantage of rare HLA supertype in HIV disease progression.** *Nat Med* 2003; **9**:928–935.
6. de Groot NG, Otting N, Doxiadis GGM, Balla-Jhagjhoorsingh SS, Heeney JL, van Rood JJ, *et al*. **Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees.** *Proc Natl Acad Sci U S A* 2002; **99**:11748–11753.
7. 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.
8. 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.
9. Masemola A, Mashishi T, Khoury G, Mohube P, Mokgotho P, Vardas E, *et al*. **Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8+ T cells: correlation with viral load.** *J Virol* 2004; **78**:3233–3243.
10. 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.
11. Geldmacher C, Currier JR, Herrmann E, Haule A, Kuta E, McCutchan F, *et al*. **CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients.** *J Virol* 2007; **81**:2440–2448.
12. 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.
13. Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, Haaksma TG, Teeuwsen VJ, Bontrop RE, *et al*. **Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees.** *J Immunol* 1999; **162**:2308–2314.
14. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, *et al*. **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** *Nature* 2004; **432**:769–775.
15. Mellors JW, Rinaldo CR Jr, Gupta P, White RM, Todd JA, Kingsley LA. **Prognosis in HIV-1 infection predicted by the quantity of virus in plasma.** *Science* 1996; **272**:1167–1170.
16. Frahm N, Adams S, Kiepiela P, Linde CH, Hewitt HS, Lichterfeld M, *et al*. **HLA-B63 presents HLA-B57/B58-restricted cytotoxic T-lymphocyte epitopes and is associated with low human immunodeficiency virus load.** *J Virol* 2005; **79**:10218–10225.
17. Honeyborne I, Prendergast A, Pereyra F, Leslie A, Crawford H, Payne R, *et al*. **Control of human immunodeficiency virus type 1 is associated with HLA-B*13 and targeting of multiple gag-specific CD8+ T-cell epitopes.** *J Virol* 2007; **81**:3667–3672.
18. Tang J, Tang S, Lobashevsky E, Myracle AD, Fideli U, Aldrovandi G, *et al*. **Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1.** *J Virol* 2002; **76**:8276–8284.
19. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, *et al*. **NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence.** *PLoS ONE* 2007; **2**:e796.
20. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, *et al*. **The design and implementation of the immune epitope database and analysis resource.** *Immunogenetics* 2005; **57**:326–336.
21. Sidney J, Asabe S, Peters B, Purton KA, Chung J, Pencille TJ, *et al*. **Detailed characterization of the peptide binding specificity of five common Patr class I MHC molecules.** *Immunogenetics* 2006; **58**:559–570.
22. Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui HH, Frahm N, *et al*. **A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection.** *J Immunol* 2007; **178**:7890–7901.
23. Yewdell JW, Bennink JR. **Mechanisms of viral interference with MHC class I antigen processing and presentation.** *Annu Rev Cell Dev Biol* 1999; **15**:579–606.
24. Sharp PM, Shaw GM, Hahn BH. **Simian immunodeficiency virus infection of chimpanzees.** *J Virol* 2005; **79**:3891–3902.
25. Borghans JAM, 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.
26. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, *et al*. **Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication.** *J Virol* 2007; **81**:12382–12393.