Features that shape $CD8^+$ T-cell responses to viruses

Ana Isabel Correia de Almeida Fontaine Costa

The cover is a remembrance of my love for my homeland, and my love for science. It is a drawing of the *Padrão dos Descobrimentos* in Lisbon, a monument that celebrates the Portuguese discoveries and evokes those involved; elements of science have been 'embedded' in it. I do believe that in both science and the discoveries, the courage to dream and try to go beyond is commendable.

ISBN: 978-90-393-6211-2

Layout: Ana Costa

Cover: Doctus Design - Filipa Loureiro

Printed by: Off Page, Amsterdam, The Netherlands

Features that shape $CD8^+$ T-cell responses to viruses

Factoren die $CD8^+$ T-cel responsen tegen virussen bepalen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties in het openbaar te
verdedigen op woensdag 15 oktober 2014 des morgens te 10.30 uur

door

Ana Isabel Correia de Almeida Fontaine Costa

geboren op 5 november 1979 te Viseu, Portugal

Promotor: Prof. dr. F. Miedema

Copromotor: Dr. D. van Baarle

To my family,
the one I left behind
And the one I built here.

Leescommissie: Prof. dr. P. de Bakker
Prof. dr. R.J. de Boer
Prof. dr. R.E. Bontrop
Prof. dr. L. Meyaard
Prof. dr. R.E.M. Toes

Paranimfen: Sabrina Oliveira
Hilde Spits

The work presented in this thesis was financially supported by a High Potential (HiPo) grant from Utrecht University, and by a fellowship (A. Costa) from the Fundação para a Ciência e Tecnologia (FCT, Portugal).

CONTENTS

1

General Introduction

General Introduction

T-cell responses are believed to be important in controlling viral replication, contributing significantly to a milder course of disease in acute infections, and delayed progression in persistent infections. To this end, two general pre-requisites must be met: the immune system of the host must have a repertoire of T cells capable of detecting and destroying virally infected cells, and infected cells must signal the abnormal state out by expression of pathogen-derived fragments at the cell surface. Central to this interaction is the major histocompatibility complex (MHC) gene family, the host genes most often associated with disease outcome. In humans, this complex is called the human leukocyte antigen (HLA) (reviewed in (1)). The HLA class I gene complex in particular is critical for the immune response against intracellular pathogens, and encodes a plethora of genes. The classical class I HLA molecules associate with β2-microglobulin and are exhibited on the surface of every nucleated cell and routinely present cytosolic peptides to cytotoxic T lymphocytes (CTLs) surveying the organism.

The general process of degradation of cytosolic proteins (Fig. 1, reviewed in (2-6) is initiated by cleavage in the proteasome, and is followed by N-terminus trimming of the cleaved peptides by cytosolic aminopeptidases. Peptides that are translocated into the endoplasmic reticulum (ER) by ATP-dependent transporter-associated proteins (TAP1 and TAP2) may further be trimmed by ER aminopeptidases (ERAP1 and 2). Ultimately, in the ER, calnexin chaperones the association of the class I heavy chains with β2-microglobulin, and the MHC class I peptide loading complex (including calreticulin, ERp57, TAP, and tapasin) is assembled after the release of calnexin. In the ER, peptides will be loaded onto the newly synthesized HLA molecules, and after editing and "quality control", HLA-peptide complexes proceed to the Golgi for HLA sialylation before they reach the cell surface.

Upon infection, self-peptides are replaced by intracellular pathogen-derived peptides at the surface of infected cells, and their recognition by cognate CTLs will target the cell for destruction. Pathogen-derived peptides are believed to rarely overlap with self-peptides (7, 8), and the immunoproteasome, constitutively expressed in infected cells and induced upon inflammatory cues (e.g. IFN-γ) in other cell types, seems to contribute to the generation of distinct pathogen-derived peptides (9-12).

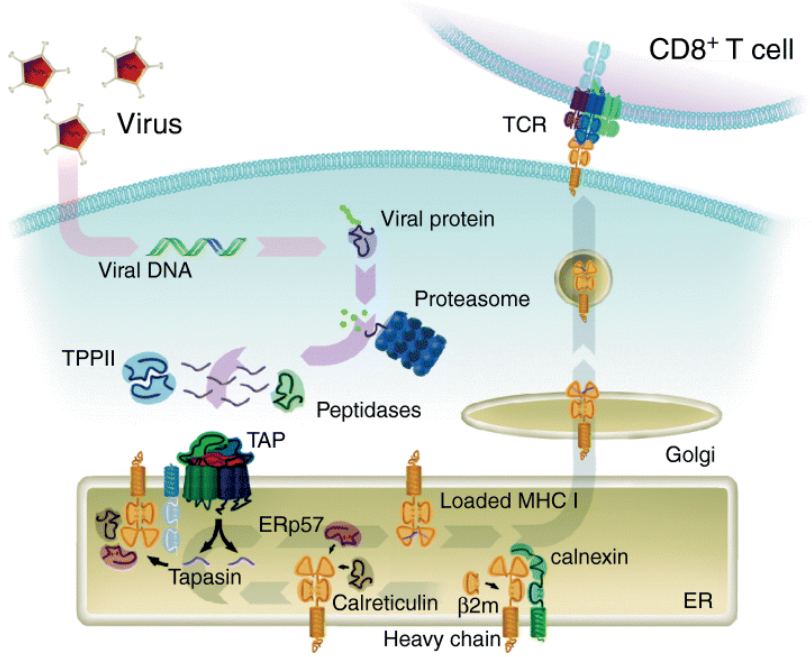


Figure 1. Classical pathway of major histocompatibility complex (MHC) class I antigen presentation.
MHC class I molecules are assembled in the endoplasmic reticulum (ER) supported by the chaperones calnexin, calreticulin, and ERp57, after which they dock onto the ER-resident peptide transporter associated with antigen processing (TAP). This docking is facilitated by the specialized chaperone tapasin. TAP pumps peptides into the ER lumen. These peptides are from cellular or viral origin and are produced in the cytosol/nucleus by the proteasome, tripeptidyl peptidase II (TPPII), and other peptidases. Once in the ER, peptides can bind to MHC class I molecules that are subsequently released form the TAP-tapasin loading complex. MHC class I-peptide complexes can then leave the ER for transport to the plasma membrane. Here, they can be inspected by the T-cell receptor of $CD8^+$ T cells. This figure and legend were reproduced from Groothuis, T.A.M. et al., (2005) MHC class I alleles and their exploration of the antigen-processing machinery. *Immunological Reviews* 207: 60–76, Copyright © Blackwell Munksgaard 2005, with permission from **John Wiley & Sons**. (2)

Pathogens, and viruses in particular, have evolved mechanisms to survive the counter-attack of the immune system by interfering with the antigen presentation pathway (13, 14) or, in the case of high mutation-rate viruses, by selecting mutant variants whose initial targeted epitopes are no longer processed, presented, or recognized by the cytotoxic T cell pool in their host (15-20). This in turn forces the host to evolve mechanisms to prevent the immune escape by the pathogens, resulting in one of the most amazing examples of co-evolution in biology.

The Human Leukocyte Antigen region

The HLA complex is located on chromosome 6 (The MHC sequence consortium (1999), quoted in (21)). The HLA region codes for class I, class II, and class III subgroup genes. A simplified representation of this region is shown in Fig. 2. In addition to the TAP genes, the class II region encodes the α and β chains constituting the polymorphic HLA-class II heterodimers (e.g., HLA-DR). Class II molecules are expressed in antigen-presenting cells and bind exogenous peptides processed via the endocytic pathway, and present these to CD4+ T cells. The class III region contains, among other immune genes, complement factor genes (e.g., C4, C2) and cytokine genes (e.g., TNF).

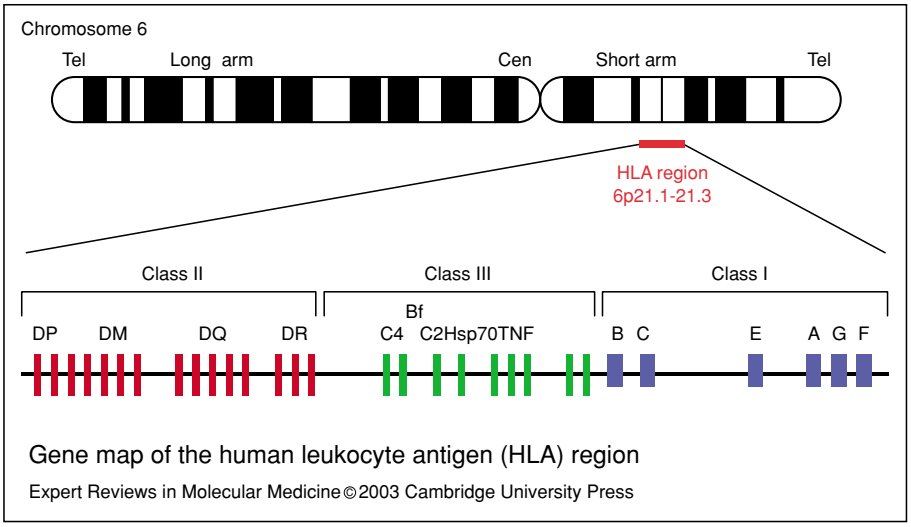


Figure 2. Gene map of the HLA region.
Schematic representation of the HLA complex on chromosome 6. Bars mark the localization of the major genes. This figure was reproduced from Mehra N.K. and Kaur G. (2003) MHC based vaccination approaches: progress and perspectives. *Expert Reviews in Molecular Medicine* 5(07): 1–17, with permission from Cambridge University Press (22).

Finally, the class I region encodes the non-classical genes (i.e., HLA-E, F, and G with limited polymorphism and restricted tissue expression), and the classical (major) loci (i.e., HLA-A, B, and C). The classical class I loci are highly polymorphic; i.e., multiple variants (alleles) of each gene have been described within the population. The distribution of the actual allele numbers per locus can be seen in Figure 3.

This high polymorphism of the three major class I HLA loci has been studied extensively (reviewed in (1, 23)). It has been shown that in the classical HLA class I genes, the rate of nonsynonymous substitution is greater at the contact amino acids of

the peptide binding site than in the remainder of the gene regions ((24, 25), reviewed in (1)), suggesting that the HLA molecules are mainly evolving to become different in the peptides they are presenting. Most importantly, Prugnolle and colleagues have reported that diversity at the HLA class I genes in a population correlates positively with local pathogen richness (26). The large polymorphism of HLA class I molecules has been explained in the light of two, not necessarily exclusive, theories that agree on pathogen-induced selection pressures as the major ground for evolution (reviewed in (1, 23, 27)). The **over-dominant selection theory**, or heterozygote advantage, relies on the fact that HLA genes are co-dominantly expressed, which means that an individual having 6 distinct HLA class I molecules will be able to present a larger variety of peptides, and elicit a broader CTL response, than a subject having less molecules. Heterozygosity in HLA molecules thus ensures a superior immune surveillance. This was first reported over thirty years ago in H-2 heterozygous mice (28). The reported associations of homozygosity at the class I HLA loci with rapid progression to acquired immunodeficiency syndrome (AIDS) (29, 30) support the importance of heterozygote advantage.

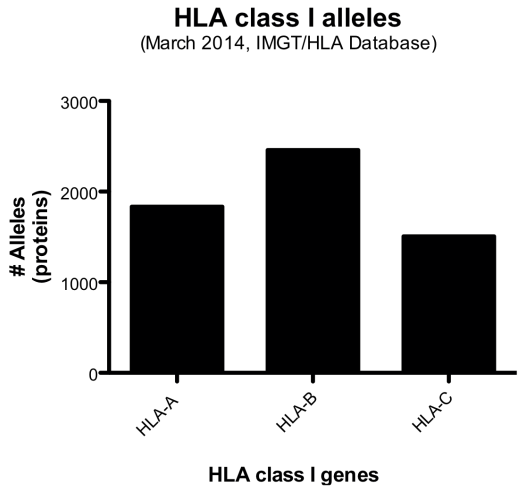


Figure 3. Distribution of the current number of described alleles (proteins) over the three major HLA class I loci.
Number of described alleles over the HLA-A, B and C loci (n=1833, 2459 and 1507, respectively), as reported in the IMGT database (http://www.ebi.ac.uk/imgt/hla/stats.html) in March 2014.

However, it was suggested that heterozygous advantage on its own is not sufficient to explain the large polymorphism observed in HLA genes when considering the different fitness contributions of HLA alleles (31). In addition, a more recent study by Leslie and coworkers (32) did not find a homozygosity disadvantage during human

immunodeficiency virus-1 (HIV-1) infection. Alternatively, the **frequency-dependent selection theory**, or rare allele advantage, suggests that pathogens are evolving to avoid the presentation of their targeted epitopes by the most prevalent HLA alleles in the host population, therefore leading to selection for hosts with rare HLA molecules (33, 34). It has actually been reported that HIV adapts to the most frequent alleles in humans (34, 35). An example of how the allele frequency in a population affects pathogenic adaption is the following. As a rare allele in a cohort infected by clade-B virus, B*1503 was associated with reduced viral load and thus seemed to be protective, although it only mediates subdominant responses (36). B*1503 lost its protective effect in a cohort of clade-C-infected individuals, a population in which this allele is present in high frequencies. Here, the differences in the virus-sequence, and consequently in the epitopes presented, are likely to play an essential role. Nonetheless, one would expect that even in the responses restricted by rare alleles, adaptation would ensue over time. Accordingly, the dynamics of HIV-1 evolution and the impact of HLA restriction have been illustrated in a study reporting the adaptation of HIV-1 to less common alleles that were thus-far associated with a low relative hazard for disease progression to AIDS (37). These examples illustrate the dynamic interplay between HLA restriction, the effectiveness of the responses elicited, and HIV-1 evolution/adaptation. In the absence of therapy and other prevention of virus transmission strategies, the rare-frequency alleles should be selected for, and subsequently increase in frequency in the general population, as has been suggested for the non-human primates (38). Several reported studies provide evidence that parasitic infections can also exert selective pressures on HLA evolution. For example, a previously rare allele in the African population, Bw53, which is associated with protection from severe malaria, was observed to increase in prevalence in those regions where malaria incidence is high (39). The same change in HLA frequencies is thought to have happened in the Amerindian tribes of South America upon the challenge of pathogens like *Trypanosoma cruzi* and *Leishmania braziliensis* (40). Taken together, the few examples listed here, and many more to be found in the literature demonstrate clearly that the large HLA diversity in the human population resulted from a long period of co-evolution with pathogens (41).

HLA-B: the most diverse HLA locus

The different HLA class I loci are believed to have arisen by gene duplications. The HLA-C locus is thought to be the youngest because no orthologue has been found in the rhesus monkeys, in contrast to HLA-A and –B (42). Most likely, HLA-B and –C were

duplicated after the divergence between apes and Old World monkeys (reviewed in (25, 43)). Curiously, even though HLA-B seems to be younger than HLA-A, it has a higher degree of polymorphism (Fig. 3). Actually, Mungall and colleagues have reported that HLA–B is the most polymorphic locus in the whole human genome (21), and is definitely the gene that is most rapidly evolving and diversifying in the MHC complex (44). For example, studies of isolated Amerindian populations have shown that novel recombinant B alleles (in relation to the alleles of the founding populations) exist in larger variety than novel A and C alleles ((40, 45), reviewed in (27)). This is probably due to the fact that the HLA-B locus possesses an exceptionally high level of recombination and therefore evolves differently from HLA-A, which mainly evolves via point mutations (46). Additionally, the elevated reassortment rate in the α2 chain, which does not alter the basic peptide-binding motifs but favours presentation of novel epitopes, is a possible factor contributing to the large allelic diversity in the HLA-B locus (47). The positive correlation between diversity of class I genes and local pathogen richness is most significant for HLA-B, which suggests a higher pathogen selective pressure on this locus (26). Indeed, HLA-B alleles have been associated with opposing disease outcomes, namely susceptibility to, or protection from, parasitic (39) and viral infections, the most striking cases being described for HIV-1.

Several studies have highlighted that, in HIV-1 infection, particular HLA-B alleles are protective (e.g.: HLA-B*27; B*57; B*5801), whereas others are related to a swift progression towards AIDS (e.g.: B*35, B*5301, B*5802) (reviewed in (48-50). HLA-B alleles influence the establishment of viral set-points, and are also most strongly associated with variation in absolute $CD4^+$ T-cell counts and, consequently, with the rate of progression to AIDS (32, 44, 51). Most importantly, a recent genome-wide association analysis in a cohort of HIV-1-infected progressors and controllers has shown that the strongest associations with significant differences in viral control were attributed to allelic variations within the peptide binding groove of HLA molecules, and most of these were located in HLA-B molecules (52). The strongest associations of HIV-1 inter-clade amino acid variability with sites of HLA-imposed selection pressure – "HLA footprints" – have been attributed to HLA-B alleles, especially those associated with viral control, e.g. B*57 (53). At the population level, the impact of HLA-B alleles on the epidemic was supported by the loss of CTL epitopes restricted by the protective HLA-B alleles, and not by HLA-A (37). This data suggests that HIV-1 will shape HLA-B allele frequencies in the future because of its strong selection pressure especially in Africa. Chimpanzees, a species naturally resistant to development of AIDS upon infection with SIV^{cpz} (simian immunodeficiency virus), the "parental" virus of HIV-1, were suggested to have gone through such a selective sweep by an ancestral HIV-like retrovirus (54).

The repertoire reduction is most pronounced in the Patr-B locus (the chimpanzee HLA-B homologue in the contemporary chimpanzee population), which is likely due to death of hosts carrying other alleles during a SIV^{cpz} epidemic, and illustrates the importance of this MHC locus in HIV-1 control (reviewed in (55, 56). Interestingly, HLA alleles associated with low viral load in HIV long-term non-progressors (LTNPs) target identical epitopes (57, 58), and share binding motifs with the chimpanzee MHC (59, 60).

Likewise, in hepatitis C virus (HCV), most HLA class I associations with disease outcome are with HLA-B. Associations between HLA-B*27 and B*57 and HCV clearance have been described in several cohorts, while B*08, Cw*04, and B*18 are associated with viral persistence (61-65). However, the role of HLA-B alleles in the outcome of other viral infections is not consistent across studies. Examples of associations between disease outcomes and HLA-A or C alleles (e.g., A*02 (and Cw*08) with protection, and B*5401 as a detrimental allele in HTLV infection (66)), as well as examples of viruses in which the HLA-A alleles take the main role (e.g., EBV (67, 68) and Dengue virus (69-71)) have been reported.

HLA-B-restricted $CD8^+$ T-cell responses and immunodominance

A likely mechanism explaining why HLA-B alleles tend to shape the infectious disease outcome might lie in their tendency to evoke immunodominant T-cell responses. An immunodominant T-cell response has been defined both as i) a response directed against an epitope that is frequently targeted/recognized at the population level, and as ii) a response of superior magnitude when compared to other responses in the same individual. Immunodominant responses have received most attention because they are believed to be associated with immune control of infections, and thus correlate with protection (e.g., the significant correlation between the magnitude of the T-cell response and HIV-1 control was demonstrated in (72, 73)).

In HIV-1 infection, HLA-B-restricted $CD8^+$ T-cell responses of C-clade infected individuals from southern Africa were significantly more frequently detected than the HLA-A restricted responses (44). Bihl and co-workers confirmed this finding, and added that the HLA-B responses were of higher magnitude (74). In fact, an older study had already shown the superiority of HLA-B57-restricted responses in acute infection: they were broader and stronger than those restricted by all other co-expressed class I alleles combined (75). This was concomitant with a lower incidence of symptomatic acute infection, suggestive of the protective effect of the elicited responses.

Upon HCV infection, spontaneous viral clearance is associated with the expression of HLA-B*27. An NS5-derived (non-structural protein 5) epitope was shown to be frequently recognized and elicit a high magnitude response, as well as to frequently bear mutations in $B27^+$ individuals (64, 76), thus highlighting the pressure induced by this dominant response. Finally, mutations in HCV epitopes restricted by B*07, B*25, B*37 (77), and B*08 (with reversion upon transmission to $B8^-$ individuals) (78, 79) stand as other indications of a strong B-allele-specific selection pressure. For HLA-A and –C alleles conflicting results were obtained with respect to their association with HCV infection outcome. For example, HLA-A*03 is suggested to be protective in the Irish population but is associated with chronic HCV infection in West India and Korea (64, 80).

For influenza infection, there is also no consensus. Although a dominant HLA-A*02 restricted matrix M1.58-specific response has been reported in very early studies (81, 82), and frequently ever since, B*35 and B*27-restricted immune responses against Influenza A, and B*08 in Influenza B have also been described to be immunodominant (83). Latter studies indicated that co-expression of A*02 in influenza A-infected individuals changed the hierarchy of the responses (83) and that there seems to exist a tendency for A*02 or B*27-restricted responses to be immunodominant, while B*08 was superior in the absence of B*27 (84). These studies illustrate that the hierarchy of immunodominant responses varies not only with the virus targeted, but also with the combination of HLA class I molecules that are co-expressed in the host.

In conclusion, most data from the literature suggest that HLA-B alleles are major determinants of the dominance of antiviral $CD8^+$ T-cell responses. Additionally, in the case of intracellular bacterial pathogens such as *M. tuberculosis,* the immunodominant $CD8^+$ T-cell responses tend to be restricted by HLA-B alleles (85). However, the mechanisms underlying the dominance of these responses remain elusive. With the interaction of peptide-MHC class-I complex and T-cell receptor (TCR) at its interface, studies on factors impacting immunodominance have naturally focused on the peptide repertoire being presented and the features of the $CD8^+$ T-cell populations that respond to it. The density of epitope presentation, as well as its stability at the cell surface, will impact not only the functional sensitivity and the chance of recognition by specific $CD8^+$ T cells, but also the downstream effects of activation, proliferation, function and transition into memory. Furthermore, dissecting the diversity of the TCR repertoire and the relative magnitude of the clones making up several responding $CD8^+$ T-cell populations might help elucidate what the requisites for a dominant response are. Finally, studying the plasticity of a given $CD8^+$ T-cell response (i.e. its ability to adapt to

antigen persistence or (re-)infection) might bring forward new information regarding the co-evolution of pathogens and their hosts.

Aim of the Thesis

All HLA class I molecules are structurally very similar, yet not similarly polymorphic. HLA-B locus is most polymorphic and tends to be often associated with i) the outcome (resistance or susceptibility) of infectious diseases, and ii) the establishment of immunodominant T-cell responses. In this thesis, we took 'immunodominance' as a starting point and 'probed' its components: the peptide repertoire that sets it in motion, and the T-cell responses that compose it.

In the first part, we made use of *in silico* approaches and experimental data to analyse whether differences in the peptide repertoire of HLA-B *versus* HLA-A loci could explain that HLA-B alleles would often induce immunodominant $CD8^+$ T-cell responses. To investigate potential contrasting features of epitope presentation via HLA-B *versus* HLA-A, in terms of amount and binding affinity of presented peptides, and relative distribution of self *versus* non-self peptides per locus, we performed an in depth analysis of experimental HLA binding data curated in the Immune Epitope Database (IEDB) and predicted ligand data (chapter 2). In addition, and using HIV as a pathogen model, we investigated the degree of conservation of HLA-B *versus* HLA-A epitopes reported in the LANL database (chapter 3). Ultimately, we explored the promiscuity of peptide binding to HLA class I molecules as an additional contributing factor to immunodominance (chapter 4). In the second part, we investigated features of HLA class I-restricted T-cell responses that could be related to immunodominance. We examined the repertoire of the T-cell response against frequently targeted epitopes in EBV and CMV, persistent viruses usually kept under control in healthy individuals (chapter 5). We looked into the plasticity of the responding $CD8^+$ T-cell repertoire upon chronic infection (HIV) or re-encounter with an acute virus (influenza A) in chapters 6 and 7, respectively. The results are put in a broader perspective in chapter 8.

NOTE: This introduction has been written within the framework of a HiPo grant that brought together the groups of Theoretical Biology (Utrecht University) and Immunology (UMC Utrecht), and has been published before in a slightly altered form, in the context of the collaboration (PhD Thesis *A detailed comparison of peptides presented by different HLA class I loci: an in silico approach*, by Xianyu Rao). Ana Costa was the major contributor in the previous version.

References

1. Cooke GS, and Hill AV. Genetics of susceptibility to human infectious disease. Nat Rev Genet. 2001;2(12):967-77.
2. Groothuis TA, Griekspoor AC, Neijssen JJ, Herberts CA, and Neefjes JJ. MHC class I alleles and their exploration of the antigen-processing machinery. Immunol Rev. 2005;207:60-76.
3. Peaper DR, and Cresswell P. Regulation of MHC class I assembly and peptide binding. Annu Rev Cell Dev Biol. 2008;24:343-68.
4. Purcell AW, and Elliott T. Molecular machinations of the MHC-I peptide loading complex. Curr Opin Immunol. 2008;20(1):75-81.
5. Rock KL, Farfan-Arribas DJ, Colbert JD, and Goldberg AL. Re-examining class-I presentation and the DRiP hypothesis. Trends Immunol. 2014;35(4):144-52.
6. Vyas JM, Van der Veen AG, and Ploegh HL. The known unknowns of antigen processing and presentation. Nat Rev Immunol. 2008;8(8):607-18.
7. Burroughs NJ, de Boer RJ, and Kesmir C. Discriminating self from nonself with short peptides from large proteomes. Immunogenetics. 2004;56(5):311-20.
8. Calis JJ, de Boer RJ, and Kesmir C. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. PLoS computational biology. 2012;8(3):e1002412.
9. Chen W, Norbury CC, Cho Y, Yewdell JW, and Bennink JR. Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens. J Exp Med. 2001;193(11):1319-26.
10. Kincaid EZ, Che JW, York I, Escobar H, Reyes-Vargas E, Delgado JC, Welsh RM, Karow ML, Murphy AJ, Valenzuela DM, et al. Mice completely lacking immunoproteasomes show major changes in antigen presentation. Nat Immunol. 2012;13(2):129-35.
11. Sijts EJ, and Kloetzel PM. The role of the proteasome in the generation of MHC class I ligands and immune responses. Cell Mol Life Sci. 2011;68(9):1491-502.
12. Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, Laplace C, Zwinderman A, Dick TP, Muller J, et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med. 2001;194(1):1-12.
13. Hansen TH, and Bouvier M. MHC class I antigen presentation: learning from viral evasion strategies. Nat Rev Immunol. 2009;9(7):503-13.
14. Horst D, Verweij MC, Davison AJ, Ressing ME, and Wiertz EJ. Viral evasion of T cell immunity: ancient mechanisms offering new applications. Curr Opin Immunol. 2011;23(1):96-103.
15. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, Le Gall S, John M, Mothe BR, Lee PK, Kalife ET, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. J Virol. 2004;78(13):7069-78.
16. Bronke C, Almeida CA, McKinnon E, Roberts SG, Keane NM, Chopra A, Carlson JM, Heckerman D, Mallal S, and John M. HIV escape mutations occur preferentially at HLA-binding sites of CD8 T-cell epitopes. AIDS. 2013;27(6):899-905.
17. Iglesias MC, Almeida JR, Fastenackels S, van Bockel DJ, Hashimoto M, Venturi V, Gostick E, Urrutia A, Wooldridge L, Clement M, et al. Escape from highly effective public CD8+ T-cell clonotypes by HIV. Blood. 2011;118(8):2138-49.
18. Kwun HJ, da Silva SR, Shah IM, Blake N, Moore PS, and Chang Y. Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen 1 mimics Epstein-Barr virus EBNA1 immune evasion through central repeat domain effects on protein processing. J Virol. 2007;81(15):8225-35.

19. Yokomaku Y, Miura H, Tomiyama H, Kawana-Tachikawa A, Takiguchi M, Kojima A, Nagai Y, Iwamoto A, Matsuda Z, and Ariyoshi K. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. J Virol. 2004;78(3):1324-32.

20. Zhang SC, Martin E, Shimada M, Godfrey SB, Fricke J, Locastro S, Lai NY, Liebesny P, Carlson JM, Brumme CJ, et al. Aminopeptidase substrate preference affects HIV epitope presentation and predicts immune escape patterns in HIV-infected individuals. J Immunol. 2012;188(12):5924-34.

21. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Beck S, and al. e. The DNA sequence and analysis of human chromosome 6. Nature. 2003;425(6960):805-11.

22. Mehra NK, and Kaur G. MHC-based vaccination approaches: progress and perspectives. Expert Rev Mol Med. 2003;5(7):1-17.

23. Jeffery KJ, and Bangham CR. Do infectious diseases drive MHC diversity? Microbes Infect. 2000;2(11):1335-41.

24. Hughes AL, and Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 1988;335(6186):167-70.

25. Lienert K, and Parham P. Evolution of MHC class I genes in higher primates. Immunol Cell Biol. 1996;74(4):349-56.

26. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, and Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol. 2005;15(11):1022-7.

27. Parham P, and Ohta T. Population biology of antigen presentation by MHC class I molecules. Science. 1996;272(5258):67-74.

28. Doherty PC, and Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature. 1975;256(5512):50-2.

29. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, and O'Brien SJ. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science. 1999;283(5408):1748-52.

30. Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E, Allen S, and Kaslow RA. HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses. 1999;15(4):317-24.

31. De Boer RJ, Borghans JA, van Boven M, Kesmir C, and Weissing FJ. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics. 2004;55(11):725-31.

32. Leslie A, Matthews PC, Listgarten J, Carlson JM, Kadie C, Ndung'u T, Brander C, Coovadia H, Walker BD, Heckerman D, et al. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. J Virol. 2010;84(19):9879-88.

33. Kubinak JL, Ruff JS, Hyzer CW, Slev PR, and Potts WK. Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. Proc Natl Acad Sci U S A. 2012;109(9):3422-7.

34. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, Funkhouser R, Fugate M, Theiler J, Hsu YS, et al. Advantage of rare HLA supertype in HIV disease progression. Nat Med. 2003;9(7):928-35.

35. Moore CB, John M, James IR, Christiansen FT, Witt CS, and Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science. 2002;296(5572):1439-43.

36. Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, Feeney ME, Addo MM, Lichterfeld M, Lahaie MP, et al. Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes. Nat Immunol. 2006;7(2):173-8.

37. Schellens IM, Navis M, van Deutekom HW, Boeser-Nunnink B, Berkhout B, Kootstra N, Miedema F, Kesmir C, Schuitemaker H, van Baarle D, et al. Loss of HIV-1 derived CTL epitopes restricted by protective HLA-B alleles during the HIV-1 epidemic. AIDS. 2011;25(14):1691-700.

38. de Groot NG, Otting N, Arguello R, Watkins DI, Doxiadis GG, Madrigal JA, and Bontrop RE. Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. Immunogenetics. 2000;51(6):398-409.

39. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, and Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991;352(6336):595-600.

40. Watkins DI, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG, Garber TL, Dogon AL, Lord CI, Ghim SH, et al. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature. 1992;357(6376):329-33.

41. Borghans JA, Beltman JB, and De Boer RJ. MHC polymorphism under host-pathogen coevolution. Immunogenetics. 2004;55(11):732-9.

42. Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, and Bontrop RE. Unparalleled complexity of the MHC class I region in rhesus macaques. Proc Natl Acad Sci U S A. 2005;102(5):1626-31.

43. Fukami-Kobayashi K, Shiina T, Anzai T, Sano K, Yamazaki M, Inoko H, and Tateno Y. Genomic evolution of MHC class I region in primates. Proc Natl Acad Sci U S A. 2005;102(26):9230-4.

44. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature. 2004;432(7018):769-75.

45. Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzl-Erler ML, and Parham P. Unusual HLA-B alleles in two tribes of Brazilian Indians. Nature. 1992;357(6376):326-9.

46. McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, and Watkins DI. A uniquely high level of recombination at the HLA-B locus. Proc Natl Acad Sci U S A. 1994;91(13):5893-7.

47. Barber LD, Percival L, Arnett KL, Gumperz JE, Chen L, and Parham P. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. J Immunol. 1997;158(4):1660-9.

48. Carrington M, and O'Brien SJ. The influence of HLA genotype on AIDS. Annu Rev Med. 2003;54:535-51.

49. Goulder PJ, and Watkins DI. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol. 2008;8(8):619-30.

50. Martin MP, and Carrington M. Immunogenetics of viral infections. Curr Opin Immunol. 2005;17(5):510-6.

51. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, et al. A whole-genome association study of major determinants for host control of HIV-1. Science. 2007;317(5840):944-7.

52. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Zhao M, and et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010;330(6010):1551-7.

53. Matthews PC, Leslie AJ, Katzourakis A, Crawford H, Payne R, Prendergast A, Power K, Kelleher AD, Klenerman P, Carlson J, et al. HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering. J Virol. 2009;83(9):4605-15.

54. de Groot NG, Otting N, Doxiadis GG, Balla-Jhagjhoorsingh SS, Heeney JL, van Rood JJ, Gagneux P, and Bontrop RE. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci U S A. 2002;99(18):11748-53.

55. Bontrop RE, and Watkins DI. MHC polymorphism: AIDS susceptibility in non-human primates. Trends Immunol. 2005;26(4):227-33.

56. de Groot NG, and Bontrop RE. The HIV-1 pandemic: does the selective sweep in chimpanzees mirror humankind's future? Retrovirology. 2013;10:53-67.

57. Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, Haaksma TG, Teeuwsen VJ, Bontrop RE, and Heeney JL. Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. J Immunol. 1999;162(4):2308-14.

58. van Deutekom HW, Hoof I, Bontrop RE, and Kesmir C. A comparative analysis of viral peptides presented by contemporary human and chimpanzee MHC class I molecules. J Immunol. 2011;187(11):5995-6001.

59. de Groot NG, Heijmans CM, Zoet YM, de Ru AH, Verreck FA, van Veelen PA, Drijfhout JW, Doxiadis GG, Remarque EJ, Doxiadis, II, et al. AIDS-protective HLA-B*27/B*57 and chimpanzee MHC class I molecules target analogous conserved areas of HIV-1/SIVcpz. Proc Natl Acad Sci U S A. 2010;107(34):15175-80.

60. Hoof I, Kesmir C, Lund O, and Nielsen M. Humans with chimpanzee-like major histocompatibility complex-specificities control HIV-1 infection. AIDS. 2008;22(11):1299-303.

61. Chuang WC, Sarkodie F, Brown CJ, Owusu-Ofori S, Brown J, Li C, Navarrete C, Klenerman P, and Allain JP. Protective effect of HLA-B57 on HCV genotype 2 infection in a West African population. J Med Virol. 2007;79(6):724-33.

62. Kim AY, Kuntzen T, Timm J, Nolan BE, Baca MA, Reyor LL, Berical AC, Feller AJ, Johnson KL, Schulze zur Wiesch J, et al. Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. Gastroenterology. 2011;140(2):686-96 e1.

63. Kuniholm MH, Kovacs A, Gao X, Xue X, Marti D, Thio CL, Peters MG, Terrault NA, Greenblatt RM, Goedert JJ, et al. Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. Hepatology. 2010;51(5):1514-22.

64. McKiernan SM, Hagan R, Curry M, McDonald GS, Kelly A, Nolan N, Walsh A, Hegarty J, Lawlor E, and Kelleher D. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. Hepatology. 2004;40(1):108-14.

65. Neumann-Haefelin C, and Thimme R. Impact of the genetic restriction of virus-specific T-cell responses in hepatitis C virus infection. Genes Immun. 2007;8(3):181-92.

66. Jeffery KJ, Siddiqui AA, Bunce M, Lloyd AL, Vine AM, Witkover AD, Izumo S, Usuku K, Welsh KI, Osame M, et al. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. J Immunol. 2000;165(12):7278-84.

67. McAulay KA, Higgins CD, Macsween KF, Lake A, Jarrett RF, Robertson FL, Williams H, and Crawford DH. HLA class I polymorphisms are associated with development of infectious mononucleosis upon primary EBV infection. J Clin Invest. 2007;117(10):3042-8.

68. Niens M, Jarrett RF, Hepkema B, Nolte IM, Diepstra A, Platteel M, Kouprie N, Delury CP, Gallagher A, Visser L, et al. HLA-A*02 is associated with a reduced risk and HLA-A*01 with an increased risk of developing EBV+ Hodgkin lymphoma. Blood. 2007;110(9):3310-5.

69. Loke H, Bethell DB, Phuong CX, Dung M, Schneider J, White NJ, Day NP, Farrar J, and Hill AV. Strong HLA class I--restricted T cell responses in dengue hemorrhagic fever: a double-edged sword? J Infect Dis. 2001;184(11):1369-73.

70. Nguyen TP, Kikuchi M, Vu TQ, Do QH, Tran TT, Vo DT, Ha MT, Vo VT, Cao TP, Tran VD, et al. Protective and enhancing HLA alleles, HLA-DRB1*0901 and HLA-A*24, for severe forms of dengue virus infection, dengue hemorrhagic fever and dengue shock syndrome. PLoS Negl Trop Dis. 2008;2(10):e304.

71. Stephens HA, Klaythong R, Sirikong M, Vaughn DW, Green S, Kalayanarooj S, Endy TP, Libraty DH, Nisalak A, Innis BL, et al. HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. Tissue Antigens. 2002;60(4):309-18.

72. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, and 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(5):2298-305.

73. Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S, Segal JP, Cao Y, Rowland-Jones SL, Cerundolo V, et al. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. Science. 1998;279(5359):2103-6.

74. Bihl F, Frahm N, Di Giammarino L, Sidney J, John M, Yusim K, Woodberry T, Sango K, Hewitt HS, Henry L, 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(7):4094-101.

75. Altfeld M, Addo MM, Rosenberg ES, Hecht FM, Lee PK, Vogel M, Yu XG, Draenert R, Johnston MN, Strick D, et al. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS. 2003;17(18):2581-91.

76. Neumann-Haefelin C, McKiernan S, Ward S, Viazov S, Spangenberg HC, Killinger T, Baumert TF, Nazarova N, Sheridan I, Pybus O, et al. Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution. Hepatology. 2006;43(3):563-72.

77. Ray SC, Fanning L, Wang XH, Netski DM, Kenny-Walsh E, and Thomas DL. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. J Exp Med. 2005;201(11):1753-9.

78. Gaudieri S, Rauch A, Park LP, Freitas E, Herrmann S, Jeffrey G, Cheng W, Pfafferott K, Naidoo K, Chapman R, et al. Evidence of viral adaptation to HLA class I-restricted immune pressure in chronic hepatitis C virus infection. J Virol. 2006;80(22):11094-104.

79. Timm J, Lauer GM, Kavanagh DG, Sheridan I, Kim AY, Lucas M, Pillay T, Ouchi K, Reyor LL, Schulze zur Wiesch J, et al. CD8 epitope escape and reversion in acute HCV infection. J Exp Med. 2004;200(12):1593-604.

80. Tripathy AS, Shankarkumar U, Chadha MS, Ghosh K, and Arankalle VA. Association of HLA alleles with hepatitis C infection in Maharashtra, western India. The Indian journal of medical research. 2009;130(5):550-5.

81. Bednarek MA, Sauma SY, Gammon MC, Porter G, Tamhankar S, Williamson AR, and Zweerink HJ. The minimum peptide epitope from the influenza virus matrix protein. Extra and intracellular loading of HLA-A2. J Immunol. 1991;147(12):4047-53.

82. Morrison J, Elvin J, Latron F, Gotch F, Moots R, Strominger JL, and McMichael A. Identification of the nonamer peptide from influenza A matrix protein and the role of pockets of HLA-A2 in its recognition by cytotoxic T lymphocytes. Eur J Immunol. 1992;22(4):903-7.

83. Boon AC, De Mutsert G, Fouchier RA, Sintnicolaas K, Osterhaus AD, and Rimmelzwaan GF. Preferential HLA usage in the influenza virus-specific CTL response. J Immunol. 2004;172(7):4435-43.

84. Boon AC, de Mutsert G, Graus YM, Fouchier RA, Sintnicolaas K, Osterhaus AD, and Rimmelzwaan GF. The magnitude and specificity of influenza A virus-specific cytotoxic T-lymphocyte responses in humans is related to HLA-A and -B phenotype. J Virol. 2002;76(2):582-90.

85. Lewinsohn DA, Winata E, Swarbrick GM, Tanner KE, Cook MS, Null MD, Cansler ME, Sette A, Sidney J, and Lewinsohn DM. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. PLoS Pathog. 2007;3(9):1240-9.

A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant $CD8^+$ T cell responses

Xiangyu Rao[1], Ana Isabel C.A. Fontaine Costa[2], Debbie van Baarle[2], Can Keşmir[1,3]

[1]*Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands.*

[2]*Department of Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands.*

[3]*Academic Biomedical Centre, Utrecht University, Utrecht, The Netherlands.*

The Journal of Immunology (2009), 182: 1526-1532.

Abstract

Conventional $CD8^+$ T cell responses against intracellular infectious agents are initiated upon recognition of pathogen-derived peptides presented at the cell surface of infected cells in the context of MHC class I molecules. Among the major MHC class I loci, HLA-B is the swiftest evolving and the most polymorphic locus. Additionally, responses restricted by HLA-B molecules tend to be dominant, and most associations with susceptibility or protection against infectious diseases have been assigned to HLA-B alleles. To assess whether the differences in responses mediated via two major HLA class I loci, HLA-B and HLA-A, may already begin at the Ag presentation level, we have analyzed the diversity and binding affinity of their peptide repertoire by making use of curated pathogen-derived epitope data retrieved from the Immune Epitope Database and Analysis Resource, as well as in silico predicted epitopes. In contrast to our expectations, HLA-B alleles were found to have a less diverse peptide repertoire, which points toward a more restricted binding motif, and the respective average peptide binding affinity was shown to be lower than that of HLA-A-restricted epitopes. This unexpected observation gives rise to new hypotheses concerning the mechanisms underlying immunodominance of $CD8^+$ T cell responses.

Introduction

MHC class I molecules play a crucial role in initiating potentially protective immune responses by presenting intracellular pathogen-derived peptides to $CD8^+$ T cells and thus targeting infected cells for elimination. The enormous polymorphism of MHC class I genes (1) is most likely pathogen-driven (2-6). Still, this selection pressure does not seem to be acting homogeneously in all major class I loci. Among the homologous major human leukocyte antigen (HLA) class I loci, HLA-B seems to possess some distinctive characteristics. HLA-B is the most diverse class I locus and the most polymorphic gene of the human genome (7). Indeed, it seems to be the most rapidly evolving locus and it has an exceptional rate of recombination (8), suggesting that it is under the strongest selective pressure. The latter has also been highlighted by Prugnolle and colleagues, who have shown that the positive correlation between variation at the classical HLA class I loci and pathogen richness is strongest for HLA-B (5). Furthermore, HLA-B-restricted pathogen-derived epitopes seem to be more frequently targeted by $CD8^+$ T cells than those presented on HLA-A. This has been shown for viral pathogens such as HIV, EBV, Influenza and CMV (9-12), as well as for an intracellular bacterial pathogen, *Mycobacterium tuberculosis* (13). Additionally, HLA-B-dictated T cell responses seem to be of higher magnitude (10, 13, 14) and have been described to influence infectious disease course and outcome. In HIV infection, a specific group of B alleles (e.g. B*3502, B*3503, B*3504, B*5301, B*5802) has been associated with fast progression toward AIDS (9, 15, 16), while others, like B27, B57 and B*5801, are commonly referred to as "protective", as they are associated with a longer AIDS-free survival period or other favourable features related to a better prognosis, such as a lower viral load (17, 18). Kiepiela and colleagues have also demonstrated a broader dominance of B alleles in influencing HIV disease outcome due to their impact on viral setpoint and $CD4^+$ T cell counts (9). Also, in hepatitis C virus (HCV) infection, HLA-B27 seems to be the allele most significantly associated with spontaneous viral clearance (19). Furthermore, the association of B53 with protection from severe malaria has most likely contributed to its high frequencies in West Africa (20). Although the associations of HLA-B molecules with disease outcome suggest a fundamental role of $CD8^+$ T cells in responding against intracellular pathogens upon recognition of peptide-HLA-B complexes in infected cells, they may also be a consequence of engagement of HLA-B by Natural Killer (NK) cell receptors and subsequent impact on NK cell function (17, 21).

The three features of HLA-B molecules mentioned above – high polymorphism, trend to induce immunodominant $CD8^+$ T cell responses, and association with disease

susceptibility or protection – are most likely related: HLA-B alleles may be major determinants, "for good or for bad", of the effectiveness of immune responses against pathogens because the responses they elicit are dominant. This immunodominance may, among others, stem from differences in presentation of antigenic peptides via HLA-B, or from a more amplified, qualitative and/or quantitatively, response from specific $CD8^+$ T cells to HLA-B-restricted epitopes (22, 23). Intrinsic differences in the amount/diversity and binding affinity (and thus stability at the cell surface) of the peptides they present might provide the grounds for their superiority. These issues on peptide presentation have been addressed in this study. We envisioned that HLA-B alleles would have a broader peptide repertoire than that of HLA-A alleles, inducing a more diverse CTL response, and that the epitopes would be of higher affinity, ensuring a longer stimulation of T cell responses and thus further accounting for the immunodominance observed. Surprisingly, our results show that, in contrast with our initial expectations, HLA-B alleles may actually have a more restricted binding motif, they may present fewer peptides than do their HLA-A homologs, and their binding affinity tends to be lower. How these particular HLA-B features may lead to immunodominance is thus far unclear, yet they open up new hypotheses for the effects of antigen presentation on immunodominance.

Materials and Methods

Epitope data

All epitopes previously shown to bind to HLA molecules were downloaded as an XML file from the Immune Epitope Database and Analysis Resource (IEDB http://www.immuneepitope.org/; downloads were made in May 2007). As the chance of two different proteins sharing a peptide larger than a 7-mer seems to be maximally 2-3% (24), all epitopes with fewer than 7 amino acids similarity were kept as unique epitopes for each HLA allele. In total, 11878 unique epitopes for 19 HLA-A alleles and 2353 unique epitopes for 16 HLA-B alleles having a length varying from 6 to 16 amino acids were obtained from IEDB as our experimental binding data.

The affinity of most (>98%) HLA-A and B epitopes in IEDB was measured by two experimental methods. The first method (competition assays where purified MHC and radioactive labelling are used) covered 94% of HLA-A epitopes and 78% of HLA-B epitopes, while the second method (association studies where purified MHC and fluorescence labelling are used) covered 5% HLA-A epitopes and 21% HLA-B epitopes.

Eighty-five epitopes were measured by both methods. Using this overlap data, we found a significant correlation of measured affinities by both methods (p=0.0001).

Genomic data

Sequences of the human proteome as a resource of self-peptides, 13 common bacteria and 17 common virus proteomes (supplementary Table S1) as resources of pathogenic (non-self) peptides were downloaded from the European Bioinformatics Institute (EBI) web site (http://www.ebi.ac.uk/; downloads were made in Oct 2006). Peptide fragments of 9 or 10 amino acids long were generated from each protein in all proteomes using all positions as possible first positions. All redundant 9-mers or 10-mers that are present more than one time in a proteome or contain any ambiguous amino acids such as "B, Z, X" were deleted from the dataset. In the end, a total of 10941519 unique human 9-mers, 11647396 unique bacterial 9-mers and 57433 unique viral 9-mers were retrieved from these proteomes.

Two types of sequence modifications in the original proteomes were applied in this study: 1) protein shuffling, that is, shuffling residues within a protein by keeping the frequency of amino acids and the length of the protein intact; and 2) generating artificial proteomes, that is, artificial proteomes were generated by using a uniform amino acid distribution (i.e., frequency of each amino acid is 0.05) without changing the size of original proteome. All of these sequence modifications were performed by Perl programming.

HLA-peptide binding predictions

There are at present several HLA-peptide binding predictors, such as the polynomial method (PM) (25), artificial neural networks (ANN) (26), the BIMAS algorithm (27), a classification and regression tree (CART) (28), an additive method (29) and the stabilized matrix method (SMM) (30). The performance of these algorithms to identify new epitopes has recently been benchmarked on experimental data (31). In general, ANN and SMM methods were found to be superior to the other ones. Moreover, the results of these large-scale benchmark studies show that the accuracy of HLA-peptide binding prediction algorithms has increased over time to such an extent that the correlation between predicted and measured binding affinity is as good as the correlation between measurements from different laboratories (31). Another demonstration of how powerful these predictions are is a study by Schellens et al. (32), who identified 18 new CTL epitopes out of a set of 22 predicted CTL epitopes using artificial neural networks-based HLA-peptide binding predictors. This suggests that the specificity of the predictors can be as as high as 80%.

SMM was chosen as the main prediction method in this project because it has the broadest HLA coverage at the time this project was initiated. For each HLA molecule a specific SMM matrix was generated using the experimental binding data. The details of this algorithm are given elsewhere (30). We downloaded SMM prediction software and training datasets from **IEDB** (http://mhcbindingpredictions.immuneepitope.org/), and generated scoring matrices ourselves. For 9-mer prediction, 19 HLA-A and 15 HLA-B SMM matrices were generated, while for 10-mer prediction 14 HLA-A and 5 HLA-B matrices were generated because of training data limitation.

As suggested by Ruppert et al. (33), an IC_{50} threshold of 500nM was used to distinguish binding peptides from nonbinding peptides.

Thus, the binding fraction of a particular HLA allele was calculated as:

$$\text{Binding fraction} = \frac{\text{Number of binding n-mers}}{\text{Number of all n-mers}} \quad (\text{n} = 9 \text{ or } 10).$$

Shannon entropy calculation and binding motif visualization

The Shannon entropy (34) was calculated to measure the degree of variability in HLA binding motifs. This entropy measure is defined as:

$$E(i) = -\sum_{L=1}^{20} q_i \log_2 q_i ,$$

where $E(i)$ is the Shannon entropy at position i, and q_i is the probability that a particular amino acid occurs at position i in the alignment. The maximum value of the Shannon entropy is obtained when all amino acids occur with the same frequency ($-20(1/20)\log_2(1/20) \approx 4.3$) and the minimum Shannon entropy is 0 if a position is fully conserved. The binding motifs of HLA molecules were visualized by sequence logo technique (35) in this study (see caption of supplementary Fig. S5 for an explanation).

Results

HLA-A alleles present a larger set of epitopes than HLA-B alleles

To explore the differences in the epitope repertoire of HLA molecules, publicly available experimental epitope data were analyzed. At present, IEDB (http://www.immuneepitope.org) is the most complete database for experimentally verified epitopes. All epitopes that can bind to HLA-A or HLA-B alleles were retrieved from IEDB (downloads were made in October 2006). The number of known epitopes per allele shows a very skewed distribution: for the common alleles, thousands of epitopes

are identified, while for the rare alleles as few as only one epitope is found. In IEDB the average number of unique epitopes for each HLA-A allele is 216 (median, 15; range, 1-2341) and the average number of unique epitopes for each HLA-B allele is 37 (median, 8; range, 1-399). Thus, the diversity of HLA-A binding epitopes seems to be larger than that of HLA-B. After excluding alleles having fewer than 20 epitopes, we calculated the epitope length distribution for 19 HLA-A and 16 HLA-B alleles (Fig. 1). The epitope length distributions of HLA-A and HLA-B alleles were very similar and showed a marked preference to bind epitopes of length 9 (9-mers) or length 10 (10-mers) (Fig. 1). For the rest of the paper this data set is used and referred to as "experimental binding data".

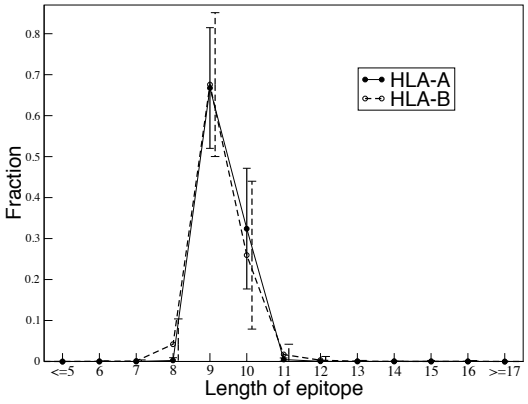


Figure 1. Length distribution of experimentally verified HLA-A and HLA-B epitopes.
11878 unique experimental epitopes for 19 HLA-A alleles and 2353 unique experimental epitopes for 16 HLA-B alleles were downloaded from IEDB and were classified according to the length. The lines show the average fraction for HLA-A (solid) and HLA-B (dashed) epitopes, while error bars represent the SD (standard deviation).

The experimental data might have some biases, for example, due to dominance of certain HLA molecules like HLA-A*02 (www.allelefrequencies.net). Alternatively, one can screen proteomes for epitopes restricted by every HLA molecule for which prediction methods are available to determine epitope repertoires. We based our analysis on stabilized matrix method (SMM) predictions because at the time we initialized our study the SMM method had the largest allele coverage. Using publicly available software and data, we generated 19 HLA-A and 15 HLA-B SMM scoring matrices and predicted the binding affinity of all possible 9-mers from the human, bacterial and viral proteomes (as explained in *Materials and Methods*). SMM matrices assign to each 9-mer-HLA pair an IC_{50} value that indicates the binding affinity. The lower IC_{50} value implies the higher binding affinity of the 9-mer to a certain HLA molecule. As suggested earlier (33), an IC_{50} threshold of 500nM was used to distinguish

binding 9-mers from nonbinding 9-mers. We defined the binding fraction as the fraction of SMM-predicted binding 9-mers on an HLA molecule among all possible 9-mers in a proteome and used this as a parameter to indicate the epitope diversity of the HLA molecule. Fig. 2 shows that on average the binding fraction of HLA-A alleles is ~5%, which is significantly higher than the average binding fraction of HLA-B alleles (2%) in the human proteome (Fig. 2, $p<0.001$, Mann-Whitney *U* test). Similar results were obtained for viral 9-mers and bacterial 9-mers (Fig. 2).

Changing the IC_{50} threshold (using 50nM and 5000nM instead of 500nM) or excluding two alleles that are either too generic (A*3002, with a binding fraction > 20%) or too specific (B*0801, which does not have a predicted binder in many viral proteomes) did not change our results (data not shown). Thus, HLA-A alleles seem to bind 2-fold more peptides than do HLA-B alleles. We repeated the same prediction scheme for 10-mers and found similar results (data not shown). Note that among HLA alleles there is a large variation in predicted binding fraction (data for individual alleles are shown in supplementary Fig. S1).

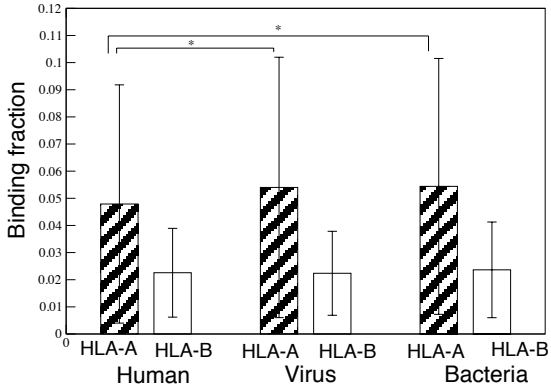


Figure 2. Average binding fraction of HLA-A and HLA-B alleles in human, viral and bacterial proteomes.
All possible 9-mers from human, viral and bacterial proteomes that can bind to certain HLA-A or HLA-B molecules were predicted by SMM matrices as described in *Material and Methods*. The average binding fractions, defined as the average fraction of 9-mers that were predicted to bind an HLA molecule with an IC_{50} value < 500 nM in a proteome, are shown for HLA-A (striped bars) and HLA-B (empty bars) molecules. In all cases (human, viral and bacterial proteomes) the average binding fraction of HLA-A molecules was higher than HLA-B molecules ($p<0.001$, Mann-Whitney *U* test). Moreover, HLA-A molecules bind significantly fewer self (human) 9-mers than non-self (viral or bacterial) 9-mers. *, $p<0.001$, Mann-Whitney *U* test.

When we shuffled the protein sequences in different proteomes, or when we created artificial proteomes with equal amino acid frequencies (i.e., when the frequency of each amino acid is 0.05), the binding fraction of HLA-A alleles remained 2-fold higher than that of HLA-B alleles (supplementary Fig. S2). This result implies that sequence

patterns or amino acid frequencies in different proteomes are not responsible for differences in epitope diversity between HLA-A and HLA-B molecules.

Surprisingly, HLA-A molecules were predicted to bind significantly fewer human 9-mers than pathogen-derived 9-mers ($p<0.001$, Mann-Whitney U test), which was not the case for HLA-B molecules (Fig. 2). This preference pattern was robust to shuffling of protein sequences, but it disappeared when we created artificial proteomes with equal amino acid frequencies. Apparently, very small differences in amino acid usage among human, viral and bacterial proteomes (supplementary Fig. S3) may create a preference for presenting non-self.

Experimentally verified epitopes of a given HLA molecule are used to construct a SMM scoring matrix that can predict new epitopes of this HLA molecule (36). The number of experimental epitopes used to construct SMM predictors, that is, the training set size, varies among different HLA molecules. To exclude the possibility that the low binding fraction of HLA-B alleles is due to small training sets, we plotted the predicted binding fraction versus training set size of each allele in supplementary Fig. S4. Neither for human nor for pathogenic proteomes did we find a significant correlation ($p>0.5$), which suggests that the higher binding specificity of HLA-B molecules is not an artifact of the data size used to develop the predictors.

In summary, both experimental data and our *in silico* predictions suggest that the binding motif of HLA-B molecules is in general more restricted than that of HLA-A molecules, and therefore HLA-B molecules present a less diverse set of epitopes than do HLA-A molecules.

The binding affinity of HLA-A epitopes is significantly higher than that of HLA-B epitopes

Another factor that may affect the immunodominance T cell responses is the binding affinity of peptides to HLA molecules (37). We analyzed both experimental and predicted binding data to compare the binding affinity of HLA-A and HLA-B molecules.

All experimentally verified HLA-A and HLA-B epitopes that have an IC_{50} value associated were retrieved from the IEDB database, and are shown in Fig. 3. *In vitro*-measured binding affinities of HLA-A epitopes are significantly higher than those of HLA-B epitopes ($p< 0.001$, Mann-Whitney *U* test). The binding affinities of most epitopes in IEDB (94% HLA-A epitopes and 78% HLA-B epitopes) were measured using a competition (or equilibrium binding) assay with purified MHC molecules and radioactive labelling. By repeating our analysis for this subset to exclude a possible bias caused by experimental methods, we found that binding affinities of HLA-A epitopes

are still significantly higher than those of HLA-B epitopes ($p < 0.001$, Mann-Whitney *U* test). Thus, the differences we found between HLA-A and HLA-B epitope affinities are not due to the use of different experimental assays.

We then analyzed the *in silico*-predicted binding affinities from the human, viral and bacterial proteomes. To exclude a potential bias introduced by the 2-fold higher amount of predicted HLA-A epitopes as compared with those restricted by HLA-B alleles, we selected the top 5 predicted epitopes, with the highest predicted binding affinity, for each HLA molecule from each proteome. The average binding affinity of predicted HLA-A epitopes is significantly higher ($p < 0.001$, Mann-Whitney test) than that of predicted HLA-B epitopes (Fig. 3). The results were confirmed by repeating the same analysis for the top 10 epitopes of each HLA molecule (results not shown).

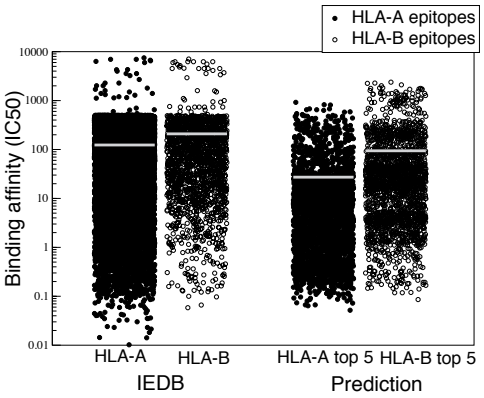


Figure 3. Binding affinities of HLA-A epitopes are significantly higher than HLA-B epitopes. *Left:* Binding affinity (IC_{50} value) for HLA-A and HLA-B epitopes measured *in vitro* (data downloaded from IEDB). *Right:* Binding affinity of top 5 predicted epitopes in human, virus and bacteria proteomes for 19 HLA-A and 15 HLA-B molecules. The gray lines show the average. In all cases the binding affinities of HLA-A molecules were significantly higher ($p<0.001$, Mann-Whitney *U* test) than those of HLA-B molecules.

Although the predicted binding affinities show the same tendency as the experimental data (Fig. 3), one can argue that the predictions are not true validations of the experimental data, because they are not independent, that is, the prediction methods were trained on the available experimental data. To separate the experimental data as much as possible from the predictions, we used the following strategy: each time we chose three HLA-A and three HLA-B alleles randomly from all of the alleles where prediction methods were available. Then, we compared the predicted binding affinities for the randomly chosen three HLA-A and HLA-B alleles, while excluding the six alleles used in the prediction analysis from the experimental binding data and compared HLA-A and HLA-B binding affinities in the rest. In this way, the alleles (and thus the

epitope data) tested using the experimental data and the predictions are nonoverlapping. We repeated this process 100 times, and in 92 cases we found that both the predicted and experimentally verified binding affinities of HLA-B alleles were significantly lower than those for HLA-A alleles. One can therefore conclude that the significant difference in binding affinities of HLA-A and HLA-B alleles found in the experimental data is further supported with our predictions.

Next we tested whether the higher binding affinity can be a result of lower specificity in HLA-A binding motifs. In Fig. 4 we plotted the binding fraction *vs* the binding affinity of the top 5 predicted epitopes for all HLA molecules in our *in silico* analysis. Interestingly, there is a significant positive correlation ($p<0.001$) between the binding fraction and the average peptide binding affinity of these alleles, which suggests that a less specific binding motif would imply the higher binding affinities.

In summary, HLA-A epitopes have significantly higher binding affinities than do HLA-B epitopes, and a significant positive correlation was found between the binding fraction and the binding affinity of HLA molecules.

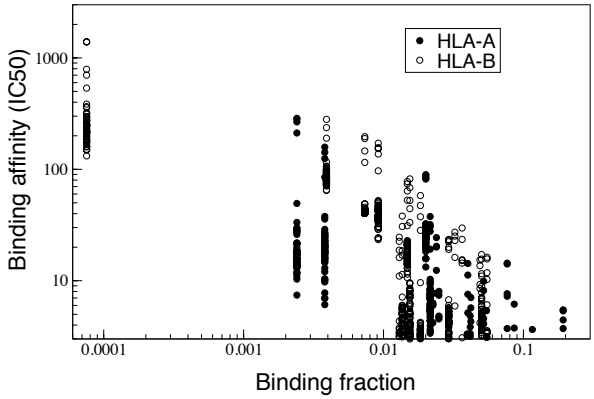


Figure 4. Correlation between binding fraction and binding affinity.
A significant correlation (correlation of -0.25; $p<0.001$, Mann-Whitney U test) was found between predicted binding fraction of 19 HLA-A and 15 HLA-B molecules and predicted binding affinity of their top 5 epitopes in bacterial proteomes. Filled circles depict HLA-A epitopes and empty circles depict HLA-B epitopes. Similar results were found in human and viral proteomes (human: correlation of -0.32, $p<0.001$; virus: correlation of -0.31, $p<0.001$).

HLA-A and HLA-B molecules have different restrictions in anchor residues

Sequence patterns in a cohort of epitopes specific for a certain HLA molecule define its binding motif (38). Some positions of HLA binding motif (e.g. P2, P3, P5, P9), which are very important for epitope binding, were defined as anchor positions (39-42).

To explore if the difference in binding specificity of HLA-A and HLA-B molecules is due to their binding motifs, all 9-mer epitopes from our experimental binding data were used to measure the degree of variability at each position in terms of the Shannon entropy(34), as explained in *Materials and Methods*. The larger Shannon entropy indicates higher level of variation in a position. Only the sequence conservation in P2 and P9 was significantly different between HLA-A and HLA-B epitopes ($p<0.001$, Mann-Whitney *U* test, Fig. 5). Interestingly, P2 is the most conserved position for HLA-B alleles, while for HLA-A alleles P9 has the lowest Shannon entropy (Fig. 5).

Classifying all HLA molecules by their 9-mer binding motif patterns, we observed three groups of binding motifs. Group 1 includes HLA alleles with binding motifs that have a similar level of conservation in P2 and P9, and contains both HLA-A and HLA-B molecules. Group 2 contains only HLA-B molecules, defined by a markedly conserved P2. Group 3 encloses only HLA-A molecules, and these share a conservative P9 in their binding motifs. Supplementary Fig. S5 gives two examples for each of these three groups using sequence logos (see figure caption for explanation).

Thus, in general both HLA-A and HLA-B molecules are more conserved in their anchor residues (e.g., P2 and P9), where P2 is more important for HLA-B alleles and P9 is more important for HLA-A alleles.

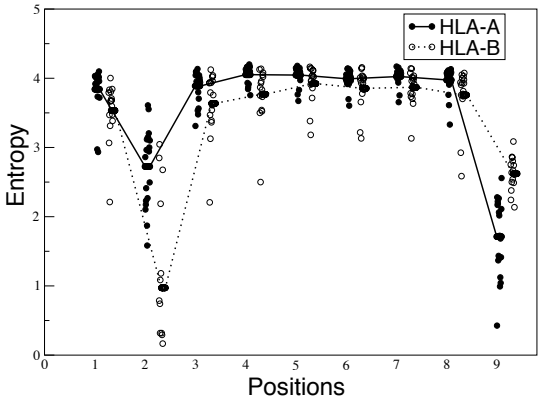


Figure 5. Variability in experimentally verified HLA-A and HLA-B epitopes.
The entropy of HLA-A (filled circles) and HLA-B (open circles) epitopes were calculated based on 7238 unique epitopes (length of 9) of 19 HLA-A alleles and 1585 unique epitopes (length of 9) of 16 HLA-B alleles downloaded from IEDB database. Only in P2 and P9 is the variability of HLA-A and HLA–B epitopes significantly different (both cases $p<0.001$, Mann-Whitney *U* test). Shannon entropy for P2 of HLA-B epitopes is significantly lower ($p=0.017$, Mann-Whitney *U* test) than that of the best conserved position of HLA-A, the P9.

Discussion

Immune responses restricted by HLA-B alleles to several pathogens have been previously shown to be immunodominant (9-11, 13, 14). Additionally, particular HLA-B alleles seem to be associated with either protection or susceptibility to infectious diseases (see reviews refs. (17-20)). It has been suggested that this association of HLA-B molecules with a favorable disease outcome may derive from the fact that they are ligands for NK cell receptors, and thus may promote either NK cell lysis of infected cells or inhibit NK cell function and diminish chronic immune activation (17, 21). Alternatively, the immunodominance of HLA-B-restricted $CD8^+$ T cell responses might be the underlying reason why particular HLA-B alleles of a given individual tend to determine the effectiveness of the immune responses that he/she mounts.

In this study, we undertook two approaches to evaluate if the immunodominance of HLA-B-restricted responses relates to the intrinsic antigen presentation features of these alleles in terms of diversity and binding affinity of the epitopes they present. First, we compared experimental data on pathogen-derived epitopes available in the IEDB database. Second, and to exclude for potential biases in the IEDB epitope set, we performed the same analysis on *in silico*-predicted peptides derived from proteomes of viruses, bacteria and human genomes. This analysis did not include epitopes restricted by HLA-C alleles, given that these have been historically less studied, and thus the data are scarcer and, consequently, no good predictors are currently available.

Surprisingly, we show that HLA-B alleles present a less diverse set of epitopes, ~5-fold lower than that of HLA-A alleles (based on experimental data). This disparity does not seem to be exclusively explained by the large amount of peptides described to be presented by HLA-A2, and is narrower (around 2.5 fold) when derived from prediction data (Fig. 2). The familiar notion that 9-mers are the most preferred MHC class I ligands (43, 44) can be confirmed in this analysis, and it holds true for both loci (Fig. 1). Curiously, our predictions suggest that HLA-A alleles have a preference for foreign over self-epitopes, which does not hold true for HLA-B molecules (Fig. 2), and is lost when the subtle differences among amino acid frequencies are excluded (see supplementary Figs. S2 and S3).

High HLA peptide binding affinity is not strictly correlated with immunodominance (45). Moreover, Assarsson *et al.* showed that dominant epitopes might have lower affinities than subdominant epitopes (37). In line with these findings, we found that the average binding affinity of epitopes restricted by HLA-B alleles is significantly lower than that of HLA-A peptides, in both experimental and predicted data (Fig. 3). This might reflect a tendency for HLA-B molecules to be less dependent on

tapasin-ERp57, a complex involved in editing the peptide cargo of MHC class I in favor of complexes with longer cell surface half-lives enclosing epitopes of higher binding affinity (46-49). In fact, thus far only B alleles have been shown to associate poorly with this chaperone, among which B27 and B*4405 (49-51). HLA dependence on tapasin seems to be dictated by polymorphisms at residues 116 and 114 located at the base of the F-pocket of the binding cleft, which accommodates the peptide C-terminus (49), and the complex has been proposed to act by disrupting peptide-MHC class I complex formation until high-affinity peptides are bound (48).

Interestingly, binding fraction and binding affinity seem to be positively correlated (Fig. 4). HLA-A alleles seem to have a more permissive binding motif, where the most conserved residue is the C-terminal (Fig. 5 and supplementary Fig. S5). Consequently, the chance of finding peptides that fit their looser requirements is higher and will account for a larger pool of possible binders, from which the high affinity peptides may be selected. Conversely, the binding motif of HLA-B alleles appears to be more restricted, especially in P2 (Fig. 5 and supplementary Fig. S5). In accordance, the information content for P2 of HLA-B epitopes is significantly higher ($p=0.017$, Mann-Whitney U test) than that of the best conserved position of HLA-A, the C-terminus. Thus, there is a narrower probability that the small proteome of pathogens will enclose sequences that would perfectly conform to HLA-B prerequisites. As a result, the peptides loaded in HLA-B complexes might frequently be suboptimal.

Another factor that determines the available peptides for each HLA is the specificity of the proteasome. Kesmir *et al*. have shown that the immunoproteasome has co-evolved with human MHC molecules, thus optimizing the process of antigen presentation (52). We, therefore, plan to test next if especially HLA-A alleles have co-evolved with the proteasome, and thus have "designed" their F-pocket as to favor the C-terminus amino acids defined by the cleavage patterns of the catalytic subunits, while HLA-B alleles, evolving faster, may have a more variable F-pocket and not so strict requirements for the C-terminal amino acids.

In summary, HLA-B alleles seem to present fewer peptides to $CD8^+$ T cells, which is consistent with a more restricted binding motif, and epitope binding affinity seems to be lower as compared to their HLA-A counterparts. Note that experimental data on binding affinity are generally determined *in vitro* using synthetic peptides derived from the protein sequence under study, rather than naturally processed peptides. The *in silico* analysis, which we herein use to exclude possible sampling biases in the experimental data, was therefore also performed for all possible 9-mers, and it does not take into account the limiting steps of antigen processing (efficiency of proteasome cleavage, TAP transport, N-terminal trimming). Despite the lack of data on naturally processed

peptides, our results, although unexpected and counterintuitive, call for novel hypotheses to explain why HLA-B alleles are associated with immunodominance.

First, if epitopes indeed bind with higher affinity to HLA-A molecules, these more stable complexes might remain longer at the cell surface. Stimulation of cognate T cells will thus occur longer and may potentially lead to exhaustion of these responses, especially in cases of chronic infection and, thus, antigen persistence. This would be in agreement with the findings of Harari and colleagues, who have shown that HLA-A-restricted epitopes drive high-avidity T cell responses by CTLs that have high expression of the exhaustion marker programmed death (PD)-1 (53). As these high-avidity T cells will be preferentially deleted (54-56), the "surviving" HLA-B-restricted responses would become dominant. In contrast, Almeida and coworkers (57) have found HLA-B T-cell responses to be of higher avidity during HIV infection. Still, the highest avidity response tested, although showing signs of higher clonal turnover and some features of senescence, was polyfunctional and still seemed to be actively involved in controlling HIV infection.

Second, because the pool of B epitopes is less diverse, the antigen "density" could be higher for B-restricted ligands, that is, HLA-B alleles would present epitopes more abundantly (assuming identical A and B expression at the cell surface). As such, the chance that a responsive B-restricted $CD8^+$ T cell would see its ligand would then be increased, and clones responding to these antigens would preferentially expand.

Third, one can envision that the difference in immunodominance may already be defined at the precursor level. In two studies comparing immunodominant and subdominant responses to LCMV (lymphocytic choriomeningitis virus) in mice, differences could be explained by a higher precursor frequency of the former (45, 58). Most interestingly, a recent paper by Obar and colleagues has quantified naïve $CD8^+$ T cells and showed that their frequencies can differ broadly, even more than 10-fold, among different specificities, and that this variation can account for a swifter response – as measured by the timing of the peak response – and may play a role in immunodominance (59). Assuming again that A and B alleles will be expressed in the same level in the thymus, the diversity, abundance and affinity of the self-epitopes presented via these molecules might determine the outcome of positive and negative selection, and thus the naïve precursor level. However, the net effect is hard to predict. It seems reasonable to think that the higher binding affinity of peptides for HLA-A will allow the complexes they form to be more stable at the cell surface, hence an increased chance of promoting survival during the positive selection, but also of being recognized with too high affinity and inducing deletion. Additionally, taking into account that HLA-B molecules will present a less diverse repertoire of epitopes, but potentially to a

higher cell surface density, it is tempting to speculate that they may select for larger epitope-specific clones.

In this study, we showed that the HLA locus more associated with immunodominant T cell responses seems to be, intriguingly, the one exhibiting the least diverse peptide-binding repertoire and presenting peptides with a lower average binding affinity. Although one might conclude that features other than epitope diversity and binding affinity might determine immunodominance, we have explored, above, plausible mechanisms on how these traits can still play a major role. More detailed studies, for example the determination of HLA-A and B restricted T cell precursor frequencies, may resolve this discrepancy.

Acknowledgements

We thank Julie Grenouillet for generating scoring matrices, Rob de Boer for valuable comments on the manuscript, and Jorg Calis and Boris Schmid for technical support.

References

1. Parham P, and Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science.* 1996;272(5258):67-74.
2. Hughes AL, and Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature.* 1988;335(6186):167-70.
3. Vogel TU, Evans DT, Urvater JA, O'Connor DH, Hughes AL, and Watkins DI. Major histocompatibility complex class I genes in primates: co-evolution with pathogens. *Immunological reviews.* 1999;167:327-37.
4. Cooke GS, and Hill AV. Genetics of susceptibility to human infectious disease. *Nature reviews Genetics.* 2001;2(12):967-77.
5. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, and Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. *Current biology : CB.* 2005;15(11):1022-7.
6. Borghans JA, Beltman JB, and De Boer RJ. MHC polymorphism under host-pathogen coevolution. *Immunogenetics.* 2004;55(11):732-9.
7. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, et al. The DNA sequence and analysis of human chromosome 6. *Nature.* 2003;425(6960):805-11.
8. McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, and Watkins DI. A uniquely high level of recombination at the HLA-B locus. *Proceedings of the National Academy of Sciences of the United States of America.* 1994;91(13):5893-7.
9. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature.* 2004;432(7018):769-75.
10. Bihl F, Frahm N, Di Giammarino L, Sidney J, John M, Yusim K, Woodberry T, Sango K, Hewitt HS, Henry L, et al. Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *Journal of immunology.* 2006;176(7):4094-101.
11. Boon AC, De Mutsert G, Fouchier RA, Sintnicolaas K, Osterhaus AD, and Rimmelzwaan GF. Preferential HLA usage in the influenza virus-specific CTL response. *Journal of immunology.* 2004;172(7):4435-43.
12. Lacey SF, Villacres MC, La Rosa C, Wang Z, Longmate J, Martinez J, Brewer JC, Mekhoubad S, Maas R, Leedom JM, et al. Relative dominance of HLA-B*07 restricted CD8+ T-lymphocyte immune responses to human cytomegalovirus pp65 in persons sharing HLA-A*02 and HLA-B*07 alleles. *Human immunology.* 2003;64(4):440-52.
13. Lewinsohn DA, Winata E, Swarbrick GM, Tanner KE, Cook MS, Null MD, Cansler ME, Sette A, Sidney J, and Lewinsohn DM. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS pathogens.* 2007;3(9):1240-9.
14. Weichold FF, Mueller S, Kortsik C, Hitzler WE, Wulf MJ, Hone DM, Sadoff JC, and Maeurer MJ. Impact of MHC class I alleles on the M. tuberculosis antigen-specific CD8+ T-cell response in patients with pulmonary tuberculosis. *Genes and immunity.* 2007;8(4):334-43.
15. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, and O'Brien SJ. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science.* 1999;283(5408):1748-52.
16. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, et al. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *The New England journal of medicine.* 2001;344(22):1668-75.
17. Martin MP, and Carrington M. Immunogenetics of viral infections. *Current opinion in immunology.* 2005;17(5):510-6.

18. Carrington M, and O'Brien SJ. The influence of HLA genotype on AIDS. *Annual review of medicine.* 2003;54:535-51.

19. Neumann-Haefelin C, McKiernan S, Ward S, Viazov S, Spangenberg HC, Killinger T, Baumert TF, Nazarova N, Sheridan I, Pybus O, et al. Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution. *Hepatology.* 2006;43(3):563-72.

20. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, and Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. *Nature.* 1991;352(6336):595-600.

21. Khakoo SI, and Carrington M. KIR and disease: a model system or system of models? *Immunological reviews.* 2006;214:186-201.

22. Yewdell JW, and Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annual review of immunology.* 1999;17:51-88.

23. Yewdell JW. Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. *Immunity.* 2006;25(4):533-43.

24. Burroughs NJ, de Boer RJ, and Kesmir C. Discriminating self from nonself with short peptides from large proteomes. *Immunogenetics.* 2004;56(5):311-20.

25. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, and Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics.* 1999;50(3-4):213-9.

26. Gulukota K, Sidney J, Sette A, and DeLisi C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *Journal of molecular biology.* 1997;267(5):1258-67.

27. Parker KC, Bednarek MA, and Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *Journal of immunology.* 1994;152(1):163-75.

28. Segal MR, Cummings MP, and Hubbard AE. Relating amino acid sequence to phenotype: analysis of peptide-binding data. *Biometrics.* 2001;57(2):632-42.

29. Doytchinova IA, Blythe MJ, and Flower DR. Additive method for the prediction of protein-peptide binding affinity. Application to the MHC class I molecule HLA-A*0201. *Journal of proteome research.* 2002;1(3):263-72.

30. Peters B, Tong W, Sidney J, Sette A, and Weng Z. Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. *Bioinformatics.* 2003;19(14):1765-72.

31. Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, et al. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS computational biology.* 2006;2(6):e65.

32. Schellens IM, Kesmir C, Miedema F, van Baarle D, and Borghans JA. An unanticipated lack of consensus cytotoxic T lymphocyte epitopes in HIV-1 databases: the contribution of prediction programs. *Aids.* 2008;22(1):33-7.

33. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, and Sette A. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell.* 1993;74(5):929-37.

34. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal* 1948;27:379.

35. Schneider TD, and Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic acids research.* 1990;18(20):6097-100.

36. Peters B, and Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinformatics.* 2005;6:132.

37. Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui HH, Frahm N, Brander C, Peters B, Grey H, and Sette A. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *Journal of immunology.* 2007;178(12):7890-901.

38. Rammensee HG, Friede T, and Stevanoviic S. MHC ligands and peptide motifs: first listing. *Immunogenetics.* 1995;41(4):178-228.

39. Falk K, Rotzschke O, Grahovac B, Schendel D, Stevanovic S, Jung G, and Rammensee HG. Peptide motifs of HLA-B35 and -B37 molecules. *Immunogenetics.* 1993;38(2):161-2.

40. Takamiya Y, Schonbach C, Nokihara K, Yamaguchi M, Ferrone S, Kano K, Egawa K, and Takiguchi M. HLA-B*3501-peptide interactions: role of anchor residues of peptides in their binding to HLA-B*3501 molecules. *International immunology.* 1994;6(2):255-61.

41. DiBrino M, Parker KC, Margulies DH, Shiloach J, Turner RV, Biddison WE, and Coligan JE. The HLA-B14 peptide binding site can accommodate peptides with different combinations of anchor residues. *The Journal of biological chemistry.* 1994;269(51):32426-34.

42. Lamas JR, Paradela A, Roncal F, and Lopez de Castro JA. Modulation at multiple anchor positions of the peptide specificity of HLA-B27 subtypes differentially associated with ankylosing spondylitis. *Arthritis and rheumatism.* 1999;42(9):1975-85.

43. Falk K, Rotzschke O, Stevanovic S, Jung G, and Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature.* 1991;351(6324):290-6.

44. Yewdell JW, Reits E, and Neefjes J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature reviews Immunology.* 2003;3(12):952-61.

45. Kotturi MF, Scott I, Wolfe T, Peters B, Sidney J, Cheroutre H, von Herrath MG, Buchmeier MJ, Grey H, and Sette A. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. *Journal of immunology.* 2008;181(3):2124-33.

46. Howarth M, Williams A, Tolstrup AB, and Elliott T. Tapasin enhances MHC class I peptide presentation according to peptide half-life. *Proceedings of the National Academy of Sciences of the United States of America.* 2004;101(32):11737-42.

47. Kienast A, Preuss M, Winkler M, and Dick TP. Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. *Nature immunology.* 2007;8(8):864-72.

48. Wearsch PA, and Cresswell P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nature immunology.* 2007;8(8):873-81.

49. Purcell AW, and Elliott T. Molecular machinations of the MHC-I peptide loading complex. *Current opinion in immunology.* 2008;20(1):75-81.

50. Turnquist HR, Thomas HJ, Prilliman KR, Lutz CT, Hildebrand WH, and Solheim JC. HLA-B polymorphism affects interactions with multiple endoplasmic reticulum proteins. *European journal of immunology.* 2000;30(10):3021-8.

51. Groothuis TA, Griekspoor AC, Neijssen JJ, Herberts CA, and Neefjes JJ. MHC class I alleles and their exploration of the antigen-processing machinery. *Immunological reviews.* 2005;207:60-76.

52. Kesmir C, van Noort V, de Boer RJ, and Hogeweg P. Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome. *Immunogenetics.* 2003;55(7):437-49.

53. Harari A, Cellerai C, Bellutti Enders F, Kostler J, Codarri L, Tapia G, Boyman O, Castro E, Gaudieri S, James I, et al. Skewed association of polyfunctional antigen-specific CD8 T cell populations with HLA-B genotype. *Proceedings of the National Academy of Sciences of the United States of America.* 2007;104(41):16233-8.

54. Van Parijs L, and Abbas AK. Homeostasis and self-tolerance in the immune system: turning lymphocytes off. *Science.* 1998;280(5361):243-8.

55. Molldrem JJ, Lee PP, Kant S, Wieder E, Jiang W, Lu S, Wang C, and Davis MM. Chronic myelogenous leukemia shapes host immunity by selective deletion of high-avidity leukemia-specific T cells. *The Journal of clinical investigation.* 2003;111(5):639-47.

56. Lichterfeld M, Yu XG, Mui SK, Williams KL, Trocha A, Brockman MA, Allgaier RL, Waring MT, Koibuchi T, Johnston MN, et al. Selective depletion of high-avidity human immunodeficiency virus type 1 (HIV-1)-specific CD8+ T cells after early HIV-1 infection. *Journal of virology.* 2007;81(8):4199-214.

57. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, Bornstein E, Asher TE, Samri A, Schnuriger A, Theodorou I, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of experimental medicine.* 2007;204(10):2473-85.

58. De Boer RJ, Homann D, and Perelson AS. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *Journal of immunology.* 2003;171(8):3928-35.

59. Obar JJ, Khanna KM, and Lefrancois L. Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection. *Immunity.* 2008;28(6):859-69.

Supplementary material

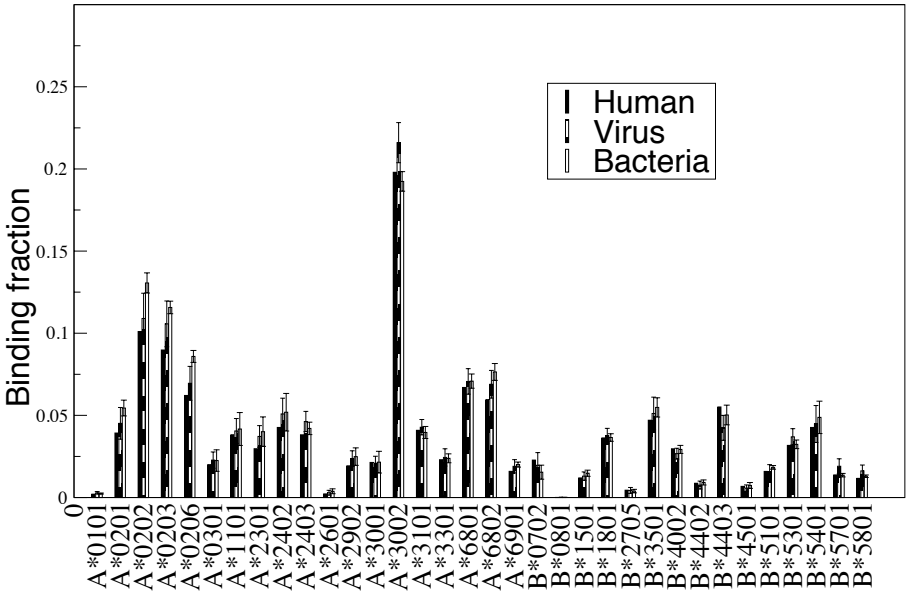


Figure S1. Binding fractions of individual HLA alleles in human, viral and bacterial proteomes. Average binding fractions, predicted by SMM method, for individual alleles were displayed for the different proteomes (human: filled bars; virus: shaded bars; bacteria: empty bars).

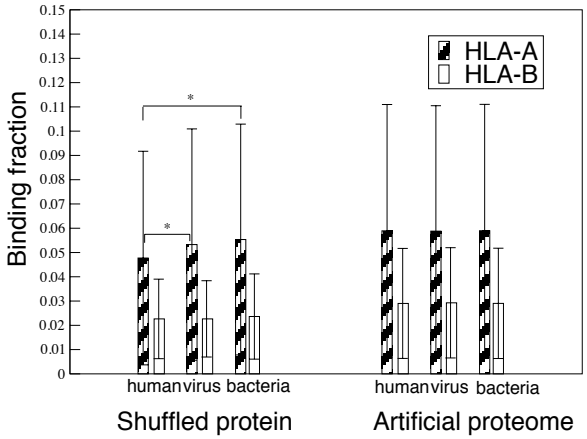


Figure S2. Average binding fraction of HLA-A and HLA-B alleles in shuffled and artificial proteomes.
Left: To test whether HLA-A and -B binding is sensitive to sequence patterns accruing in natural proteins, every protein in human, viral and bacterial proteomes was shuffled and average binding fraction of HLA-A (shaded bars) and -B (open bars) alleles was calculated after shuffling. HLA-A alleles still present significantly more non-self ($p<0.001$, Mann-Whitney test) *Right:* Similarly, to see effect of amino acid frequencies on HLA binding, we generated artificial human, viral and bacterial proteomes, where every amino acid has equal frequency. In these artificial proteomes neither HLA-A nor –B alleles have a preference for presenting non-self.

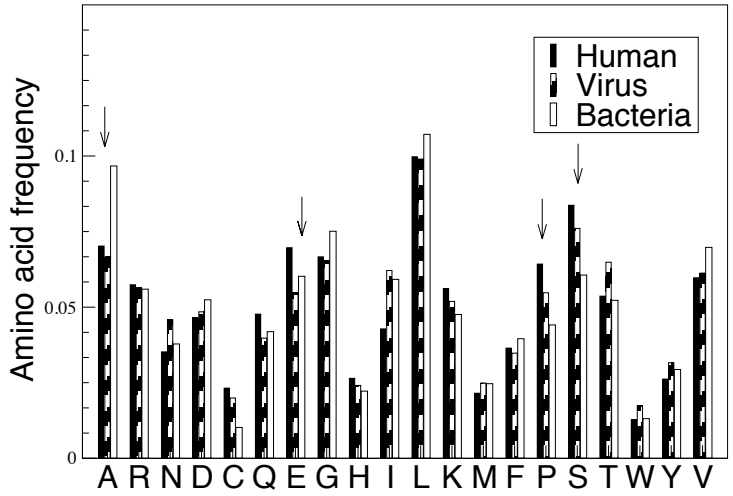


Figure S3. Amino acid frequencies in human, viral and bacterial proteomes.
Filled bars indicate the frequencies of 20 amino acids in human proteome, shaded bars present the amino acids frequencies in a combined set of 17 different virus proteomes and similarly empty bars depict the amino acids frequencies in a combined set of 13 bacterial proteomes (see Table S1 for list of viral and bacterial proteomes). It seems that Alanine (A) and Isoleucine (I) occur in general substantially more in viral and bacterial proteomes, while Glutamic acid (E), Proline (P) and Serine (S) are to some extent abundant in the human proteome.

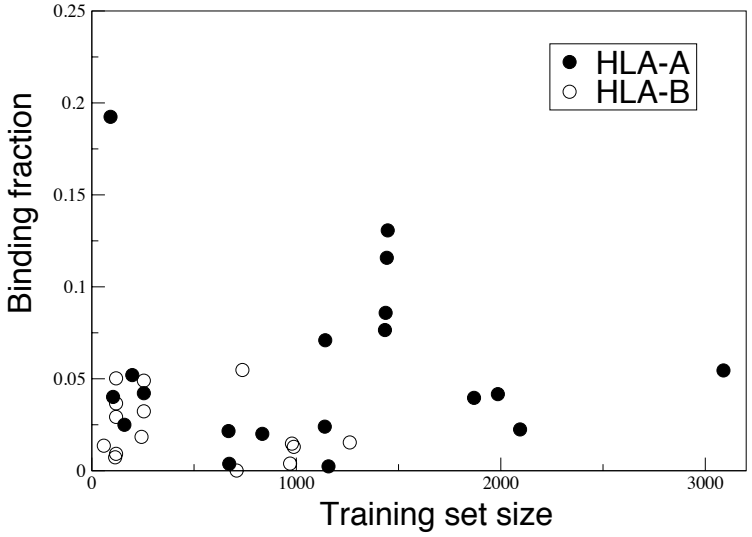


Figure S4. Lack of Correlation between predicted binding fraction and training set size for HLA-A and B.
The number of 9-mers used to train HLA-A and HLA-B predictors (in Peters et al, 2006) is plotted against the average binding fraction for the respective HLA alleles for bacterial proteomes. Similar results are obtained for human and viral proteomes (data not shown). No significant correlation was found between training set size and average binding fraction of the HLA alleles ($p>0.50$ for all proteomes tested). Separating HLA alleles into HLA-A and HLA-B did not change this result.

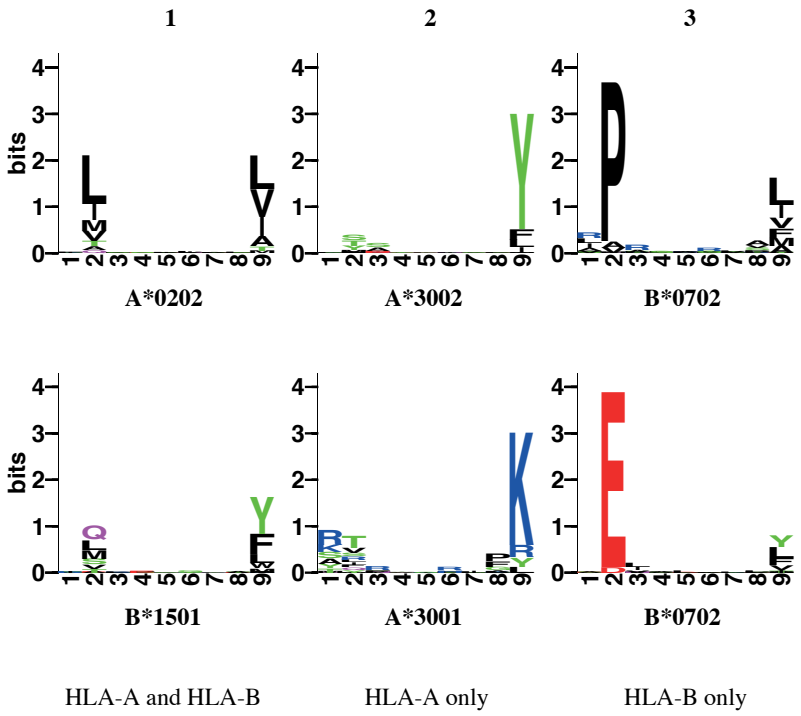


Figure S5. Examples of HLA-A and HLA-B binding motifs.
The binding motifs of HLA molecules were visualized by sequence logo technique. In the logo plot, the height of each bar gives the Shannon information content at that position which can be used as a measure of amino acid conservation. The higher Shannon information content represents a more conserved position. Amino acids are color coded according to their physicochemical characteristics. For example neutral and polar amino acids are green, basic amino acid is blue, acidic amino acid is marked in red and neutral or hydrophobic amino acid is black.
HLA-A and HLA-B molecules can be classified into three subgroups by their binding motifs. *Left panel:* Two examples of HLA molecules whose binding motif has similar conservation at P2 and P9. *Middle panel:* Two examples of HLA molecules whose binding motif is most conserved in P9. This group contains only HLA-A molecules. *Right panel:* Two examples of HLA molecules whose binding motif is most conserved in P2. This group contains only HLA-B molecules.

Table S1. List of viral and bacterial proteomes used in this study.

Species	Accession number
VIRUS	
Human immunodeficiency virus 1	X01762
Dengue virus 1	U88536
Reston ebolavirus	AB050936
Hepatitis A virus	M14707
Hepatitis B virus	X51970
Hepatitis C virus	AJ132997
Human T-lymphotropic virus 1	D13784
Influenza A virus (A/Goose/Guangdong/1/96(H5N1)) segment 1-8	AF144300-AF144307
Measles virus	K01711
Mumps virus	AB040874
H-1 parvovirus	X01457
Human poliovirus 1	AJ132961
Rabies virus	M31046
Human respiratory syncytial virus	AF013254
Rubella virus	AF188704
Sendai virus	M69046
Yellow fever virus	X03700
BACTERIA	
Chlamydophila pneumoniae AR39	AE002161
Rickettsia conorii str. Malish 7	AE006914
Escherichia coli 536	CP000247
Salmonella typhimurium LT2	AE006468
Helicobacter pylori 26695	AE000511
Staphylococcus aureus RF122	AJ938182
Mycobacterium tuberculosis CDC1551	AE000516
Streptococcus pneumoniae CGSP14	CP001033
Neisseria meningitidis 053442	CP000381
Vibrio cholerae O1 biovar eltor str. N16961	AE003852
Pseudomonas aeruginosa plasmid pMATVIM-7	AM778842
Yersinia pestis Angola	CP000901
Bacillus subtilis subsp. subtilis str. 168	AL009126

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

Ana I. Fontaine Costa[1,*], Xiangyu Rao[2,*], Emmanuelle LeChenadec[2], Debbie van Baarle[1,4], Can Keşmir[2,3]

[1]*Department of Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands.*

[2]*Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands.*

[3]*Academic Biomedical Centre, Utrecht University, Utrecht, The Netherlands.*

[4]*Department of Internal Medicine and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands.*

*These authors contributed equally.

AIDS (2010), 24:211-215.

Abstract

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.

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 (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).

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 vs 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 black 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 frameshift mutation in Vpr (position 5772).

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 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 to HLA-A counterparts (reversions:

HLA-A = 5; HLA-B = 22; escapes: HLA-A=6, HLA-B=26) ($p<0.01$, Chi-Square; 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 over-expressed 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.

References

1. Goulder PJ, and Watkins DI. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature reviews Immunology.* 2008;8(8):619-30.
2. Bihl F, Frahm N, Di Giammarino L, Sidney J, John M, Yusim K, Woodberry T, Sango K, Hewitt HS, Henry L, et al. Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *Journal of immunology.* 2006;176(7):4094-101.
3. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal* 1948;27:379-423.
4. Borghans JA, Molgaard A, de Boer RJ, and Kesmir C. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PloS one.* 2007;2(9):e920.
5. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, Treurnicht F, de Rosa DA, Hide W, Karim SA, et al. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. *PLoS pathogens.* 2008;4(3):e1000033.
6. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, and 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. *Journal of virology.* 2002;76(5):2298-305.
7. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, Derdeyn CA, Tang J, Kaslow RA, Bansal A, et al. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *The Journal of experimental medicine.* 2008;205(5):1009-17.
8. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine.* 2007;13(1):46-53.
9. Novitsky V, Gilbert P, Peter T, McLane MF, Gaolekwe S, Rybak N, Thior I, Ndung'u T, Marlink R, Lee TH, et al. Association between virus-specific T-cell responses and plasma viral load in human immunodeficiency virus type 1 subtype C infection. *Journal of virology.* 2003;77(2):882-90.
10. Rolland M, Heckerman D, Deng W, Rousseau CM, Coovadia H, Bishop K, Goulder PJ, Walker BD, Brander C, and Mullins JI. Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads. *PloS one.* 2008;3(1):e1424.
11. Zuniga R, Lucchetti A, Galvan P, Sanchez S, Sanchez C, Hernandez A, Sanchez H, Frahm N, Linde CH, Hewitt HS, et al. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *Journal of virology.* 2006;80(6):3122-5.
12. Frahm N, Korber BT, Adams CM, Szinger JJ, Draenert R, Addo MM, Feeney ME, Yusim K, Sango K, Brown NV, et al. Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. *Journal of virology.* 2004;78(5):2187-200.
13. Altfeld M, Addo MM, Shankarappa R, Lee PK, Allen TM, Yu XG, Rathod A, Harlow J, O'Sullivan K, Johnston MN, 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. *Journal of virology.* 2003;77(13):7330-40.
14. Betts MR, Ambrozak DR, Douek DC, Bonhoeffer S, Brenchley JM, Casazza JP, Koup RA, and Picker LJ. Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection. *Journal of virology.* 2001;75(24):11983-91.

15. Addo MM, Yu XG, Rathod A, Cohen D, Eldridge RL, Strick D, Johnston MN, Corcoran C, Wurcel AG, Fitzpatrick CA, 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. *Journal of virology.* 2003;77(3):2081-92.

16. Yusim K, Kesmir C, Gaschen B, Addo MM, Altfeld M, Brunak S, Chigaev A, Detours V, and Korber BT. 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. *Journal of virology.* 2002;76(17):8757-68.

17. Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, Schneidewind A, Power KA, Toth I, Frahm N, 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. *Journal of virology.* 2009;83(4):1845-55.

18. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, and Allen TM. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *Journal of virology.* 2007;81(1):193-201.

19. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature.* 2009;458(7238):641-5.

20. Rapin N, Hoof I, Lund O, and Nielsen M. MHC motif viewer. *Immunogenetics.* 2008;60(12):759-65.

21. Schellens IM, Kesmir C, Miedema F, van Baarle D, and Borghans JA. An unanticipated lack of consensus cytotoxic T lymphocyte epitopes in HIV-1 databases: the contribution of prediction programs. *Aids.* 2008;22(1):33-7.

HLA-B class I allele promiscuity revisited

Xiangyu Rao[1], Ilka Hoof[1], Ana Isabel C.A. Fontaine Costa[2], Debbie van Baarle[2], Can Keşmir[1]

[1]*Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands.*

[2]*Department of Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands.*

Immunogenetics (2011), 63:691-701.

Abstract

The peptide repertoire presented on human leukocyte antigen (HLA) class I molecules is largely determined by the structure of the peptide binding groove. It is expected that the molecules having similar grooves (i.e., belonging to the same supertype) might present similar/overlapping peptides. However, the extent of promiscuity among HLA class I ligands remains controversial: while in many studies T cell responses are detected against epitopes presented by alternative molecules across HLA class I supertypes and loci, peptide elution studies report minute overlaps between the peptide repertoires of even related HLA molecules. To get more insight into the promiscuous peptide binding by HLA molecules, we analyzed the HLA peptide binding data from the large epitope repository, Immune Epitope Database (IEDB), and further performed *in silico* analysis to estimate the promiscuity at the population level. Both analyses suggest that an unexpectedly large fraction of HLA ligands (>50%) bind two or more HLA molecules, often across supertype or even loci. These results suggest that different HLA class I molecules can nevertheless present largely overlapping peptide sets, and that "functional" HLA polymorphism on individual and population level is probably much lower than previously anticipated.

Introduction

The repertoire of the cytotoxic T cell response is shaped by the peptides presented on human leukocyte antigen (HLA) class I molecules. HLA class I genes are the most polymorphic loci known in the human genome: more than 2000 HLA-A and -B allelic variants have been reported (1-3). Most polymorphism is accumulated in the peptide binding groove of these molecules, giving rise to specific binding motifs for every HLA molecule, which allow for selective binding of a set of peptides forming the so called ligand-binding repertoire. HLA class I molecules may be grouped into several supertypes based on their potential binding motif similarity (4-10).

The specificity of HLA binding has been studied extensively in the last 30 years. It has become clear early on that some HLA molecules may significantly overlap in peptide binding specificity (11-15), meaning that one peptide ligand has the ability to bind to several HLA molecules. The majority of ligand sharing was observed among molecules that have similar binding motifs and therefore would be assigned to the same HLA supertype, i.e., a form of promiscuity that may be considered as "expected" (16-22). Few other reports showed promiscuity across supertypes or even loci; these findings were considered as "exceptions" (23, 24). Recently, however, two systematic studies challenged this general view on promiscuity of HLA class I peptides and reported unexpected but also conflicting results. Frahm *et al.* (25) tested T cell responses to 242 well-defined viral epitopes from HIV and EBV in 100 subjects and found that 95% of these epitopes elicited a T cell response in at least one individual not expressing the original restricting HLA molecule. The majority of potential alternative HLA molecules were not matched to the same HLA supertype or even the same locus as the original restricting HLA molecule. Shortly after this study, Hillen *et al.* (26) reported only minute overlaps (3%) between the epitope repertoires of HLA molecules belonging to the B44 supertype, based on several hundreds of eluted peptides from nine members of the HLA-B44 supertype. This result was very surprising also because Sidney *et al.* (19) reported largely overlapping peptide-binding repertoires for HLA molecules belonging to HLA-B44 family, based on *in vitro* MHC binding experiments. Of note, the experiments of Hillen *et al.* resulted in less than 30 peptides for some of the HLA molecules belonging to this supertype (e.g., B*5001, B*4701 and B*4501, see Hillen *et al.* (26) and Table 3 of this article), suggesting that the peptide elution approach might underestimate the peptide binding repertoire of an HLA molecule.

Here we study the ligand sharing among HLA class I molecules by carrying out a systematic study, in which we analyze the data from Frahm *et al.* and Hillen *et al.*

together with a large amount of data available from the IEDB (www.iedb.org; (27)) database. Although the experimental data in IEDB on MHC binding is extensive, it nevertheless does not provide a reliable estimate of promiscuous binding for every HLA molecule, because the number of HLA molecules that can bind the same peptide depends largely on the number of HLA molecules for which *in vitro* binding data is available for the peptide in question. To avoid this problem, we repeated the same analysis using state-of-the-art MHC class I binding predictors (28-30), where we estimate the extent of promiscuous peptide binding by taking into account every common HLA molecule in the population. In all cases, our results suggest that more than 60% of HLA ligands show promiscuous binding. Finally, we discuss consequences of the extensive ligand sharing among HLA class I molecules in the context of immunodominance and infectious diseases.

Materials and Methods

Experimental MHC binding and T cell response data

The experimental data used in our analysis was extracted from the Immune Epitope Database and Analysis Resource (IEDB; www.immuneepitope.org; downloads were made in March 2010). The first data set included all peptides for which the HLA class I binding affinity was determined by *in vitro* MHC binding assays; the second data set consisted of peptides with measured T cell responses. We considered only peptides that were tested on at least six HLA class I molecules and with an IC_{50} value lower than 500 nM for at least one of these molecules (i.e. the peptide has to be a binder for at least one HLA molecule). In addition, to make sure that all HLA-peptide associations were well defined, only the data with four-digit HLA class I identifiers were included. These selection criteria resulted in a set of 3738 non-redundant peptides obtained from the MHC binding assay database of IEDB and is here referred to as "IEDB MHC binding data". The T cell response data, filtered using the same criteria, resulted in a much smaller data set. Therefore, we relaxed the requirements on four-digit HLA identifiers by including T cell response data for which only one- or two-digit HLA identifiers were available. In total, filtering of the IEDB T cell assay data resulted in 135 non-redundant T cell epitopes.

Quantifying HLA binding promiscuity

We divided the ligands of a particular HLA class I molecule into two groups: i) unique ligands, which are exclusively presented by this HLA class I molecule; ii) promiscuous

ligands, which are capable to bind to at least one other HLA class I molecule. We define the fraction of the unique ligands as, $F_u = \frac{N_u}{N_{all}}$, where N_u is the number of unique ligands, and N_{all} is the total number of ligands. In order to estimate this fraction as reliably as possible, we calculated it only for the HLA molecules for which more than 10 peptides were experimentally tested on alternative HLA molecules. Changing this arbitrary threshold of 10 (to 20 or 30) did not change the results reported in the text (data not shown).

Promiscuity at supertype level

Throughout our analysis, we followed allele-supertype associations defined by Sidney *et al.* (10) except in a single case: HLA-B*4901 is not classified into any supertype by Sidney *et al.*, and therefore, we assigned it to the B44 supertype, as was done by Hillen et al. (26).

In silico analysis

HLA allele selection

HLA allele frequencies were obtained from the National Marrow Donor Program (NMDP) website (bioinformatics.nmdp.org) for four predominant US census categories of race and ethnicity: African Americans, Asians, European Americans and Hispanics (31, 32). We included the 20 most frequent HLA-A and 20 most frequent HLA-B alleles for each ethnic group into our *in silico* analysis (Table S4A, S4B). The peptide-MHC binding predictions for majority of these molecules are of high quality (30).

HLA class I ligand prediction

To have an as large as possible population coverage, we used NetMHCpan (30) to predict peptide-HLA binding affinity. NetMHCpan assigns to each peptide-HLA pair a predicted IC_{50} value, indicative of the predicted binding affinity. An IC_{50} threshold of 500 nM was used to discriminate HLA binding ligands from nonbinding peptides. NetMHCpan is not an allele-specific method: it has been trained on peptide binding data for many different MHC molecules (also from non-human species), and its prediction relies on intra- and extrapolation from characterized to uncharacterized HLA alleles. Thus, NetMHCpan may overestimate the promiscuity of HLA class I peptides. In order to check this issue, we compared NetMHCpan with an allele-specific predictor NetMHC3.2 (29, 33, 34). A total of 42 HLA molecules (21 HLA-A and 21 HLA-B) have NetMHC3.2 predictions available. For each HLA molecule, we calculated the fraction of unique ligands to estimate the promiscuity of its ligands, by using the prediction results

obtained from NetMHC3.2 and NetMHCpan. Predicted promiscuity of HLA class I binding by both predictors is highly correlated ($p < 0.001$, $r=0.86$, Pearson correlation test). Moreover, both predictors estimate the fraction of promiscuous ligands restricted by these 42 HLA alleles to be around 58-60% (data not shown and Fig. 1). These results suggest that using NetMHCpan, which has a broader population coverage than NetMHC, would not result in an overestimation of promiscuity of HLA ligands.

Using a fixed threshold of 500 nM IC_{50} to define predicted binders may result in differences in predicted repertoire sizes between HLA molecules, which in turn may introduce a bias into the promiscuity analysis (35). To avoid this, we repeated the analysis by defining the top 1% of the peptides as candidate binders for each HLA molecule, thereby ensuring the same ligand repertoire size for each HLA molecule. With this alternative scaled threshold approach, all *in silico* results reported in this paper remain unchanged. For example, for the European subpopulation, the predicted promiscuity of HLA class I binders is 57% with the fixed threshold of 500 nM and 54% with the scaled threshold.

Predicting antigen processing

The Stabilized Matrix Method (SMM) was applied to predict the TAP (transporter associated with antigen processing) transport efficiency and proteasomal cleavage, which are the two main steps of antigen processing (36). Applying an alternative predictor of antigen processing, NetChop (37, 38), did not affect our results.

Viral data

The proteomes of 17 common human viruses were downloaded from the European Bioinformatics Institute website (www.ebi.ac.uk; downloads were made in Oct 2006, listed in Table S1) as the source of potential HLA ligands. We used the HLA, TAP and proteasome predictors to screen all possible unique virus-derived 9-mer peptides for potential HLA ligands. This data set is later extended with the viruses given in Table S3 to test the dependence of our results on the initial set of viral proteomes.

Results

HLA class I binding shows a high degree of promiscuity

To our knowledge, Frahm *et al.* (25) were the first to study HLA class I binding promiscuity systematically. In short, a total of 242 known HIV-1 and EBV epitopes were tested in a cohort of 100 (50 HIV-1 infected and 50 healthy) subjects regardless of the individual's HLA type. This cohort had a diverse HLA distribution, covering 46

(common) HLA-A, -B, and -C molecules. Almost all of the tested epitopes, 95%, elicited a response in at least one individual not expressing the original restricting HLA molecule. Using two independent statistical approaches, Frahm *et al.* predicted the alternative HLA molecules. Surprisingly, the majority of potential alternative HLA molecules were outside the original restricting molecule's supertype or even the locus (25). Using the pan-specific MHC class I binding predictor NetMHCpan (28, 30) we confirmed 91% of these alternative HLA restrictions among the most significant associations and 75% of all significant associations. This result suggests that the responses identified by Frahm *et al.* are largely due to promiscuous presentation of the same epitope via two or more HLA class I molecules, instead of possible T cell (CD4 and CD8) cross-reactivity to different (embedded) epitopes presented by HLA class I and II molecules. We observed that the predicted affinity for the alternative HLA molecules in the data of Frahm *et al*. is significantly lower than that for the original restricting HLA ($p=0.001$, Mann-Whitney *U* test, for all associations). This may explain why the responses elicited by alternative HLA molecules could have been overseen so far, even though MHC-peptide binding at lower affinity does not necessarily result in lower T cell responses (39-41).

To test the HLA class I binding promiscuity in an independent data set, we analyzed HLA class I binding data from the IEDB database (27) (details are given in Materials and Methods). This database covers approximately 99% of all publicly available information on peptide epitopes mapped in infectious agents. Obviously, the promiscuity of HLA binding depends on the number of different HLA alleles for which peptide binding is tested. To provide a realistic estimate of promiscuous HLA class I binding, we selected IEDB peptide epitopes for which *in vitro* binding assays were performed on at least six different HLA class I molecules. We will refer to this data set as "IEDB MHC binding data". With this criterion, a total of 3738 HLA class I binding peptides were retrieved, among which 72% were promiscuous, i.e., reported to bind to at least two HLA class I molecules (Table 1). Using a more stringent criterion, e.g., when including only the peptides which were tested on eight or ten HLA molecules, the average promiscuity remained high (>65%, results not shown). In line with the results of Frahm *et al*., 68% of promiscuous HLA class I binding was observed across serotypes, 47% across HLA supertypes, and 23% across HLA loci (Table 1). Although being a much smaller data set, CTL response data from IEDB suggests similar levels of promiscuity: Out of 135 non-redundant CTL epitopes, each of which was tested on at least six HLA alleles, 82 (60%) elicited responses in the context of two or more HLA molecules.

Table 1. Summary of the promiscuity analysis of HLA class I ligands based on IEDB MHC binding data.

Category	Number of ligands	Percentage of ligands (%)
All HLA class I ligands from IEDB (Tested on at least six HLA class I alleles)	3738	
Unique ligands	1062	28
Promiscuous ligands[a]	2676	72
Promiscuous ligands across serotypes[b]	2526	68
Promiscuous ligands across supertypes[c]	1751	47
Promiscuous ligands across loci[d]	864	23

[a] The number of ligands that bind at least two different HLA alleles.
[b] The number of ligands that bind at least two different serotypes.
[c] The number of ligands that bind at least two different HLA supertypes.
[d] The number of ligands that bind at least two different HLA loci.

Finally, to estimate the promiscuous peptide binding on the population level, i.e., to estimate the chance of a peptide being presented by two or more HLA molecules in a population, we repeated a similar analysis using HLA binding predictors and focused on the most frequent 20 HLA-A and 20 HLA-B alleles in four US subpopulations with different ethnicity (European, Hispanic, African and Asian ethnicities, data extracted from the National Marrow Donor Program resource, http://bioinformatics.nmdp.org/ (32)). We predicted potential 9-mer binders to these HLA molecules within common viral proteomes (n=17, see Table S1) using NetMHCpan (28, 30). This prediction method was demonstrated to be the best one in a recent large benchmark performance test (42). The most frequent HLA molecules (top 20 for A- and B-locus, respectively, listed in Table S4) in all four ethnic groups have, on average, a fraction of predicted promiscuous ligands around 60%, of which almost half are predicted to be presented by multiple HLA supertypes (Fig. 1). As expected, using a more stringent threshold to define the peptide binding (e.g., a predicted IC_{50} value of 50 nM instead of 500 nM) decreases the level promiscuity to 35-40%, as the ligand repertoire for each HLA molecule is severely reduced (results not shown). These results were reproducible with another neural network predictor, NetMHC3.2 (29) (see Materials and Methods for a discussion on the choice of peptide-MHC binding predictors). Moreover, defining top 1-2% ranking binders as predicted binders changed only slightly the values reported in Table 1 and Fig. 1 (results not shown). In order to evaluate whether antigen processing has an impact on ligand promiscuity, we then added TAP and proteasomal cleavage predictions (36-38) to our MHC binding predictions. The level of promiscuity remained very similar (data not shown), implying that (predicted) antigen processing does not significantly influence the ligand sharing among HLA class I molecules.

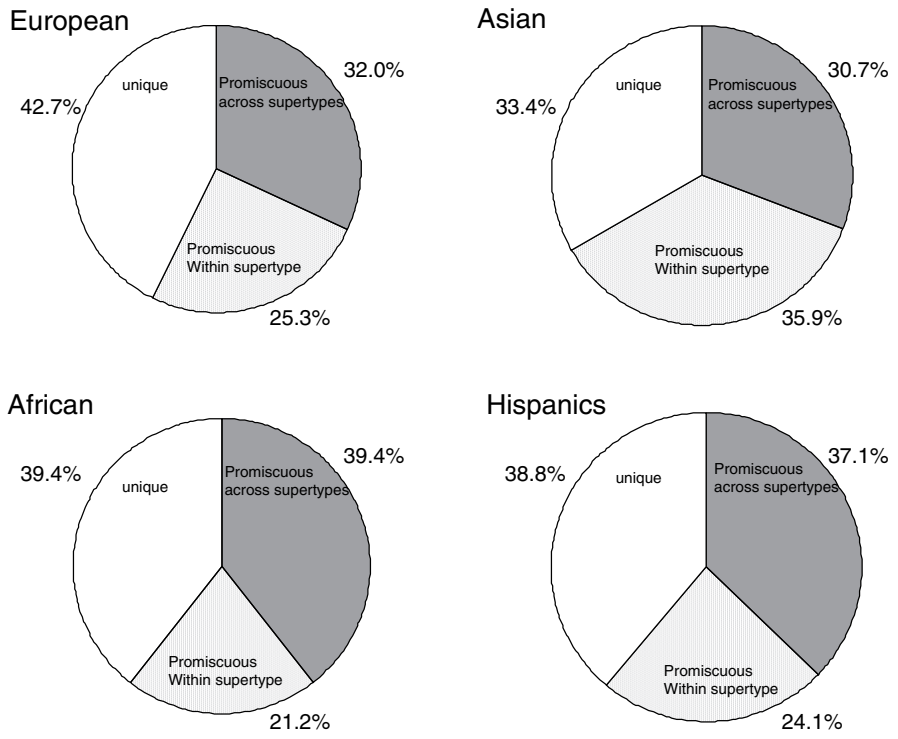


Figure 1. **Distribution of predicted HLA class I ligands of viral origin.**
All predicted ligands of the 20 most frequent HLA-A and HLA-B molecules in US subpopulations of a certain ethnic background (European, African, Asian and Hispanic) were classified into three categories: unique ligands (exclusively presented by one HLA class I molecule), within-supertype promiscuous ligands (exclusively targeted by one HLA supertype, but presented by at least two class I HLA molecules within this supertype) and across-supertype promiscuous ligands (targeted by HLA molecules belonging to at least two different HLA supertypes).

Taken together, not only the data reported by Frahm *et al*., but also HLA binding data available in the large IEDB repository and the analysis of HLA binding predictions on population level strongly suggest that a high fraction of HLA class I ligands (>60%) can bind to two or more HLA molecules and, frequently, the observed promiscuity occurs across different HLA class I supertypes.

Next we detailed our analysis at the supertype level to pinpoint possible differences on ligand sharing among different HLA class I supertypes. Analyzing IEDB MHC binding data on per-supertype basis, we found that every supertype has exceptionally many peptides that exhibit promiscuous HLA binding (Table 2). The HLA-B44 supertype, analyzed by Hillen *et al*. (26) is somewhat an "exception" among other supertypes: while 74% of the B44-ligands reported in IEDB exhibit promiscuous binding, across supertypes promiscuity is much lower at 17% (Table 2). A similar result

was obtained in *in silico* analysis where we estimate the HLA peptide binding promiscuity on the population level (see Table S2).

Table 2. Promiscuity of different HLA supertype ligands based on IEDB MHC binding data

HLA supertype	N[a]	Promiscuous[b]	Promiscuous only in another supertype[c]	Supertype members
HLA-A01	1132	954 (84%)	438 (39%)	10
HLA-A02	868	797 (92%)	462 (53%)	14
HLA-A03	886	783 (88%)	289 (33%)	9
HLA-A24	887	815 (92%)	637 (72%)	4
HLA-B07	601	476 (79%)	375 (62%)	8
HLA-B08	273	200 (73%)	189 (69%)	3
HLA-B27	406	267 (66%)	197 (48%)	11
HLA-B44	624	460 (74%)	103 (17%)	6
HLA-B58	402	346 (86%)	258 (64%)	3
HLA-B62	701	637 (91%)	566 (81%)	3

[a] Total number of unique ligands.
[b] The number of ligands reported to bind at least two different HLA molecules. Within parenthesis, the fraction with respect to the number of ligands is given.
[c] The number (and fraction with respect to the number of ligands) of promiscuous ligands that exclusively bind to HLA molecules outside of the supertype.

Comparison of different experimental approaches for HLA peptide binding promiscuity

Hillen *et al*. (26) undertook a different approach to study HLA class I binding promiscuity by directly comparing ligand repertoires based on peptides eluted from HLA molecules. As this is a very labor-intensive approach, they focused on a single supertype: HLA-B44, of which nine members were included in the elution study (listed in Table 3). Only a very small fraction (25 out of 670, 3%) of the "natural" ligands were found to bind two or more HLA molecules within the HLA-B44 supertype (26). The binding promiscuity of ligands from different allele varies between 18% to none, with an average of 10% (Table 3). This unique data set allows us to compare the estimates of ligand sharing (for B44 supertype only) using three different experimental approaches: i) T cell binding (25), ii) eluted peptides (26) and iii) HLA-peptide binding measurements performed *in vitro* (IEDB data).

Frahm *et al*. tested the promiscuity of 20 CTL epitopes restricted by four HLA-B44 supertype members (Table 3). The far majority of those (17 epitopes) elicited a response in at least one individual negative for the original restricting allele. Five of these CTL epitopes elicited T cell responses in the context of another member of the HLA-B44

supertype (25% within-supertype promiscuity, Table 3). This result is similar to the one reported by Hillen *et al*., where out of four $CD8^+$ T cell responses restricted by HLA-B44 molecules, only one epitope induced minor T cell responses in individuals negative for the restricting allele, but positive for a different HLA molecule within the B44 supertype (26). The remaining 12 promiscuous B44-restricted CTL epitopes identified by Frahm *et al.* elicited responses in individuals that do not carry any HLA molecule belonging to the HLA-B44 supertype (60% outside-supertype promiscuity, Table 3).

Table 3. HLA-B44 alleles used by Hillen et al. (26) and HLA-B44 epitopes tested by Frahm et al. (25).

Hillen et al.			Frahm et al.			
HLA	N[a]	Promiscuous[b]	HLA	N[c]	Promiscuous Within B44	Promiscuous elsewhere
B*1801	121 (18%)	7 (5.7%)	B18	4	0	3
B*37	60 (9%)	0	B37	1	0	1
B*4001	60 (9%)	8 (13.3%)	B40	7	4	2
B*4101	38 (6%)	7 (18.4%)	B44	8	1	6
B*4402	142 (21%)	14 (10%)				
B*4501	29 (4%)	4 (13.8%)				
B*4901	184 (27%)	8 (4%)				
B*5001	18 (3%)	3 (16.7%)				
B*4701	27 (4%)	2 (7%)				
Total	670	25				

[a] The number of peptides eluted from a specific allele. Within parenthesis, the fraction with respect to the total number of eluted peptides is given.
[b] The number of peptides eluted at least from two alleles within B44 supertype. Within parenthesis, the fraction with respect to the number of eluted peptides from a specific allele is given.
[c] The number of CTL responses tested.

T cell response data extracted from IEDB show a similar trend: out of 14 T cell epitopes restricted by members of the B44 supertype and tested on at least six HLA molecules, only two (14%) show promiscuous binding within the supertype, while six epitopes (43%) elicit T cell responses when presented by an HLA molecule belonging to a different (non-B44) supertype.

Taken together, the results discussed above suggest that the estimates of HLA binding promiscuity for B44 supertype based on T cell responses (14-25% within supertype) are higher than the estimates based on eluted peptides (3% within supertype), and *in vitro* analysis of HLA binding provides the highest estimate of HLA binding promiscuity (see Table 2, 57% within supertype).

HLA-A and HLA-B molecules have similar levels of ligand promiscuity

So far we did not distinguish between different HLA loci in the promiscuity analysis. In the context of several infectious diseases, immune responses to epitopes restricted by HLA-B alleles were shown to be immunodominant (see e.g., (43, 44)). Moreover, particular HLA-B alleles seem to be associated with either protection or susceptibility to infectious diseases, best documented for HIV-1 infection (e.g., (21, 43, 45, 46)). In order to see whether these features of HLA-B restricted T cell responses may be due to the promiscuous binding of HLA-B restricted epitopes, we compared their binding promiscuity to HLA-A restricted epitopes. Since Hillen *et al*. focused on HLA-B44 epitopes only, we used the data from Frahm *et al*. to perform this analysis. Frahm *et al.* tested 242 CTL epitopes, of which 148 and 181 epitopes were inferred to be presented by at least one HLA-A and -B molecule, respectively. The number of epitopes that were exclusively presented by a single HLA molecule was slightly higher for HLA-A than for HLA-B (HLA-A: 37 out of 148, HLA-B: 28 out of 181, p=0.04 Chi-square test), suggesting that HLA-B restricted epitopes might exhibit higher binding promiscuity. However, this difference between HLA-A and HLA-B epitopes was not to be found in IEDB T cell assay data (HLA-A: 45 out of 117, and HLA-B: 8 out of 33, p=0.19 Chi-square test), and surprisingly, IEDB MHC binding data analysis suggested that HLA-A ligands have a higher level of binding promiscuity (p=0.02, Mann-Whitney *U* test, Fig. S1).

Due to these conflicting results on the experimental data, we addressed the difference in the promiscuity of HLA-A and HLA-B ligands also by using HLA binding predictions. The fraction of peptides binding exclusively a single HLA molecule remained similar in predicted HLA-A and HLA-B ligands from viral proteomes (see Fig. S2, p=0.34, Mann-Whitney *U* test). We repeated this analysis using a different set of criteria to define "binders" (i.e., using the top 1-2% percentile, see Materials and Methods) and by extending our viral data set (Table S3), but in all cases we obtained similar results.

Taken together, since the available experimental data yield conflicting results and our *in silico* predictions suggest no significant difference in promiscuous peptide binding of HLA-A and HLA-B, we conclude that the ligand binding promiscuity probably does not play a major role in generating dominant HLA-B restricted responses.

Functional Consequences of HLA peptide binding promiscuity in the context of HIV-1 infection

Our studies (see above) and others in the field provide solid evidence showing that HLA class I ligands show a high level of promiscuity. But why would HLA molecules

have promiscuous ligand binding? After all, it is believed that the extensively polymorphic MHC has evolved due to a selective advantage of being able to present epitopes on rare MHC molecules in cases where the pathogens are (fully) adapted to common MHC molecules (47). In search of a clue to explain functional aspects of such a high degree of promiscuity, we studied the effect of promiscuity on the disease outcome using HIV-1 as a case study. We speculated that the individuals carrying HLA molecules with largely overlapping repertoires can be considered "functionally homozygous" and may therefore progress more rapidly to AIDS (45, 48). We calculated the fraction of uniquely presented HIV-1 peptides for the frequent HLA alleles (based on the predicted peptide repertoires of the top 20 HLA-A and HLA-B molecules) in the Caucasian population.

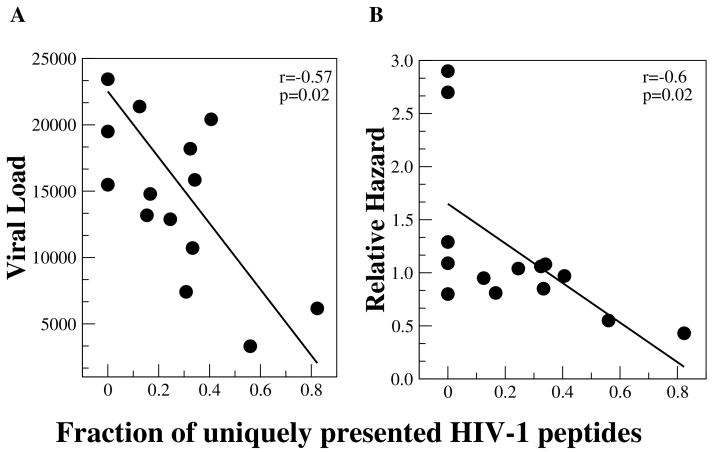


Figure 2. Correlation between the predicted fraction of unique ligands for HLA-B molecules and mean set point viral load associated with the same molecule, or the relative hazard (RH).
The fraction of uniquely presented HIV-1 peptides for an allele was calculated by comparing the predicted HIV-1 peptides for a particular allele with all the predicted HIV-1 peptides for the most frequent HLA alleles (top 20 HLA-A and HLA-B as listed in Table S4) in the Caucasian population. The predictions were performed with NetMHCpan (30) to obtain data for as many as possible alleles. The correlation between the fraction of unique ligands and (A) mean set point viral load per HLA molecule taken from (51), and (B) RH taken from (52) are shown. For (B), whenever available, we have used allele-specific RH; in other cases, we used the fraction of HIV-1 peptides estimated for the most dominant HLA-B allele to correlate with the relative hazard assigned to two-digit HLA-B identifier (e.g., the relative hazard associated with B*40 is correlated with the fraction of unique HIV-1 peptides presented by HLA-B*4001). The Spearman correlation coefficients and corresponding significance values are reported in each figure. The data used to generate these graphs is given in Table S5.

In line with the studies demonstrating that HLA-B alleles show the strongest association with disease outcome in HIV-1 infection (43, 49, 50), we found a strong negative correlation between the fraction of uniquely presented epitopes of HLA-B

molecules and median viral loads reported by Fellay *et al.* (51) (Fig. 2A, r=-0.57, p=0.02, Spearman correlation test) or the relative hazard (RH) reported by Gao *et al.* (52) (Fig. 2B, r=-0.60, p=0.02, Spearman correlation test). Remarkably, some protective alleles, like B*2705 and B*5701, which have low RH and are associated with low viral load, have more unique ligand repertoires than other alleles (82% and 56%, respectively, see Table S5), implying that having less promiscuous peptide presentation may contribute to viral control. However, when we repeated this analysis with data from the Durban cohort, infected mostly with HIV-1 clade C (50), the fraction of uniquely presented epitopes no longer significantly correlated with the median viral loads (results not shown).

Discussion

The genes of the human major histocompatibility complex belong to the most polymorphic loci in the human population. However, it is not yet clear whether this large diversity at genotype level is reflected at the phenotype level by distinct ligand repertoires. It has been known for a long time that some HLA molecules have very similar binding motifs, and thus these molecules can be grouped into HLA supertypes (4-10). More recently, Frahm *et al*. demonstrated that promiscuous HLA class I binding reaches beyond the supertype: the far majority of HLA pairs that can elicit T cell responses to the same peptide belong to different supertypes. Following Frahm's study on HIV-1 and EBV epitopes, it was demonstrated that human papillomavirus (HPV) and *Mycobacterium tuberculosis* (TB) epitopes also show extensive promiscuity of HLA class I binding when tested systematically (53, 54). These findings were challenged by Hillen *et al*. (26) who showed that even within a supertype, eluted HLA ligands can show as little as 3% promiscuity.

We have taken another approach to estimate HLA-peptide binding promiscuity by using *in vitro* binding measurements reported in the IEDB database (27). Moreover, in order to be able to estimate the promiscuity of binding at the population level, i.e., by testing peptide binding to all major HLA molecules in a population, we performed *in silico* HLA-peptide binding predictions. In both cases we found extensive promiscuity in HLA class I ligand binding: 72% in the IEDB HLA binding data and 60% in our predicted HLA ligands. In addition, a high fraction of promiscuous ligands are found to be ligands for at least two different HLA supertypes (see Table 1, Fig. 1). As expected, HLA supertype pairs with similar binding motifs share more ligands (e.g., A2 and B62 supertype ligands in IEDB overlap by 21%), than the pairs with dissimilar motifs (e.g., only 4.4% B27-supertype ligands in IEDB are also reported as binders for at least one

allele belonging to B44 supertype). Since our *in silico* analysis covers different viruses and HLA molecules common in different ethnic groups, we believe that our results provide solid evidence for a high level of promiscuity being an intrinsic characteristic of HLA binding, regardless of the source of the ligand and the HLA molecule.

Why then was the fraction of shared ligands in the study by Hillen *et al*. as low as 3%? Unfortunately, our efforts did not produce a concrete answer to this question. Remarkably, the number of peptides eluted per allele by Hillen *et al*. was low (summarized in Table 2), considering that around 100,000 MHC molecules are expected to be on the cell surface at a time (55). This might be (among others) due to degradation of presented peptides under the rather harsh conditions necessary for the elution. If the elution studies underestimate the peptide repertoires, then the overlaps between the peptide repertoires of different MHC molecules might be underestimated as well. Indeed, when we use a stringent threshold to define binders in our *in silico* analysis, the predicted peptide repertoires of individual HLA molecules are reduced and as a consequence the average promiscuity decreases (data not shown).

By sampling eluted peptides, Hillen *et al*. may have biased their data to high affinity MHC binders. When we predicted the MHC binding affinity for the eluted peptides from the B44 supertype, we found that the median predicted binding affinity was lower than 50 nM, which is generally used as a cutoff to discriminate high binders (e.g., for B*1801 the median affinity is 9nM, and for B*4001 the median affinity is 27nM). Following this, promiscuity of high affinity binders may be lower than of MHC ligands in general. However, this explanation is not in line with the earlier studies, which suggest that high affinity binders also tend to be the most promiscuous binders (56-58). Similarly, we found a significant (but weak) negative correlation between ligand binding affinity and promiscuity ($r=-0.27$, $p<0.0001$, Spearman correlation test), suggesting that the promiscuous binding among high affinity binders should be even higher than on average. Taken together, earlier studies and our present study suggest that the lower promiscuity observed by Hillen *et al.* might be due to other mechanisms than MHC binding per se.

The functional consequences of the extensive ligand sharing among HLA class I molecules remain to be discovered. In order to see whether promiscuous ligand presentation might be the underlying reason of immunodominance by HLA-B restricted T cell responses, we compared promiscuity between different HLA-A and HLA-B ligands. However, our numerous attempts did not result in a consistent picture (see section on HLA-A and -B), suggesting that there is not a direct association between (non-) promiscuous ligand presentation and dominant T cell responses. On the other hand, we have found a relationship between the fraction of uniquely presented peptides

and HIV-1 disease progression, where HLA molecules associated with slow disease progression are also the ones that have the lowest degree of promiscuity (see Fig. 2). We believe that carrying HLA molecules with unique peptide repertoires increase the heterozygous advantage, based on the principle that individuals heterozygous at HLA loci are able to present a greater diversity of antigenic peptides than are homozygotes (59). The heterozygous advantage was suggested to generate a more effective immune response and therefore resulted in better control of HIV-1 infection (60). In addition, the individuals with HLA molecules having unique binding motifs have lower chances of transmission of a pre-adapted virus. More data on HLA associations and disease outcome will help to resolve the functional aspects of the high level of promiscuity among HLA class I epitopes and especially how it affects an individual's fitness.

Acknowledgements

We thank Rob de Boer for valuable comments on the manuscript, Hanneke van Deutekom, Jorg Calis and Boris Schmid for technical support. This work was financially supported by a HIPO (High Potential) grant from Universiteit Utrecht.

References

1. Bjorkman PJ, and Parham P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual review of biochemistry.* 1990;59:253-88.
2. Parham P, and Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science.* 1996;272(5258):67-74.
3. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernandez-Vina M, Geraghty DE, Holdsworth R, Hurley CK, et al. Nomenclature for factors of the HLA system, 2010. *Tissue antigens.* 2010;75(4):291-455.
4. Sette A, and Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics.* 1999;50(3-4):201-12.
5. Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, Sylvester-Hvid C, Lamberth K, Roder G, Justesen S, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics.* 2004;55(12):797-810.
6. Doytchinova IA, Guan P, and Flower DR. Identifiying human MHC supertypes using bioinformatic methods. *Journal of immunology.* 2004;172(7):4314-23.
7. Kangueane P, Sakharkar MK, Rajaseger G, Bolisetty S, Sivasekari B, Zhao B, Ravichandran M, Shapshak P, and Subbiah S. A framework to sub-type HLA supertypes. *Frontiers in bioscience : a journal and virtual library.* 2005;10:879-86.
8. Reche PA, and Reinherz EL. Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. *Methods in molecular biology.* 2007;409:163-73.
9. Hertz T, and Yanover C. Identifying HLA supertypes by learning distance functions. *Bioinformatics.* 2007;23(2):e148-55.
10. Sidney J, Peters B, Frahm N, Brander C, and Sette A. HLA class I supertypes: a revised and updated classification. *BMC immunology.* 2008;9:1.
11. Sidney J, del Guercio MF, Southwood S, Engelhard VH, Appella E, Rammensee HG, Falk K, Rotzschke O, Takiguchi M, Kubo RT, et al. Several HLA alleles share overlapping peptide specificities. *Journal of immunology.* 1995;154(1):247-59.
12. Barber LD, Gillece-Castro B, Percival L, Li X, Clayberger C, and Parham P. Overlap in the repertoires of peptides bound in vivo by a group of related class I HLA-B allotypes. *Current biology : CB.* 1995;5(2):179-90.
13. Doolan DL, Hoffman SL, Southwood S, Wentworth PA, Sidney J, Chesnut RW, Keogh E, Appella E, Nutman TB, Lal AA, et al. Degenerate cytotoxic T cell epitopes from P. falciparum restricted by multiple HLA-A and HLA-B supertype alleles. *Immunity.* 1997;7(1):97-112.
14. Bertoni R, Sidney J, Fowler P, Chesnut RW, Chisari FV, and Sette A. Human histocompatibility leukocyte antigen-binding supermotifs predict broadly cross-reactive cytotoxic T lymphocyte responses in patients with acute hepatitis. *The Journal of clinical investigation.* 1997;100(3):503-13.
15. Threlkeld SC, Wentworth PA, Kalams SA, Wilkes BM, Ruhl DJ, Keogh E, Sidney J, Southwood S, Walker BD, and Sette A. Degenerate and promiscuous recognition by CTL of peptides presented by the MHC class I A3-like superfamily: implications for vaccine development. *Journal of immunology.* 1997;159(4):1648-57.
16. Brusic V, Petrovsky N, Zhang G, and Bajic VB. Prediction of promiscuous peptides that bind HLA class I molecules. *Immunology and cell biology.* 2002;80(3):280-5.
17. Ueno T, Tomiyama H, and Takiguchi M. Single T cell receptor-mediated recognition of an identical HIV-derived peptide presented by multiple HLA class I molecules. *Journal of immunology.* 2002;169(9):4961-9.

18. Burrows SR, Elkington RA, Miles JJ, Green KJ, Walker S, Haryana SM, Moss DJ, Dunckley H, Burrows JM, and Khanna R. Promiscuous CTL recognition of viral epitopes on multiple human leukocyte antigens: biological validation of the proposed HLA A24 supertype. *Journal of immunology.* 2003;171(3):1407-12.

19. Sidney J, Southwood S, Pasquetto V, and Sette A. Simultaneous prediction of binding capacity for multiple molecules of the HLA B44 supertype. *Journal of immunology.* 2003;171(11):5964-74.

20. Takedatsu H, Shichijo S, Katagiri K, Sawamizu H, Sata M, and Itoh K. Identification of peptide vaccine candidates sharing among HLA-A3+, -A11+, -A31+, and -A33+ cancer patients. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2004;10(3):1112-20.

21. Frahm N, Adams S, Kiepiela P, Linde CH, Hewitt HS, Lichterfeld M, Sango K, Brown NV, Pae E, Wurcel AG, et al. HLA-B63 presents HLA-B57/B58-restricted cytotoxic T-lymphocyte epitopes and is associated with low human immunodeficiency virus load. *Journal of virology.* 2005;79(16):10218-25.

22. Leslie A, Price DA, Mkhize P, Bishop K, Rathod A, Day C, Crawford H, Honeyborne I, Asher TE, Luzzi G, et al. Differential selection pressure exerted on HIV by CTL targeting identical epitopes but restricted by distinct HLA alleles from the same HLA supertype. *Journal of immunology.* 2006;177(7):4699-708.

23. Sabbaj S, Bansal A, Ritter GD, Perkins C, Edwards BH, Gough E, Tang J, Szinger JJ, Korber B, Wilson CM, et al. Cross-reactive CD8+ T cell epitopes identified in US adolescent minorities. *Journal of acquired immune deficiency syndromes.* 2003;33(4):426-38.

24. Masemola AM, Mashishi TN, Khoury G, Bredell H, Paximadis M, Mathebula T, Barkhan D, Puren A, Vardas E, Colvin M, et al. Novel and promiscuous CTL epitopes in conserved regions of Gag targeted by individuals with early subtype C HIV type 1 infection from southern Africa. *Journal of immunology.* 2004;173(7):4607-17.

25. Frahm N, Yusim K, Suscovich TJ, Adams S, Sidney J, Hraber P, Hewitt HS, Linde CH, Kavanagh DG, Woodberry T, et al. Extensive HLA class I allele promiscuity among viral CTL epitopes. *European journal of immunology.* 2007;37(9):2419-33.

26. Hillen N, Mester G, Lemmel C, Weinzierl AO, Muller M, Wernet D, Hennenlotter J, Stenzl A, Rammensee HG, and Stevanovic S. Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. *European journal of immunology.* 2008;38(11):2993-3003.

27. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, and Peters B. The immune epitope database 2.0. *Nucleic acids research.* 2010;38(Database issue):D854-62.

28. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Roder G, Peters B, Sette A, Lund O, 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(8):e796.

29. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, and Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic acids research.* 2008;36(Web Server issue):W509-12.

30. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, and Nielsen M. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* 2009;61(1):1-13.

31. Mori M, Beatty PG, Graves M, Boucher KM, and Milford EL. HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. *Transplantation.* 1997;64(7):1017-27.

32. Maiers M, Gragert L, and Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Human immunology.* 2007;68(9):779-88.

33. Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, and Brunak S. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue antigens.* 2003;62(5):378-84.

34. Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, and Lund O. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS computational biology.* 2008;4(7):e1000107.

35. MacNamara A, Kadolsky U, Bangham CR, and Asquith B. T-cell epitope prediction: rescaling can mask biological variation between MHC molecules. *PLoS computational biology.* 2009;5(3):e1000327.

36. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, and Holzhutter HG. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and molecular life sciences : CMLS.* 2005;62(9):1025-37.

37. Kesmir C, Nussbaum AK, Schild H, Detours V, and Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Protein engineering.* 2002;15(4):287-96.

38. Nielsen M, Lundegaard C, Lund O, and Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics.* 2005;57(1-2):33-41.

39. Feltkamp MC, Vierboom MP, Kast WM, and Melief CJ. Efficient MHC class I-peptide binding is required but does not ensure MHC class I-restricted immunogenicity. *Molecular immunology.* 1994;31(18):1391-401.

40. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *Journal of immunology.* 1994;153(12):5586-92.

41. Fortier MH, Caron E, Hardy MP, Voisin G, Lemieux S, Perreault C, and Thibault P. The MHC class I peptide repertoire is molded by the transcriptome. *The Journal of experimental medicine.* 2008;205(3):595-610.

42. Zhang H, Lundegaard C, and Nielsen M. Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. *Bioinformatics.* 2009;25(1):83-9.

43. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature.* 2004;432(7018):769-75.

44. Bihl F, Frahm N, Di Giammarino L, Sidney J, John M, Yusim K, Woodberry T, Sango K, Hewitt HS, Henry L, et al. Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *Journal of immunology.* 2006;176(7):4094-101.

45. Carrington M, and O'Brien SJ. The influence of HLA genotype on AIDS. *Annual review of medicine.* 2003;54:535-51.

46. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Zhao M, and et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* 2010;330(6010):1551-7.

47. Borghans JA, Beltman JB, and De Boer RJ. MHC polymorphism under host-pathogen coevolution. *Immunogenetics.* 2004;55(11):732-9.

48. Jeffery KJ, Siddiqui AA, Bunce M, Lloyd AL, Vine AM, Witkover AD, Izumo S, Usuku K, Welsh KI, Osame M, et al. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. *Journal of immunology.* 2000;165(12):7278-84.

49. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine.* 2007;13(1):46-53.

50. Leslie A, Matthews PC, Listgarten J, Carlson JM, Kadie C, Ndung'u T, Brander C, Coovadia H, Walker BD, Heckerman D, et al. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *Journal of virology.* 2010;84(19):9879-88.

51. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, Urban TJ, Zhang K, Gumbs CE, Smith JP, et al. Common genetic variation and the control of HIV-1 in humans. *PLoS genetics.* 2009;5(12):e1000791.

52. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, et al. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *The New England journal of medicine.* 2001;344(22):1668-75.

53. Nakagawa M, Kim KH, Gillam TM, and Moscicki AB. HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein. *Journal of virology.* 2007;81(3):1412-23.

54. Axelsson-Robertson R, Weichold F, Sizemore D, Wulf M, Skeiky YA, Sadoff J, and Maeurer MJ. Extensive major histocompatibility complex class I binding promiscuity for Mycobacterium tuberculosis TB10.4 peptides and immune dominance of human leucocyte antigen (HLA)-B*0702 and HLA-B*0801 alleles in TB10.4 CD8 T-cell responses. *Immunology.* 2010;129(4):496-505.

55. Yewdell JW, Reits E, and Neefjes J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature reviews Immunology.* 2003;3(12):952-61.

56. Sidney J, Southwood S, del Guercio MF, Grey HM, Chesnut RW, Kubo RT, and Sette A. Specificity and degeneracy in peptide binding to HLA-B7-like class I molecules. *Journal of immunology.* 1996;157(8):3480-90.

57. Sidney J, Southwood S, Mann DL, Fernandez-Vina MA, Newman MJ, and Sette A. Majority of peptides binding HLA-A*0201 with high affinity crossreact with other A2-supertype molecules. *Human immunology.* 2001;62(11):1200-16.

58. Sette A, Sidney J, Livingston BD, Dzuris JL, Crimi C, Walker CM, Southwood S, Collins EJ, and Hughes AL. Class I molecules with similar peptide-binding specificities are the result of both common ancestry and convergent evolution. *Immunogenetics.* 2003;54(12):830-41.

59. Dean M, Carrington M, and O'Brien SJ. Balanced polymorphism selected by genetic versus infectious human disease. *Annual review of genomics and human genetics.* 2002;3:263-92.

60. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, and O'Brien SJ. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science.* 1999;283(5408):1748-52.

Supplementary Material

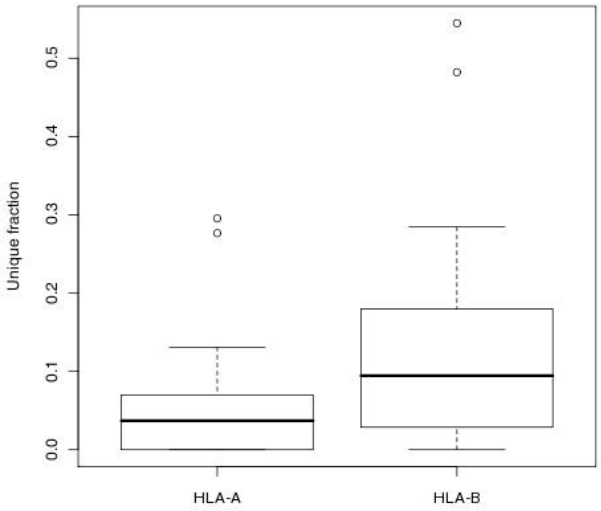


Figure S1. Promiscuity of HLA-A and HLA-B ligands based on IEDB MHC binding data.
30 HLA-A and 26 HLA-B molecules (each of which with at least 10 ligands in IEDB MHC binding data) were selected. The fraction of unique ligands was calculated for each allele. HLA-B ligands are significantly less promiscuous than HLA-A ligands (p=0.02, Mann-Whitney *U* test).

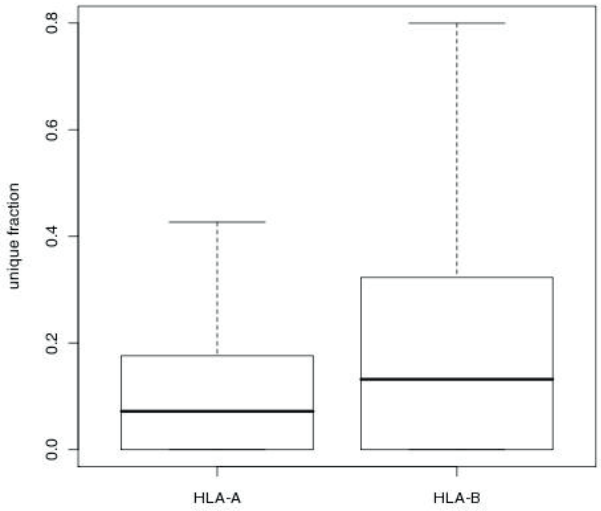


Figure S2. Promiscuity of predicted HLA-A and HLA-B ligands based on NetMHCpan predictions.
All possible 9-mer viral ligands from viral proteomes (listed in Table S1 and S3) were predicted for the 20 most frequent HLA-A and HLA-B alleles in USA subpopulation of European ethnic background. The fraction of unique ligands (in percentages) was calculated for each allele (median of the group is indicated by a horizontal line). There is no significant difference between the fraction of unique HLA-A and HLA-B ligands (p=0.34, Mann-Whitney *U* test).

Table S1. List of viral proteomes used as the resource of NetMHCpan predicted ligands.

VIRUS	*Accession number*
Human immunodeficiency virus 1	X01762
Dengue virus 1	U88536
Reston ebolavirus	AB050936
Hepatitis A virus	M14707
Hepatitis B virus	X51970
Hepatitis C virus	AJ132997
Human T-lymphotropic virus 1	D13784
Influenza A virus (A/Goose /Guuangdong/1/96(H5N1)) segment 1-8	AF144300-AF144307
Measles virus	K01711
Mumps virus	AB040874
H-1 parvovirus	X01457
Human poliovirus 1	AJ132961
Rabies virus	M31046
Human respiratory syncytial virus	AF013254
Rubella virus	AF188704
Sendai virus	M69046
Yellow fever virus	X03700

Table S2. Ligand promiscuity of HLA supertypes for different ethnic groups based on NetMHCpan prediction data. We obtained very similar results with NetMHC3.2 predictor.

Ethnic group	HLA Supertype	N[a]	Promiscuous	Promiscuous only in another supertypes[b]	Supertype members
European	HLA-A01	5444	5041	3356 (62%)	8
	HLA-A02	4342	2791	621 (14%)	3
	HLA-A03	5480	4575	1998 (36%)	8
	HLA-A24	1689	1562	96 (6%)	4
	HLA-B07	1766	1173	835 (47%)	6
	HLA-B08	877	583	583 (66%)	1
	HLA-B27	583	126	124 (21%)	3
	HLA-B44	1451	902	306 (21%)	7
	HLA-B58	453	305	305 (67%)	1
	HLA-B62	1682	1405	1405 (83%)	1
African	HLA-A01	5220	5139	3774 (72%)	4
	HLA-A02	4785	3677	790 (16%)	4
	HLA-A03	5177	4561	2085 (40%)	9
	HLA-A24	1689	1577	1116 (66%)	3
	HLA-B07	1818	1402	799 (44%)	6
	HLA-B08	877	631	631 (72%)	1
	HLA-B27	5369	3555	3455 (64%)	3
	HLA-B44	797	463	143 (18%)	5
	HLA-B58	4562	3552	2623 (57%)	4
	HLA-B62	21	21	21 (100%)	1
Asian	HLA-A01	6095	5599	4101 (67%)	6
	HLA-A02	6688	4968	618 (9%)	5
	HLA-A03	5374	4636	1906 (35%)	7
	HLA-A24	1466	1343	1300 (89%)	2
	HLA-B07	1823	1420	764 (42%)	6
	HLA-B27	106	106	106 (100%)	1
	HLA-B44	1416	1049	148 (10%)	4
	HLA-B58	945	767	318 (34%)	2
	HLA-B62	1884	1707	764 (41%)	4
Hispanic	HLA-A01	5444	5178	3671 (67%)	7
	HLA-A02	4884	3617	676 (14%)	1
	HLA-A03	5383	4672	1912 (35%)	8
	HLA-A24	1689	1575	1114 (66%)	3
	HLA-B07	1803	1441	469 (26%)	5
	HLA-B08	877	630	630 (72%)	1
	HLA-B27	5984	3951	3315 (55%)	7
	HLA-B44	1410	933	505 (36%)	5
	HLA-B62	1698	1691	1686 (99%)	2

[a] Total number of unique ligands. [b] The fraction of promiscuous ligands that are exclusively bind to the HLA molecules outside of the supertype is given within parenthesis.

Table S3. List of human viral proteomes used in the extension of the viral set.

HUMAN VIRUSES	*Accession number*
Aichi_virus	AB010145
Chapare_virus	EU260464
Duvenhage_virus	EU293119
Hepatitis_E_virus	X98292
Human_coxsackievirus_A16	U05876
Human_papillomavirus_type_43	AJ620205
Merkel_cell_polyomavirus	EU375803
Norwalk-like_virus	AB084071
Small_anellovirus_1	AY622908
Torque_teno_midi_virus B	AB290919
TTV-like_mini_virus	AB026929

Table S4A. Most frequent 20 HLA-A alleles in four ethnic groups selected for NetMHCpan predictions.

Allele	European		African		Asian		Hispanic	
	Freq.	*Rank*	*Freq.*	*Rank*	*Freq.*	*Rank*	*Freq.*	*Rank*
A*0201	0.296	1	0.125	1	0.095	3	0.194	1
A*0101	0.171	2	0.047	8	0.051	5	0.067	4
A*0301	0.143	3	0.081	3	0.026	11	0.079	3
A*2402	0.087	4	0.022	15	0.182	1	0.123	2
A*1101	0.056	5	0.016	18	0.179	2	0.046	7
A*2902	0.033	6	0.036	12			0.042	8
A*3201	0.031	7	0.014	20	0.013	18	0.027	13
A*2601	0.030	8			0.039	8	0.029	11
A*6801	0.025	9	0.037	11	0.019	13	0.047	6
A*3101	0.024	10			0.032	9	0.048	5
A*2501	0.019	11					0.009	20
A*2301	0.017	12	0.108	2			0.037	10
A*3001	0.013	13	0.069	4	0.021	12	0.021	15
A*3301	0.010	14	0.021	16			0.019	16
A*3002	0.009	15	0.062	6			0.028	12
A*6802	0.008	16	0.065	5			0.025	14
A*0205	0.008	17	0.019	1			0.015	17
A*0302	0.003	18						
A*6601	0.003	19	0.015	19				
A*2901	0.002	20			0.014	17		
A*7401			0.052	7				
A*3303			0.045	9	0.094	4	0.013	19
A*0202			0.042	10				
A*3402			0.033	13				
A*3601			0.024	14				
A*0206					0.048	6	0.039	9
A*0207					0.044	7		
A*0203					0.032	10		
A*2407					0.018	14		
A*3401					0.016	15		
A*1102					0.015	16		
A*0211					0.012	19		
A*2602					0.007	20		
A*6803							0.014	18

Table S4B. Most frequent 20 HLA-B alleles in four ethnic groups selected for NetMHCpan predictions.

Allele	European		African		Asian		Hispanic	
	Freq.	*Rank*	*Freq.*	*Rank*	*Freq.*	*Rank*	*Freq.*	*Rank*
B*0702	0.140	1	0.073	2	0.026	15	0.055	4
B*0801	0.125	2	0.038	9			0.045	6
B*4402	0.090	3	0.021	17			0.033	9
B*1501	0.067	4	0.030	13	0.035	11	0.029	10
B*3501	0.057	5	0.065	3	0.043	5	0.064	1
B*4001	0.056	6			0.080	1		
B*4403	0.050	7	0.054	6	0.042	6	0.061	2
B*1801	0.046	8	0.036	10			0.040	8
B*5101	0.045	9	0.022	16	0.063	2	0.058	3
B*5701	0.038	10			0.021	18		
B*2705	0.033	11					0.017	18
B*1402	0.031	12	0.022	15			0.041	7
B*1302	0.026	13			0.023	17		
B*3801	0.022	14					0.019	16
B*5501	0.017	15						
B*3503	0.016	16			0.024	16		
B*3701	0.013	17						
B*4901	0.013	18	0.028	14			0.024	12
B*3502	0.011	19						
B*4002	0.010	20			0.031	14	0.049	5
B*5301			0.112	1				
B*1503			0.062	4			0.016	20
B*4201			0.055	5				
B*5802			0.041	8				
B*5801			0.035	11	0.058	4		
B*5703			0.034	12				
B*8101			0.020	18				
B*1516			0.017	19				
B*5201			0.014	20	0.037	8	0.027	11
B*4501			0.045	7				
B*4601					0.061	3		
B*4006					0.037	7		
B*3802					0.037	9		
B*1502					0.036	10		
B*5401					0.033	12		
B*1301					0.033	13		
B*4801					0.020	19	0.021	14
B*0705					0.020	20		
B*3905							0.023	13
B*3906							0.020	15
B*3512							0.019	17
B*3517							0.016	19

Table S5. Fraction of uniquely presented HIV-1 peptides by frequent HLA-B molecules, and the corresponding set point viral loads (mean) as reported by Fellay et al., 2009, and relative hazard (RH) values reported by Gao et al., 2001.

HLA-B molecule	Fraction of unique HIV-1 presentation	Set point VL (mean, log)	RH
B0702	0.32	4.26	1.06
B0801	0.41	4.31	0.97
B1501	0.25	4.11	1.04
B1801	0.17	4.17	0.81
B2705	0.82	3.79	0.43
B3501	0.34	4.20	1.08
B3502	0	4.68	2.7
B3503	0	4.19	2.9
B3801	0	4.29	0.8
B4001	0.125	4.33	0.95
B4002	0.31	3.87	NA
B4402	0	4.19	1.09
B4403	0.15	4.12	NA
B5101	0.33	4.03	0.85
B5501	0	4.37	1.29
B5701	0.56	3.52	0.55

The predictions reported here were made with NetMHCpan, however, similar results are obtained using allele specific predictor NetMHC3.2. Only the alleles where a reliable peptide binding predictor is available (defined as an estimated correlation coefficient between predicted and measured peptide binding affinities larger than 0.6) are included in the analysis. This data is used to make the analysis summarized in Figure 2 of the main text.

CD8$^+$ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex

Dan Koning[1], Ana I. Costa[1], Ilka Hoof[2], John J. Miles[3,4], Nening M. Nanlohy[1], Kristin Ladell[3], Katherine K. Matthews[3,4], Vanessa Venturi[5], Ingrid M.M. Schellens[1], José A.M. Borghans[1], Can Keşmir[2], David A. Price[3] and Debbie van Baarle[1,6]

[1]*Department of Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands.*

[2]*Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands.*

[3]*Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK*

[4]*Human Immunity Laboratory, Queensland Institute of Medical Research, Brisbane, Australia.*

[5]*Computational Biology Group, Centre for Vascular Research, University of New South Wales, Kensington, Australia.*

[6]*Department of Internal Medicine and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands.*

The Journal of Immunology (2013), 190(3): 931-939.

Abstract

CD8$^+$ T-cells recognize infected or dysregulated cells via the clonotypically expressed αβ TCR, which engages antigen in the form of peptide bound to MHC class I (MHCI) on the target cell surface. Previous studies have indicated that a diverse antigen-specific TCR repertoire can be beneficial to the host, yet the determinants of clonotypic diversity are poorly defined. To better understand the factors that govern TCR repertoire formation, we conducted a comprehensive clonotypic analysis of CD8$^+$ T-cell populations directed against epitopes derived from EBV and CMV. Neither pathogen source nor the restricting MHCI molecule were linked with TCR diversity; indeed, both HLA-A and HLA-B molecules were observed to interact with an overlapping repertoire of expressed *TRBV* genes. Peptide specificity, however, markedly impacted TCR diversity. In addition, distinct peptides sharing HLA restriction and viral origin mobilized TCR repertoires with distinct patterns of *TRBV* gene usage. Notably, no relationship was observed between immunodominance and TCR diversity. These findings provide new insights into the forces that shape the antigen-specific TCR repertoire *in vivo* and highlight a determinative role for the peptide component of the peptide-MHCI complex on the molecular frontline of CD8$^+$ T cell-mediated immune surveillance.

Introduction

$CD8^+$ T cells are critical for effective control of viral infections because they enable the immune system to detect intracellular abnormalities. Surface presentation of short, cytosol-derived peptides in association with MHC class I (MHCI) molecules provides the molecular basis through which $CD8^+$ T cells discriminate healthy from diseased cells. Peptide-MHCI (pMHCI) complex antigen recognition is mediated by the clonotypically expressed αβ TCR, a dimeric glycoprotein expressed on the surface of $CD8^+$ T cells that scans the surface of MHC molecules. Although it has been shown that TCRs interact with self-antigens to maintain a basal level of activation (1), it is the high-affinity interplay with pathogen-derived, foreign peptides that typically results in immune activation, clonal expansion and eventual lysis of infected cells. To cope with the almost limitless number of foreign antigens that exist in nature, a vast repertoire of TCRs is required for adequate coverage. Such diversity is generated by somatic rearrangement of V, (D) and J genes across the TCRα and β loci, combined with junctional modifications and differential αβ-chain pairing (2). In humans, it has been estimated that this process results in the peripheral expression of ~10^7 different TCRs at any given time (3). In addition, the plasticity of these receptors allows them to recognize multiple ligands (4,5), although the need to prevent autoimmunity acts as a selective counterbalance to limit T-cell crossreactivity (6).

Previous studies of TCRβ diversity have revealed substantial variation in the clonality of $CD8^+$ T-cell populations. Some responses are characterized by a restricted and sometimes biased repertoire, whereas other responses can be highly polyclonal (7-9). An elegant study performed by Messaoudi and coworkers (10) highlighted the importance of a broad antigen-specific T-cell repertoire in protection from viral infection. C57BL/6 mice from an H-$2K^b$ or an H-$2K^{bm8}$ background were both able to present the immunodominant HSV-8p epitope from herpesvirus homini type 1, but mice with an H-$2K^{bm8}$ background selected a more diverse TCRβ repertoire that conferred increased resistance to infection. Several reports have since extended these observations in other systems (11,12), although it is clear that additional characteristics also have an impact on the overall efficacy of the antigen-specific repertoire (13,14). However, the factors that govern diverse repertoire formation are poorly understood. Either extrinsic influences or the nature of the pMHCI antigen itself could be determinative, with the latter potentially guided primarily by either the bound peptide or the restricting MHCI molecule. The MHCI molecule typically engages the germline gene-encoded CDR1 and CDR2 loops of the TCR as well as the CD8 coreceptor, which

functions as a key molecule in T-cell engagement, activation and cross-reactivity (15). Although the three major HLA class I loci (A, B, and C) are closely related in evolutionary terms, HLA-B appears to play the defining role during an immune response. This locus is often associated with immunodominance (16,17), polyfunctional $CD8^+$ T-cell responses (18) and disease outcome (19-21). To date, however, the impact of HLA class I molecules on the breadth of the antigen-specific T-cell repertoire has not been investigated. In addition, the effects of antigenic origin and the bound peptide epitope on the breadth of the responding T-cell repertoire are not well characterized.

To address these questions and to achieve a more complete understanding of TCR repertoire biology in general, we conducted a comprehensive survey of TCRβ repertoire architecture across a spectrum of $CD8^+$ T-cell populations specific for different epitopes derived from EBV and CMV. These model viruses establish persistent infection in the majority of the human population and generate T-cell responses of substantial magnitude that are maintained throughout life (22,23). Our data suggest that both HLA-A and HLA-B molecules restrict antigen-specific $CD8^+$ T-cell populations with comparable TCRβ diversity and similar patterns of *TRBV* gene usage. Interestingly, both clonotypic diversity at the CDR3β level as well as *TRBV* gene usage appeared to be determined by the MHCI-bound peptide and not by immunodominance hierarchy or antigenic origin. These findings demonstrate that the antigenic peptide acts as the principal determinant of the responding TCRβ repertoire.

Materials and Methods

Study population

Healthy donors (n=54) were selected based on the occurrence of HLA alleles with either a high prevalence in the Caucasian population (A*02: 29.6%; A*11: 5.6%; B*07: 14.0%; B*08: 12.5%; www.allelefrequencies.net) or a clearly described influence on disease outcome (B*35, B*57) (Table I). Informed consent was obtained in all cases. Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll-Paque density centrifugation and cryopreserved following standard protocols.

Flow cytometric sorting of antigen-specific $CD8^+$ T cells

Antigen-specific $CD8^+$ T cells were labeled at room temperature with pre-titrated concentrations of cognate fluorochrome-conjugated pMHCI tetrameric complexes and sorted at >98% purity using a FACSAria flow cytometer (BD Biosciences). The pMHCI

tetramers used in this study, detailed in Table II, were produced as described previously with minor modifications (24). Live/Dead Violet Viability dye (Invitrogen) was used to exclude nonviable cells from the analysis. Subsequently, cells were washed and surface stained with the following monoclonal antibodies (mAbs): (i) anti-CD3-PerCP and anti-CD8-AmCyan/V500 (BD Biosciences); and, (ii) anti-CD14-Pacific Blue and anti-CD19-Pacific Blue (Biolegend). After exclusion of nonviable/$CD14^+$/$CD19^+$ cells in a single dump channel, $CD3^+CD8^+$tetramer$^+$ cells were sorted directly into RNAlater (Ambion Inc., Applied Biosystems) and stored at -80°C for subsequent TCRβ clonotype analysis.

T-cell receptor clonotype analysis

Clonotype analysis was performed as described previously with minor modifications (25). Briefly, mRNA from sorted populations was extracted with the Poly(A)Tract system 1000 (Promega) according to the manufacturer's instructions. An anchored template-switch RT-PCR was performed to amplify all expressed TCRβ chains linearly. Amplified products were ligated into the pGEMT-Easy vector (Promega) and used to transform chemically competent *E. coli* bacteria. A colony PCR was performed using M13 primers and the resulting products were sequenced using Big Dye Terminator V3.1 (Invitrogen) with capillary electrophoresis. Analysis of TCRβ sequences was performed using web-based software from ImMunoGeneTics (IMGT; http://www.imgt.org), acknowledged as the international reference for immunoglobulin and TCRβ sequence analysis (WHO/IUIS). At least 50 TCRβ sequences were analyzed for each sample and gene usage was assigned using the IMGT nomenclature (26).

Clonotype diversity

Sample clonality was estimated using both the number of different clonotypes and Simpson's diversity index (D_S) (27), which is defined as

$$D_S = 1 - \sum_{i=1}^{c} \frac{n_i(n_i - 1)}{n(n-1)},$$

where n_i is the clonal size of the ith clonotype (i.e. the amount of copies of a specific clonotype), c is the number of different clonotypes and n is the total number of TCRβ sequences analyzed. This index uses the relative frequency of each clonotype to calculate a diversity index ranging between 0 and 1, with 0 being minimal and 1 being maximal diversity.

To calculate sample heterogeneity, CDR3β amino acid sequences were aligned using web-based MUSCLE alignment software (http://www.ebi.ac.uk/Tools/msa/muscle/). A heterogeneity score per CDR3β position was calculated using Shannon's entropy (28), which is defined as

$$H(i) = -\sum_{L=1}^{21} q_i \log_2 q_i ,$$

where *H(i)* is the Shannon index at position *i*, and q_i is the probability that a certain amino acid (or gap) will occur at position *i* in the alignment. A gap was considered as a position as well and was weighed similar to the amino acids. Each position in the alignment has a maximum heterogeneity of $-\log_2 21 = 4.39$ (if all 20 amino acids and a gap are equally distributed) and a minimum heterogeneity of 0 (if a position is fully conserved). Heterogeneity scores were averaged for the complete CDR3β (excluding the conserved cysteine at the beginning and the phenylalanine residue at the terminus) to determine the sequence variation between individual clonotypes. For all three diversity measures, to correct for any bias that could arise due to the amount of analyzed sequences, data were normalized by randomly drawing 50 TCRβ readouts 1000 times per sample, and using the median value to estimate TCRβ diversity.

IFN-γ ELISpot

IFN-γ enzyme-linked immunospot (ELISpot) assays were performed as described previously (29). Briefly, 10^6 PBMCs/ml were stimulated with 20 μg/ml of the relevant peptide and incubated for 16-20 hours at 37°C on 96-well membrane-bottomed plates coated with αIFN-γ mAbs. PHA (10 μg/ml) served as a positive control and medium alone served as a negative control. Samples were considered positive if the number of spots exceeded 50 + 3 x medium per 10^6 PBMCs.

Statistical analysis

Sample normalization was performed using *R* Development Core Team (*R* Foundation for Statistical Computing). All statistical analyses were conducted with the software program SPSS 15.0.1 (SPSS Inc.) and graphics were obtained using GraphPrism 5.04 (Graphpad Software Inc.). The dependency between some data, due to multiple epitope-specific TCRβ repertoires being obtained from some of the donors, could not be taken into account in the statistical comparisons. Nonparametric tests were applied for the comparison of groups, and a p-value <0.05 was considered statistically significant.

Results

Efficient isolation of CD8⁺ T-cell populations specific for epitopes derived from EBV and CMV

To gain a better understanding of TCRβ diversity, antigen-experienced memory CD8⁺ T-cell populations specific for the persistent herpesviruses EBV and CMV were sorted by flow cytometry using pMHCI tetramers (Table I). Healthy donors (n=54) were selected based on the presence of prevalent HLA alleles (A*02, A*11, B*07, B*08) or well-studied HLA alleles with a known role in specific infections (B*35, B*57) (19,30). Donors were routinely screened for epitope-specific responses and cognate CD8⁺ T-cell populations were sorted from 26 donors in total (Table I).

Table I. Study population.

Donor	HLA type	EBV-specific sorted CD8⁺ T cells[a]	CMV-specific sorted CD8⁺ T cells[a]
47-3904	A*01, A*02, B*07, B*08	A*02-GL9, B*08-RL8, B*08-FL9	-
47-3970	A*01, A*02, B*08, B*44	A*02-GL9, A*02-YV9, B*08-RL8, B*08-FL9	A*02-NV9, A*02-VL9
47-5008	A*02, A*11, B*35, B*44	A*11-AK10, A*02-GL9	A*02-NV9, B*35-IY9
47-5010	A*02, A*11, B*35, B*44	A*02-GL9, A*11-IK9, B*35-YM9	-
47-5017	A*02, A*03, B*07, B*18	A*02-GL9, B*07-RL9	-
47-5018	A*02, A*24, B*08, B*57	B*08-RL8, B*08-FL9, B*57-VW9	-
47-5022	A*02, A*24, B*07, B*62	-	A*02-NV9, A*02-VL9, B*07-TM10
47-5023	A*02, A*24, B*08, B*57	B*08-RL8	-
47-5024	A*02, A*03, B*07, B*08	B*08-RL8	-
47-5025	A*02, A*11, B*35, B*57	A*11-AK10, A*11-IK9, A*02-GL9, B*35-YM9	A*02-NV9, B*35-IY9
47-5027	A*02, A*11, B*07, B*35	A*11-IK9, A*11-AK10, B*07-RL9	-
47-5032	A*02, A*24, B*07, B*60	A*02-YV9	-
47-5033	A*03, A*11, B*35	A*11-IK9, A*11-AK10, B*35-YM9	B*35-IY9
47-5036	A*01, A*02, B*08, B*57	A*02-YV9, B*08-FL9, B*57-VW9	-
47-5039	A*02, A*03, B*07, B*44	A*02-YV9, B*07-RL9	-
47-5040	A*11, A*32, B*07, B*35	A*11-IK9, A*11-AK10, B*35-YM9	-
47-5041	A*01, A*02, B*08, B*27	B*08-FL9	-
47-5042	A*24, B*07, B*27	-	B*07-TM10
47-5043	A*01, A*02, B*35, B*57	B*57-VW9	-
47-5048	A*02, A*28, B*07, B*27	A*02-YV9, B*07-RL9	-
47-5052	A*02, A*03, B*07, B*44	-	B*07-TM10
47-5053	A*01, A*03, B*07, B*08	B*07-RL9	B*07-TM10
47-5054	A*02, A*03, B*08, B*35	-	B*35-IY9
47-5055	A*02, A*29, B*44, B*57	B*57-VW9	-
47-5060	A*03, A*11, B*27, B*35	-	B*35-IY9
328376	A*02, A*03, B*07, B*35	B*07-RL9	B*07-TM10

[a] - , no virus-specific CD8⁺ T-cell populations were sorted.

Response magnitudes varied between 0.02 and 7.16% of the CD8$^+$ T-cell population (median: 0.28%; Table II); a mean of 3,690 cells (range: 467-10,000 cells) were sorted per sample (data not shown). To measure the array of distinct TCRβ sequences in each antigen-specific population, clonotype analysis was performed as described previously (25).

Table II. Overview of analyzed CD8$^+$ T-cell populations.

Peptide	Peptide sequence	HLA restriction	Virus, protein	Donors analyzed (n)	Median CD8$^+$ T-cell response, % (range)	Fraction of recognition[a]
A*02-GL9	GLCTLVAML	A*0201	EBV, BMLF1	6	0.17 (0.02-4.45)	0.65
A*02-YV9	YVLDHLIVV	A*0201	EBV, BRLF1	5	0.28 (0.11-0.61)	0.39
A*02-NV9	NLVPMVATV	A*0201	CMV, pp65	4	0.50 (0.14-3.59)	0.50
A*02-VL9	VLEETSVML	A*0201	CMV, IE1	2	1.06 (0.14-1.98)	0.25
A*11-IK9	IVTDFSVIK	A*1101	EBV, EBNA3B	5	0.32 (0.23-1.10)	0.75
A*11-AK10	AVFDRKSDAK	A*1101	EBV, EBNA3B	5	0.11 (0.07-0.45)	1.00
B*07-RL9	RPPIFIRRL	B*0702	EBV, EBNA3A	6	0.25 (0.06-0.86)	0.87
B*07-TM10	TPRVTGGGAM	B*0702	CMV, pp65	5	0.43 (0.27-1.11)	0.67
B*08-RL8	RAKFKQLL	B*0801	EBV, BZLF1	5	1.66 (0.62-7.16)	1.00
B*08-FL9	FLRGRAYGL	B*0801	EBV, EBNA3A	5	0.52 (0.08-0.8)	0.67
B*35-YM9	YPLHEQHGM	B*3501	EBV, EBNA3A	4	0.17 (0.07-0.66)	0.64
B*35-IY9	IPSINVHHY	B*3501	CMV, pp65	5	0.86 (0.09-1.19)	0.70
B*57-VW9	VSFIEFVGW	B*5701	EBV, EBNA3B	4	0.09 (0.04-0.60)	0.57

[a] Fraction of recognition represents the relative proportion of HLA-matched individuals with a detectable tetramer-positive CD8$^+$ T-cell response against the epitope of interest.

TCRβ repertoire diversity is independent of pathogen source

We first investigated whether clonotypic diversity is influenced by the source of viral stimulation. The TCRβ repertoire was analyzed for a total of 45 EBV-specific CD8$^+$ T-cell populations (A*02-GL9, A*02-YV9, A*11-IK9, A*11-AK10, B*07-RL9, B*08-RL8, B*08-FL9, B*35-YM9 and B*57-VW9) and 16 CMV-specific CD8$^+$ T-cell populations (A*02-NV9, A*02-VL9, B*07-TM10 and B*35-IY9) (Supplementary Figs. S1 and S2). As there is no universal formula to assess TCRβ diversity, different approaches were undertaken to compare these CD8$^+$ T-cell repertoires comprehensively. First, the number of distinct clonotypes present in each CD8$^+$ T-cell population was determined (Fig. 1A). After normalization of each sample size to 50 sequences, the median relative number of distinct TCRβ clonotypes was 3.0 and 3.8 for EBV-specific and CMV-specific CD8$^+$ T-cell populations, respectively, which was not significantly different ($p = 0.98$, Mann-Whitney *U* test). To control for any potential bias arising from the normalization procedure, a Chao1 estimate of diversity (31) was calculated for all samples. This

estimation of total diversity correlated strongly with the relative number of clonotypes (data not shown), indicating that normalization introduced little bias in this study.

Although straightforward, counting the number of different clonotypes used in a response lacks sensitivity in that it neglects the frequency of individual clonotypes. To correct for this, Simpson's diversity index (D_S) was used as a second measure of TCRβ diversity (Fig. 1B). This index uses the relative frequency of each clonotype in the population to derive a score between 0 and 1, resembling a highly focused or a highly diverse repertoire, respectively. It has been shown that Simpson's diversity index is a good standard for the comparison of different $CD8^+$ T-cell repertoires (27). Using this formula, similar indices for EBV-specific and CMV-restricted TCRβ repertoires emerged (median EBV: $D_S = 0.29$; median CMV: $D_S = 0.46$; p=0.90, Mann-Whitney *U* test), again indicating that the viral origin of the target antigen does not appear to influence clonotypic diversity.

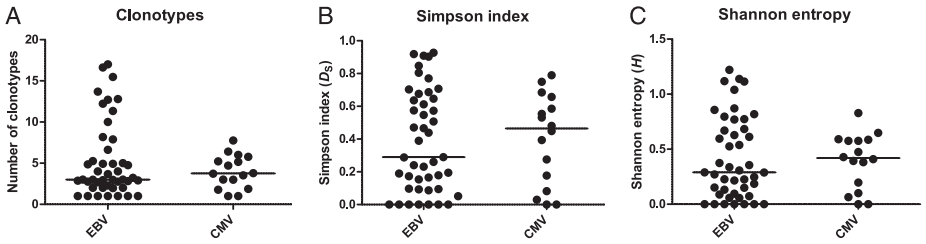


Figure 1. TCRβ diversity is not restricted by pathogen source.
Epitope-specific $CD8^+$ T-cell populations were sorted using pMHCI tetrameric complexes and TCRβ sequencing was performed as outlined in the *Materials and Methods*. Three distinct analytic approaches were used to compare TCRβ diversity across 45 EBV-specific and 16 CMV-specific $CD8^+$ T-cell populations from 26 healthy donors. The relative number of distinct clonotypes **(A)** represents a simple and straightforward measurement of diversity**.** Simpson's diversity index (D_S) **(B)** was used to examine diversity based on the clonal dominance of each clonotype. Shannon's entropy (*H*) **(C)** was applied to measure CDR3β heterogeneity. All samples were normalized to 50 TCRβ sequences to correct for any bias introduced by sample size. Bars represent median values. Significance testing was performed using the Mann-Whitney *U* test.

To take into account any sequence conservation between TCR CDR3β regions, a Shannon entropy (*H*) score was calculated for each $CD8^+$ T-cell population (Fig. 1C) (28). This score is based on sequence variation within the hypervariable CDR3β loop. The CDR3β sequences in each sample were first aligned, and the heterogeneity at each position was subsequently calculated and averaged. The conserved cysteine residue at the beginning and the phenylalanine residue at the end of the CDR3β sequence were excluded from the analysis. As CDR3 length variation can be interpreted as a parameter for sequence heterogeneity, a gap was considered as a unique position, akin to each of

the 20 amino acids. The resulting index is an indication of CDR3β sequence variation within a T-cell population. Again, our data showed similar Shannon entropy values for EBV-specific and CMV-specific $CD8^+$ T-cell populations (median EBV: $H = 0.29$; median CMV: $H = 0.42$; p=0.94, Mann-Whitney U test). These results indicate that the $CD8^+$ T-cell populations elicited by each virus are similarly heterogeneous across their CDR3β regions and, together with the number of clonotypes and Simpson's diversity index, demonstrate that TCRβ repertoire diversity is not determined by pathogen source.

TCRβ repertoire diversity is not governed by the HLA locus

Next, we examined the role of the HLA locus in repertoire formation. The TCRβ diversity of 27 $CD8^+$ T-cell populations restricted by HLA-A molecules (A*02-GL9, A*02-YV9, A*02-NV9, A*02-VL9, A*11-IK9 and A*11-AK10) and 34 $CD8^+$ T-cell populations restricted by HLA-B molecules (B*07-RL9, B*07-TM10, B*08-RL8, B*08-FL9, B*35-YM9, B*35-IY9 and B*57-VW9) was compared. Both the relative number of clonotypes (HLA-A: 3.0; HLA-B: 4.7; p = 0.19, Mann-Whitney U test; Fig. 2A), as well as Simpson's diversity index (HLA-A: $D_S = 0.28$; HLA-B: $D_S = 0.43$; p = 0.43, Mann-Whitney U test; Fig. 2B) and Shannon's entropy (HLA-A: $H = 0.32$; HLA-B: $H = 0.38$; p = 0.53, Mann-Whitney U test; Fig. 2C) were all comparable across both loci. In addition, TCR CDR3β length showed a comparable, Gaussian-shaped distribution for both HLA groups (data not shown). These data demonstrate that $CD8^+$ T-cell populations restricted by HLA-A and HLA-B molecules share similar degrees of TCRβ diversity, thereby indicating that repertoire diversity is not confined by the HLA locus.

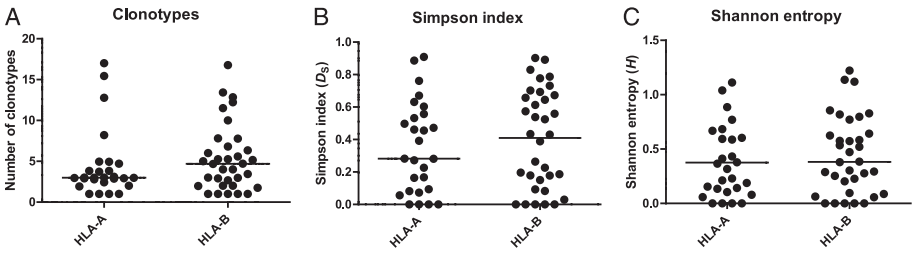


Figure 2. TCRβ diversity is not governed by the HLA locus.
The relative number of clonotypes **(A)**, Simpson's diversity index **(B)** and Shannon's entropy **(C)** were used to compare TCRβ diversity across 27 HLA-A-restricted and 34 HLA-B-restricted $CD8^+$ T-cell populations specific for EBV-derived or CMV-derived epitopes from 26 healthy donors. Bars represent median values. Significance testing was performed using the Mann-Whitney U test.

CD8⁺ T-cell populations restricted by HLA-A and HLA-B use overlapping TRBV genes

We further analyzed *TRBV* gene usage across EBV- and CMV-specific CD8⁺ T-cell populations restricted by HLA-A and HLA-B. First, *TRBV* gene usage was calculated per epitope by assembling all unique clonotypes (identified at the CDR3β nucleotide level) across all donors, regardless of clonotype frequency. Next, epitopes were combined based on their HLA restriction. Similar to the individual *TRBV* genes, we also calculated the distribution of groups of *TRBV* genes. To this end, individual *TRBV* genes were categorized into six groups according to their mutual homology (Supplementary Table S1) (32). A comparably diverse usage of individual *TRBV* genes and *TRBV* gene groups was observed across the HLA-A and HLA-B loci (Fig. 3A, 3C). Furthermore, a significant correlation was found for *TRBV* gene usage between the HLA-A and HLA-B loci ($r = 0.74$, $p < 0.001$, Spearman rank correlation; Fig. 3B).

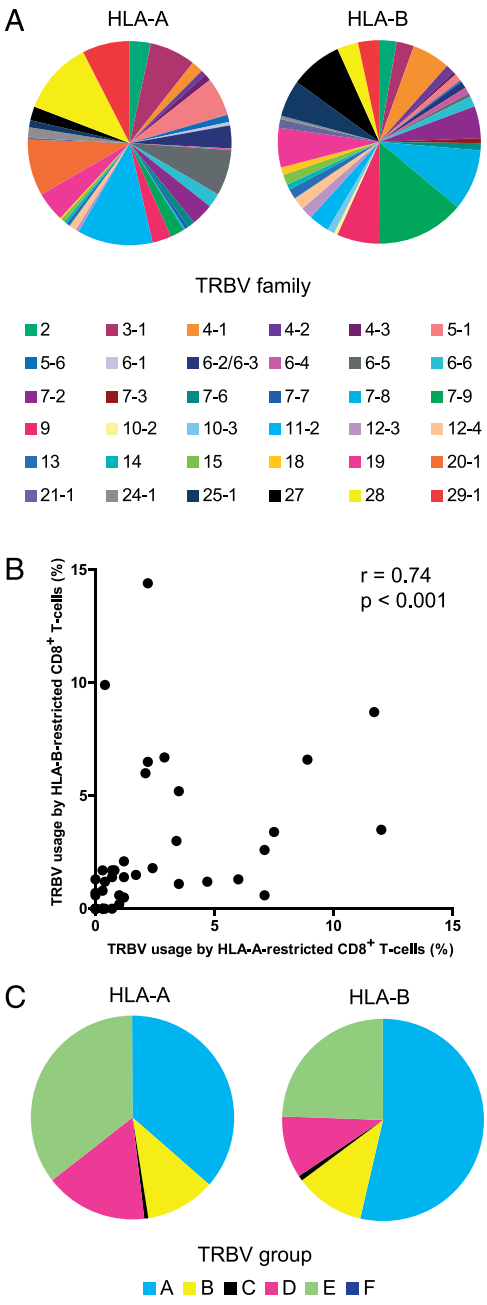


Figure 3. HLA-A and HLA-B molecules interact with an overlapping TRBV repertoire. *TRBV* gene usage **(A)** and *TRBV* gene group distribution **(C)** in CD8⁺ T-cell populations specific for epitopes restricted by HLA-A and HLA-B. *TRBV* gene and group usage were calculated across all unique TCRβ clonotypes specific for each peptide and then pooled together with other specificities restricted by HLA-A or HLA-B. Each clonotype was weighed equally, irrespective of clonal dominance. The usage of individual *TRBV* genes across all CD8⁺ T-cell populations restricted by HLA-A and HLA-B was correlated using the Spearman rank test **(B)**. The IMGT nomenclature was used for the *TRBV* genes, and *TRBV* gene family groups were created based on the homology between individual families (Supplementary Table S1).

Together, these data suggest that CD8$^+$ T-cell populations restricted by HLA-A and HLA-B use an overlapping array of *TRBV* genes and evolutionary gene groups, and reveal no clear preferences across these HLA loci in terms of their interplay with the *TRBV* gene repertoire.

TCRβ diversity does not correlate with immunodominance

We then examined whether the antigenic peptide shapes repertoire diversity as a function of immunodominance. This immunological phenomenon is somewhat loosely defined, but can refer either to the epitope-specific CD8$^+$ T-cell population with the highest frequency within an individual or to the most targeted epitope on a population level. For both definitions of immunodominance, it is unclear how this relates to the breadth of the responding repertoire. We first analyzed whether there was an association between the percentage of pMHCI tetramer-positive cells (Table II) in the CD8$^+$ T-cell compartment and TCRβ diversity. Neither the number of clonotypes ($r = 0.12$, $p = 0.37$, Spearman rank correlation; Fig. 4A) nor Simpson's diversity index ($r = 0.10$, $p = 0.43$, Spearman rank correlation; Fig. 4B) was significantly correlated with the magnitude of the CD8$^+$ T-cell response. This also applied when HLA-A and HLA-B molecules were analyzed separately, and further held true when CD8$^+$ T-cell response magnitude was quantified functionally using IFN-γ ELISpot analysis (data not shown).

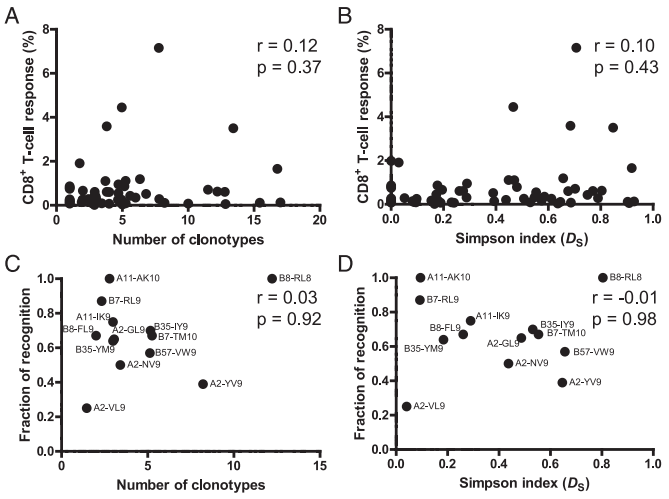


Figure 4. TCRβ repertoire diversity does not correlate with immunodominance. A total of 27 HLA-A- and 34 HLA-B-restricted CD8$^+$ T-cell populations specific for epitopes derived from EBV and CMV were quantified for response magnitude and correlated with TCRβ diversity. Repertoire diversity for each CD8$^+$ T-cell population was determined by the relative number of clonotypes **(A)** and Simpson's diversity index **(B).** The fraction of epitope recognition on a population level was calculated and correlated with TCRβ diversity using the relative number of clonotypes **(C)** and Simpson's diversity index **(D).** Recognition frequency for each epitope was based on measurable pMHCI tetramer-positive CD8$^+$ T-cell populations across a cohort of HLA-matched, EBV/CMV-reactive donors. The y-axis depicts the fraction of recognition for a peptide on a population level, with 0 representing epitopes that were never recognized and 1 representing epitopes that were recognized in all HLA-matched, EBV/CMV-reactive individuals tested (Table II, Supplementary Table S2). Median TCRβ diversity values per epitope are shown. Correlation testing was performed using the Spearman rank test.

In further analyses, we determined whether there was a correlation between TCRβ diversity and epitope dominance on a population level. To this end, we used pMHCI tetramer analysis to examine how frequently a specific epitope was recognized by $CD8^+$ T cells in our cohort of healthy individuals (Table II and Supplementary Table S2). The fraction of recognition for each peptide was used as a measure of population dominance. No significant correlation was found between the median TCRβ diversity per peptide-specific repertoire and the frequency of peptide recognition in the population (number of clonotypes: $r = 0.03$, $p = 0.92$, Spearman rank correlation; Simpson's diversity index: $r = -0.01$, $p = 0.98$, Spearman rank correlation; Fig. 4C and Fig. 4D). When using IFN-γ ELISpot analysis to determine the frequency of recognition of each epitope, again no significant correlation was observed between population dominance and TCRβ diversity (data not shown). Of note, there was no significant correlation between the magnitude of a $CD8^+$ T-cell response against an individual epitope and the frequency of recognition of that specific epitope on a population level (data not shown).

Taken together, these data indicate that a high-magnitude response can be the end product of a few highly proliferative clonotypes or the sum of several small-magnitude clonotypes. In addition, frequent recognition of an epitope on a population level is not correlated with the diversity of the $CD8^+$ T-cell response. In other words, a dominant epitope does not necessarily elicit a diverse repertoire and, vice versa, a subdominant epitope is not necessarily associated with an oligoclonal repertoire.

The antigen-specific TCRβ repertoire is shaped in a peptide-dependent manner

After excluding viral source, HLA restriction and immunodominance as driving factors behind T-cell repertoire diversity, we investigated the role of the epitope in repertoire formation. Samples were grouped according to pMHCI specificity, and TCRβ diversity was assessed using the different approaches described above. Interestingly, whereas some peptides elicited predominantly diverse or moderately diverse T-cell repertoires (e.g. A*02-YV9, B*07-TM10, B*08-RL8), others recruited a clonotypically much more skewed $CD8^+$ T-cell response (e.g. A*02-VL9, A*11-AK10, B*08-FL9) (Fig. 5). On a population level, significant differences were apparent between different peptides, regardless of the method used to quantify TCR diversity (number of clonotypes: $p = 0.005$; Simpson's diversity index: $p = 0.005$; Shannon's entropy; $p = 0.006$, Kruskal-Wallis test). Furthermore, a number of isolated $CD8^+$ T-cell populations were also analyzed for TCRα chain usage (n=16). TCRα diversity appeared to correlate positively with TCRβ diversity, i.e., little diversity in the TCRβ chain coincided with little TCRα

diversity and a more diverse TCRβ population was seen in conjunction with higher TCRα diversity (data not shown).

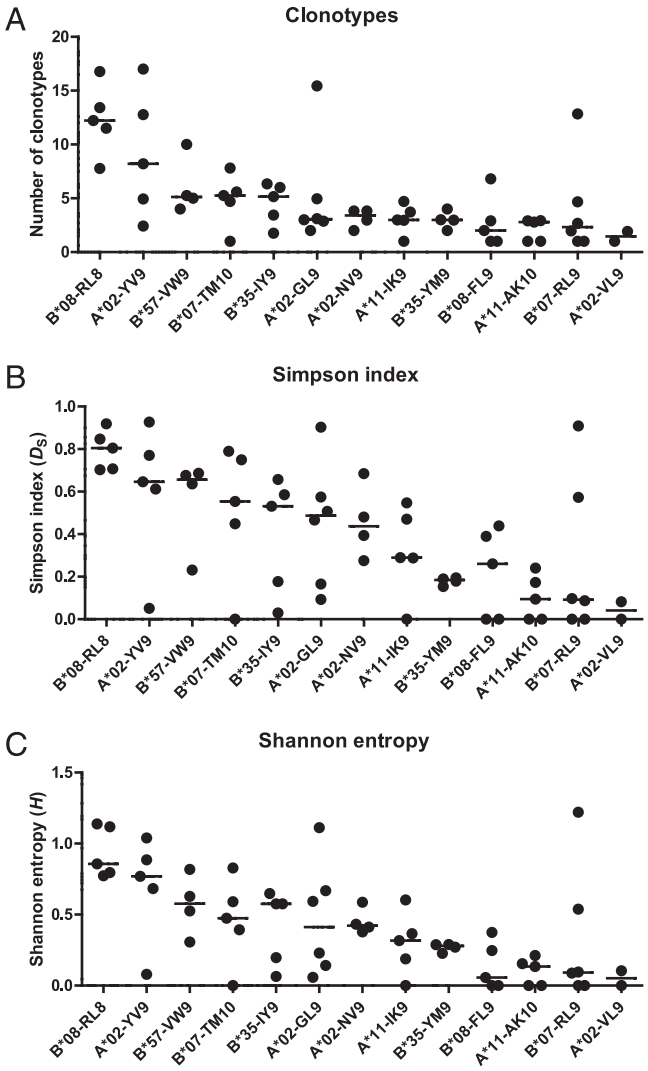


Figure 5. TCRβ repertoire diversity is guided in a peptide-dependent manner.
The relative number of clonotypes **(A)**, Simpson's diversity index **(B)** and Shannon's entropy **(C)** were examined for $CD8^+$ T-cell populations directed against a variety of epitopes derived from EBV and CMV, restricted by different HLA-A and HLA-B molecules. Epitopes are arranged on the x-axis in order of decreasing TCRβ diversity. Bars represent median values.

The EBV BZLF1-derived peptide RAKFKQLL, presented by HLA-B*08:01, selected $CD8^+$ T-cells with the highest level of TCRβ diversity, yielding a median of 12.3 clonotypes (range: 7.7-16.6), a Simpson's index of 0.80 (range: 0.70-0.92) and a Shannon's entropy score of 0.91 (range: 0.81-1.19). The CMV IE1-derived peptide VLEETSVML, presented by HLA-A*02:01, gave rise to the least diverse $CD8^+$ T-cell pool, with a median of 1.5 clonotypes (range: 1.0-1.9), a Simpson's index of 0.04 (range: 0-0.08) and a Shannon's entropy of 0.07 (range: 0-0.14). Of note, public [type III TCR bias (33)] and near-public [type IV TCR bias (8)] clonotypes were present in a number of $CD8^+$ T-cell populations, although no preferential selection within broad or narrow repertoires was observed (Supplementary Figs. S1 and S2).

To assess the role of the presented peptide epitope in the selection of specific *TRBV* gene families, we sampled distinct peptides sharing HLA restriction and viral background. Interestingly, each pMHCI complex appeared to select its own unique set of *TRBV* families and very little overlap was observed between peptides with shared HLA restriction (Fig. 6). For instance, the peptides GLCTLVAML and YVLDHLIVV, which are both presented by HLA-A*02:01 and derived from EBV, had neither a clear overlap in *TRBV* gene family usage nor a shared preference for a *TRBV* gene family group. Similar patterns were observed for the viral epitopes A*11-AVFDRKSDAK and A*11-IVTDFSVIK (EBV), B*08-RAKFKQLL and B*08-FLRGRAYGL (EBV), and A*02-NLVPMVATV and A*02-VLEETSVML (CMV), although $CD8^+$ T-cell responses to the latter epitope were only analyzed in two individuals. Together with the peptide-dependent pattern of TCRβ diversity, these results highlight a dominant role for the MHCI-bound peptide in antigen-driven repertoire selection.

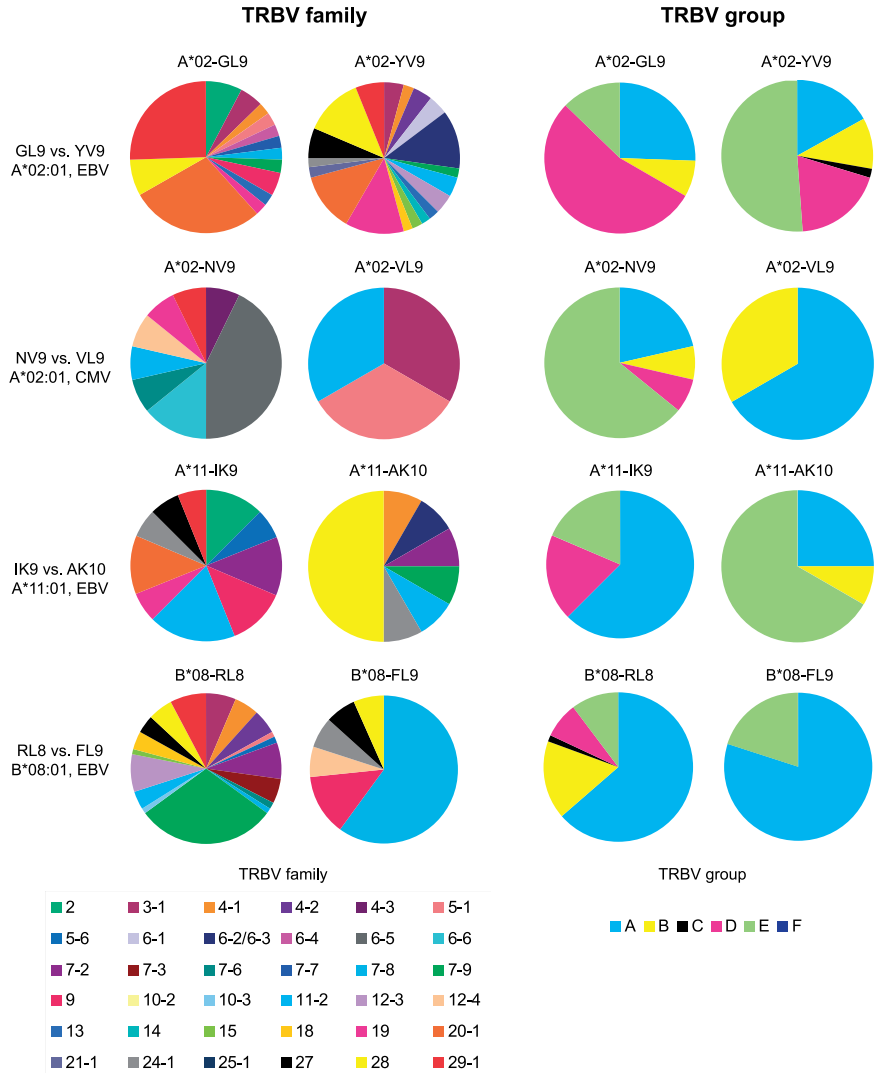


Figure 6. *TRBV* gene usage is peptide-driven.
TRBV gene usage and *TRBV* gene group usage were calculated for each peptide epitope by combining all unique TCRβ clonotypes from multiple individuals. Each unique clonotype was weighed equally, irrespective of clonal dominance. Repertoires for individual *TRBV* genes (left panel) and groups of homologous *TRBV* genes (right panel; Supplementary Table S1) were compared for peptides sharing viral origin and HLA background: A*02-GL9 and A*02-YV9 (EBV), A*02-NV9 and A*02-VL9 (CMV), A*11-IK9 and A*11-AK10 (EBV), and B*08-RL8 and B*08-FL9 (EBV).

Discussion

Previous studies have shown that epitope-specific $CD8^+$ T-cell populations can exhibit highly diverse or restricted patterns of TCR usage (7-9), but the factors that determine these contrasting outcomes are not well understood. Here, we studied the persistent viruses EBV and CMV to better understand the factors that govern TCRβ diversity. Analysis of multiple different epitope-specific $CD8^+$ T-cell populations excluded pathogen source as a principal determinant of TCRβ diversity, as judged by the relative number of clonotypes used, Simpson's diversity index and Shannon's entropy. Usage of shared, or so-called public, clonotypes was also found not to correlate with TCRβ diversity. This is in accordance with previous studies showing the presence of public TCRs in both diverse (14) and highly skewed antigen-specific T-cell repertoires (34). Furthermore, no significant differences in TCRβ diversity were observed between $CD8^+$ T-cell responses restricted by HLA-A and HLA-B molecules, which also showed overlapping patterns of *TRBV* gene and *TRBV* gene family group usage. These findings suggest that the frequently reported associations of HLA-B subtypes with immunodominance and disease outcome are largely repertoire-independent (16,17,20,21). On a global scale, there is evidence that some *TR* genes have a preference for either MHCI or MHCII molecules (35). Moreover, some early studies suggested a correlation between HLA haplotype and the expression of specific *TRBV* gene segments (36,37). Our data, however, are consistent with more recent literature showing that the mature human TCR repertoire is shaped, to a substantial extent at least, independently of the HLA locus (38,39).

The major driving factor behind antigen-specific TCRβ repertoire formation appeared to be the presented peptide. Accordingly, we observed that some peptides preferentially recruited a diverse T-cell repertoire, whereas others were much more focused in their clonotypic architecture. In addition, distinct epitopes sharing HLA restriction elicited dissimilar epitope-specific repertoires. These differences related neither to $CD8^+$ T-cell avidity for pMHCI nor to predicted peptide affinity for MHCI (A.I. Costa, unpublished observations), and is presumably caused by structural determinants of the MHCI-bound peptide, as has been shown previously in mice (40,41). These studies revealed that single amino acid mutants can have a profound impact on T-cell repertoire diversity, both in an antigen-specific situation (40) and during selection of clonotypes in the thymus (41). Furthermore, antigenic peptides that bulge out of the MHC binding groove tend to select for highly biased, narrow T-cell repertoires (42,43). Shorter antigenic peptides that form flatter, more featureless

landscapes can recruit either diverse (43) or biased T-cell repertoires (40). However, the precise structural factor(s) that underlie this functional divide remain yet to be determined. Interestingly, although MHCI-bound peptide has been observed to influence T-cell repertoire diversity in murine settings and/or studies focusing on a few epitopes, this message holds true and is even enforced by the data in this unbiased and comprehensive study in which a heterogeneous human population is investigated. Furthermore, crystallographic studies are in agreement with the notion that *TR* gene usage is shaped around specific pMHC antigens rather than more broadly around the MHC molecule itself. Human TCRs sharing common *TRBV* (or *TRAV*) gene usage, but not necessarily the same antigen specificity, engage the same MHC molecule using different anchoring sites (44,45). In addition, two human TCRs with similar *TRAV* and *TRBV* gene usage have been shown to dock HLA-B*44 with dramatically different orientations (46,47). However, one of these complexes is alloreactive, which may confound interpretation. Although in mice there is evidence that different TCRβ chains with different antigen specificity can engage the same MHC molecule using structurally superimposable interactions (48,49), the evidence for TCR-MHC interaction "codons" in humans is lacking at present. This suggests that other forces may shape the phenomenon of TCR bias, and MHC-bound peptide is an obvious candidate. Mutagenesis mapping work showing that TCR binding energy is predominantly peptide-focused is consistent with this hypothesis (50).

It should be noted that not all TCRβ repertoires against a given pMHCI antigen were equally diverse, as highlighted by several outliers, for example in the A*02-YV9 and B*07-RL9 analyses. Several explanations for this phenomenon can be proposed, but none are easy to address. For instance, in the case of CMV, the occurrence of superinfection might distort the pre-existing TCRβ repertoire (51). In addition, an acute, unrelated infection could temporarily deflate the memory compartment against other specificities, potentially eroding clonotypic diversity for previously encountered epitopes (52). It is also possible that heterologous immunity could play a role. In this scenario, epitope-specific $CD8^+$ T-cell populations that are significantly less diverse than expected might consist of cross-reactive clonotypes that were originally recruited in response to a different pathogen. Several studies have shown the existence of such heterologous immunity for a variety of viruses, including EBV (53,54). This phenomenon mostly resulted in a skewing of the epitope-specific TCRβ repertoire, although a recent study posed the possibility of a diverse heterologous repertoire (55). Finally, disparate nonrestricting MHC alleles have been shown to alter the available repertoire of cognate clonotypes through selective thymic deletion (56). This latter effect likely has far-reaching consequences for antigen-specific repertoire formation.

Regarding immunodominance, our data showed no evidence for a significant correlation between TCRβ diversity and either $CD8^+$ T-cell response magnitude or the dominance of peptide antigens on a population level. Interestingly, a recent study in HIV-1 infected individuals found a positive correlation between the magnitude of a response and TCRβ diversity (57). This discrepancy might be caused by qualitative differences between the viruses studied. It is not known to what extent the continuously active level of viral replication in HIV-1 infection and the progressive type of disease (neither of which are seen in EBV and CMV infections) might influence the immunodominance of epitopes and the breadth of the responding TCRβ repertoire. Nevertheless, in the context of persistent DNA viruses, dominant $CD8^+$ T-cell responses appear to be driven independently of T-cell repertoire diversity.

In summary, our study is the first, to our knowledge, to analyze the factors that shape the human antigen-specific T-cell repertoire in an unbiased and comprehensive setting. Despite the broad nature of genetic background in human populations and the broad range of environmental factors that can influence many immunological components, MHCI-bound peptide was found to shape the T-cell repertoire rather than the viral source, HLA restriction or immunodominance. These findings enhance our understanding of the factors involved in TCR biology and may have ramifications when attempting to manipulate the TCR repertoire for the purposes of rational vaccine design and therapeutic intervention.

Acknowledgements

We thank Gerrit Spierenburg and Koos Gaiser for assistance with fluorescence-activated cell sorting, and Jeffrey Beekman for the contribution to optimal figure design. This work was supported by a High Potential (HiPo) grant from the University of Utrecht.

References

1. Hochweller, K., G. H. Wabnitz, Y. Samstag, J. Suffner, G. J. Hammerling, and N. Garbi. 2010. Dendritic cells control T cell tonic signaling required for responsiveness to foreign antigen. *Proc. Natl. Acad. Sci. U. S. A* 107: 5931-5936.

2. Market, E., and F. N. Papavasiliou. 2003. V(D)J recombination and the evolution of the adaptive immune system. *PLoS. Biol.* 1:E16.

3. Arstila, T. P., A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, and P. Kourilsky. 1999. A direct estimate of the human alphabeta T cell receptor diversity. *Science* 286: 958-961.

4. Wooldridge, L., J. Ekeruche-Makinde, H. A. van den Berg, A. Skowera, J. J. Miles, M. P. Tan, G. Dolton, M. Clement, S. Llewellyn-Lacey, D. A. Price, M. Peakman, and A. K. Sewell. 2012. A single autoimmune T cell receptor recognizes more than a million different peptides. *J. Biol. Chem.* 287:1168-1177.

5. Wilson, D. B., D. H. Wilson, K. Schroder, C. Pinilla, S. Blondelle, R. A. Houghten, and K. C. Garcia. 2004. Specificity and degeneracy of T cells. *Mol. Immunol.* 40: 1047-1055.

6. Borghans, J. A., A. J. Noest, and R. J. De Boer. 1999. How specific should immunological memory be? *J. Immunol.* 163: 569-575.

7. Campos-Lima, P. O., V. Levitsky, M. P. Imreh, R. Gavioli, and M. G. Masucci. 1997. Epitope-dependent selection of highly restricted or diverse T cell receptor repertoires in response to persistent infection by Epstein-Barr virus. *J. Exp. Med.* 186: 83-89.

8. Miles, J. J., D. C. Douek, and D. A. Price. 2011. Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol. Cell Biol.* 89: 375-387.

9. Silins, S. L., S. M. Cross, S. L. Elliott, S. J. Pye, J. M. Burrows, D. J. Moss, and I. S. Misko. 1997. Selection of a diverse TCR repertoire in response to an Epstein-Barr virus-encoded transactivator protein BZLF1 by CD8+ cytotoxic T lymphocytes during primary and persistent infection. *Int. Immunol.* 9: 1745-1755.

10. Messaoudi, I., J. A. Guevara Patino, R. Dyall, J. LeMaoult, and J. Nikolich-Zugich. 2002. Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science* 298: 1797-1800.

11. Price, D. A., S. M. West, M. R. Betts, L. E. Ruff, J. M. Brenchley, D. R. Ambrozak, Y. Edghill-Smith, M. J. Kuroda, D. Bogdan, K. Kunstman, N. L. Letvin, G. Franchini, S. M. Wolinsky, R. A. Koup, and D. C. Douek. 2004. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. *Immunity.* 21: 793-803.

12. Meyer-Olson, D., N. H. Shoukry, K. W. Brady, H. Kim, D. P. Olson, K. Hartman, A. K. Shintani, C. M. Walker, and S. A. Kalams. 2004. Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape. *J. Exp. Med.* 200: 307-319.

13. Dong, T., G. Stewart-Jones, N. Chen, P. Easterbrook, X. Xu, L. Papagno, V. Appay, M. Weekes, C. Conlon, C. Spina, S. Little, G. Screaton, M. A. van der, D. D. Richman, A. J. McMichael, E. Y. Jones, and S. L. Rowland-Jones. 2004. HIV-specific cytotoxic T cells from long-term survivors select a unique T cell receptor. *J. Exp. Med.* 200: 1547-1557.

14. Price, D. A., T. E. Asher, N. A. Wilson, M. C. Nason, J. M. Brenchley, I. S. Metzler, V. Venturi, E. Gostick, P. K. Chattopadhyay, M. Roederer, M. P. Davenport, D. I. Watkins, and D. C. Douek. 2009. Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. *J. Exp. Med.* 206: 923-936.

15. Laugel, B., D. K. Cole, M. Clement, L. Wooldridge, D. A. Price, and A. K. Sewell. 2011. The multiple roles of the CD8 coreceptor in T cell biology: opportunities for the selective modulation of self-reactive cytotoxic T cells. *J. Leukoc. Biol.* 90: 1089-1099.

16. Bihl, F., N. Frahm, G. L. Di, J. Sidney, M. John, K. Yusim, T. Woodberry, K. Sango, H. S. Hewitt, L. Henry, C. H. Linde, J. V. Chisholm, III, T. M. Zaman, E. Pae, S. Mallal, B. D. Walker, A. Sette, B. T. Korber, D. Heckerman, and C. Brander. 2006. Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *J. Immunol.* 176: 4094-4101.

17. Hollsberg, P. 2002. Contribution of HLA class I allele expression to CD8+ T-cell responses against Epstein-Barr virus. *Scand. J. Immunol.* 55: 189-195.

18. Harari, A., C. Cellerai, F. B. Enders, J. Kostler, L. Codarri, G. Tapia, O. Boyman, E. Castro, S. Gaudieri, I. James, M. John, R. Wagner, S. Mallal, and G. Pantaleo. 2007. Skewed association of polyfunctional antigen-specific CD8 T cell populations with HLA-B genotype. *Proc. Natl. Acad. Sci. U. S. A* 104: 16233-16238.

19. Kiepiela, P., S. Chetty, L. Hilton, S. Moore, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, C. Day, B. Korber, H. M. Coovadia, B. D. Walker, and P. J. Goulder. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432: 769-775.

20. Neumann-Haefelin, C., S. McKiernan, S. Ward, S. Viazov, H. C. Spangenberg, T. Killinger, T. F. Baumert, N. Nazarova, I. Sheridan, O. Pybus, W. F. von, M. Roggendorf, D. Kelleher, H. E. Blum, and R. Thimme. 2006. Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution. *Hepatology* 43: 563-572.

21. Hill, A. V., C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett, D. Brewster, A. J. McMichael, and B. M. Greenwood. 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352: 595-600.

22. Sylwester, A. W., B. L. Mitchell, J. B. Edgar, C. Taormina, C. Pelte, F. Ruchti, P. R. Sleath, K. H. Grabstein, N. A. Hosken, F. Kern, J. A. Nelson, and L. J. Picker. 2005. Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. *J. Exp. Med.* 202: 673-685.

23. Hislop, A. D., G. S. Taylor, D. Sauce, and A. B. Rickinson. 2007. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. *Annu. Rev. Immunol.* 25: 587-617.

24. Kostense, S., G. S. Ogg, E. H. Manting, G. Gillespie, J. Joling, K. Vandenberghe, E. Z. Veenhof, Baarle. D. van, S. Jurriaans, M. R. Klein, and F. Miedema. 2001. High viral burden in the presence of major HIV-specific CD8(+) T cell expansions: evidence for impaired CTL effector function. *Eur. J. Immunol.* 31: 677-686.

25. Quigley, M. F., J. R. Almeida, D. A. Price, and D. C. Douek. 2011. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. *Curr. Protoc. Immunol.* Chapter 10:Unit10.

26. Lefranc, M. P. 2001. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* 29: 207-209.

27. Venturi, V., K. Kedzierska, S. J. Turner, P. C. Doherty, and M. P. Davenport. 2007. Methods for comparing the diversity of samples of the T cell receptor repertoire. *J. Immunol. Methods* 321: 182-195.

28. Shannon, C. E. 1997. The mathematical theory of communication. 1963. *MD Comput.* 14: 306-317.

29. Navis, M., I. Schellens, Baarle. D. van, J. Borghans, Swieten. P. van, F. Miedema, N. Kootstra, and H. Schuitemaker. 2007. Viral replication capacity as a correlate of HLA B57/B5801-associated nonprogressive HIV-1 infection. *J. Immunol.* 179: 3133-3143.

30. Carrington, M., G. W. Nelson, M. P. Martin, T. Kissner, D. Vlahov, J. J. Goedert, R. Kaslow, S. Buchbinder, K. Hoots, and S. J. O'Brien. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283: 1748-1752.

31. Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. Bohannan. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67: 4399-4406.

32. Su, C., and M. Nei. 2001. Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. *Mol. Biol. Evol.* 18: 503-513.

33. Turner, S. J., P. C. Doherty, J. McCluskey, and J. Rossjohn. 2006. Structural determinants of T-cell receptor bias in immunity. *Nat. Rev. Immunol.* 6: 883-894.

34. Argaet, V. P., C. W. Schmidt, S. R. Burrows, S. L. Silins, M. G. Kurilla, D. L. Doolan, A. Suhrbier, D. J. Moss, E. Kieff, T. B. Sculley, and I. S. Misko. 1994. Dominant selection of an invariant T cell antigen receptor in response to persistent infection by Epstein-Barr virus. *J. Exp. Med.* 180: 2335-2340.

35. Sim, B. C., D. Lo, and N. R. Gascoigne. 1998. Preferential expression of TCR V alpha regions in CD4/CD8 subsets: class discrimination or co-receptor recognition? *Immunol. Today* 19: 276-282.

36. Akolkar, P. N., B. Gulwani-Akolkar, R. Pergolizzi, R. D. Bigler, and J. Silver. 1993. Influence of HLA genes on T cell receptor V segment frequencies and expression levels in peripheral blood lymphocytes. *J. Immunol.* 150: 2761-2773.

37. Gulwani-Akolkar, B., D. N. Posnett, C. H. Janson, J. Grunewald, H. Wigzell, P. Akolkar, P. K. Gregersen, and J. Silver. 1991. T cell receptor V-segment frequencies in peripheral blood T cells correlate with human leukocyte antigen type. *J. Exp. Med.* 174: 1139-1146.

38. Robins, H. S., S. K. Srivastava, P. V. Campregher, C. J. Turtle, J. Andriesen, S. R. Riddell, C. S. Carlson, and E. H. Warren. 2010. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci. Transl. Med.* 2:47ra64.

39. Melenhorst, J. J., M. D. Lay, D. A. Price, S. D. Adams, J. Zeilah, E. Sosa, N. F. Hensel, D. Follmann, D. C. Douek, M. P. Davenport, and A. J. Barrett. 2008. Contribution of TCR-beta locus and HLA to the shape of the mature human Vbeta repertoire. *J. Immunol.* 180: 6484-6489.

40. Turner, S. J., K. Kedzierska, H. Komodromou, N. L. La Gruta, M. A. Dunstone, A. I. Webb, R. Webby, H. Walden, W. Xie, J. McCluskey, A. W. Purcell, J. Rossjohn, and P. C. Doherty. 2005. Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific CD8+ T cell populations. *Nat. Immunol.* 6: 382-389.

41. Fukui, Y., T. Oono, J. P. Cabaniols, K. Nakao, K. Hirokawa, A. Inayoshi, T. Sanui, J. Kanellopoulos, E. Iwata, M. Noda, M. Katsuki, P. Kourilsky, and T. Sasazuki. 2000. Diversity of T cell repertoire shaped by a single peptide ligand is critically affected by its amino acid residue at a T cell receptor contact. *Proc. Natl. Acad. Sci. U. S. A* 97: 13760-13765.

42. Miles, J. J., D. Elhassen, N. A. Borg, S. L. Silins, F. E. Tynan, J. M. Burrows, A. W. Purcell, L. Kjer-Nielsen, J. Rossjohn, S. R. Burrows, and J. McCluskey. 2005. CTL recognition of a bulged viral peptide involves biased TCR selection. *J. Immunol.* 175: 3826-3834.

43. Wynn, K. K., Z. Fulton, L. Cooper, S. L. Silins, S. Gras, J. K. Archbold, F. E. Tynan, J. J. Miles, J. McCluskey, S. R. Burrows, J. Rossjohn, and R. Khanna. 2008. Impact of clonal competition for peptide-MHC complexes on the CD8+ T-cell repertoire selection in a persistent viral infection. *Blood* 111: 4283-4292.

44. Cole, D. K., F. Yuan, P. J. Rizkallah, J. J. Miles, E. Gostick, D. A. Price, G. F. Gao, B. K. Jakobsen, and A. K. Sewell. 2009. Germ line-governed recognition of a cancer epitope by an immunodominant human T-cell receptor. *J. Biol. Chem.* 284: 27281-27289.

45. Garboczi, D. N., P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384: 134-141.

46. Archbold, J. K., W. A. Macdonald, S. Gras, L. K. Ely, J. J. Miles, M. J. Bell, R. M. Brennan, T. Beddoe, M. C. Wilce, C. S. Clements, A. W. Purcell, J. McCluskey, S. R. Burrows, and J. Rossjohn. 2009. Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. *J. Exp. Med.* 206: 209-219.

47. Macdonald, W. A., Z. Chen, S. Gras, J. K. Archbold, F. E. Tynan, C. S. Clements, M. Bharadwaj, L. Kjer-Nielsen, P. M. Saunders, M. C. Wilce, F. Crawford, B. Stadinsky, D. Jackson, A. G. Brooks, A. W. Purcell, J. W. Kappler, S. R. Burrows, J. Rossjohn, and J. McCluskey. 2009. T cell allorecognition via molecular mimicry. *Immunity.* 31: 897-908.

48. Feng, D., C. J. Bond, L. K. Ely, J. Maynard, and K. C. Garcia. 2007. Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. *Nat. Immunol.* 8: 975-983.

49. Dai, S., E. S. Huseby, K. Rubtsova, J. Scott-Browne, F. Crawford, W. A. Macdonald, P. Marrack, and J. W. Kappler. 2008. Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. *Immunity.* 28: 324-334.

50. Borg, N. A., L. K. Ely, T. Beddoe, W. A. Macdonald, H. H. Reid, C. S. Clements, A. W. Purcell, L. Kjer-Nielsen, J. J. Miles, S. R. Burrows, J. McCluskey, and J. Rossjohn. 2005. The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. *Nat. Immunol.* 6: 171-180.

51. Meyer-Konig, U., K. Ebert, B. Schrage, S. Pollak, and F. T. Hufert. 1998. Simultaneous infection of healthy people with multiple human cytomegalovirus strains. *Lancet* 352: 1280-1281.

52. Selin, L. K., M. Y. Lin, K. A. Kraemer, D. M. Pardoll, J. P. Schneck, S. M. Varga, P. A. Santolucito, A. K. Pinto, and R. M. Welsh. 1999. Attrition of T cell memory: selective loss of LCMV epitope-specific memory CD8 T cells following infections with heterologous viruses. *Immunity.* 11: 733-742.

53. Urbani, S., B. Amadei, P. Fisicaro, M. Pilli, G. Missale, A. Bertoletti, and C. Ferrari. 2005. Heterologous T cell immunity in severe hepatitis C virus infection. *J. Exp. Med.* 201: 675-680.

54. Clute, S. C., L. B. Watkin, M. Cornberg, Y. N. Naumov, J. L. Sullivan, K. Luzuriaga, R. M. Welsh, and L. K. Selin. 2005. Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. *J. Clin. Invest* 115: 3602-3612.

55. Clute, S. C., Y. N. Naumov, L. B. Watkin, N. Aslan, J. L. Sullivan, D. A. Thorley-Lawson, K. Luzuriaga, R. M. Welsh, R. Puzone, F. Celada, and L. K. Selin. 2010. Broad cross-reactive TCR repertoires recognizing dissimilar Epstein-Barr and influenza A virus epitopes. *J. Immunol.* 185: 6753-6764.

56. Burrows, S. R., S. L. Silins, D. J. Moss, R. Khanna, I. S. Misko, and V. P. Argaet. 1995. T cell receptor repertoire for a viral epitope in humans is diversified by tolerance to a background major histocompatibility complex antigen. *J. Exp. Med.* 182: 1703-1715.

57. Balamurugan, A., H. L. Ng, and O. O. Yang. 2010. Rapid T cell receptor delineation reveals clonal expansion limitation of the magnitude of the HIV-1-specific CD8+ T cell response. *J. Immunol.* 185: 5935-5942.

Supplementary material

Table S1. Distribution of individual *TRBV* genes into evolutionarily conserved groups.

TRBV group[a]	TRBV gene(s)
A	2, 5.1, 5.3, 5.4, 5.5, 5.6, 5.8, 7.2, 7.3, 7.6, 7.7, 7.8, 7.9, 9, 11.1, 11.2, 11.3, 12.3, 12.4, 13, 14, 18, 23.1
B	3.1, 4.1, 4.2, 4.3
C	15
D	20.1, 29.1
E	6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.8, 10.1, 10.2, 10.3, 19, 24.1, 25.1, 27, 28
F	30

[a] *TRBV* gene groups were based on mutual homology between functional *TRBV* genes (32).

(32) Su, C., and M. Nei. 2001. Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. *Mol. Biol. Evol.* 18: 503-513.

Table S2. The frequency of recognition of each epitope on a population level was determined by IFN-γ ELISPOT analysis or pMHCI tetramer staining.

	IFN-γ ELISPOT analysis		pMHCI tetramer analysis	
Peptide	Donors positive / donors tested	Fraction	Donors positive / donors tested	Fraction
A*02-GL9	14/24	0.58	15/23	0.65
A*02-YV9	8/24	0.33	9/23	0.39
A*02-NV9	11/17	0.65	8/16	0.50
A*02-VL9	4/17	0.24	4/16	0.25
A*11-IK9	6/8	0.75	6/8	0.75
A*11-AK10	8/8	1.00	8/8	1.00
B*07-RL9	11/15	0.73	13/15	0.87
B*07-TM10	7/10	0.70	6/9	0.67
B*08-RL8	13/13	1.00	12/12	1.00
B*08-FL9	4/13	0.31	8/12	0.67
B*35-YM9	11/15	0.73	9/14	0.64
B*35-IY9	9/11	0.82	7/10	0.70
B*57-VW9	4/8	0.50	4/7	0.57

Figure S1

A*02-YV9

47-3970			
Vβ	**CDR3**	**Jβ**	%
28	CASSSLTGSGQPQHF	1.5	35.7
28	CASSSLRGSNQPQHF	1.5	31.4
12.3	CASSLGTSFRTDTQYF	2.3	5.7
20.1	CSASNQESYGYTF	1.2	4.3
20.1	CSAPIPPYNEQFF	2.1	4.3
6.2/6.3	CASSYMGVEQFF	2.1	2.9
19	CASMGGASYNEQFF	2.1	2.9
11.2	CASSSPGDYTYEQYF	2.7	2.9
19	CASSMLLGGGQPQHF	1.5	1.4
19	CASRGLLGGITEAFF	1.1	1.4
3.1	CASSPSLVGSADTQYF	2.3	1.4
6.1	CATSWTGTYEQYF	2.7	1.4
20.1	CSARGAGFSGANVLTF	2.6	1.4
3.1	CASSQTLAGGGETQYF	2.5	1.4
28	CASSLYSDTQYF	2.3	1.4

47-5032			
Vβ	**CDR3**	**Jβ**	%
4.1	CASSQGFGGGETQYF	2.5	16.0
7.9	CASSLGVSQETQYF	2.5	14.0
28	CAGPQGRETQYF	2.5	12.0
4.2	CASSQGVLLNEQFF	2.1	8.0
27	CASRPLMGGASEAFF	1.1	6.0
29.1	CSVVLAPVQEQFF	2.1	6.0
6.2/6.3	CASTPMGLSSGETQYF	2.5	4.0
14	CASSQVSGQGALRALEQFF	2.1	4.0
12.3	CASSLGDRGLGYTF	1.2	4.0
29.1	CSVVLAPVQEQFF	2.1	4.0
20.1	CSASPPSYNEQFF	2.1	4.0
19	CASTALLGGGYGYTF	1.2	4.0
24.1	CATSDTGSFTDTQYF	2.5	4.0
6.2/6.3	CASSLFGERNEQYF	2.7	2.0
6.2/6.3	CASTSGFGSGETQYF	2.5	2.0
20.1	CSAGQALYNEQFF	2.1	2.0
15	CATSTGLAGNHEQYF	2.7	2.0
21.1 (P)	CASSKEESSYNSPLHF	1.6	2.0

47-5036			
Vβ	**CDR3**	**Jβ**	%
6.2/6.3	CASTVQGRETQYF	2.5	57.1
18	CASSPFQRSSGNTIYF	1.3	23.2
28	CASSPLPGTTGTDTQYF	2.3	10.7
19	CASRGFLGGGQPQHF	1.5	7.1
29.1	CSVEVQRETEQFF	2.1	1.8

47-5048			
Vβ	**CDR3**	**Jβ**	%
6.2/6.3	CASSLQGSNQPQHF	1.5	57.6
27	CASQRDYTGELFF	2.2	13.6
28	CASSSSLGFGELFF	2.2	7.6
13	CASTGGSGYTF	1.2	7.6
4.2	CASSPGVTSGRMGEQYF	2.7	6.1
20.1	CSASRALDQPQHF	1.5	3.0
6.4	CASSESLAVHEQFF	2.1	1.5
5.6	CASSPASSGSWETQYF	2.5	1.5
19	CATSQGSYGYTF	1.2	1.5

47-5039			
Vβ	**CDR3**	**Jβ**	%
27	CASRTLLGGASEQYF	2.7	97.3
6.1	CASSEEDGQAYEQYF	2.7	1.3
11.2	CASSSGNTEAFF	1.1	1.3

Figure S1 (cont.)

A*02-GL9

47-3904			
Vβ	**CDR3**	**Jβ**	*%*
29.1	CSVGADGTNEKLFF	1.4	16.1
20.1	CSARDRTGNGYTF	1.2	11.3
20.1	CSARDRTGNGYTF	1.2	9.7
20.1	CSARDQTGNGYTF	1.2	8.1
20.1	CSARDGTGNGYTF	1.2	8.1
5.1	CASSFSTDTQYF	2.3	8.1
28	CASSFFLPTQREYEQYF	2.7	8.1
3.1	CASSPTSGSIYEQYF	2.7	4.8
7.9	CASSEPGQGGRGEQYF	2.7	4.8
20.1	CSARDRVGNTIYF	1.3	3.2
20.1	CSARDGTGNGYTF	1.2	3.2
20.1	CSARDRVGNTIYF	1.3	1.6
3.1	CASSQDPRRSYNEQFF	2.1	1.6
4.1	CASSQEGFREQFF	2.1	1.6
29.1	CSVGSGGTNEKLFF	1.4	1.6
29.1	CSVGTGGTNEKLFF	1.4	1.6
29.1	CSVGSQGTNEKLFF	1.4	1.6
20.1	CSARDRVGNGYTF	1.2	1.6
28	CASSLRNGYEQYF	2.7	1.6
9	CASGTGQLTNTEAFF	1.1	1.6

47-3970			
Vβ	**CDR3**	**Jβ**	*%*
20.1	CSARGEAITEKLFF	1.4	71.7
19	CASSIDGTAYEQYF	2.7	15.0
7.7	CASTREGLNTGELFF	2.2	5.0
7.8	CASSLGLENEQFF	2.1	5.0
2	CAGGTGETSPGELFF	2.2	3.3

47-5008			
Vβ	**CDR3**	**Jβ**	*%*
7.8	CASIPEGGRETQYF	2.5	60.0
20.1	CSARSETGNTIYF	1.3	20.0
20.1	CSAEIIAEAIYNEQFF	2.1	20.0

47-5017			
Vβ	**CDR3**	**Jβ**	*%*
29.1	CSVG**S**GGTNEKLFF	1.4	43.5
29.1	CSVG**T**GGTNEKLFF	1.4	35.5
29.1	CSVG**T**GGTNEKLFF	1.4	9.7
29.1	CSVG**S**GGTNEKLFF	1.4	8.1
29.1	CSVG**T**GGTNEKLFF	1.4	3.2

47-5010			
Vβ	**CDR3**	**Jβ**	*%*
2	CASSDGKIAPSEQYF	2.7	95.6
9	CASSTGQLTNTEAFF	1.1	1.5
29.1	CSVGTGGTNEKLFF	1.4	1.5
28	CASSFIGLNTEAFF	1.1	1.5

47-5025			
Vβ	**CDR3**	**Jβ**	*%*
14	CASSQ**S**PGGIQYF	2.4	91.2
2	CASTPGSKLPDTQYF	2.3	7.0
6.4	CASREGSGDNNEQFF	2.1	1.8

A*02-NV9

47-3970			
Vβ	**CDR3**	**Jβ**	*%*
11.2	CASSFTWTSGGTTDTQYF	2.3	84.4
4.3	CASSPTSGVAGELFF	2.2	7.8
19	CASSIPDTEAFF	1.1	6.3
6.6	CASSYGFGPAQGEKLFF	1.4	1.6

47-5008			
Vβ	**CDR3**	**Jβ**	*%*
7.6	CASSLAPGATNEKLFF	1.4	35.9
6.5	CASSYSTGVPGELFF	2.2	34.4
6.5	CASSEVTGTAWGYTF	1.2	28.1
29.1	CSVVSQFVTDTYF	2.3	1.6

47-5022			
Vβ	**CDR3**	**Jβ**	*%*
6.5	CASSYSTGTHLEQYF	2.7	76.3
12.4	CASSSAYYGYTF	1.2	15.3
6.5	CASSLVTGTGKYGYTF	1.2	8.5

47-5025			
Vβ	**CDR3**	**Jβ**	*%*
6.5	CASSYQTGAAYGYTF	1.2	65.6
6.6	CASSHWDRETSGNTIYF	1.3	31.1
6.5	CASSCQTGAACGYTF	1.2	3.3

Figure S1 (cont.)

A*02-VL9

47-3970			
Vβ	**CDR3**	**Jβ**	*%*
5.1	CASSLDSQSSGNTIYF	1.3	96.0
3.1	CASSPVGGSMAKNIQYF	2.4	4.0

47-5022			
Vβ	**CDR3**	**Jβ**	*%*
11.2	CASSLQRGRTDTQYF	2.3	100

A*11-IK9

47-5010			
Vβ	**CDR3**	**Jβ**	*%*
7.2	CASSLNGGHYEQYF	2.7	100

47-5025			
Vβ	**CDR3**	**Jβ**	*%*
7.2	CASSWGQGSNYGYTF	1.2	51.4
9	CASREGDINTGELFF	2.2	44.3
19	CASSIVGAHTEAFF	1.1	4.3

47-5027			
Vβ	**CDR3**	**Jβ**	*%*
20.1	CSARDGLGLAYEQYF	2.7	84.1
11.2	CASSFGGGSNYGYTF	1.2	6.3
20.1	CSARDCLGLAYEQYF	2.7	4.8
11.2	CASSWGAGSNYGYTF	1.2	3.2
25.1	CASSARDRGIGGYTF	1.2	1.6

47-5033			
Vβ	**CDR3**	**Jβ**	*%*
2	CASSEAAGGFYNEQFF	2.1	67.2
2	CASSASAGGYYNEQFF	2.1	29.5
27	CASSPTGLGGGYTF	1.2	1.6
5.6	CASSFRDRGSSGGIKGEKLFF	1.4	1.6

47-5040			
Vβ	**CDR3**	**Jβ**	*%*
11.2	CASSWGGGSNYGYTF	1.2	84.1
29.1	CSVEGGPGTAYNEQFF	2.1	10.1
9	CASTELGDTHNEQFF	2.1	5.8

A*11-AK10

47-5008			
Vβ	**CDR3**	**Jβ**	*%*
7.2	CASSLNRGAREKLFF	1.4	100

47-5025			
Vβ	**CDR3**	**Jβ**	*%*
7.9	CASGWGSQPQHF	1.5	100

47-5027			
Vβ	**CDR3**	**Jβ**	*%*
28	CASSRTSATGELFF	2.2	86.1
6.2/6.3	CASSQEGTGSYEQYF	2.7	9.0
28	CASSLTSATGELFF	2.2	4.1
28	CASSRTSATGELFF	2.2	0.8

47-5033			
Vβ	**CDR3**	**Jβ**	*%*
28	CASTQTSAAGELFF	2.2	90.9
28	CASSLSSATGELFF	2.2	7.3
11.2	CASTPGTEYNEQYFF	2.1	1.8

47-5040			
Vβ	**CDR3**	**Jβ**	*%*
28	CASSLTSAAGELFF	2.2	95.2
24.1	CATSDPDSPTDTQYF	2.5	3.2
4.1	CASSQDAGYEAFF	1.1	1.6

Figure S1 - TCRβ repertoires for virus-specific CD8$^+$ T-cell populations restricted by HLA-A molecules. TRBV and TRBJ usage, CDR3β amino acid sequence and relative frequency are shown for HLA-A-restricted clonotypes specific for epitopes derived from EBV and CMV.

Figure S2

B*07-RL9

47-5017

Vβ	CDR3	Jβ	%
4.1	CASSQETGSYEQYF	2.7	95.1
4.1	CASSQDRLTGGYTF	1.2	4.9

47-5027

Vβ	CDR3	Jβ	%
4.1	CASSQETGSYEQFF	2.1	95.7
7.2	CASSLTGGTYEQYF	2.7	2.9
24.1	CATATGTGSNTEAFF	1.1	1.4

47-5039

Vβ	CDR3	Jβ	%
4.1	CASSLETGSYEQYF	2.7	55.4
6.6	CASSAAGELETQYF	2.5	35.7
12.4	CASLYRSSTDTQYF	2.3	5.4
3.1	CASSPGTLYGYTF	1.2	1.8
2	CASSDSQRMNTEAFF	1.1	1.8

328376

Vβ	CDR3	Jβ	%
28	CASYWGLAVNEQFF	2.1	17.0
4.1	CASSQDRLTGGYTF	1.2	15.1
7.2	CASRTLAGPNTGELFF	2.2	13.2
9	CASSEILGMNTEAFF	1.1	11.3
4.1	CASSQETGSASQHF	2.1	9.4
7.8	CASSPTNPAGDTETQYF	2.5	7.5
29.1	CSVGGAYEQYF	2.7	5.7
14	CASSGTHNEQFF	2.1	5.7
27	CASSLSQGIQNIQYF	2.4	5.7
5.1	CASSFRLAVKNEQFF	2.1	3.8
10.3	CATLAGEGEQFF	2.1	1.9
7.6	CASSLEAGVSYEQYF	2.7	1.9
7.9	CASSLGASTDTQYF	2.3	1.9

47-5048

Vβ	CDR3	Jβ	%
4.1	CASSQETSLYNEQFF	2.1	100

47-5053

Vβ	CDR3	Jβ	%
4.1	CASSQETGASGTQYF	2.3	100

B*07-TM10

47-5022

Vβ	CDR3	Jβ	%
27	CASRVGTGNLYEQYF	2.7	33.3
7.9	CASSLALGVAKNIQYF	2.4	23.3
7.9	CASSSHDRDGGYSPLHF	1.6	15.0
7.9	CASSSHERQGQNSPLHF	1.6	8.3
27	CASRVGAGNTEAFF	1.1	8.3
7.9	CASSQHDRGFSTEAFF	1.1	3.3
7.9	CASSLALGVAKNIQYF	2.4	3.3
7.9	CASSSHDRQGQNSPLHF	1.6	3.3
3.1	CASSLLPYNALVSGTDTQYF	2.3	1.7

47-5053

Vβ	CDR3	Jβ	%
7.9	CASSLQTGLNTEAFF	1.1	63.8
7.9	CASSLIGQGGPDTQYF	2.3	18.8
7.9	CASSLRQGIDTGELFF	2.2	11.6
7.9	CASSLGEGGAQSEQFF	2.1	4.3
6.2/6.3	CASSYSGNTEAFF	1.1	1.4

47-5042

Vβ	CDR3	Jβ	%
7.9	CASSFRDRGHYEQYF	2.7	100

328376

Vβ	CDR3	Jβ	%
6.2/6.3	CASSYSGNTEAFF	1.1	72.0
27	CASSLSPPTYEQYF	2.7	17.3
27	CASKLGPAAYEQYF	2.7	5.3
4.3	CASGRGTGTRGRNEQFF	2.1	2.7
10.3	CAISDRGGEDEQYF	2.7	1.3
7.9	CASSLLGQDNSPLHF	1.6	1.3

47-5052

Vβ	CDR3	Jβ	%
7.9	CASSLVGEGLRDEQFF	1.1	32.4
7.9	CASSLLQGARTEAFF	1.1	30.9
7.9	CASSLIGVSSYNEQFF	2.1	23.5
28	CASSLQTGQETQYF	2.5	8.8
7.9	CASSNRDRGNYEQFF	2.1	2.9
5.1	CASSLSGLGAAGVYEQYF	2.7	1.5

Figure S2 (cont.)

B*08-RL8

47-3904

Vβ	CDR3	Jβ	%
4.1	CASSQGVGENTEAFF	1.1	15.9
7.3	CASSLAVLSSYNEQFF	2.1	15.9
7.3	CASSRGTGEGYEQYF	2.7	10.1
18	CASSPPGPTGRGWYDTF	1.2	8.7
4.2	CASSQAQGMGNSPLHF	1.6	7.2
7.9	CASSLTGPGDQPQHF	1.5	5.8
7.9	CASSLTGAGDQPQHF	1.5	4.3
7.9	CASSLTAPGIQPQHF	1.5	4.3
27	CASSPLTDTQYF	2.3	4.3
7.9	CASSSVGPGDQPQHF	1.5	2.9
7.9	CASSKQGRGTDTQYF	2.3	2.9
7.3	CASSLTLLGNNEQYF	2.7	2.9
7.6	CASSLLIGANNEQFF	2.1	2.9
28	CASSLEGSYEQYF	2.7	2.9
11.2	CASSPITRIVGDTQYF	2.3	2.9
7.3	CASSLIASGGYNEQFF	2.1	1.4
7.9	CASSLAVLSSYNEQFF	2.1	1.4
29.1	CSVGAGDYEQYF	2.7	1.4
3.1	CASSQVRGGYEQFF	2.1	1.4

47-3970

Vβ	CDR3	Jβ	%
27	CASSDLNSPLHF	1.6	44.3
3.1	CASSQVRGRGEQYF	2.7	23.3
7.2	CASSVLAGDTQYF	2.3	18.6
7.9	CASSLTGPGDQPQHF	1.5	3.3
7.9	CASSLTEAGNQPHHF	1.5	2.9
7.9	CASLRGLGTDTQYF	2.3	2.4
4.2	CASSPLIGTSGTEAFF	1.1	1.4
4.2	CASSPGQVTPLHF	1.6	1.0
29.1	CSGGQGEGEQYF	2.7	1.0
28	CAGRPRLAGGNNEQF	2.1	0.5
7.9	CASSLSSPGNQPQH	1.5	0.5
7.2	CASSQIVTSAAQY	2.7	0.5
29.1	CSTGTGEGGTQY	2.5	0.5

47-5018

Vβ	CDR3	Jβ	%
10.3	CAISATGAVTEAFF	1.1	26.2
7.9	CASSSTEQATYGYTF	1.2	24.6
5.1	CASRAGLAEVSYEQYF	2.7	13.1
29.1	CSVGAGDYEQYF	2.7	11.5
12.3	CASSLIGGDYEQYF	2.7	4.9
7.8	CASSPGTGEGYEQYF	2.7	3.3
4.1	CASSQAGGTMNTEAFF	1.1	3.3
4.1	CASSPGQGSSYEQYF	2.7	1.6
7.9	CASSSLGGVDQPQHF	1.5	1.6
7.9	CASSLNTLSTEAFF	1.1	1.6
7.9	CASSHIRDYGAFF	1.1	1.6
4.1	CASSQAGATRTDTQYF	2.3	1.6
29.1	CSDGTGDFGTQYF	2.5	1.6
18	CASLASGQGSHEQFF	2.1	1.6
3.1	CASSQEGRVVGGEQFF	2.1	1.6

47-5023

Vβ	CDR3	Jβ	%
7.9	CASSLVSSGANVLTF	2.6	38
12.3	CASSFSGSLTTEAFF	1.1	19.7
18	CASKMRGDVYEQYF	2.7	9.9
3.1	CASSQAGQGNSPLHF	1.6	8.5
28	CASSFQGSNEKLFF	1.4	5.6
15	CATTQYGTDTQYF	2.3	2.8
7.2	CASSLLGHPSGELFF	2.2	2.8
7.9	CASSLPGTSGISQDTQYF	2.3	1.4
7.9	CASSYLGDSTNSPLHF	1.6	1.4
7.9	CASSLAPSYEQYF	2.7	2.8
7.9	CASSRTDDTEAFF	1.1	1.4
12.3	CASLFTGTSPTDTQYF	2.3	2.8
12.3	CASSPGTSGGGDEETQYF	2.5	1.4
12.3	CASSLLGEKAYEQYF	2.7	1.4

47-5024

Vβ	CDR3	Jβ	%
7.2	CASSLVLSSPTYYEQYF	2.7	51.8
3.1	CASSQTTSVNTEAFF	1.1	15.3
27	CASSSLNTEAFF	1.1	5.9
11.2	CASSHVINQFF	2.1	4.7
7.9	CASSLPRGRDNEQFF	2.1	4.7
11.2	CASSLGTGHNEQFF	2.1	3.5
5.6	CASSSNRDRNTIYF	1.3	2.4
7.9	CASSLGLGVNNEQFF	2.1	2.4
7.9	CASSSTGPGNSPLHF	1.6	2.4
29.1	CSVSAGEEDTQYF	2.3	1.2
4.2	CASSQVQGTSGGEQYF	2.7	1.2
12.3	CASSSMVAGEYEQFF	2.1	1.2
7.2	CASSLVVIQETQYF	2.5	1.2
7.9	CASSPSKPGDNEQFF	2.1	1.2
7.2	CASSPSKPGDNEQFF	2.1	1.2

Figure S2 (cont.)

B*08-FL9

47-3904			
Vβ	**CDR3**	**Jβ**	%
7.8	CASSSGQAYEQYF	2.7	84.7
9	CASSVASGTDTQYF	2.3	12.9
9	CASSAASGTDTQYF	2.3	2.4

47-3970			
Vβ	**CDR3**	**Jβ**	%
7.8	CASSVGQAYEQYF	2.7	77.1
7.8	CASSRFFPGQGDHAKNIQYF	2.4	11.4
7.8	CASSVGQAYELYF	2.7	2.9
27	CASSPGQPGPVRETQYF	2.5	2.9
7.8	CASSSVNEQYF	2.7	1.4
24.1	CATSDPTLAAYEQYF	2.7	1.4
12.4	CASSFPTSPPITQYF	2.3	1.4
28	CASSLWGEGLTPEAFF	1.1	1.4

47-5018			
Vβ	**CDR3**	**Jβ**	%
7.8	CASSVGQAYEQYF	2.7	100

47-5036			
Vβ	**CDR3**	**Jβ**	%
7.8	CASSLGQAYEQYF	2.7	100

47-5041			
Vβ	**CDR3**	**Jβ**	%
7.8	CASSIGQAYEQYF	2.7	68.6
7.8	CASSLGQAYEQYF	2.7	31.4

B*35-IY9

47-5008			
Vβ	**CDR3**	**Jβ**	%
15	CATSDRTGRHEQYF	2.7	60.0
28	CASSPSHGPNEKLFF	1.4	24.0
9	CASSYDWHEQFF	2.1	8.0
20.1	CTARTGSGLAGGNEQFF	2.1	4.0
6.5	CASSRTGGVDAFF	1.1	2.0
14	CASSQDPGAAGYGTF	1.2	2.0

47-5025			
Vβ	**CDR3**	**Jβ**	%
6.6	CASKGQGARDGYTF	1.2	50.7
29.1	CSVEEGVLYEQYF	2.7	46.5
18	CASSRENSNQPQHF	1.5	1.4
7.2	CASILERGSSYNEQFF	2.1	1.4

47-5033			
Vβ	**CDR3**	**Jβ**	%
7.9	CASSLELAMGETQYF	2.5	47.0
11.2	CASSLDPAGQGYEQYF	2.7	34.8
20.1	CSATGGGEQYF	2.7	7.6
11.2	CASSIPGLAGQETQYF	2.5	4.5
27	CASRSDKYEQYF	2.7	3.0
7.2	CASSLESGHYEQYF	2.7	1.5
28	CASRQLPRDSLYEQYF	2.7	1.5

47-5054			
Vβ	**CDR3**	**Jβ**	%
28	CASSSSSAPNEKLFF	1.4	90.5
2	CASSTQGSPDEQYF	2.7	3.2
14	CASSQDWGTSGNNEQFF	2.1	1.6
28	CASTTSTAPNEKLFF	1.4	1.6
28	CASSSETAPNEKLFF	1.4	1.6
7.2	CASIEERGSYNEQFF	2.1	1.6

47-5060			
Vβ	**CDR3**	**Jβ**	%
7.9	CASSRELSMGETQYF	2.5	98.6
20.1	CSVGQGVVYGYTF	1.2	1.4

Figure S2 (cont.)

B*35-YM9

47-5010				47-5025			
Vβ	**CDR3**	**Jβ**	%	**Vβ**	**CDR3**	**Jβ**	%
28	CASSLRTTNEKLFF	1.4	90.5	27	CASSFSPTAQGLEQYF	2.7	92.0
19	CASSIQDSLYNEQFF	2.1	6.3	4.1	CASSQGALTGQPQHF	1.5	4.0
29.1	CSVEEVEAFF	1.1	3.2	11.2	CASSLASSGYTGELFF	2.2	4.0

47-5033				47-5040			
Vβ	**CDR3**	**Jβ**	%	**Vβ**	**CDR3**	**Jβ**	%
20.1	CSAPEVRGVWGYTF	1.2	89.1	2	CASSLRTLDGYTF	1.2	90.0
28	CASSELTGGPYEQYF	2.7	10.9	9	CASSVELGSGRETQYF	2.5	6.0
				20.1	CSANTGGTYEQYF	2.7	2.0
				6.4	CASSDSSLTEAFF	1.1	2.0

B*57-VW9

47-5018				47-5036			
Vβ	**CDR3**	**Jβ**	%	**Vβ**	**CDR3**	**Jβ**	%
20.1	CSARQEGGSYNEQFF	2.1	44.4	7.8	CASSMLGGGRDEQFF	2.1	57.1
20.1	CSARGEGGSYNEQFF	2.1	33.3	7.9	CASSPLGHRDTEAFF	1.1	15.9
7.9	CASTGWVTDTQYF	2.3	15.3	12.4	CASSPTGTSHTGELFF	2.2	14.3
20.1	CSAREIGGRHQPQHF	1.5	6.9	2	CASRGMTPPTQYF	2.3	6.3
				10.2	CASSDTGGNQPQHF	1.5	6.3

47-5043				47-5055			
Vβ	**CDR3**	**Jβ**	%	**Vβ**	**CDR3**	**Jβ**	%
18	CASSPDGELFF	2.2	87.7	25.1	CASSEGEGQGHYGYTF	1.2	54.0
7.9	CASSSARGHHTQYF	2.3	4.6	3.1	CASSQGPGQEHYGYTF	1.2	12.0
4.2	CASSQDQASEGSETQYF	2.5	3.1	12.4	CATLISQGSRAGELFF	2.2	10.0
11.2	CASSFTWTSGETTDTQYF	2.3	1.5	7.2	CASSQGSGATNEQFF	2.1	8.0
9	CASSARGGADEQFF	2.1	1.5	7.2	CASSQGSGATNGQFF	2.1	4.0
28	CASSFLAGGPPADTQYF	2.3	1.5	7.9	CASSSARTGANYGYTF	1.2	4.0
				20.1	CSARDGARGNTIYF	1.3	2.0
				12.3	CASSLVGGQETQYF	2.5	2.0
				6.6	CASSRRSQGLNTEAFF	1.1	2.0
				7.6	CASSLGRGRGNEQFF	2.1	2.0

Figure S2 - TCRβ repertoires for virus-specific CD8$^+$ T-cell populations restricted by HLA-B molecules. TRBV and TRBJ usage, CDR3β amino acid sequence and relative frequency are shown for HLA-B-restricted clonotypes specific for epitopes derived from EBV and CMV.

Complex TCR repertoire dynamics underlie the CD8$^+$ T-cell response to HIV-1

Ana I. Costa[1,*], Dan Koning[1,*], Kristin Ladell[2], James E. McLaren[2], Bart P.X. Grady[3], Ingrid M. M. Schellens[1], Petra van Ham[4], Monique Nijhuis[4], Jose A. M. Borghans[1], Can Kesmir[5], David A. Price[2] and Debbie van Baarle[1,3]

[1]*Laboratory of Translational Immunology, Department of Immunology, University Medical Center Utrecht, Utrecht, The Netherlands.*

[2]*Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK.*

[3]*Department of Internal Medicine and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands.*

[4]*Department of Virology, University Medical Center Utrecht, Utrecht, The Netherlands.*

[5]*Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands.*

* These authors contributed equally.

Submitted for publication.

Abstract

Although $CD8^+$ T cells are important for the control of HIV-1 *in vivo*, the precise correlates of immune efficacy remain unclear. In this study, we conducted a comprehensive analysis of the T-cell receptor (TCR) repertoire across multiple epitope specificities together with viral sequence information in a group of antiretroviral treatment-naïve individuals chronically infected with HIV-1. A negative correlation was detected between changes in TCR repertoire diversity and in $CD8^+$ T-cell response magnitude, reflecting clonotype-specific expansions and contractions related to alterations in cognate viral epitope sequences. These patterns were independent of the patient, evidenced by discordant TCR evolution against different epitopes in single individuals. Moreover, long-term asymptomatic HIV-1 infection was characterized by parallel evolution of the TCR repertoire and viral replication. Collectively, these data suggest a continuous bidirectional process of adaptation between HIV-1 and virus-specific $CD8^+$ T-cell clonotypes orchestrated at the TCR/antigen interface.

Introduction

$CD8^+$ T cells are key determinants of immune efficacy in HIV-1 infection (reviewed in (1)). Although simple quantitative correlates of protection are generally lacking, previous studies have identified specific qualities that typically associate with effective HIV-specific $CD8^+$ T-cell responses, including targeting specificity and breadth, antigen sensitivity, recall proliferation and polyfunctionality (reviewed in (2-4)). Nonetheless, these and other potential determinants of efficacy cannot fully explain the different disease outcomes associated with infection.

The inherent quality of a $CD8^+$ T-cell response depends on the arsenal of T-cell receptor (TCR) clonotypes deployed to engage the targeted peptide-human leukocyte antigen class I (pHLA-I) complex. Current paradigms hold that diverse and/or cross-reactive TCR repertoires are beneficial in the face of rapidly evolving RNA viruses because they enable early recognition of emerging epitope variants (5-7). Indeed, restricted TCR diversity predisposes to immune escape in HCV infection (8). Moreover, diverse but highly biased repertoires can facilitate escape due to a lack of variant recognition (9). On the other hand, $CD8^+$ T-cell repertoires that incorporate highly cross-reactive clonotypes are associated with delayed disease progression in SIV and HIV-1 infection (10-12). Nonetheless, a broad TCR repertoire *per se* is not necessarily protective, implicating additional attributes as clonotypic determinants of $CD8^+$ T-cell efficacy (13). In this light, it has been shown previously that superior viral control can associate with enhanced antigen-specific clonal turnover, reflecting continual replenishment of the response with effective T-cell clonotypes (14-16). However, repertoire evolution is a variable phenomenon, even within $CD8^+$ T-cell responses directed against the same viral epitope (16). Moreover, clonotype persistence has also been linked with long-term asymptomatic HIV-1 infection (17). These contrasting observations underpin the fact that the HIV-specific $CD8^+$ T-cell response is highly heterogeneous.

Most antigen-specific repertoire studies to date in the HIV-1 field have focused on a single epitope with limited information on the circulating viral quasispecies. In contrast, we conducted a comprehensive analysis of the TCR repertoire and epitope sequence variation of up to four specificities in a group of antiretroviral treatment-naïve individuals with chronic HIV-1 infection. All patients carried the highly prevalent *HLA-I* alleles *A*02* and *B*08*, enabling simultaneous analysis of up to four targeted epitope-specific $CD8^+$ T-cell responses over time from seroconversion. Our results suggest a delicate balance between HIV-1 variation and the cognate TCR repertoire, whereby only a few clonotypes reacted to changes in the viral milieu. These so-called 'clonotypic

shifts' markedly impacted $CD8^+$ T-cell response magnitude in an antigen-driven manner. Moreover, long-term asymptomatic HIV-1 infection was characterized by co-evolution of the TCR repertoire with viral replication.

Materials and methods

Study population

Eight initial participants with known seroconversion dates were selected from the Amsterdam Cohort Studies on HIV-1 infection and AIDS based on the presence of both *HLA-A*02* and *HLA-B*08*; three individuals also carried the protective *HLA-B*27* allele. All subjects were antiretroviral therapy-naïve prior to, and during, the time of sample collection. Peripheral blood mononuclear cells (PBMC) and serum samples were drawn from at least two time points per patient: (i) early (t1), median 218 days post-seroconversion (range 169-568 days); and, (ii) late (t2), median 1133 days post-seroconversion (range 986-1226 days) (Table I). Extra PBMC samples were collected from three patients: pt #5, pt #6, and an additional, seroprevalent subject exclusively selected for these extra analyses, pt #9 (Supplementary Table S1). All individuals were in the asymptomatic, chronic phase of infection.

Flow-assisted cell sorting of antigen-specific $CD8^+$ T cells

As no pre-screening information was available on the presence/absence of measurable epitope-specific $CD8^+$ T-cell responses, we selected well-defined, dominant epitopes for the three *HLA-I* alleles of interest: B*08-FLKEKGGL (B8-FL8), B*08-EIYKRWII (B8-EI8), A*02-SLYNTVATL (A2-SL9) and B*27-KRWIILGLNK (B27-KK10). Antigen-specific $CD8^+$ T-cells were labelled with pre-titrated concentrations of the respective fluorochrome-conjugated pHLA-I tetrameric complexes: (i) B*0801-FL8 and B*0801-EI8 (monomers produced in-house as described previously with minor modifications (18), conjugated with QD705 and QD605 (Life Technologies), respectively; (ii) A*02:01-SL9-APC (Sanquin); and, as applicable, (iii) B*27:05-KK10-PE (Sanquin). Non-viable cells were eliminated from the analysis using Live/Dead Aqua (LifeTechnologies). Cells were then washed and surface-stained with the following monoclonal antibodies (mAbs): anti-CD3-APC-H7, anti-CD4-PE-Cy5.5, anti-CD8-PE-Cy7, anti-CD14-Alexa Fluor 700 and anti-CD19-AmCyan (Caltag/Invitrogen). After exclusion of non-viable/$CD14^+$/$CD19^+$ cells, up to four $CD3^+CD8^+$tetramer$^+$ populations were sorted in parallel at >98% purity directly into RNAlater (Life Technologies) using a customized FACSAria II flow cytometer (BD Biosciences) and stored at -80°C for subsequent TCRβ clonotype analysis.

Table I. Subject characteristics.

Patient	ACS #	HLA-I	t=1 (days post-SC)	t=2 (days post-SC)	$CD4^+$ T-cell counts (cells/µl)		$CD8^+$ T-cell counts (cells/µl)		Viral load (copies/ml)	
					t=1	t=2	t=1	t=2	t=1	t=2
1	**19957**	A*01, A*02, B*0801, B*15	202	1226	470	330	650	790	12000	~58000* (38000-98000)
2	**19885**	A*01, A*02, B*0801,B*27	234	1188	550	410	750	730	~49000* (86000-5800)	1200
3	**19861**	A*01, A*02, B*0801, B*51 or B*52	184	1092	620	380	500	960	26000	12000
4	**19453**	A*01, A*02, B*0801, B*38	169	986	1020	920	650	640	15000	<1000*
5	**19342**	A*01, A*02, B*0801, B*40	536	1166	490	800	670	860	~17000* (22000-7600)	~7500* (9600-<1000)
6	**18840**	A*02, A*02, B*0801,B*27	568	1169	480	690	600	910	~1400* (<1000-1640)	5400
7	**18839**	A*0207, B*0801,B*27	266	1100	360	420	440	850	~33000* (48000-19000)	15000
8	**18785**	A*01, A*02, B*0801, B*07	189	1080	400	380	510	690	40000	210000
	TOTAL	**Median (range)**	**218 (169-568)**	**1133 (986-1226)**	**485 (360-1020)**	**415 (330-920)**	**625 (440-750)**	**820 (640-960)**		

* Trend viral load, estimated from the closest adjacent time points.

TCRβ clonotype analysis

Clonotype analysis was performed as described previously with minor modifications (19). Briefly, mRNA from sorted $CD8^+$ T-cell populations was extracted using the μMACS mRNA Isolation Kit (Miltenyi Biotec). An anchored template-switch RT-PCR was then used to amplify all expressed TCRβ chains linearly. Amplified products were ligated into the pGEMT-Easy vector (Promega) and transformed into chemically competent *E. coli* bacteria. Subcloned products were amplified using M13 primers and sequenced via capillary electrophoresis with the Big Dye Terminator v3.1 Cycle Kit (Life Technologies). Analysis of each TCRβ sequence and assignment of gene usage was performed using the international reference, web-based software from ImMunoGeneTics (20). At least 50 TCRβ sequences were successfully analyzed for each sample.

Sequence analysis of HIV-1 epitopes

For Gag, viral RNA was isolated from serum using a Viral RNA Mini Kit (Qiagen) or silica particles as described previously (21). A combined cDNA synthesis and first-round PCR was then performed in 30 μl reactions using a Titan One Tube RT-PCR Kit (Roche). The following parameters were used: (i) 50°C for 30 min to synthesize cDNA; (ii) 95°C for 2 min to melt; (iii) 40 cycles of 95°C for 15 sec, 57°C for 30 sec and 68°C for 2.5 min (increased by 5 sec per cycle for the last 30 cycles) to amplify; and, (iv) 72°C for 10 min to complete extension. The second, nested, PCR was performed using 5 μl of the first-round product in 30 μl reactions with the Expand High Fidelity PCR System (Roche). The following parameters were used: (i) 95°C for 2 min to melt; (ii) 30 cycles of 95°C for 15 sec, 58°C for 30 sec and 68°C for 2.5 min to amplify; and (iii) 72°C for 10 min to complete extension. Primers KVL064 (forward, 5'-GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGA-3') and NCrev-2 (reverse, 5'-CCTTCCTTTCCACATTTCCAACAG-3') were used for the combined cDNA synthesis/first-round PCR, and primers KVL066 (forward, 5'-TCTCTAGCAGTGGCGCCCGAACAG-3') and NCrev-3 (reverse, 5'-CTTTTTCCTAGGGGCCCTGCAATTT-3') were used for the second, nested, PCR. For Nef, viral RNA was isolated from serum using a Viral RNA Mini Kit (Qiagen). cDNA was synthesized with SuperScript III Reverse Transcriptase (Invitrogen) using a Nef-specific primer (Nef rv1: 5'-GCTTATATGCAGGATCTGAGG-3') and purified on silica-based columns (Macherey-Nagel). Template-specific amplification was performed as described previously (22). Amplified Gag and Nef products were gel-purified (Macherey-Nagel), A-tailed and ligated using the pGEM-T Easy Vector System (Promega). Ligated products were then transformed into chemically competent *E.coli* bacteria and sequenced as described above (4-48 clones per sample).

TCRβ diversity analysis

A T-cell clonotype was defined as a TCRβ chain encoded by a unique nucleotide sequence. Sample clonality was estimated by counting the relative number of distinct clonotypes and by using Simpson's diversity index (D_S) (23). This index is defined as:

$$D_S = 1 - \sum_{i=1}^{c} \frac{n_i(n_i - 1)}{n(n-1)},$$

where n_i is the clonal size of the ith clonotype (i.e., the amount of copies of a specific clonotype), c is the number of different clonotypes and n is the total number of analyzed TCRβ sequences. This index uses the relative frequency of each clone to calculate a diversity index ranging between 0 and 1, indicating minimal and maximal diversity, respectively. To account for differences in sample size (i.e. the number of successfully analyzed TCRβ sequences), all samples were normalized by random sampling (without replacement) to an equal number of sequences (n=50) prior to the calculation of TCR diversity (i.e., the relative number of unique clonotypes and Simpson's diversity index). This process was repeated 1000 times, after which median values of TCR diversity were determined and used for subsequent analyses.

Statistical analysis

Sample normalization and statistical analyses were performed using SPSS 20.0.0 (SPSS Inc.). A p-value ≤0.05 was considered statistically significant. Graphics were generated using GraphPrism 5.04 (GraphPad Software Inc.). Note that in some analyses (Fig. 2A, Fig. 3) data on multiple T-cell populations per individual and different epitope specificities were pooled, rendering these data not fully independent of each other.

Results

Isolation and analysis of antigen-specific CD8⁺ T cells

Eight treatment-naïve individuals with chronic HIV-1 infection were selected for co-expression of the *HLA-A*02* and *HLA-A*B08* alleles. Each patient was studied at two time points, approximately one (0.5-1.5) and three (2.7-3.5) years post-seroconversion (Table I). Initially, we used pHLA-I tetramers to characterize $CD8^+$ T cell responses directed against the frequently targeted epitopes A*02-SLYNTVATL (A2-SL9, p17-Gag), B*08-EIYKRWII (B8-EI8, p24-Gag), B*08-FLKEKGGL (B8-FL8, Nef) and B*27-KRWIILGLNK (B27-KK10, p24-Gag). Response magnitude varied as a function of

specificity, with B27-KK10 and A2-SL9 eliciting the biggest and smallest CD8$^+$ T-cell responses, respectively (Fig. 1A).

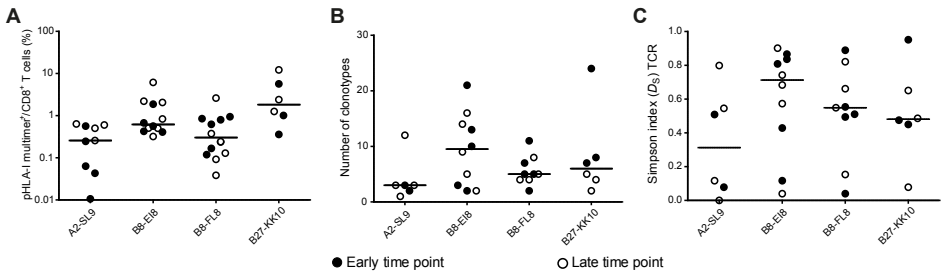


Figure 1. Analysis of CD8$^+$ T-cell populations directed against A2-SL9, B8-EI8, B8-FL8 and B27-KK10.
Antigen-specific CD8$^+$ T-cells were labelled with pHLA-I tetramers and quantified by flow cytometry. Response magnitude **(A)** is shown as the frequency of tetramer$^+$ events in the total CD8$^+$ T-cell population. TCRβ diversity was quantified using the relative number of clonotypes **(B)** and Simpson's diversity index **(C)**. Data are shown for each patient at both the early (closed circles) and late (open circles) time points.

To compare the clonotypic composition of distinct antigen-specific CD8$^+$ T-cell populations, we used a template-switch anchored RT-PCR to amplify all expressed TCRβ chains from pHLA-I tetramer$^+$ cells sorted by flow cytometry to high levels of purity. Two measures of diversity were calculated: (i) the relative number of clonotypes (i.e. each unique TCRβ nucleotide sequence after normalization): and, (ii) Simpson's diversity index (D_S), which accounts for clonotype frequency (23). CD8$^+$ T-cell populations directed against B8-EI8 and A2-SL9 selected repertoires with the highest and lowest degrees of diversity, respectively (Fig. 1B and 1C; Supplementary Figure S1). Of note, highly focused and polyclonal responses were observed within each specificity; these patterns were not associated with either the time of sampling (closed *vs* open symbols) or the magnitude of the CD8$^+$ T-cell population (data not shown).

CD8$^+$ T-cell repertoire diversity does not correlate with viral epitope variation

Previous studies from our group and others have suggested that the T-cell repertoire is shaped primarily by the antigenic epitope (24,25). To assess the relationship between TCRβ diversity and viral epitope mutation, we conducted an extensive analysis of autologous HIV-1 sequences in targeted regions of the viral genome. Modest variations were detected across viral populations. A single epitope sequence was usually dominant, deviating from wildtype (WT) most prominently in the B8-FL8 and B27-KK10

epitope regions (Table II). Of note, the majority of these variants were predicted to bind their respective HLA-I molecules (Supplementary Table S2) (26,27).

Table II-A. HIV-1 sequence analysis of epitopes restricted by HLA-A*02 and HLA-B*08.

Patient	ACS #	HLA type	Time point	A2-SL9		B8-EI8		B8- FL8	
				SLYNTVATL	%Seqs	EIYKRWII	%Seqs	FLKEKGGL	%Seqs
1	19957	A*01, A*02, B*0801, B*15	T=1	---------	100%	--------	100%	----T---	79.2%
								--------	19.2%
								----R---	1.6%
			T=2	-----I---	100%	--------	100%	----M---	93.4%
								--------	6.6%
3	19861	A*01, A*02, B*0801, B*51 or B*52	T=1	---------	100%	--------	100%	--------	100%
			T=2	ND		ND		ND	
4	19453	A*01, A*02, B*0801, B*38	T=1	--F------	51.7%	ND		ND	
				--------	48.3%				
			T=2	---------	100%	--------	100%	----M---	86.7%
								--------	13.3%
5	19342	A*01, A*02, B*0801, B*40	T=1	ND		ND		ND	
			T=2	--F------	100%	D-------	100%	----E---	100%
8	18785	A*01, A*02, B*0801, B*07	T=1	---------	100%	--------	100%	--N-----	86.1%
								----E---	14.0%
			T=2	---------	100%	ND		--N-----	93.9%
								L-N-----	3.0%
								--------	3.0%

ND = not determined

Table II-B. HIV-1 sequence analysis of epitopes restricted by HLA-A*02, HLA-B*08, and HLA-B*27.

Patient	ACS #	HLA type	Time point	A2-SL9		B8-EI8		B8- FL8		B27-KK10	
				SLYNTVATL	%Seqs	EIYKRWII	%Seqs	FLKEKGGL	%Seqs	KRWIILGLNK	%Seqs
2	19885	A*01, A*02, B*0801,B*27	T=1	--F-A--V-	100%	-------- ---E----	96.8% 3.2%	--R-----	100%	---------- E--------- -----P----	93.6% 3.2% 3.2%
			T=2	--F----V- --F--A-V-	93.8% 6.3%	-------- -------V	85.7% 14.3%	ND		---------- ----VM----	85.7% 14.3%
6	18840	A*02, A*02, B*0801,B*27	T=1	--------- P---------	94.7% 5.3%	--------	100%	-------- --R-----	48.9% 51.1%	-----M---- ---------- -----I----	68% 28% 4%
			T=2	ND		ND		ND		ND	
7	18839	A*0207, A*0207, B*0801,B*27	T=1	---------	100%	--------	100%	-------- ----R---	64% 36%	-----M----	100%
			T=2	--------- -----I---	90.3% 9.7%	-------- ---R----	94.7% 5.3%	----R--- ----E---	97.6% 2.4%	---------- -----M---- R---------	52.6% 42.1% 5.3%

ND = not determined

No significant correlations were detected between the frequency of the WT epitope and the diversity of the corresponding TCRβ repertoire, either in terms of Simpson's diversity index (Fig. 2A) or the relative number of clonotypes (data not shown). Moreover, there were no associations between TCRβ diversity and either the presence of epitope variants or the number of different epitope sequences (Fig. 2B). These data indicate that the composition of the viral epitope population does not predict cognate TCRβ repertoire diversity at any given time point.

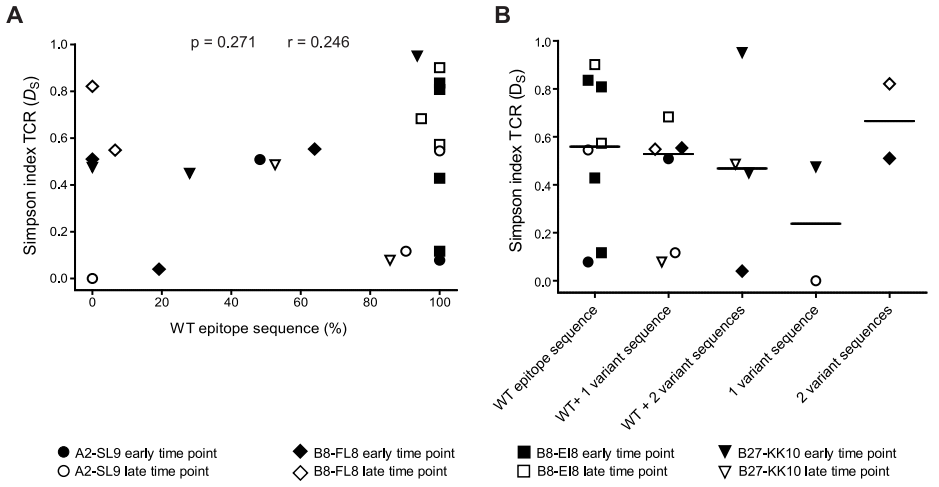


Figure 2. CD8⁺ T-cell repertoire diversity is not related to viral epitope variation.
Antigen-specific TCRβ repertoire diversity and viral epitope sequences were determined in parallel for a subset of samples (n=22). The percentage of wildtype (WT) epitope sequences **(A)** and the number of variant epitopes present at the time of analysis **(B)** were used as measures of epitope composition in the autologous viral quasispecies. All plots include data derived from the early (closed symbols) and late (open symbols) time points. Symbol shape denotes epitope specificity: A2-SL9 (circle), B8-EI8 (square), B8-FL8 (diamond) and B27-KK10 (inverted triangle). Correlation testing in **(A)** was performed using the Spearman Rank test. Note that Simpson's TCR diversity index was determined after data normalization for appropriate diversity comparisons (see *Materials and methods* for details).

Parallel evolution of the TCR repertoire, viral quasispecies and CD8⁺ T-cell responses

It is well established that CD8⁺ T-cell response magnitude, TCR diversity and viral epitope sequences can evolve significantly during the course of HIV-1 infection (16). Accordingly, we examined a subset of HIV-specific CD8⁺ T-cell responses (n=10) over time. Repertoire diversity varied between the early and late time points without a common tendency to increase or decrease (Fig. 3A and 3B). Next, we studied how changes in TCRβ diversity (measured as the ratio of the number of normalized clonotypes or the Simpson's index at timepoint 2 divided by that at timepoint 1) related

with longitudinal changes in CD8$^+$ T-cell response magnitude (measured as the ratio of the response magnitude at timepoint 2 divided by that at timepoint 1) (Fig. 3C and 3D). A significant negative correlation was observed between these two parameters, indicating that the repertoire became less diverse with CD8$^+$ T-cell expansion and more diverse with CD8$^+$ T cell contraction. This observation suggests that shifts in response magnitude over time were associated with inflation and deflation of particular clonotypes.

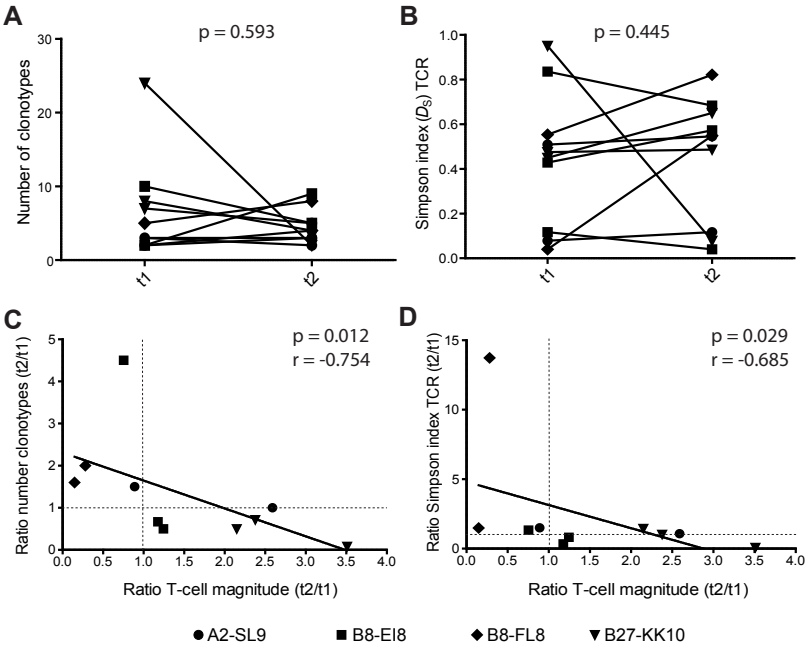
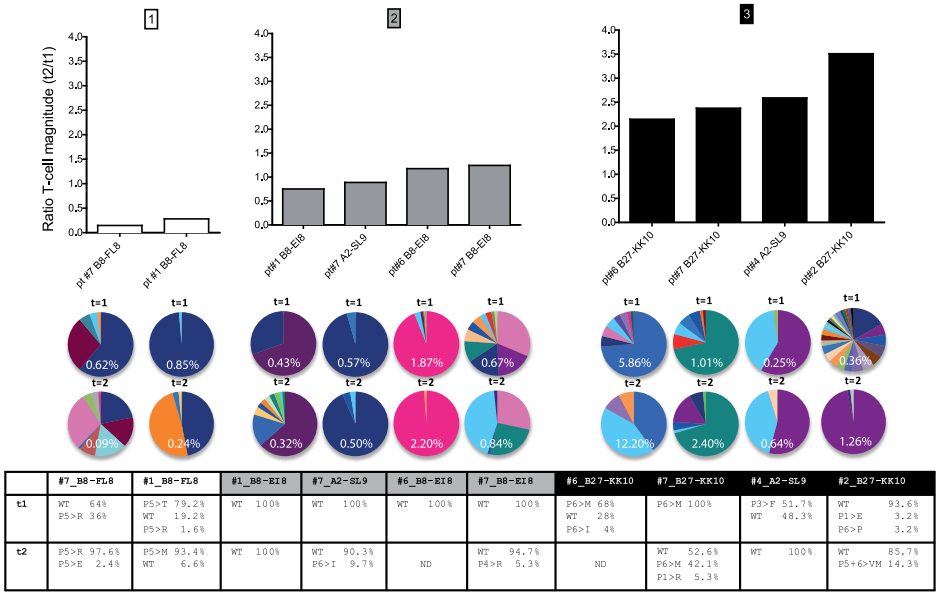


Figure 3. Longitudinal variations in CD8$^+$ T-cell response magnitude correlate negatively with TCRβ diversity.
TCRβ diversity was quantified at the early (t1) and late (t2) time points for a subset of antigen-specific CD8$^+$ T-cell responses (n=10) using the relative number of clonotypes **(A)** and Simpson's diversity index **(B)**. Statistical analyses were performed using the Wilcoxon Signed-Rank test. **(C, D)**. Changes in CD8$^+$ T-cell response magnitude (ratio t2 to t1) were related to differences in TCRβ diversity (ratio t2 to t1) determined by the relative number of clonotypes (**C**) and Simpson's diversity index (**D**). Each dot represents a CD8$^+$ T-cell response analyzed at two time points. Symbol shape denotes epitope specificity: A2-SL9 (circle), B8-EI8 (square), B8-FL8 (diamond) and B27-KK10 (inverted triangle). Correlation testing was performed using the Spearman Rank test.

On this basis, we examined the relationship between clonotypic stability and viral epitope diversity. All longitudinal CD8$^+$ T-cell responses were first stratified according to changes in magnitude over time (ratio t2 to t1): (i) large decreases in magnitude (ratio ≤0.5); (ii) conservation of magnitude (ratio >0.5 and <2.0); and, (iii) large increases in magnitude (ratio ≥2.0) (Fig. 4, upper panel). The corresponding TCRβ repertoires (Fig. 4,

middle panel) and circulating viral epitopes (Fig. 4, lower panel) were then compared across categories. Interestingly, the observed shifts in CD8$^+$ T-cell response magnitude were often linked with changes in the TCRβ repertoire and viral epitope over time. For example, the two decreasing responses (pt #1 and pt #7, B8-FL8) were accompanied by deflation of one (pt #1) or two (pt #7) dominant clonotypes (Fig 4, panel 1). Similarly, two increasing responses (pt #2 and pt #6, B27-KK10) were paralleled to inflation of previously subdominant clonotypes (Fig. 4, panel 3). In both scenarios, viral epitope sequences changed over time. For CD8$^+$ T-cell responses that remained relatively stable over time, however, fewer mutations were detected in the targeted viral epitopes and TCRβ repertoire composition remained largely unchanged (Fig. 4, panel 2). Together, these results suggest that changes in CD8$^+$ T-cell response magnitude are associated with viral epitope mutations and preferential inflation or deflation of specific clonotypes within the available repertoire.



	#7_B8-FL8	#1_B8-FL8	#1_B8-EI8	#7_A2-SL9	#6_B8-EI8	#7_B8-EI8	#6_B27-KK10	#7_B27-KK10	#4_A2-SL9	#2_B27-KK10
t1	WT 64% P5>R 36%	P5>T 79.2% WT 19.2% P5>R 1.6%	WT 100%	WT 100%	WT 100%	WT 100%	P6>M 68% WT 28% P6>I 4%	P6>M 100%	P3>F 51.7% WT 48.3%	WT 93.6% P1>E 3.2% P6>P 3.2%
t2	P5>R 97.6% P5>E 2.4%	P5>M 93.4% WT 6.6%	WT 100%	WT 90.3% P6>I 9.7%	ND	WT 94.7% P4>R 5.3%	ND	WT 52.6% P6>M 42.1% P1>R 5.3%	WT 100%	WT 85.7% P5+6>VM 14.3%

Figure 4. The relationship between CD8$^+$ T-cell response magnitude, TCRβ diversity and viral epitope variation.
Antigen-specific CD8$^+$ T-cell responses were stratified according to changes in magnitude over time (ratio t2 to t1). Panel 1 (left, white bars) depicts CD8$^+$ T-cell responses that subsided over time, panel 2 (middle, grey bars) depicts CD8$^+$ T-cell responses that remained stable over time and panel 3 (right, black bars) depicts CD8$^+$ T-cell responses that increased over time. The clonotypic composition of each CD8$^+$ T-cell population is illustrated in the pie charts (middle panel) and the respective viral epitope sequences are shown (lower panel). Response magnitudes are indicated in the pie charts as the frequency of tetramer$^+$ events in the total CD8$^+$ T-cell population. Pie charts colours match clonotypes for each epitope pair, but do not correspond between pairs.

Discordant evolution of CD8⁺ T-cell responses within patients

To extend these findings, we conducted similar analyses stratified on a patient basis (Fig. 5). Discordant evolutionary patterns were observed with respect to CD8⁺ T-cell response magnitude and TCRβ repertoire composition across specificities. For example, the B8-EI8 response in patient #1 remained relatively stable in terms of magnitude and clonotypic representation, whereas the B8-FL8 response showed substantial changes over time. Similar differences across specificities were observed in patients #6 and #7.

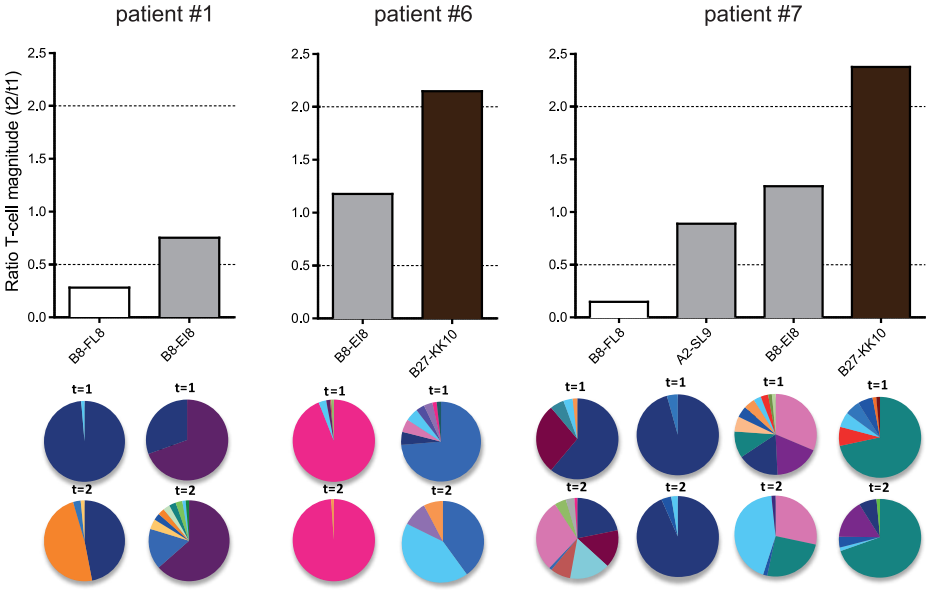


Figure 5. CD8⁺ T-cell response evolution is a process not dependent on the patient.
Antigen-specific CD8⁺ T-cell responses were stratified according to patient origin. Bar colours in the upper panel are adapted from Figure 4 and represent decreasing (white), stable (grey) and increasing (black) response magnitudes over time. The clonotypic composition of each CD8⁺ T-cell population is illustrated in the pie charts with response magnitudes indicated as the frequency of tetramer⁺ events in the total CD8⁺ T-cell population (lower panel). Pie chart colours match clonotypes for each epitope pair, but do not correspond between pairs.

Thus, CD8⁺ T-cell response magnitude and repertoire composition show distinct trends in evolution within-patient, suggesting an antigen-driven evolutionary process.

TCR repertoire evolution and viral load dynamics

A minority of individuals infected with HIV-1 maintain control of viral load at low or undetectable levels. To determine the long-term impact of such low-level viral replication and antigen presentation, we analyzed TCRβ repertoire composition in CD8⁺ T-cell populations specific for B8-FL8 and B8-EI8 using additional samples from patients #5, #6 and #9 (Fig. 6), all of whom showed signs of delayed disease progression

(asymptomatic with stable viral loads and CD4$^+$ T-cell counts >300 cells/μl at least 7 years after seroconversion).

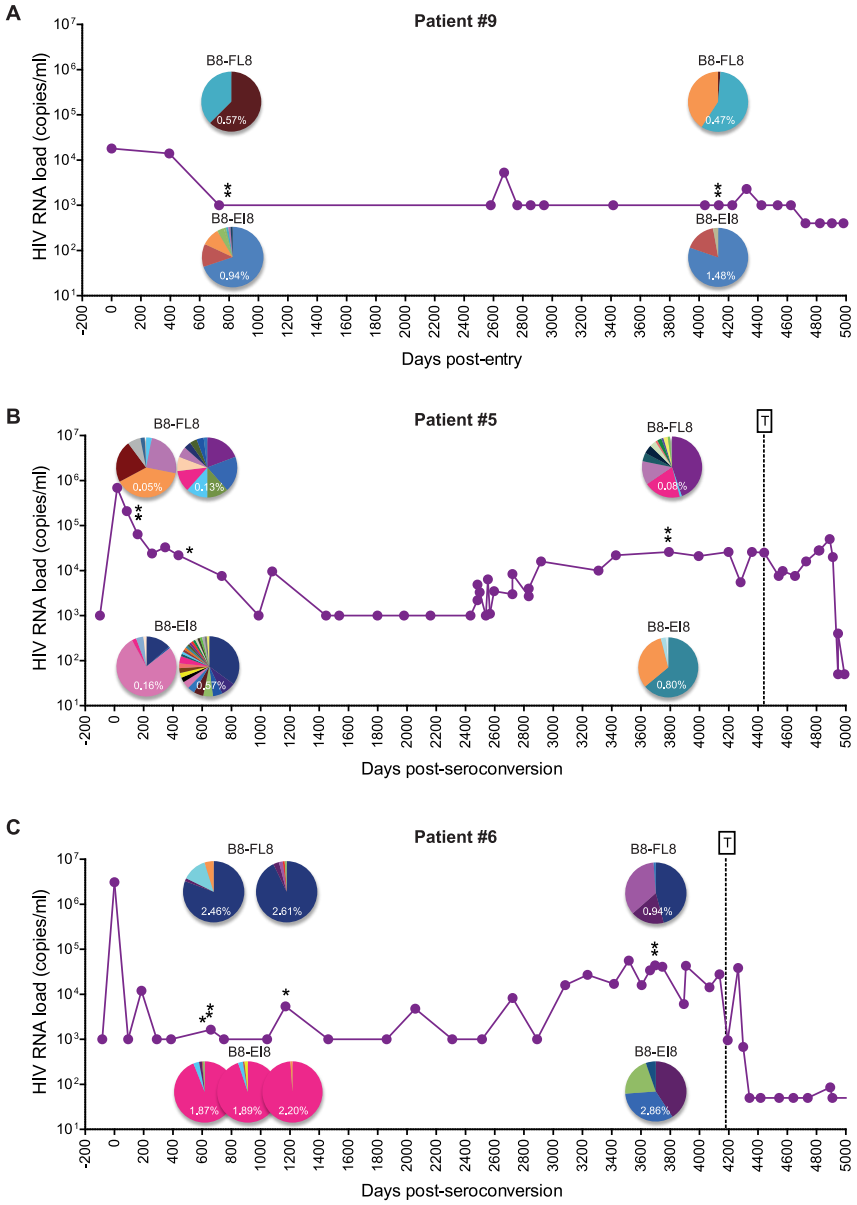


Figure 6. TCR repertoire evolution and viral load dynamics during long-term asymptomatic infection. Viral load trajectories and CD8$^+$ T-cell repertoires specific for B8-FL8 and B8-EI8 are shown for patient #9 **(A)**, patient #5 **(B)** and patient #6 **(C)**. Pie chart colours match clonotypes for each epitope pair in each patient, but do not correspond between pairs or patients. Single asterisks correspond to the time points described in Table I. Double asterisks correspond to additional time points (Supplementary Table S1). T, antiretroviral therapy commencement.

Patient #9 maintained undetectable viral loads 14 years post-entry into the cohort (Fig. 6A). Patients #5 and #6 similarly controlled viral loads to low or undetectable levels after acute infection, although progressive increases approximately seven years after seroconversion warranted subsequent antiretroviral therapy (Fig. 6B and 6C). Different patterns of TCRβ repertoire evolution were observed in these patients. Thus, clonotypic representation remained stable in some epitope-specific $CD8^+$ T-cell populations (B8-EI8 in pt #9; B8-FL8 and B8-EI8 in pt #6 during early infection), whereas considerable changes in the constituent TCRβ clonotypes were observed in others (B8-FL8 and B8-EI8 in pt #5; B8-FL8 and B8-EI8 in pt #6 during late infection). Moreover, these clonotypic characteristics often paralleled viral load trajectories. Patient #6 displayed stable viral loads during early infection in conjunction with largely constant TCRβ repertoires specific for B8-FL8 and B8-EI8. As viral loads increased during late infection, however, dramatic changes in clonotypic composition were apparent for both specificities. Similar patterns were observed in patient #5. In this case, epitope-specific TCRβ repertoire instability mirrored viral load fluctuations during both early and late infection. Conversely, the B8-EI8-specific TCRβ repertoire in patient #9 remained stable in the presence of undetectable viral loads, although clear clonotypic shifts were observed in the B8-FL8-specific $CD8^+$ T-cell population. Collectively, these data indicate that the HIV-specific TCR repertoire evolves as a function of viral load. Thus, viremic control is associated with relatively conserved repertoires, whereas higher levels of viral replication drive clonotypic turnover.

Discussion

Although it is widely accepted that HIV-1 evades $CD8^+$ T-cell immunity via epitope mutation, the clonotypic correlates of this phenomenon remain poorly understood. Accordingly, we investigated antigen-specific $CD8^+$ T-cell repertoire dynamics in relation to viral epitope mutation in antiretroviral therapy-naïve seroconverters with asymptomatic HIV-1 infection. In accordance with previous work, we found no correlation between viral epitope composition and clonotypic diversity within the cognate HIV-specific $CD8^+$ T-cell response (28). However, longitudinal analyses revealed a more nuanced picture. A negative correlation across multiple specificities was initially detected between changes in TCR repertoire diversity and in $CD8^+$ T-cell response magnitude. More detailed investigations then showed that this association reflected clonotype-specific expansions and contractions related to alterations in cognate viral epitope sequences. These patterns were discordant within patients,

suggesting an epitope-driven process. Moreover, turnover of clonotypes was related to viral load, as noted previously (16). Collectively, these data suggest a continuous bidirectional process of adaptation between HIV-1 and virus-specific $CD8^+$ T-cell clonotypes that could ultimately govern immune efficacy and the outcome of infection.

Previous studies have highlighted such dynamic interplay between clonotypic adaptation and lentiviral pathogens. This is perhaps best exemplified in the B27-KK10 system, where the early mobilization of public TRBV4-3/TRBJ1-3 clonotypes drives the emergence of TCR escape mutations, which can subsequently be controlled by cross-reactive TRBV6-5/TRBJ1-1 clonotypes in some patients (11,29). Adaptive plasticity in the B27-KK10-specific repertoire may even underlie the protective phenotype conferred by this *HLA-I* allele (30). Similarly cross-reactive clonotypes may also confer preferential outcomes in the context of non-protective *HLA-I* alleles. Indeed, a B8-FL8-specific TCR previously associated with long-term non-progressive disease was detected in this study (12). Nonetheless, it is possible that clonotypic adaptation represents a double-edged sword, in some cases exhausting immune resources without demonstrable benefit. Further detailed studies spanning multiple specificities will be required to clarify these issues in relation to specific epitopes and restriction elements.

Despite the primary roles of antigen quantity and quality as determinants of TCR repertoire dynamics, it is important to note that other factors are implicated by the heterogeneous patterns observed in our study. For example, the A2-SL9-specific TCR repertoire in patient #4 remained stable despite substantial sequence variation in the cognate viral epitope and changes in response magnitude. In this case, it seems likely that both dominant clonotypes were equally responsive to the emerging variant, suggesting the operation of other selection pressures during viral evolution. Conversely, the B8-FL8-specific TCR repertoire in patient #9 shifted substantially over time despite a consistently undetectable viral load. Moreover, the corresponding B8-EI8-specific response remained clonotypically stable over the same prolonged time period. Thus, antigen drive alone does not fully explain the evolutionary patterns observed across distinct epitope-specific TCR repertoires in this study.

In summary, our data show that the antigen-specific $CD8^+$ T-cell repertoire is intimately linked with viral load and epitope variation during chronic HIV-1 infection. These findings establish a framework for the identification of clonotype-specific correlates of disease outcome across epitope specificities with attendant implications for vaccine design.

Acknowledgements

This work was supported by a grant from the AIDS Fonds Netherlands (number 2011033) and a fellowship from the Fundação para a Ciência e Tecnologia. The authors are grateful to Gerrit Spierenburg and Koos Gaiser for technical assistance.

References

1. Golder PJ, Watkins DI. 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat. Rev. Immunol. 8:619-630.
2. Seder RA, Darrah PA, Roederer M. 2008. T-cell quality in memory and protection: implications for vaccine design. Nat. Rev. Immunol. 8:247-258.
3. Appay V, Douek DC, Price DA. 2008. CD8+ T cell efficacy in vaccination and disease. Nat. Med. 14:623-628.
4. Bangham CR. 2009. CTL quality and the control of human retroviral infections. Eur. J. Immunol. 39:1700-1712.
5. Turner SJ, La Gruta NL, Kedzierska K, Thomas PG, Doherty PC. 2009. Functional implications of T cell receptor diversity. Curr. Opin. Immunol. 21:286-290.
6. Nikolich-Zugich J, Slifka MK, Messaoudi I. 2004. The many important facets of T-cell repertoire diversity. Nat. Rev. Immunol. 4:123-132.
7. Davenport MP, Price DA, McMichael AJ. 2007. The T cell repertoire in infection and vaccination: implications for control of persistent viruses. Curr. Opin. Immunol. 19:294-300.
8. Meyer-Olson D, Shoukry NH, Brady KW, Kim H, Olson DP, Hartman K, Shintani AK, Walker CM, Kalams SA. 2004. Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape. J. Exp. Med. 200:307-319.
9. Price DA, West SM, Betts MR, Ruff LE, Brenchley JM, Ambrozak DR, Edghill-Smith Y, Kuroda MJ, Bogdan D, Kunstman K, Letvin NL, Franchini G, Wolinsky SM, Koup RA, Douek DC. 2004. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. Immunity. 21:793-803.
10. Price DA, Asher TE, Wilson NA, Nason MC, Brenchley JM, Metzler IS, Venturi V, Gostick E, Chattopadhyay PK, Roederer M, Davenport MP, Watkins DI, Douek DC. 2009. Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. J. Exp. Med. 206:923-936.
11. Ladell K, Hashimoto M, Iglesias MC, Wilmann PG, McLaren JE, Gras S, Chikata T, Kuse N, Fastenackels S, Gostick E, Bridgeman JS, Venturi V, Arkoub ZA, Agut H, van Bockel DJ, Almeida JR, Douek DC, Meyer L, Venet A, Takiguchi M, Rossjohn J, Price DA, Appay V. 2013. A Molecular Basis for the Control of Preimmune Escape Variants by HIV-Specific CD8(+) T Cells. Immunity. 38:425-436.
12. Dong T, Stewart-Jones G, Chen N, Easterbrook P, Xu X, Papagno L, Appay V, Weekes M, Conlon C, Spina C, Little S, Screaton G, van der MA, Richman DD, McMichael AJ, Jones EY, Rowland-Jones SL. 2004. HIV-specific cytotoxic T cells from long-term survivors select a unique T cell receptor. J. Exp. Med. 200:1547-1557.
13. Mendoza D, Royce C, Ruff LE, Ambrozak DR, Quigley MF, Dang T, Venturi V, Price DA, Douek DC, Migueles SA, Connors M. 2012. HLA B*5701-positive long-term nonprogressors/elite controllers are not distinguished from progressors by the clonal composition of HIV-specific CD8+ T cells. J. Virol. 86:4014-4018.
14. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, Bornstein E, Asher TE, Samri A, Schnuriger A, Theodorou I, Costagliola D, Rouzioux C, Agut H, Marcelin AG, Douek D, Autran B, Appay V. 2007. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. J. Exp. Med. 204:2473-2485.
15. Price DA, Brenchley JM, Ruff LE, Betts MR, Hill BJ, Roederer M, Koup RA, Migueles SA, Gostick E, Wooldridge L, Sewell AK, Connors M, Douek DC. 2005. Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. J. Exp. Med. 202:1349-1361.

16. Meyer-Olson D, Brady KW, Bartman MT, O'Sullivan KM, Simons BC, Conrad JA, Duncan CB, Lorey S, Siddique A, Draenert R, Addo M, Altfeld M, Rosenberg E, Allen TM, Walker BD, Kalams SA. 2006. Fluctuations of functionally distinct CD8+ T-cell clonotypes demonstrate flexibility of the HIV-specific TCR repertoire. Blood 107:2373-2383

17. van Bockel DJ, Price DA, Munier ML, Venturi V, Asher TE, Ladell K, Greenaway HY, Zaunders J, Douek DC, Cooper DA, Davenport MP, Kelleher AD. 2011. Persistent survival of prevalent clonotypes within an immunodominant HIV gag-specific CD8+ T cell response. J. Immunol. 186:359-371.

18. Altman JD, Moss PA, Goulder PJ, Barouch DH, Heyzer-Williams MG, Bell JI, McMichael AJ, Davis MM. 1996. Phenotypic analysis of antigen-specific T lymphocytes. Science 274:94-96.

19. Quigley MF, Almeida JR, Price DA, Douek DC. 2011. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. Curr. Protoc. Immunol. Chapter 10:Unit10.

20. Giudicelli V, Brochet X, Lefranc MP. 2011. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb. Protoc. 2011:695-715.

21. Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der NJ. 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. 28:495-503.

22. Navis M, Schellens IM, van SP, Borghans JA, Miedema F, Kootstra NA, van BD, Schuitemaker H. 2008. A nonprogressive clinical course in HIV-infected individuals expressing human leukocyte antigen B57/5801 is associated with preserved CD8+ T lymphocyte responsiveness to the HW9 epitope in Nef. J. Infect. Dis. 197:871-879.

23. Venturi V, Kedzierska K, Turner SJ, Doherty PC, Davenport MP. 2007. Methods for comparing the diversity of samples of the T cell receptor repertoire. J. Immunol. Methods 321:182-195.

24. Koning D, Costa AI, Hoof I, Miles JJ, Nanlohy NM, Ladell K, Matthews KK, Venturi V, Schellens IM, Borghans JA, Kesmir C, Price DA, van BD. 2013. CD8+ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex. J. Immunol. 190:931-939.

25. Turner SJ, Kedzierska K, Komodromou H, La Gruta NL, Dunstone MA, Webb AI, Webby R, Walden H, Xie W, McCluskey J, Purcell AW, Rossjohn J, Doherty PC. 2005. Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific CD8+ T cell populations. Nat. Immunol. 6:382-389.

26. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res. 36:W509-W512.

27. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Roder G, Peters B, Sette A, Lund O, Buus S. 2007. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS. One. 2:e796.

28. Douek DC, Betts MR, Brenchley JM, Hill BJ, Ambrozak DR, Ngai KL, Karandikar NJ, Casazza JP, Koup RA. 2002. A novel approach to the analysis of specificity, clonality, and frequency of HIV-specific T cell responses reveals a potential mechanism for control of viral escape. J. Immunol. 168:3099-3104.

29. Iglesias MC, Almeida JR, Fastenackels S, van Bockel DJ, Hashimoto M, Venturi V, Gostick E, Urrutia A, Wooldridge L, Clement M, Gras S, Wilmann PG, Autran B, Moris A, Rossjohn J, Davenport MP, Takiguchi M, Brander C, Douek DC, Kelleher AD, Price DA, Appay V. 2011. Escape from highly effective public CD8+ T-cell clonotypes by HIV. Blood 118:2138-2149.

30. Chen H, Ndhlovu ZM, Liu D, Porter LC, Fang JW, Darko S, Brockman MA, Miura T, Brumme ZL, Schneidewind A, Piechocka-Trocha A, Cesa KT, Sela J, Cung TD, Toth I, Pereyra F, Yu XG, Douek DC, Kaufmann DE, Allen TM, Walker BD. 2012. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. Nat. Immunol. 13:691-700.

Supplementary Material

Table S1. Additional PBMC samples collected from patients #5, # 6 and# 9.

Patient	ACS #	HLA-I	Days post-SC		CD4+T-cell counts (cells/µl)		CD8+ T-cell counts (cells/ µl)		Viral load (RNA copies/ml)	
5	19342	A*01, A*02, B*0801, B*40	*pre-t1*	*post-t2*	*pre-t1*	*post-t2*	*pre-t1*	*post-t2*	*pre-t1*	*post-t2*
			158	3791	810	700	430	1810	64000	26000
6	18840	A*02, A*02, B*0801, B*27	*post-t1*	*post-t2*	*post-t1*	*post-t2*	*post-t1*	*post-t2*	*post-t1*	*post-t2*
			660	*3696*	*540*	*280*	*730*	*1250*	*1640*	*44000*
9	19956	A*01, A*02, B*07, B*08	*t1*	*t2*	*t1*	*t2*	*t1*	*t2*	*t1*	*t2*
			825	4134	630	370	870	1020	<1000* (<1000-<1000)	<1000

*Trend viral load, estimated from the closest adjacent time points.

Table S2. *In silico* binding predictions for wildtype and variant epitope sequences of A2-SL9, B8-EI8, B8-FL8 and B27-KK10 with respect to the most representative *HLA-I* allele.

Peptide	HLA-I	NetMHC 3.4*		NetMHCpan 2.8* (supertype)		
		Affinity (nM)		*Affinity (nM)*	*RANK*	
SLYNTVATL	A*02:01	94	WB	256.45	4.00	WB
SL**F**NTVATL	A*02:01	68	WB	146.74	3.00	WB
SLYNT**I**ATL	A*02:01	60	WB	111.11	3.00	WB
SL**F**NTVA**V**L	A*02:01	73	WB	245.48	4.00	WB
SL**F**N**A**VA**V**L	A*02:01	38	SB	58.94	2.00	WB
SL**F**NT**A**A**V**L	A*02:01	121	WB	447.10	4.00	WB
EIYKRWII	B*08:01	251	WB	398.64	1.00	WB
DIYKRWII	B*08:01	662		902.23	2.00	WB
EIY**E**RWII	B*08:01	756		1923.36	4.00	
EIY**R**RWII	B*08:01	327	WB	235.43	0.80	WB
EIYKRWI**V**	B*08:01	405	WB	517.70	1.50	WB
FLKEKGGL	B*08:01	64	WB	129.60	0.40	SB
FLKE**R**GGL	B*08:01	45	SB	55.82	0.17	SB
FLKE**E**GGL	B*08:01	850		1986.87	4.00	
FLKE**M**GGL	B*08:01	433	WB	630.09	1.50	WB
FLKE**T**GGL	B*08:01	603		602.12	1.50	WB
FL**N**EKGGL	B*08:01	578		1871.43	4.00	
LLN**E**KGGL	B*08:01	1890		3128.12	5.00	
FL**R**EKGGL	B*08:01	72	WB	98.94	0.30	SB
KRWIILGLNK	B*27:05	14	SB	31.85	0.05	SB
ERWIILGLNK	B*27:05	67	WB	458.93	1.00	WB
RRWIILGLNK	B*27:05	8	SB	16.60	0.03	SB
KRWII**P**GLNK	B*27:05	20	SB	38.40	0.08	SB
KRWI**VM**GLNK	B*27:05	19	SB	33.39	0.08	SB
KRWII**M**GLNK	B*27:05	18	SB	32.66	0.05	SB
KRWII**I**GLNK	B*27:05	14	SB	33.75	0.08	SB

* Net MHC 3.4 classifies peptides as strong binders if the affinity is <50 nM, weak binders if the affinity lies between 50-500 nM and non-binders if the affinity is >500 nM (26). NetMHCpan, in addition to the predicted binding affinity, takes peptide rank into consideration; the threshold for strong binders is 0.5, whereas weak binders may rank up to 2.0. The %-Rank is relative to a set of 200,000 random natural peptides (27). Grey shaded areas indicate peptide mutants that are predicted not to bind HLA-I by one (light grey) or both (dark grey) prediction programs.

Figure S1

A*02-SL9

Pt	TRBV	CDR3	TRBJ	%	Counts
#1					
t=2	2	CASSTPGEGGEQFF	2.1	100.00	83
#4					
t=1	**5.1**	**CASSFDAEQFF**	**2.1**	58.62	51
	28	**CASSDRTGGGETQYF**	**2.5**	39.08	34
	11.2	CASSLEHEQYF	2.7	2.30	2
t=2	**5.1**	**CASSFDAEQFF**	**2.1**	54.12	46
	28	**CASSDRTGGGETQYF**	**2.5**	41.18	35
	27	CASSLFGGSGNTIYF	1.3	4.71	4
#6					
t=2	6.5	CASSYGMGVGDRDTQYF	2.3	35.53	27
	27	CASSRTAGTYYNEQFF	2.1	26.32	20
	5.5	CASSLPTENTDTQYF	2.3	7.89	6
	12.4	CASSASTGGGYTF	1.2	5.26	4
	5.1	CASSFLAGGITDTQYF	2.3	5.26	4
	11.2	CASSLGAGRTEAFF	1.1	3.95	3
	20.1	CSAVPVSGVDEQFF	2.1	3.95	3
	19	CASSLERSNTEAFF	1.1	2.63	2
	4.3	CASSQEGAGITEAFF	1.1	2.63	2
	25.1	CASSEYRGKNTEAFF	1.1	1.32	1
	23.1	CASSGWTLRIYEQYF	2.7	1.32	1
	10.3	CAISEPGQQLYGYTF	1.2	1.32	1
	9	CASSVGVETGEQYF	2.7	1.32	1
	14	CASSQDRETQYF	2.5	1.32	1
#7					
t=1	**24.1**	**CATSDPIRASSHEQFF**	**2.1**	95.65	66
	24.1	CATNDPIRASSHEQFF	2.1	4.35	3
t=2	**24.1**	**CATSDPIRASSHEQFF**	**2.1**	93.42	71
	24.1	CATSDPIRASSHEQFF	2.1	3.95	3
	24.1	CATSDPIRASSHGQFF	2.1	2.63	2

B*08-EI8

Pt	TRBV	CDR3	TRBJ	%	Counts
#1					
t=1	**27**	**CASSPNGDRVFDQPQHF**	**1.5**	69.49	41
	2	**CASSENSLGRGLVKTQYF**	**2.5**	30.51	18
t=2	**27**	**CASSPNGDRVFDQPQHF**	**1.5**	63.51	47
	2	CASSPPMGRAEGYTF	1.2	16.22	12
	7.9	CASSFLRLAGGRDEQFF	2.1	4.05	3
	2	**CASSENSLGRGLVKTQYF**	**2.5**	2.70	2
	7.8	CASSLLDGTRETQYF	2.5	2.70	2
	9	CASSVAGDDRETQYF	2.5	2.70	2
	9	CASSVVGDGRETQYF	2.5	2.70	2
	9	CASSEGQGTTYEQYF	2.5	2.70	2
	2	CASSEAATGRGNQPQHF	1.5	1.35	1
	9	CASSVLHGRQETQYF	2.5	1.35	1
#3					
t=2	2	CASSPLAVGKETQYF	2.5	48.78	40
	27	CASSLSGRGATEAFF	1.1	12.20	10
	7.3	CASSLMDRTEAFF	1.1	6.10	5
	10.3	CAITDGNGPRIGYTF	1.2	6.10	5
	29.1	CSVGAGLTWANYGYTF	1.2	4.88	4
	7.2	CASSLPGQGRTPLHF	1.6	4.88	4
	9	CASSVVGGGRTDTQYF	2.3	2.44	2
	6.2/6.3	CASSWDTSRTEAFF	1.1	2.44	2
	11.2	CASRTTSGGTDTQYF	2.3	1.22	1
	9	CASSVAGDDRETQYF	2.5	1.22	1
	20.1	CSARDSRLEWDTQYF	2.3	1.22	1
	9	CASSVVGDTRETQYF	2.5	1.22	1
	9	CASSVVGDSYNEQFF	2.1	1.22	1
	9	CASSVVGDLRETQYF	2.5	1.22	1
	9	CASSVASYNSPLHF	1.6	1.22	1
	12.4	CASSLDSRGAFF	1.1	1.22	1
	7.2	CASSLPSQGRTEAFF	1.1	1.22	1
	13	CASSARDTLRGTGELFF	2.2	1.22	1
#4					
t=2	18	CASSPSGQGARTGELFF	2.2	25.00	14
	4.1	CASSPAGSYTEAFF	1.1	12.50	7
	7.2	CASSLTAGARVWQFF	2.1	10.71	6
	29.1	CSVRQGYEQYF	2.7	7.14	4
	9	CASSVLGGNYGYTF	1.2	7.14	4
	6.2/6.3	CASSYLWGQYEQYF	2.7	7.14	4
	12.3	CASSFLEDGYTF	1.2	5.36	3
	11.2	CASRGLAGGDTQYF	2.3	5.36	3
	9	CASSPRRGGPLQETQYF	2.5	3.57	2
	5.1	CASSFLPSYEQYF	2.7	3.57	2
	10.3	CAISADRVGYEQYF	2.7	1.79	1
	2	CASSSGTGVPYTF	1.2	1.79	1
	2	CASSELGNPNYGYTF	1.2	1.79	1
	4.3	CASSQEYRATTDTQYF	2.3	1.79	1
	4.3	CASSQGMFDYGYTF	1.2	1.79	1
	7.3	CASSRLAGGTDTQYF	2.3	1.79	1
	6.6	CASSYGGTEAFF	1.1	1.79	1

Figure S1 (cont.)

B*08-EI8

Pt	TRBV	CDR3	TRBJ	%	Counts
#5					
t=1	**9**	**CASSVLGGRTGELFF**	**2.2**	35.06	27
	2	CASSEVGSKSRQGAYEQYF	2.7	7.79	6
	9	**CASSTPGRRTSNQPQHF**	**1.5**	5.19	4
	5.1	CASSLNGFRGDTEAFF	1.1	5.19	4
	4.1	CASSQDRRDTYTYEQYF	2.7	5.19	4
	9	CASSVDAGGELFF	2.2	3.90	3
	7.8	**CASSFQGERTEAFF**	**1.1**	3.90	3
	7.8	**CASSLLGNTGELFF**	**2.2**	3.90	3
	14	CASSQDRRKQFF	2.1	2.60	2
	6.1	CASSPLDGFRDEQFF	2.1	2.60	2
	12.4	CASSSGPTRPQETQYF	2.5	2.60	2
	10.2	CASSESILSYEQYF	2.7	2.60	2
	3.1	CASSQEAGDSPQSTDTQYF	2.3	1.30	1
	9	CASSPFTGVGQPQHF	1.5	1.30	1
	4.1	CASSQTSASEQYF	2.7	1.30	1
	4.1	CASSQWDTGNEQFF	2.1	1.30	1
	3.1	**CASRNRAAGRFRETQYF**	**2.5**	1.30	1
	5.1	CASSSAPDGRNTGELFF	2.2	1.30	1
	2	CASKLAGQSLYNEQYFF	2.1	1.30	1
	12.3	CASSPWDTRGEKLFF	1.4	1.30	1
	20.1	CSARQGGNYEQYF	2.7	1.30	1
	7.9	CASSLAPYSNYEQYF	2.7	1.30	1
	11.2	CASSFQGNYEQYF	2.7	1.30	1
	20.1	CSVDRGPTGELFF	2.2	1.30	1
	28	CASSFLFRGRAGETQYF	2.5	1.30	1
	28	CASSDSTGDSGANVLTF	2.6	1.30	1
	14	CASSLDRETQYF	2.3	1.30	1
#6					
t=1	**7.2**	**CASSLIPGNQPQHF**	**1.5**	**94.0**	188
	12.5	**CASGLGRNQPQHF**	**1.5**	**3.0**	6
	7.8	**CASSLDDGRRPLHF**	**1.6**	**1.5**	3
	7.2	**CASSLPGQGRTEAFF**	**1.1**	**0.5**	1
	27	CASSFTGTGARTDTQYF	2.3	1.0	2
t=2	**7.2**	**CASSLIPGNQPQHF**	**1.5**	98.9	91
	20.1	CSARLSYEQYF	2.7	1.1	1

B*08-EI8

Pt	TRBV	CDR3	TRBJ	%	Counts
#7					
t=1	**2**	**CASSEASPGRSYGYTF**	**1.2**	31.34	21
	7.8	CASSMLDGGRDEQFF	2.1	17.91	12
	20.1	CSARWDYEQYF	2.7	16.42	11
	5.1	**CASSPLTSGGARDTQYF**	**2.3**	10.45	7
	27	CASSPTGTGVRNTEAFF	1.1	5.97	4
	20.1	**CSARYSYEQYF**	**2.7**	4.48	3
	27	CASSLYPGVTEAFF	1.1	4.48	3
	20.1	**CSARENYEQYF**	**2.7**	2.99	2
	9	CASSVVGGGRTDTQYF	2.3	2.99	2
	6.1	CASSGRLGLGTEAFF	1.1	1.49	1
	6.2/6.3	CASSGSGGRTTEAFF	1.1	1.49	1
t=2	**20.1**	**CSARENYEQYF**	**2.7**	43.33	26
	2	**CASSEASPGRSYGYTF**	**1.2**	28.33	17
	5.1	**CASSPLTSGGARDTQYF**	**2.3**	25.00	15
	20.1	**CSARYSYEQYF**	**2.7**	1.67	1
	7.8	CASSLLDGTRDEQYF	2.7	1.67	1
#8					
t=1	23.1	CASSQSPTRLAGTPHNEQFF	2.1	39.71	27
	5.5	CASSLDLNRERTGELFF	2.2	16.18	11
	2	CASGGQGDRQPQHF	1.5	8.82	6
	7.8	CASSLVEIRNEQFF	2.1	5.88	4
	23.1	CASSQSPTRLAGTPHNEQFF	2.1	4.41	3
	6.5	CASSYEVRTEAFF	1.1	4.41	3
	2	CASSISERESYEQYF	2.7	4.41	3
	6.1	CASSEILLRGAETQYF	2.5	4.41	3
	9	CASSLVGDARETQYF	2.5	2.94	2
	14	CASSPGVSQPQHF	1.5	1.47	1
	9	CASSVVGDSRETQYF	2.5	1.47	1
	9	CASSVVGDGRETQYF	2.5	1.47	1
	7.8	CASSLQGERTEAFF	1.1	1.47	1
	7.8	CASSLLNGQSSGNTIYF	1.3	1.47	1
	7.8	CASSLLDAMRNEQFF	2.1	1.47	1

Figure S1 (cont.)

B*08-FL8

Pt	TRBV	CDR3	TRBJ	%	Counts
#1					
t=1	**2**	**CASSVLAGLGNEQFF**	**2.1**	98.67	74
	2	CASSVLAGHGNGQFF	2.1	1.33	1
t=2	11.2	CASSLLAGLTDTQYF	2.3	48.48	32
	2	**CASSVLAGLGNEQFF**	**2.1**	**46.97**	31
	2	CASSVLAGLGNEQFF	2.1	3.03	2
	11.2	CASSLLAGLADTQYF	2.3	1.52	1
#3					
t=2	14	CASSHSPGTSGSPGVTQYF	2.5	45.71	32
	27	CASSQGQGIALYGYTF	1.2	35.71	25
	4.3	CASSPNSGRVTGELFF	2.2	10.00	7
	27	CASSLGQGLVLSGYTF	1.2	5.71	4
	2	CASSELAGLGDTQYF	2.3	2.86	2
#4					
t=1	20.1	CSARDRGRALNEQFF	2.1	71.15	37
	27	CASSFLHPSTEAFF	1.1	11.54	6
	6.1	CASRDFPSGGSYEQYF	2.7	7.69	4
	12.3	CASTYTGGAGGELFF	2.2	5.77	3
	20.1	CSAKDRLAYSYEQYF	2.7	3.85	2
#5					
t=1	**6.2/6.3**	**CASRYLRGEGAHPSNTEAFF**	**1.1**	19.23	10
	9	**CASSTLASGRYEQYF**	**2.7**	19.23	10
	7.8	CASSPIEGLAKNIQYF	2.4	11.54	6
	9	**CASSLMESVWVEQFF**	**2.1**	11.54	6
	6.1	**CASSEFGQGFYEQYF**	**2.7**	11.54	6
	3.1	CASSQWGLPTDTQYF	2.3	7.69	4
	6.1	**CASSEFPAGVYEQFF**	**2.1**	5.77	3
	9	CASSLGQGTTYEQYF	2.7	3.85	2
	7.9	CASSPPSGDYEQFF	2.7	3.85	2
	9	CASSTLASGRYEQYF	2.7	3.85	2
	9	CASSPGEGGRYEQYF	2.7	1.92	1

B*08-FL8

Pt	TRBV	CDR3	TRBJ	%	Counts
#6					
t=2	**7.9**	**CASSLGDLYTGELFF**	**2.2**	92.86	91
	6.2/6.3	**CASSYLAGQGAPYSNQPQHF**	**1.5**	3.06	3
	6.4	**CASSDSGAGTPNGYTF**	**1.2**	2.04	2
	12.3	CASSFTGENSPLHF	1.6	1.02	1
	29.1	CSVWTGEKHEAFF	1.1	1.02	1
#7					
t=1	**27**	**CASRVGQGVVGELFF**	**2.2**	61.11	33
	25.1	**CASGVLREAFF**	**1.1**	27.78	15
	11.2	CASSLDHQAGHEQYF	2.7	5.56	3
	27	CASRVGQGVVGGLFF	2.2	3.70	2
	20.1	CSVYREIPEAFF	1.1	1.85	1
t=2	6.5	CASSYQTGTGTYGYTF	1.2	28.74	25
	27	**CASRVGQGVVGELFF**	**2.2**	21.84	19
	28	CASSLLDRVREDTQYF	2.3	16.09	14
	25.1	**CASGVLREAFF**	**1.1**	14.94	13
	20.1	CSARDHLRANEKLFF	1.4	8.05	7
	30	CAWSSRDRVLGQSEQYF	2.7	4.60	4
	5.1	CASSLYSIGSGNTIYF	1.3	3.45	3
	27	CASSLLDRVREDTQYF	2.3	1.15	1
	6.5	CASSYQTGTGTYDHTF	1.2	1.15	1
#8					
t=1	2	CASSELAGLGTEAFF	1.1	67.89	74
	5.6	CASSWGIGSQQETQYF	2.5	18.35	20
	11.2	CASSFTTSGTIPNEQFF	2.1	4.59	5
	28	CASSPRDRRTWNEQFF	1.4	1.83	2
	27	CASGIGQGIALRELFF	1.4	1.83	2
	2	CASSEFGAGVYEQYF	2.7	0.92	1
	5.6	CTSSAGTGSAQETQYF	2.5	0.92	1
	7.9	CASSLLAGLSYNEQFF	2.1	0.92	1
	20.1	CSARDFLADSTDTQYF	2.3	0.92	1
	6.6	CASRSTSGDRDTQYF	2.3	0.92	1
	11.2	CASSLIPGITGELFF	2.2	0.92	1

Figure S1 (cont.)

B*27-KK10

Pt	TRBV	CDR3	TRBJ	%	Counts
#2					
t=1	27	CASSSRTGELFF	2.2	16.90	12
	4.3	**CASSPGRLSHEQFF**	**2.1**	5.63	4
	27	CASSAQTGFLFF	1.4	5.63	4
	4.3	CASSPGRTSHEQFF	2.1	5.63	4
	7.9	CASSWDRNEQFF	2.1	5.63	4
	4.3	CASSPGQLGNTIYF	1.3	5.63	4
	4.3	CASSQGAGGIEQFF	2.1	5.63	4
	27	CASSQRTGELFF	2.2	4.23	3
	20.1	CSAREEARGYGYTF	1.2	4.23	3
	7.9	CASSLDRDEQFF	2.7	4.23	3
	7.9	CASSGGTGVFYEQYF	2.7	4.23	3
	27	CASSSRTGELFF	2.2	4.23	3
	6.5	CASSPGQFAYEQYF	2.7	2.82	2
	28	CASSLGLAGTHNEQFF	2.1	2.82	2
	7.9	CASSSDRDEQYF	2.7	2.82	2
	27	CASSPRTGELFF	2.2	1.41	1
	27	CASLGGDPNYGYTF	1.2	1.41	1
	27	CASSRVTGELFF	2.2	1.41	1
	28	CASSSRQNTEAFF	1.1	1.41	1
	10.1	CASSSTRNGEQFF	2.1	1.41	1
	30	CAWRRSSTEAFF	1.1	1.41	1
	27	CASSQRTGELFF	2.2	1.41	1
	27	CASSQRTGELFF	2.2	1.41	1
	4.3	CASSPGRTSHEQFF	2.1	1.41	1
	27	CASRITGWSGVNTEAFF	1.1	1.41	1
	20.1	CSARDRPESGYTF	1.2	1.41	1
	7.9	CASSYDRDEQFF	2.1	1.41	1
	4.3	CASSPGTSSSYNEQFF	2.1	1.41	1
	14	CASSQEFLAVEQFF	2.1	1.41	1
t=2	**4.3**	**CASSPGRLSHEQFF**	**2.1**	96.77	120
	27	CASSAQTGELFF	1.4	1.61	2
	4.3	CASSPGLLSNEQYF	2.7	1.61	2

B*27-KK10

Pt	TRBV	CDR3	TRBJ	%	Counts
#6					
t=1	**4.3**	**CASSPGQVAYEQYF**	**2.7**	73.68	42
	4.3	CASSQGVLAYEQYF	2.7	5.26	3
	7.9	CASSLDRGEQFF	2.1	5.26	3
	4.3	**CASSAGISAYEQYF**	**2.7**	5.26	3
	20.1	CSARARSDSYGYTF	1.2	3.51	2
	20.1	**CSAREGVAGALYEQYF**	**2.7**	3.51	2
	7.9	CASSLDSYEQYF	2.7	1.75	1
	21.1	CASKNRPTGNTIYF	1.3	1.75	1
t=2	**4.3**	**CASSAGISAYEQYF**	**2.7**	43.08	28
	4.3	**CASSPGQVAYEQYF**	**2.7**	40.00	26
	20.1	**CSAREGVAGALYEQYF**	**2.7**	9.23	6
	15	CATSMGQTTYEQYF	2.7	7.69	5
#7					
t=1	**9**	**CASSVVSDNEQFF**	**2.1**	71.64	48
	12.5	CASGFGGTEAFF	1.1	7.46	5
	4.3	**CASSQGTSDYEQYF**	**2.7**	5.97	4
	4.3	CASSQGRGDYGYTF	1.2	5.97	4
	4.3	**CASATGIFNNEQFF**	**2.1**	5.97	4
	6.1	CASSPGQFSHEQYF	2.7	1.49	1
	5.1	CASSPDVGGAFFYGYTF	1.2	1.49	1
t=2	**9**	**CASSVVSDNEQFF**	**2.1**	70.42	50
	20.1	CSAREGQGSIEQYF	2.7	15.49	11
	4.3	CASSTGLAGGKEQYF	2.7	7.04	5
	4.3	**CASATGIFNNEQFF**	**2.1**	4.23	3
	4.3	**CASSQGTSDYEQYF**	**2.7**	1.41	1
	9	CAGSVVSDNEQFF	2.1	1.41	1

Figure S1 (cont.)

Additional T-cell repertoires analyzed upon long-term viral control (*Fig. 6*)

B*08-EI8

Pt	TRBV	CDR3	TRBJ	%	Counts
#5					
pre-t1	**7.8**	**CASSFQGERTEAFF**	**1.1**	77.60	97
	9	**CASSVLGGRTGELFF**	**2.2**	13.60	17
	7.9	CASSQGSGLLPGEQFF	2.1	4.00	5
	7.8	**CASSLLGNTGELFF**	**2.2**	2.40	3
	7.2	CASSSRGVRSLDTQYF	2.3	1.60	2
	9	**CASSTPGRRTSNQPQHF**	**1.5**	0.80	1
post-t2	**3.1**	**CASRNRAAGRFRETQYF**	**2.5**	64.00	48
	5.4	CASSLPTALTGELFF	2.2	32.00	24
	3.1	CAGRNRAAGRFRETQYF	2.2	2.67	2
	3.1	CANRNRAAGRFRETQYF	2.2	1.33	1
#6					
post-t1	**7.2**	**CASSLIPGNQPQHF**	**1.5**	94.87	111
	12.5	**CASGLGRNQPQHF**	**1.5**	2.56	3
	7.2	**CASSLPGQGRTEAFF**	**1.1**	0.85	1
	7.3	CASSLLGSTDTQYF	2.3	1.71	2
post-t2	**7.8**	**CASSLDDGRRPLHF**	**1.6**	40.77	53
	7.3	CASSLIDGARDEQFF	2.1	33.08	43
	27	CASSLSRDRNEQFF	2.1	20.77	27
	29	CSGGPLETQYF	2.5	5.38	7
#9					
t1	**9**	**CASSVGGDARETQYF**	**2.5**	69.88	58
	2	**CASSASTKDTQYF**	**2.3**	12.05	10
	21.1	CASSKEHRGTEDYEQYF	2.7	9.64	8
	12.4	CASSFLTRQPQHF	1.5	3.61	3
	12.4	CASSLGTEWGYDNTQYF	2.3	1.20	1
	28	CASSYQPGGSGELFF	2.2	1.20	1
	28	CASSLPPPGIKGELFF	2.2	1.20	1
	7.8	CASSTLPGTPRNEQYF	2.7	1.20	1
t2	**9**	**CASSVGGDARETQYF**	**2.5**	80.28	57
	2	**CASSASTKDTQYF**	**2.3**	16.90	12
	4.1	CASSQEPAQSSYEQYF	2.7	2.82	2

B*08-FL8

Pt	TRBV	CDR3	TRBJ	%	Counts
#5					
pre-t1	3.1	CASSQEWSYEQYF	2.7	39.06	50
	6.1	**CASSEFPAGVYEQFF**	**2.1**	25.00	32
	28	CASTVARDGGELFF	2.2	22.66	29
	12.3	CASSYGLNTEAFF	1.1	7.03	9
	9	**CASSTLASGRYEQYF**	**2.7**	3.13	4
	9	CASSALSGRVSTDTQYF	2.3	2.34	3
	5.1	CASSPAGVGETQYF	2.5	0.78	1
post-t2	**6.2/6.3**	**CASRYLRGEGAHPSNTEAFF**	**1.1**	44.83	39
	6.1	**CASSEFGQGFYEQYF**	**2.7**	19.54	17
	6.1	**CASSEFPAGVYEQFF**	**2.1**	12.64	11
	19	CASSISGIGTGELFF	2.2	4.60	4
	9	CASSVDGFLVGYNEQFF	2.1	4.60	4
	7.9	CASSRAGADYNEQFF	2.1	3.45	3
	29	CSGPGPVLLGDQPQHF	1.5	2.30	2
	7.6	CASSLGTGGNQPQHF	1.5	2.30	2
	9	**CASSLMESVWVEQFF**	2.1	1.15	1
	6.1	CASRYLRGEGAHPSNTEAFF	1.1	1.15	1
	5.1	CASSLGLIYEQYF	2.7	1.15	1
	6.5	CASRNRPGQASYEQYF	2.7	1.15	1
	7.6	CASSQGSGILDSPLHF	1.6	1.15	1
#6					
post-t1	**7.9**	**CASSLGDLYTGELFF**	**2.2**	80.65	50
	6.**2/6.3**	**CASSYLAGQGAPYSNQPQHF**	**1.5**	1.61	1
	7.2	CASSSLGGGRGINSPLHF	1.6	12.90	8
	4.3	CASSQGLGQGSYEQYF	2.7	4.84	3
post-t2	**7.9**	**CASSLGDLYTGELFF**	**2.2**	45.88	39
	6.2/6.3	**CASSYLAGQGAPYSNQPQHF**	**1.5**	17.65	15
	6.4	**CASSDSGAGTPNGYTF**	**1.2**	35.29	30
	7.9	CASSLVPDTQYF	2.1	1.18	1
#9					
t1	**7.9**	**CASSLAPGTSGSPYNEQFF**	**2.1**	62.50	45
	14	**CASSLGTGIANYGYTF**	**2.1**	37.50	27
t2	**7.9**	**CASSLAPGTSGSPYNEQFF**	**2.1**	1.32	1
	14	**CASSLGTGIANYGYTF**	**2.1**	57.89	44
	24.1	CATKGTGLYNEQFF	2.1	40.79	31

Figure S1. Overview of TCRβ sequences.
TRBV usage, CDR3 amino acid sequence, TRBJ usage, percent frequency and clone count are shown for each TCRβ clonotype. **Bold** sequences indicate nucleotide-identical TCRs within individuals at different time points. Grey bars indicate amino acid-identical TCRs between unrelated individuals (public TCRs).

Evolution of a pre-existing influenza A T-cell response upon infection during the 2009-2010 flu season

Ana I. Costa[1], Ingrid M. M. Schellens[1], Nening M. Nanlohy[1], Can Keşmir[3], José A. M. Borghans[1] and Debbie van Baarle[1,2,*]

[1]Laboratory of Translational Immunology, Department of Immunology, University Medical Center Utrecht, Utrecht, The Netherlands.

[2]Department of Internal Medicine and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands.

[3]Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands.

**Current affiliation: Center for Infectious Disease Control, National Institute for Public Health and the environment (RIVM), Bilthoven, The Netherlands.*

Manuscript in preparation.

Abstract

Pre-existing $CD8^+$ T cells to conserved antigens are believed to be important in cross-recognition of divergent influenza strains and to contribute to milder disease and lower mortality from influenza A virus infection.

During the 2009-2010 pandemic-flu season, we observed a high magnitude $CD8^+$ T-cell response in an individual who had recently exhibited flu-like symptoms, and from whom we also had a pre-infection blood sample. This provided the unique opportunity to study functional and compositional aspects of an immune response upon natural influenza infection. The T-cell response targeted a conserved flu-epitope (HLA-A2-$M1_{58-66}$) in the matrix 1 protein that is frequently elicited by different influenza-A subtypes. Longitudinal analysis of the pre- and post-infection T-cell response revealed that the HLA-A2-$M1_{58-66}$-responsive $CD8^+$ T cells exhibited a transiently more polyfunctional *in vitro* profile upon infection. The A2-$M1_{58-66}$ T-cell receptor (TCR) repertoire prior to infection was strongly dominated by commonly observed TRBV19 clonotypes, while post-infection there was a permanent shift in clonotype usage, which was dominated by a TRBV20.1 clonotype.

These data show that even in the case of a constrained, highly conserved influenza A epitope, recall influenza episodes can elicit transitory changes in functionality and permanent changes in the TCR repertoire of the responding $CD8^+$ T-cell population.

Introduction

Influenza viruses cause annual outbreaks of respiratory infections with attack rates of 5–10%. Thus, humans are infected repeatedly with intervals of, on average, 10–20 years (1). B-cell memory against hemagglutinin (HA) and neuraminidase (NA) provides protection from infection with viruses of the same subtype, in which these surface proteins have not diverged considerably. In case of heterosubtypic infection, T cells are believed to be major players in determining the severity of the outcome. By recognizing epitopes derived from relatively conserved, internal viral proteins, T cells show cross-reactivity to divergent flu subtypes (2-10) and contribute significantly to viral clearance and milder disease (11-14). This makes them attractive candidates as targets for universal vaccine formulations.

The detection of a new H1N1 reassortant in the 2009-2010 flu season (pH1N1(2009)) (15, 16) raised the fear for a new pandemic. In retrospect, even though pH1N1(2009) quickly became the main circulating strain, the pandemic was moderate in incidence and severity, suggesting that pre-existing immunity played a protective role (1, 17-21).

Evidence of a beneficial role of $CD8^+$ T cells in heterosubtypic immunity stems mostly from animal studies (reviewed in (22)). In humans, McMichael and colleagues (13) have previously reported that virus-specific cytotoxicity was inversely correlated with virus shedding in experimentally infected individuals with no antibodies to the challenge virus subtype. More recently, it was shown that the frequency of influenza-specific $CD8^+$ T cells inversely correlated with associated flu symptoms (14). The remainder human data corroborating a beneficial role of $CD8^+$ T cells in the clinical course of influenza A infection relies mainly on epidemiological analyses and indirect evidence, such as efficient *in vitro* lysis of cells infected with a divergent virus subtype by cross-reactive $CD8^+$ T cells and mutational indications of $CD8^+$ T-cell pressure (18, 22-25).

The abovementioned potential role of cross-reactive $CD8^+$ T-cell responses in reducing the severity of heterosubtypic infection is strongly related to the degree of conservation of the epitopes they recognize (26). One of the epitopes that seems to be conserved across subtypes and strains is the immunodominant, A*0201-restricted, Matrix1-derived (structural) $M1_{58-66}$ epitope (18, 24, 27). In an influenza genome-wide analysis combining *in silico* and *in vitro* approaches for epitope discovery, $M1_{58-66}$ was shown to be the most frequently targeted epitope (48%, 10/21 individuals) (28). It was also shown to elicit a cross-reactive response: *in vitro* expanded $M1_{58-66}$-specific bulk $CD8^+$ T-cell lines were able to kill virus-infected cells, whether the infecting strain was from subtypes H3N2, seasonal H1N1 or pandemic H1N1 (8). Responses to $M1_{58-66}$ have

been reported as protective in an HLA-A2-transgenic mouse model (29). Finally, the TCR repertoire of HLA-A2-$M1_{58-66}$ specific $CD8^+$ T-cell responses has been described as strongly dominated by $TRBV19^+$ cells (*old nomenclature (Wei): Vb17* (30-33)), with particular TRBV19 clonotypes being consistently identified in different individuals (34-36).

Despite the abovementioned *in vitro* indications of the efficiency of $CD8^+$ T-cell responses, there is only few data on the evolution of *in vivo* T-cell responses in humans experiencing acute infection. In this study, we had the unique opportunity to characterize the HLA-A*02-$M1_{58-66}$-specific $CD8^+$ T-cell response before and after an acute flu episode in an otherwise healthy individual that presented with symptoms during the 2009-2010 flu season. This longitudinal study assessed how the recall response directed toward this highly conserved epitope evolved in terms of repertoire and functionality after likely exposure to a different influenza-A subtype.

Materials and methods

Donor

Blood from a 22-year old HLA-typed donor (HLA-A*01, A*02, B*08, B*55) was collected at different time points (-7, 1.5, 4, 9, 23, 34, 48 and 98 weeks) after subsiding of flu symptoms. The subject gave written informed consent. Peripheral blood mononuclear cells (PBMC) were isolated from blood samples by Ficoll-Paque density gradient centrifugation and cryopreserved for further experiments.

Peptides and Tetramer

PBMC collected at 1.5 weeks after symptom disappearance were screened with HLA-matched influenza A-derived peptides curated as influenza-A epitopes in the immune epitope database IEDB (www.immuneepitope.org): $M1_{58-66}$ (GILGFVFTL, HLA-A2-restricted), $NA_{213-221}$ (CVNGSCFTV, HLA-A2), NP_{44-52} (CTELKLSDY, HLA-A1), $PB1_{591-599}$ (VSDGGPNLY, HLA-A1), and $NP_{380-388}$ (ELRSRYWAI, HLA-B8). HLA-A2-$M1_{58-66}$ monomers were produced in-house as previously described (37, 38), with the exception of monomer folding by dialysis method as replacement for dilution. HLA-A2-$M1_{58-66}$ monomers were subsequently tetramerized upon addition of PE-labeled streptavidin.

Screening of T-cell responses by IFN-γ ELIspot assay

Cryopreserved PBMCs were tested for reactivity against influenza A-derived peptides of known HLA allele restriction. Briefly, PBMCs were thawed and tested for the ability

to produce interferon (IFN)-γ in an ELIspot assay. Cells were incubated at a final concentration of 10^5 cells/well in the presence or absence of 20μg/ml peptide for 20–24 h at 37°C, in 96-well multiscreen filter plates (Millipore, Volketswil, Switzerland) coated overnight with a capture anti-IFNγ monoclonal antibody at a concentration of 15mg/ml in PBS (MABTECH, Stockholm, Sweden). Cells incubated in medium alone served as negative control, and phytohemagglutinin (Murex, Dartford, UK) was used as a positive control. IFNγ production was detected using a biotinylated anti-IFNγ monoclonal antibody (MABTECH), followed by incubation with poly-streptavidin-horseradishperoxidase (Sanquin), and developed with TMB substrate (Sanquin). IFNγ-producing cells were analyzed using an automated spot reader (AELVIS GmbH, Hannover, Germany). Responses to the peptides were considered positive when higher than 50 spot-forming cells (SFC) per 10^6 PBMCs after subtraction of the mean of negative wells (background), and over 3 times the mean of negative wells.

Ex vivo staining for assessment of phenotype and activation profiles and flow-assisted cell sorting of HLA-A2-M1$_{58\text{-}66}^{+}$ CD8^{+} T cells

For phenotypic analysis, thawed PBMC were stained with live/dead violet viability dye (Invitrogen), after which they were labeled at room temperature with the respective cognate pMHC-class I tetrameric complex (HLA-A*0201-M1$_{58\text{-}66}$) at pre-titrated concentrations. Subsequently, cells were stained with the following monoclonal antibodies: CD3-PerCP, CD8-AmCyan/V500, CD45RO-PE-Cy7 and CD57-FITC (BD Biosciences), CD27-APC-Cy7 (Biolegend), and CD14 and CD19, both Pacific Blue-labeled (Biolegend), for the exclusion of monocytes and B cells. For activation marker analysis, CD38-PE-Cy7 and HLA-DR-FITC (eBioscience) replaced the CD45RO and CD57 antibodies. The staining procedure for flow-assisted cell sorting was similar to the one done for phenotyping, with the exception that monoclonals for CD27, CD45RO and CD57 were excluded. HLA-A2-M1$_{58\text{-}66}^{+}$ CD8^{+} T cells were sorted using a FACSAria flow cytometer (BD Biosciences) directly into RNAlater (Ambion Inc., (Applied Biosystems)), and stored at -80°C for subsequent TCR clonotype analysis.

Two-phase exponential decay

The height of the HLA-A2-M1$_{58\text{-}66}^{+}$ CD8^{+} T-cell response (Y) was described by fitting a two-phase exponential decay model to the experimental data:

$$Y = (Y_0 - P)xe^{-k_f t} + (Y_0 - P)(1 - x)e^{-k_s t} + P$$

where Y_0 is the height of the response at t=0, P is the height of the response at $t \rightarrow \infty$, x is

the fraction of the response declining at a fast rate k_f and (1-x) the fraction of the response declining at a slower rate k_s. The corresponding half-lives were computed as ln(2)/k. The model was fitted using GraphPad Prism 5.0.

Functional profile upon antigen stimulation

For assessment of the functional profile of cells responsive to $M1_{58-66}$ peptide, cryopreserved PBMC were thawed and rested at 37°C for 2 hours in complete RPMI medium (RPMI 1640 supplemented with 10% heat-inactivated fetal calf serum, l-glutamine, and antibiotics) at a concentration of 2×10^6 cells/ml. For stimulation, cells were incubated with $M1_{58-66}$ at a concentration of 2μM in the presence of co-stimulation (1 μg/ml αCD28 (Sanquin Reagents, Amsterdam, The Netherlands) and 1 μg/ml αCD49d (Pharmingen, San José, California, USA)) and anti-CD107a antibodies for 1 h at 37°C in a 5% CO2 incubator, followed by an additional 5 h in the presence of the secretion inhibitor monensin (Golgistop; BD Biosciences, San Jose, CA, USA). As a negative control, cells were incubated in the absence of peptide, and as a positive control, cells were stimulated with a combination of PMA and ionomycin (Sigma-Aldrich, Zwijndrecht, The Netherlands) (both included costimulation). Cells were stained with directly conjugated mAbs specific for the cell surface antigens CD3-APC-eFluor780 (eBioscience) and CD8-V500 (BD). After fixation and permeabilisation (permeabilisation reagents BD), cells were stained for intracellular cytokines/chemokine (IFNγ-FITC and MIP1β-PE (BD), and IL-2-Pacific Blue and TNFα-APC (eBioscience), and then fixed in Cellfix (BD). Data was analyzed using FlowJo software (version 9.3; TreeStar, Inc.). Frequencies of responding cells were reported after subtraction of the frequencies in medium controls.

T-cell receptor clonotype analysis

Clonotype analysis was performed as described previously (39, 40). Briefly, the Poly(A)Tract system 1000 (Promega) was used to extract mRNA from the sorted populations, according to the manufacturer's instructions. All expressed TCRβ chains were amplified linearly by anchored template-switch RT-PCR. Amplified products were ligated into the pGEMT-Easy vector (Promega), used to transform chemically competent *E. coli* bacteria, and positive colonies were selected for sequencing using Big Dye Terminator V3.1 with capillary electrophoresis. Analysis of TCRβ sequences was performed using web-based software from ImMunoGeneTics (IMGT) (http://imgt.cines.fr/), and gene usage was assigned using IMGT nomenclature (41).

IMGT is acknowledged as the international reference for immunoglobulin and TCRβ sequence analysis (WHO/IUIS).

Results

We identified a healthy individual exhibiting mild symptoms of a respiratory infection, malaise, fever and muscle pain, in November 2009, the peak of the yearly influenza season in the European region (19). As both symptoms and timing were highly suggestive of flu infection, we investigated whether influenza-A responsive $CD8^+$ T cells could be detected in PBMC collected shortly (approximately 1.5 weeks) after recovery of symptoms. To this end, we screened PBMC for responses to known, IEDB-curated influenza A-specific peptides restricted to the patient's HLA background (see materials and methods) in an IFNγ-ELIspot. The response of highest magnitude was elicited by the $M1_{58-66}$ epitope of the matrix 1 protein (Fig. 1A). A minor response to this epitope was already measurable in a PBMC sample collected 7 weeks prior to the flu episode (Fig. 1B), suggesting that the observed response was a recall response. We followed the A2-$M1_{58-66}$ specific $CD8^+$ T cell response longitudinally, and studied whether there were changes in phenotype, functionality and TCR repertoire of the responding T cells upon antigen re-exposure during natural infection.

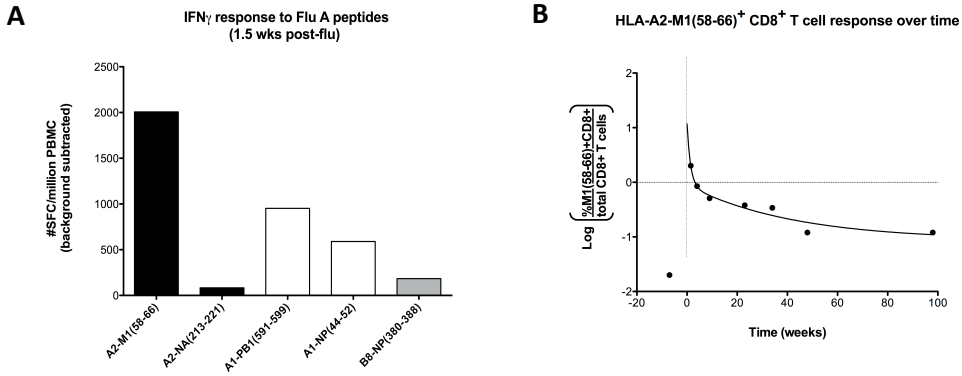


Figure 1. Dominance and kinetics of the $CD8^+$ T-cell response towards HLA-A2-$M1_{58-66}$ during influenza infection..
PBMC collected shortly after flu symptom disappearance (~ 1.5 weeks) were screened with influenza A-curated peptides restricted by 3 of the individual's HLA class I alleles in an IFNg-ELIspot assay (A), as described in Materials and Methods. The mean number of spot-forming cells per million PBMC (after subtraction of background mean) is plotted for each peptide. The fraction of A2-$M1_{58-66}$-specific $CD8^+$ T cells was followed in time, by *ex vivo* pMHC multimer staining (B). The contraction of the response (i.e. decline) was fitted using a two-phase decay equation.

We tracked the magnitude of the dominant A2-$M1_{58-66}$-specific T-cell response by tetramer staining in samples collected longitudinally at approximately 1.5, 4, 9, 23, 34, 48 and 98 weeks after recovery. Upon infection, the magnitude of the response increased considerably, from 0.02% prior to infection to 2.01% of the total $CD8^+$ T-cell population by 1.5 weeks after symptoms disappeared (Fig. 1B). During the contraction phase, the A2-$M1_{58-66}$-specific population decreased from 2.01% to 0.85% of the total $CD8^+$ T-cell population (2.4-fold) between the 1.5wk and 4wk time points, and further to 0.51% at 9 weeks post-recovery. By week 98, the percentage of $CD8^+$ memory T cells recognizing the A2-$M1_{58-66}$ complex was 0.12%, still slightly higher than pre-infection. Based on the acquired data, we used a two-phase exponential decay equation to estimate the decline of the response (Fig. 1B). The fraction of responding cells declined with a half-life of around 1 week in the fast phase and 29 weeks in the slow decay phase. Published studies have reported the CD8+ T cell response to peak around 7 days post-infection, and/or day 3-7 after onset of symptoms (42-45). Taking into account that influenza infection has an average incubation period of 2 days and 4-5 symptomatic days, we assume that the response peaked around the time of symptom disappearance (i.e.: t=0) and estimated the peak response to have been approximately 11% of the total circulating $CD8^+$ T-cell population.

We then analyzed the phenotype and function of the T cells recognizing the A2-$M1_{58-66}$ complex, by assessing a) surface markers for phenotype and activation of the tetramer-binding cells, and b) the $M1_{58-66}$ peptide-specific response upon *in vitro* short-term stimulation of PBMC. Prior to infection, the A2-$M1_{58-66}$ response consisted predominantly of $CD27^+CD45RO^+$ central memory T cells. After infection, an increase was observed in the $CD27^-CD45RO^+$ effector memory phenotype subset (Fig. 2A), which remained dominant until week 48. By week 98, the T-cell phenotype had returned to the pre-infection pattern. Activated cells were detected only in the first time point after infection (t=1.5 wks, Fig. 2B). The loss of the activation phenotype was paralleled by an increase in CD57 expression, a marker that has been used as an indicator of the proliferative history of cells, suggesting that that fraction of specific cells could be senescent after recent antigen-driven activation and proliferation.

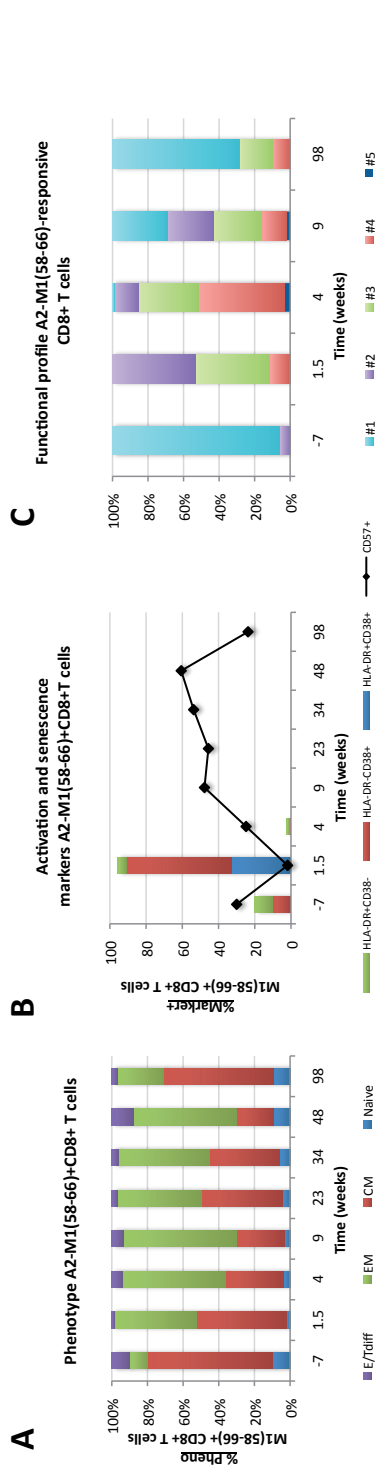


Figure 2. Post-flu A2-M1$_{58\text{-}66}$-specific CD8$^+$ T cells show a transient change in phenotype and functional profile after influenza infection.
A2-M1$_{58\text{-}66}$-specific, pMHC-multimer$^+$ CD8$^+$ T cells present at the different time points were phenotypically categorized (A) as: Naïve (CD27$^+$CD45RO$^-$), Central memory (CM, CD27$^+$CD45$^+$), Effector memory (EM, CD27$^-$CD45RO$^+$) and Effector/Terminally differentiated (E/Tdiff, CD27$^-$CD45RO$^-$). In addition, A2-M1$_{58\text{-}66}$-specific, pMHC-multimer$^+$ CD8$^+$ T cells were defined as HLA-DR$^+$CD38$^+$, HLA-DR$^+$CD38$^-$, and HLA-DR$^-$CD38$^+$, in terms of activation status (B). In parallel, the percentage of CD57$^+$ cells among the A2-M1$_{58\text{-}66}$-specific CD8$^+$ T cells is plotted. The ability of CD8$^+$ T-cells to respond to the cognate peptide by degranulation (CD107a mobilization), and cytokine (IFNγ, IL2, TNFα) and chemokine (MIP1β) production was assessed upon *in vitro* stimulation (C). PBMC of the different time points were stimulated for 6 hours in the presence of the M1$_{58\text{-}66}$. Responding cells were classified according to their ability to perform up to 5 functions (background-adjusted). Cells were gated on single live CD3$^+$ CD8$^+$ T lymphocytes, gating as in 3A, top).

Next, we studied the functional profile of the A2-M1$_{58-66}^{+}$ CD8^{+} T cell response after natural infection. To this end, we stimulated PBMC of different time points *in vitro* with the M1$_{58-66}$ peptide, and analyzed their ability to respond to the cognate peptide by degranulation (CD107a mobilization), and cytokine (IFNγ, IL2, TNFα) and chemokine (MIP1β) production (Fig. 2C). Prior to the flu episode, the majority of the CD8^{+} T cells responding to M1$_{58-66}$ displayed only one function. After infection, the response became more polyfunctional. At 1.5 weeks, almost all reacting CD8^{+} T cells were multifunctional, and at t=4 weeks an even higher fraction of cells displayed 3 or more functions. However, over time, the functional profile of the responsive cells became closer to that of pre-infection.

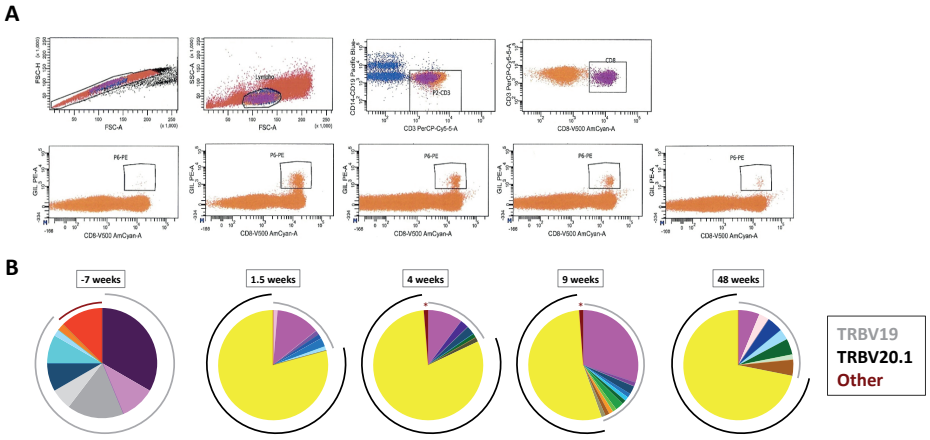


Figure 3. Changes in the A2-M1$_{58-66}$-specific CD8^{+} T-cell repertoire after influenza infection. Viable HLA-A2-M1$_{58-66}$-specific CD8^{+} T-cells were sorted by flow cytometry using the respective pMHCI tetramer, and the T-cell repertoire was determined by TCRβ clonotype analysis, as described in Materials and methods. (A) Gating strategy for the selection and sorting of antigen-specific cells. Top row, from left to right: selection of i) singlets, ii) lymphocytes, iii) live CD3^{+} T cells (dump channel: exclusion Violet dye, CD14 and CD19), and iv) CD8^{+} T cells. Lower row: FACS plots of the A2-M1$_{58-66}$-specific CD8^{+} T cells in chronological order (-7, 1.5, 4, 9 and 48 weeks, respectively). (B) Clonotypic composition of the A2-M1$_{58-66}$-specific CD8^{+} T cell pool in the different time points: distinct clonotypes, as determined by different Vβ nucleotide sequence, are color-coded. Clonotypes with the same or similar CDR3β amino acid sequence have been coded by a similar color. Clonotypes of the TRBV19 family, the TRBV 20.1 clonotype, or clonotypes of other families are marked with a grey line, black line, or dark-red asterisk/line, respectively.

Finally, we investigated the TCR usage of A2-M1$_{58-66}$-specific CD8^{+} T cells upon infection. We followed the evolution of the repertoire over time by TCR Vβ clonotypic analysis of tetramer-sorted specific cells (Gating in Fig. 3A). Interestingly, the composition of the epitope-specific population changed upon re-infection (Table I, and Fig. 3B). Prior to infection, clonotypes carrying TRBV19 were the major contributors to

the total T cell response directed against A2-$M1_{58-66}$. Influenza infection was marked by the appearance of a novel clonotype belonging to a different family, TRBV20.1. This clonotype dominated the post-flu A2-$M1_{58-66}$-specific response, and remained as the hierarchically highest in the pool at all time points post-infection. The dominant clonotype prior to infection (TRBV19-TRBJ2.7 CASSSRSSYEQYF, dark purple) was not detected in any sample post-flu (Fig. 3B). The second biggest contributor to the response elicited upon infection (TRBV19-TRBJ2.7 CASSIRSSYEQYF, pink; see Table I) is a public clonotype often described in the literature. This particular CDR3β was encoded by 4 different nucleotide sequences in this patient (Table I) (i.e. convergent recombination), one of which was present before antigen re-exposure. Thus, we observed a clear change in the clonal composition of the T-cell pool recognizing the HLA-A2-$M1_{58-66}$ complex.

Table I – TCR Repertoire of A2-M1$_{58-66}$-specific CD8$^+$ T cells over time.

TRBV	CDR3	TRBJ	-7 wk %(count)	1,5 wk %(count)	4 wk %(count)	9 wk %(count)	48 wk %(count)
19	CASSSRSSYEQYF	2.7	33.33 (16)	0	0	0	0
19	CASSIRSSYEQYF	2.7	10.42 (5)	0	0	0	0
19	CASSIRSSYEQYF	2.7	0	1.30 (1)	0	0	0
19	CASSIRSSYEQYF	2.7	0	12.99 (10)	10.26 (8)	30.12 (25)	6.25 (4)
19	CASSIRSSYEQYF	2.7	0	0	0	0	3.13 (2)
19	CASSIRSAYEQYF	2.7	0	1.30 (1)	0	0	0
19	CASSIRSTYEQYF	2.7	0	0	2.56 (2)	0	0
19	CASSSRSAYEQYF	2.7	0	0	0	1.20 (1)	0
19	CASSQSNQPQHF	1.5	16.67 (8)	0	0	0	0
19	CASSSISNQPQHF	1.5	6.25 (3)	0	0	0	0
19	CASSTRSTDTQYF	2.3	8.33 (4)	0	2.56 (2)	2.41 (2)	0
19	CASSIRSTDTQYF	2.3	0	0	0	0	4.69 (3)
19	CASSMRSTDTQYF	2.3	0	1.30 (1)	0	0	0
19	CASSIRSGDTQYF	2.3	0	2.60 (2)	0	1.20 (1)	0
19	CASSSFGTDTQYF	2.3	0	0	0	1.20 (1)	0
19	CASSGRSADTQYF	2.3	8.33 (4)	0	0	0	0
19	CASSIHGADTQYF	2.3	2.08 (1)	1.30 (1)	0	0	3.13 (2)
19	CASSSRSTGELFF	2.2	0	0	1.28 (1)	1.20 (1)	4.69 (3)
19	CASSGRATGELFF	2.2	0	0	0	2.41 (2)	0
19	CASSARSTGELFF	2.2	0	0	0	1.20 (1)	0
19	CASSIRSTGELFF	2.2	0	0	0	0	1.56 (1)
19	CASSIHSLGEQFF	2.1	0	0	0	1.20 (1)	0
19	CASSIVSYNEQFF	2.1	2.08 (1)	0	0	0	0
19	CASSIRWTALVEQFF	2.1	0	0	0	1.20 (1)	4.69 (3)
19	CASSPSAAGEPQHF	1.5	0	0	1.28 (1)	0	0
19	CASSASGVPNSPLHF	1.6	0	0	0	1.20 (1)	0
20.1	CKRPNGEPQHF	1.5	0	79.22 (61)	80.77 (63)	54.22 (45)	71.88 (46)
9	CASSAGTHSQETQYF	2.5	12.5 (6)	0	0	0	0
7.3	CASSRLGGLAGVDTQYF	2.3	0	0	1.28 (1)	1.20 (1)	0

Discussion

In the present study, we investigated how the functional profile and TCR repertoire of a pre-existing $CD8^+$ T-cell response to an immunodominant influenza-specific epitope evolved upon natural infection. We show that, upon infection, $M1_{58-66}$-responsive cells expanded and displayed i) a transient increase in polyfunctionality; ii) a permanent change in clonal composition, with a shift from a pool dominated by commonly observed TRBV19 clonotypes to one in which a novel TRBV20.1 clonotype is the major constituent.

Peripheral blood samples were collected from an individual showing symptoms characteristic of a flu infection in November 2009, during the flu season. pH1N1(2009) was the main influenza-circulating strain in 2009-2010 (98.9% was influenza A, of which 99.5% was pH1N1(2009)), in contrast to the previous seasons (H3-influenza A, or influenza B) (19). Post-flu samples had a high *in vitro* reactivity to A/California/07/2009 (H1N1) (data not shown). The availability of a sample collected prior to this flu episode allowed us to follow the A2-$M1_{58-66}$-specific $CD8^+$ T cell response upon likely infection with influenza A virus of a different subtype.

In particular, A2-$M1_{58-66}$ post-flu cells analyzed at time points closer to the flu episode had a higher fraction of polyfunctional, effector memory-phenotype cells. Upon recognition of cognate antigen, specific $CD8^+$ T cells were activated and proliferated, while differentiating into a more effector phenotype; when the contraction phase took place, surviving cells that had undergone significant rounds of division upregulated the expression of CD57. A majority of $M1_{58-66}$-specific T cells have been described as $CD45RA^-CD27^+$ when analysed *ex vivo* in healthy donors (46), while cultured $M1_{58-66}$-specific $CD8^+$ T-cell lines, or *in vitro* stimulated virus-specific $IFN\gamma^+$ $CD8^+$ T cells, seem to display a more effector-memory phenotype ($CD45RO^+CD62L^{lo}CCR7^-$ or $CD45RA^-CCR7^-$, respectively; (8, 47)). In contrast, Roos and colleagues (48) have described an increase in the number of $CD45RO^+CD27^+CD8^+$ T cells during the acute phase of HIV and EBV infections, though these are viruses that establish a persistent infection in the host. The ability to react with a broader range of functions after infection likely reflects recent contact with the antigen in an inflammatory environment: upon re-infection, cells 'turn-on' their ability to respond to the antigen with a wider range of functions, and slowly 'shut it off' when antigen is cleared. In accordance, Wagar et al. (49) showed, also in an infected individual, a similarly transient increase in $IFN\gamma^+GzB^+CD107^+$ virus-specific cells. This suggests that circulating cells that have not seen their cognate antigen in an inflammatory context for a longer period of time might be more senescent as a

pool (higher percentage of $CD57^+$tetramer$^+$ cells), and thus react less functionally in an *in vitro* re-stimulation.

In contrast to the transient changes in phenotype and functional profiles, the alteration in the responding TCR repertoire was permanent. Using an unbiased approach, we identified a novel Vβ clonotype/family recognizing A2-$M1_{58-66}$ that replaced the typical TRBV19 clonotypes as main contributor to the specific response upon re-infection. This non-TRBV19 clonotype remained dominant in the long-term memory pool. The gene usage of the A2-$M1_{58-66}$-restricted $CD8^+$ T cell repertoire has been historically described as very narrow, biased, predominantly made of TRBV19 clonotypes, frequently possessing RS-containing CDR3βs of 11 amino-acid length, and often shared among individuals, i.e.: public (30-36, 50-52). This dominance of TRBV19 clonotypes was suggested to relate to a higher availability of thymocytes bearing $TCR\beta19^+$ 'RS'-CDR3β-carrying sequences than expected by chance (53). Though the prevalence of TRBV19 clonotypes in $CD8^+$ T cell responses to A2-$M1_{58-66}$ is undeniable, the traditional way of identification of antigen-specific cells, i.e. *in vitro* culture of clones and $CD8^+$ T-cell lines, may have overlooked the involvement of other Vβ families. More recently, TRBV19 family-specific primers have also been used to discriminate Vβ clonotypes targeting A2-$M1_{58-66}$-restricted $CD8^+$ T cells (52, 54, 55). However, subdominant Vβ families can be detected directly *ex vivo* or upon *in vitro* culturing, though their contribution to the total pool decreased and/or remained minimal upon culture (30, 31, 50, 51, 56), suggesting $TRBV19^+$ cells might have a competitive advantage *in vitro*. A recent T-cell vaccine study (MVA-NP+M1, in which the expressed influenza proteins are derived from A/Panama/2007/99 (H3N2)), making use of the same unbiased TCR clonotyping approach, reported a polyclonal, TRBV19-dominated A2-$M1_{58-66}$-repertoire in 5 subjects (57). Nevertheless, and in contrast to our patient, in the 3 vaccinees analyzed pre- and post-vaccination, the pre-vaccination dominant clonotype remained major. Interestingly, one of the individuals had a non-TRBV19 clonotype (TRBV27) as major contributor to the A2-$M1_{58-66}$-specific pool, confirming that, even though TRBV19 clonotypes are more widely found in memory responses to the A2-$M1_{58-66}$-complex, other clonotypes may be hierarchically superior.

As we were not able to isolate the infecting virus strain, we cannot exclude that it carried a variant epitope, which could have elicited this shift in responding TCR repertoire. However, this epitope is believed to be highly conserved, due to structural and/or functional constraints in M1, as mutant viruses have poorer replicative fitness (58). Moreover, we could only find two epitope variants (P1 G>A; P5 F>L) in the 994 full-length human influenza-A Matrix1 sequences collected during both 2009 and 2010 in Europe, and made available in the NCBI Influenza Virus resource database

(http://www.ncbi.nlm.nih.gov/genomes/FLU/Database; data not shown). Even in a persistently infected, immunocompromised infant this epitope was conserved, while other epitopes showed sequential evolution of mutations (25). This highlights the high degree of conservation of the $M1_{58-66}$ epitope, and the improbability of the presence of a variant. Thus, despite the fact that the targeted A2-$M1_{58-66}$ epitope is very constrained and thus most often conserved, the responding $CD8^+$ T-cell repertoire had the plasticity to change, even in the presence of several previously described public clonotypes. Seven of the TRBV19 clonotypes we observed were also found in the subjects of the abovementioned vaccine study (57). Among them were the TRBV19 clonotypes CASSIRSSYEQYF and CASSSRSSYEQYF, frequently reported in other studies (e.g.: (31, 32, 56, 57)). The latter made up the biggest fraction of the pre-infection repertoire, but probably failed to expand, as it was not detected in any of the post-infection samples.

In conclusion, this study suggests that the functional and phenotypic features of the A2-$M1_{58-66}$ flu-specific changed transiently upon infection, while the T-cell repertoire may change more permanently. The transitory improvement in functional performance of the A2-$M1_{58-66}$-responding $CD8^+$ T cells is likely a natural change occurring upon acute infection, though the presence of the new TRBV20.1 clonotype may have contributed. The shift in TCR repertoire usage of responsive $CD8^+$ T cells after antigen re-encounter highlights the need for an unbiased approach when analysing potential changes in T-cell repertoires elicited by natural infection. Despite the strong prevalence of TRBV19 clonotypes, cells from other Vβ families, thus far potentially underestimated and scarcely studied, may contribute significantly to the HLA-A2-$M1_{58-66}$ response and potentially play a major role *in vivo*.

Acknowledgements

We thank Gerrit Spierenburg, Koos Gaiser and Dan Koning for technical assistance. This work was supported by a High Potential (HiPo) grant from the University of Utrecht. A.I. Costa was also supported by a fellowship from the Fundação para a Ciência e Tecnologia (FCT, Portugal).

References

1. Kreijtz JH, Fouchier RA, and Rimmelzwaan GF. Immune responses to influenza virus infection. *Virus Res.* 2011;162(1-2):19-30.

2. Gras S, Kedzierski L, Valkenburg SA, Laurie K, Liu YC, Denholm JT, Richards MJ, Rimmelzwaan GF, Kelso A, Doherty PC, et al. Cross-reactive CD8+ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. *Proc Natl Acad Sci U S A.* 2010;107(28):12599-604.

3. Greenbaum JA, Kotturi MF, Kim Y, Oseroff C, Vaughan K, Salimi N, Vita R, Ponomarenko J, Scheuermann RH, Sette A, et al. Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population. *Proc Natl Acad Sci U S A.* 2009;106(48):20365-70.

4. Hillaire ML, Vogelzang-van Trierum SE, Kreijtz JH, de Mutsert G, Fouchier RA, Osterhaus AD, and Rimmelzwaan GF. Human T-cells directed to seasonal influenza A virus cross-react with 2009 pandemic influenza A (H1N1) and swine-origin triple-reassortant H3N2 influenza viruses. *J Gen Virol.* 2013;94(Pt 3):583-92.

5. Kreijtz JH, de Mutsert G, van Baalen CA, Fouchier RA, Osterhaus AD, and Rimmelzwaan GF. Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. *J Virol.* 2008;82(11):5161-6.

6. Lee LY, Ha do LA, Simmons C, de Jong MD, Chau NV, Schumacher R, Peng YC, McMichael AJ, Farrar JJ, Smith GL, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. *The Journal of clinical investigation.* 2008;118(10):3478-90.

7. Quinones-Parra S, Grant E, Loh L, Nguyen TH, Campbell KA, Tong SY, Miller A, Doherty PC, Vijaykrishna D, Rossjohn J, et al. Preexisting CD8+ T-cell immunity to the H7N9 influenza A virus varies across ethnicities. *Proc Natl Acad Sci U S A.* 2014;111(3):1049-54.

8. Tu W, Mao H, Zheng J, Liu Y, Chiu SS, Qin G, Chan PL, Lam KT, Guan J, Zhang L, et al. Cytotoxic T lymphocytes established by seasonal human influenza cross-react against 2009 pandemic H1N1 influenza virus. *J Virol.* 2010;84(13):6527-35.

9. van de Sandt CE, Kreijtz JH, de Mutsert G, Geelhoed-Mieras MM, Hillaire ML, Vogelzang-van Trierum SE, Osterhaus AD, Fouchier RA, and Rimmelzwaan GF. Human cytotoxic T lymphocytes directed to seasonal influenza A viruses cross-react with the newly emerging H7N9 virus. *J Virol.* 2014;88(3):1684-93.

10. Zweerink HJ, Courtneidge SA, Skehel JJ, Crumpton MJ, and Askonas BA. Cytotoxic T cells kill influenza virus infected cells but do not distinguish between serologically distinct type A viruses. *Nature.* 1977;267(5609):354-6.

11. Forrest BD, Pride MW, Dunning AJ, Capeding MR, Chotpitayasunondh T, Tam JS, Rappaport R, Eldridge JH, and Gruber WC. Correlation of cellular immune responses with protection against culture-confirmed influenza virus in young children. *Clinical and vaccine immunology : CVI.* 2008;15(7):1042-53.

12. McElhaney JE, Xie D, Hager WD, Barry MB, Wang Y, Kleppinger A, Ewen C, Kane KP, and Bleackley RC. T cell responses are better correlates of vaccine protection in the elderly. *J Immunol.* 2006;176(10):6333-9.

13. McMichael AJ, Gotch FM, Noble GR, and Beare PA. Cytotoxic T-cell immunity to influenza. *N Engl J Med.* 1983;309(1):13-7.

14. Sridhar S, Begom S, Bermingham A, Hoschler K, Adamson W, Carman W, Bean T, Barclay W, Deeks JJ, and Lalvani A. Cellular immune correlates of protection against symptomatic pandemic influenza. *Nat Med.* 2013;19(10):1305-12.

15. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science.* 2009;325(5937):197-201.

16. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature.* 2009;459(7250):1122-5.

17. Peiris JS, Tu WW, and Yen HL. A novel H1N1 virus causes the first pandemic of the 21st century. *Eur J Immunol.* 2009;39(11):2946-54.

18. Valkenburg SA, Rutigliano JA, Ellebedy AH, Doherty PC, Thomas PG, and Kedzierska K. Immunity to seasonal and pandemic influenza A viruses. *Microbes Infect.* 2011;13(5):489-501.

19. WHO. Situation update in the European region: overview of influenza surveillance data week 40/2009 to 07/2010. 2010.

20. WHO. Summary of the 1st post-pandemic influenza season in the WHO European region: 2010-2011. 2011.

21. Wielders CC, van Lier EA, van 't Klooster TM, van Gageldonk-Lafeber AB, van den Wijngaard CC, Haagsma JA, Donker GA, Meijer A, van der Hoek W, Lugner AK, et al. The burden of 2009 pandemic influenza A(H1N1) in the Netherlands. *European journal of public health.* 2012;22(1):150-7.

22. Grant EJ, Chen L, Quinones-Parra S, Pang K, Kedzierska K, and Chen W. T-cell immunity to influenza A viruses. *Critical reviews in immunology.* 2014;34(1):15-39.

23. Epstein SL. Prior H1N1 influenza infection and susceptibility of Cleveland Family Study participants during the H2N2 pandemic of 1957: an experiment of nature. *J Infect Dis.* 2006;193(1):49-53.

24. Hillaire ML, Osterhaus AD, and Rimmelzwaan GF. Induction of virus-specific cytotoxic T lymphocytes as a basis for the development of broadly protective influenza vaccines. *J Biomed Biotechnol.* 2011;2011:939860.

25. Valkenburg SA, Quinones-Parra S, Gras S, Komadina N, McVernon J, Wang Z, Halim H, Iannello P, Cole C, Laurie K, et al. Acute emergence and reversion of influenza A virus quasispecies within CD8+ T cell antigenic peptides. *Nature communications.* 2013;4:2663.

26. Hertz T, Oshansky CM, Roddam PL, DeVincenzo JP, Caniza MA, Jojic N, Mallal S, Phillips E, James I, Halloran ME, et al. HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection. *Proc Natl Acad Sci U S A.* 2013;110(33):13492-7.

27. Liu J, Wu B, Zhang S, Tan S, Sun Y, Chen Z, Qin Y, Sun M, Shi G, Wu Y, et al. Conserved epitopes dominate cross-CD8+ T-cell responses against influenza A H1N1 virus among Asian populations. *Eur J Immunol.* 2013;43(8):2055-69.

28. Assarsson E, Bui HH, Sidney J, Zhang Q, Glenn J, Oseroff C, Mbawuike IN, Alexander J, Newman MJ, Grey H, et al. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. *J Virol.* 2008;82(24):12241-51.

29. Plotnicky H, Cyblat-Chanal D, Aubry JP, Derouet F, Klinguer-Hamour C, Beck A, Bonnefoy JY, and Corvaia N. The immunodominant influenza matrix T cell epitope recognized in human induces influenza protection in HLA-A2/K(b) transgenic mice. *Virology.* 2003;309(2):320-9.

30. Lawson TM, Man S, Williams S, Boon AC, Zambon M, and Borysiewicz LK. Influenza A antigen exposure selects dominant Vbeta17+ TCR in human CD8+ cytotoxic T cell responses. *Int Immunol.* 2001;13(11):1373-81.

31. Lehner PJ, Wang EC, Moss PA, Williams S, Platt K, Friedman SM, Bell JI, and Borysiewicz LK. Human HLA-A0201-restricted cytotoxic T lymphocyte recognition of influenza A is dominated by T cells bearing the V beta 17 gene segment. *J Exp Med.* 1995;181(1):79-91.

32. Moss PA, Moots RJ, Rosenberg WM, Rowland-Jones SJ, Bodmer HC, McMichael AJ, and Bell JI. Extensive conservation of alpha and beta chains of the human T-cell antigen receptor recognizing HLA-A2 and influenza A matrix peptide. *Proc Natl Acad Sci U S A.* 1991;88(20):8987-90.

33. Prevost-Blondel A, Lengagne R, Letourneur F, Pannetier C, Gomard E, and Guillet JG. In vivo longitudinal analysis of a dominant TCR repertoire selected in human response to influenza virus. *Virology.* 1997;233(1):93-104.

34. Miles JJ, Douek DC, and Price DA. Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunology and cell biology.* 2011;89(3):375-87.

35. Stewart-Jones GB, McMichael AJ, Bell JI, Stuart DI, and Jones EY. A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol.* 2003;4(7):657-63.

36. Turner SJ, Doherty PC, McCluskey J, and Rossjohn J. Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol.* 2006;6(12):883-94.

37. Altman JD, Moss PA, Goulder PJ, Barouch DH, McHeyzer-Williams MG, Bell JI, McMichael AJ, and Davis MM. Phenotypic analysis of antigen-specific T lymphocytes. *Science.* 1996;274(5284):94-6.

38. Kostense S, Ogg GS, Manting EH, Gillespie G, Joling J, Vandenberghe K, Veenhof EZ, van Baarle D, Jurriaans S, Klein MR, et al. High viral burden in the presence of major HIV-specific CD8(+) T cell expansions: evidence for impaired CTL effector function. *Eur J Immunol.* 2001;31(3):677-86.

39. Koning D, Costa AI, Hoof I, Miles JJ, Nanlohy NM, Ladell K, Matthews KK, Venturi V, Schellens IM, Borghans JA, et al. CD8+ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex. *J Immunol.* 2013;190(3):931-9.

40. Quigley MF, Almeida JR, Price DA, and Douek DC. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. *Curr Protoc Immunol.* 2011;Chapter 10:Unit 10.33.

41. Lefranc MP. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* 2001;29(1):207-9.

42. Heidema J, Rossen JW, Lukens MV, Ketel MS, Scheltens E, Kranendonk ME, van Maren WW, van Loon AM, Otten HG, Kimpen JL, et al. Dynamics of human respiratory virus-specific CD8+ T cell responses in blood and airways during episodes of common cold. *J Immunol.* 2008;181(8):5551-9.

43. Hillaire ML, van Trierum SE, Bodewes R, van Baalen CA, van Binnendijk RS, Koopmans MP, Fouchier RA, Osterhaus AD, and Rimmelzwaan GF. Characterization of the human CD8(+) T cell response following infection with 2009 pandemic influenza H1N1 virus. *J Virol.* 2011;85(22):12057-61.

44. Subbarao K, Murphy BR, and Fauci AS. Development of effective vaccines against pandemic influenza. *Immunity.* 2006;24(1):5-9.

45. Wilkinson TM, Li CK, Chui CS, Huang AK, Perkins M, Liebner JC, Lambkin-Williams R, Gilbert A, Oxford J, Nicholas B, et al. Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. *Nat Med.* 2012;18(2):274-80.

46. Hoji A, and Rinaldo CR, Jr. Human CD8+ T cells specific for influenza A virus M1 display broad expression of maturation-associated phenotypic markers and chemokine receptors. *Immunology.* 2005;115(2):239-45.

47. Sridhar S, Begom S, Bermingham A, Ziegler T, Roberts KL, Barclay WS, Openshaw P, and Lalvani A. Predominance of heterosubtypic IFN-gamma-only-secreting effector memory T cells in pandemic H1N1 naive adults. *Eur J Immunol.* 2012;42(11):2913-24.

48. Roos MT, van Lier RA, Hamann D, Knol GJ, Verhoofstad I, van Baarle D, Miedema F, and Schellekens PT. Changes in the composition of circulating CD8+ T cell subsets during acute epstein-barr and human immunodeficiency virus infections in humans. *J Infect Dis.* 2000;182(2):451-8.

49. Wagar LE, Rosella L, Crowcroft N, Lowcock B, Drohomyrecky PC, Foisy J, Gubbay J, Rebbapragada A, Winter AL, Achonu C, et al. Humoral and cell-mediated immunity to pandemic H1N1 influenza in a Canadian cohort one year post-pandemic: implications for vaccination. *PLoS One.* 2011;6(11):e28063.

50. Lawson TM, Man S, Wang EC, Williams S, Amos N, Gillespie GM, Moss PA, and Borysiewicz LK. Functional differences between influenza A-specific cytotoxic T lymphocyte clones expressing dominant and subdominant TCR. *Int Immunol.* 2001;13(11):1383-90.

51. Naumov YN, Naumova EN, Clute SC, Watkin LB, Kota K, Gorski J, and Selin LK. Complex T cell memory repertoires participate in recall responses at extremes of antigenic load. *J Immunol.* 2006;177(3):2006-14.

52. Naumova EN, Gorski J, and Naumov YN. Two compensatory pathways maintain long-term stability and diversity in CD8 T cell memory repertoires. *J Immunol.* 2009;183(4):2851-8.

53. Yassai M, Bosenko D, Unruh M, Zacharias G, Reed E, Demos W, Ferrante A, and Gorski J. Naive T cell repertoire skewing in HLA-A2 individuals by a specialized rearrangement mechanism results in public memory clonotypes. *J Immunol.* 2011;186(5):2970-7.

54. Petrova GV, Naumova EN, and Gorski J. The polyclonal CD8 T cell response to influenza M158-66 generates a fully connected network of cross-reactive clonotypes to structurally related peptides: a paradigm for memory repertoire coverage of novel epitopes or escape mutants. *J Immunol.* 2011;186(11):6390-7.

55. Zhou V, Yassai MB, Regunathan J, Box J, Bosenko D, Vashishath Y, Demos W, Lee F, and Gorski J. The functional CD8 T cell memory recall repertoire responding to the influenza A M1(58-66) epitope is polyclonal and shows a complex clonotype distribution. *Human immunology.* 2013;74(7):809-17.

56. Naumov YN, Hogan KT, Naumova EN, Pagel JT, and Gorski J. A class I MHC-restricted recall response to a viral peptide is highly polyclonal despite stringent CDR3 selection: implications for establishing memory T cell repertoires in "real-world" conditions. *J Immunol.* 1998;160(6):2842-52.

57. Antrobus RD, Lillie PJ, Berthoud TK, Spencer AJ, McLaren JE, Ladell K, Lambe T, Milicic A, Price DA, Hill AV, et al. A T cell-inducing influenza vaccine for the elderly: safety and immunogenicity of MVA-NP+M1 in adults aged over 50 years. *PLoS One.* 2012;7(10):e48322.

58. Berkhoff EG, de Wit E, Geelhoed-Mieras MM, Boon AC, Symons J, Fouchier RA, Osterhaus AD, and Rimmelzwaan GF. Functional constraints of influenza A virus epitopes limit escape from cytotoxic T lymphocytes. *J Virol.* 2005;79(17):11239-46.

8

General Discussion

General Discussion

The major loci of the HLA class I complex are the highly polymorphic HLA-A, B and C. Even though the HLA-B locus is evolutionarily younger than HLA-A, it possesses the largest number of alleles in the population (see Fig. 3 in Introduction) and is the one most rapidly diversifying (1-3). A stronger pathogen-induced selective pressure on the HLA-B locus could drive this higher increase in the number of alleles in a shorter timeframe, relative to HLA-A. Indeed, HLA-B alleles seem to be frequently associated with the outcome (resistance or susceptibility) of infectious diseases (3-15), and the T-cell responses restricted to them are often immunodominant (more frequently targeted – at the population level – and/or of higher magnitude – within an individual) (3, 14-19). HLA alleles restricting immunodominant T-cell responses likely impact disease outcome more significantly, and those increasing the host's fitness will be kept in the genetic pool of the population. In the first part of this thesis, we performed an in depth comparison of functional properties of HLA-B and HLA-A molecules, in order to assess whether they are diverging from each other on the functional level, despite their structural similarities. We did so by comparing their peptide repertoires with respect to diversity, binding affinity (chapter 2), degree of conservation (chapter 3), and binding promiscuity (chapter 4). To do this we worked with *in silico* predicted and experimentally identified peptide repertoires. In the second part, we made use of known, frequently targeted viral epitopes, and analyzed features of the T-cell responses they elicit, including the T-cell repertoire diversity of specific responses against latent, chronic and acute infections (chapters 5, 6, and 7).

Do HLA-B molecules differ from HLA-A in the features of the epitopes they present?

HLA class I molecules have the important task of mirroring the translational output of a cell, either by presenting peptides derived from self-proteins and confirming the healthy status of the cell, or by presenting peptides derived from a pathogenic proteome, and thus targeting it for destruction. As such, some features of the peptide repertoire presented by different HLA molecules may contribute to the dominance of certain immune responses, for example via an HLA's preference for presenting pathogen-derived peptides at higher diversity, binding affinity or copy number, or a tendency to present more conserved regions of a pathogen proteome. Below, we summarize the observed impact of these factors, described in the first chapters of this thesis.

Diversity of the presented peptide repertoire

As mentioned above, HLA-B alleles may diversify faster if under stronger selection pressure, to keep up with rapidly evolving pathogens. One mechanism could be that HLA-B molecules tend to be specialized at the presentation of pathogen-derived peptides, while HLA-A molecules retain the function of sampling intracellular proteins and flagging the cell as self. The 'footprints' of this road to specialization could be in the source of presented epitopes: higher presentation of human-derived peptides by HLA-A and of pathogen-derived peptides by HLA-B. Predominant presentation of pathogen-derived peptides by HLA-B alleles could cause more HLA-B-restricted $CD8^+$ T-cell responses to be immunodominant. However, when analyzing both experimentally defined and *in silico* predicted human and pathogen epitope sets, we found, rather counterintuitively, that HLA-A presents a higher fraction of both human- and pathogen-derived peptides than HLA-B (chapter 2). The finding that HLA-B molecules have a more restricted binding motif can contribute to this result (chapter 2). Paul and colleagues (20) have confirmed that HLA-A alleles are, in general, associated with broader peptide repertoires. HLA-A alleles might actually present more pathogen-derived peptides due to a widespread preference for amino acids coded by low G+C content-codons, a common feature of the genome of pathogenic agents (21), although this offers no explanation for the higher presentation of self. Thus, the frequently reported immunodominance of HLA-B-restricted responses seems neither due to a difference in the source of proteins sampled (self *vs* non-self), nor to a higher diversity of pathogen-derived peptides.

Alternatively, we hypothesized that HLA-B-restricted epitopes may bind with higher affinity than HLA-A-restricted epitopes. In principle, a higher binding affinity would allow more stable pHLA complexes at the surface of cells, and thus increase the chance of detection by specific $CD8^+$ T cells. HLA-B-restricted T-cell responses would therefore have a higher probability to become immunodominant. However, and again unexpectedly, we observed that HLA-B actually binds peptides with a lower binding affinity than HLA-A (chapter 2). In accordance, Granados and colleagues have reported that peptides eluted from HLA-B molecules were predicted to bind with lower affinity than those eluted from HLA-A (22).

The results summarized so far made use of predicted epitopes, or epitopes for which the binding affinity has been measured in *in vitro* assays. An important difference between HLA-A and -B epitopes could lie in how efficiently they are generated in a cell. To compare the naturally processed peptide repertoires of HLA-A and –B molecules, we analyzed the peptidome of 4 distinct B-cell lines co-expressing HLA alleles of different supertypes (i.e.: different binding motif of co-expressed alleles) in an elution

study (Fig. 1). We found that the set of peptides conforming to the HLA-B-binding motif was more diverse than that of HLA-A alleles, and was presented at similar copy numbers per peptide (Schellens et al, submitted). Other studies on naturally processed ligands have also reported a higher number of peptides restricted by HLA-B than HLA-A (22-25), and have suggested that peptides presented at higher densities do not necessarily elicit a T-cell response of higher magnitude (26, 27).

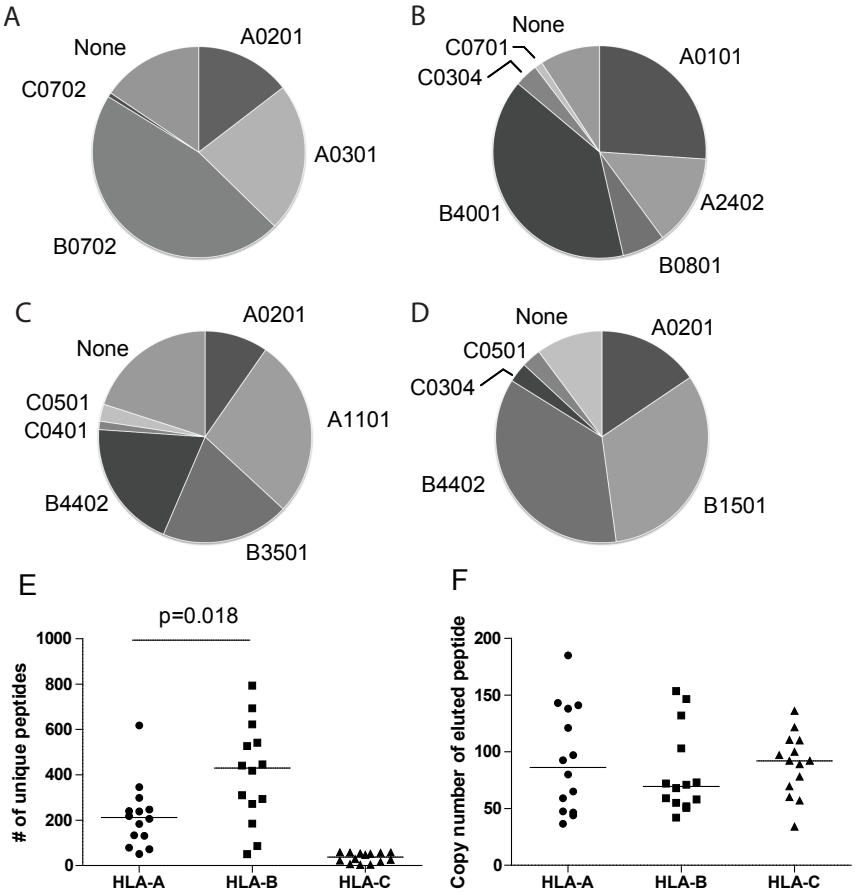


Figure 1. HLA-B alleles present a more diverse peptide repertoire than HLA-A alleles – the 'eluted' peptidome of B-cell lines (BLCLs).
The distribution of unique self-peptides eluted from uninfected cells is shown from four different BLCLs (A-D). Each slice of the pie represents the fraction of unique peptides eluted from that specific BLCL presented by the specific HLA molecule. Peptides that could not be reliably assigned to any of the HLA molecules expressed are depicted as None. (E) represents the number of unique self peptides eluted from HLA-A, HLA-B or HLA-C molecules. Each dot represents one HLA molecule on one BLCL in one infection state. In (F) the median copy number of eluted peptides is depicted per HLA molecule per BLCL per infection state (Schellens et al., submitted).

At first glance, the abovementioned analyses of HLA ligands predicted *in silico* or reported as binders in the Immune epitope database (IEBD) seem to contradict the results of the *in vitro* elution studies: while the former point toward HLA-A alleles presenting a more diverse peptide repertoire, elution studies indicate the opposite. However, the findings are not necessarily mutually exclusive. The total pool of potential binders can be higher for HLA-A alleles (chapter 2), yet the fraction of peptides presented at the cell surface can still be higher for HLA-B alleles. In fact, from our elution study one can infer that, to accommodate a more varied peptide repertoire at similar density per ligand, HLA-B molecules should be expressed in higher amounts at the cell surface than HLA-A molecules. Studies quantifying the expression levels of HLA alleles are, however, not conclusive on whether HLA-A or B molecules are expressed at higher levels ((23, 28-30), Keşmir et al, unpublished).

There are at least two scenarios that may account for a higher number of pHLA-B complexes at the cell surface: i) the extent of peptide editing, and ii) the stability of the pHLA complexes. HLA-A molecules seem to associate more efficiently with the peptide-loading complex than HLA-B molecules (31), thus tapasin editing is believed to be more efficient for pHLA-A. If there is a bigger pool of potential binders, but the optimal binding affinity is higher than what is required for HLA-B alleles, peptide editing and loading might allow fewer pHLA-A complexes to be presented at the cell surface. If this is the case, HLA-B molecules, with a more permissive binding affinity range and lower editing, might be able to reach the surface at higher densities. Alternatively, the higher numbers of pHLA complexes at the cell surface may be the result of a higher stability (slower off-rate).

On the other hand, the apparent inconsistencies between the *in silico* and elution studies may be due to a bias in the two sources of data: ligands defined by binding affinity to a wide group of HLA-alleles (predicted or determined *in vitro*) present in the human population *versus* naturally processed and presented ligands in the context of a given HLA background consisting of a maximum of two HLA-A and two HLA-B alleles. It is important to note that HLA-A peptide repertoires have been studied more extensively, and this may contribute to the abovementioned discrepancy. The most studied allele is HLA-A*02, because it is a highly prevalent allele. Investigation of the peptide repertoire of highly prevalent alleles is fuelled by T-cell vaccine design studies aiming at maximal population coverage. As such, there is more binding data available for HLA-A alleles in the IEDB database (results not shown). As the IEDB data is used as the training set for binding predictors, performance of HLA-A predictors may be more robust than that of HLA-B predictors. After all, the absolute predicted binding affinity, or the ranking (relative binding affinity) of the ligands define the threshold to

discriminate binders from non-binders. Alternatively, a bias induced by the elution methodology could contribute to the disparate results of the *in silico* and elution studies. Considering that pHLA-B complexes may be more stable, they may be less prone to dissociate during the immunoprecipitation of complexes, and thus more HLA-B-restricted peptides may ultimately be eluted. Also, a broad anti-HLA antibody (i.e.: W6/32) is often used for immunoprecipitation of pHLA complexes: one cannot exclude that the antibody could have different reactivity to HLA-A *versus* HLA-B molecules, and the ratio of precipitated complexes would consequently be influenced. Unfortunately, there is no reasonable alternative on highly specific anti-HLA antibodies for different alleles, and of equivalent avidity for antigen. A final source of bias could result from the combination of the analysis of eluted peptides and prediction of their binding. We have shown that promiscuous binding of peptides occurs even between different HLA loci (chapter 4). We have purposely chosen for B-cell lines in which co-expressed alleles have distinct binding motifs so the attribution of the restricting allele using prediction methods will most likely be accurate. Still, for the cases in which a given peptide is predicted to bind to two HLA alleles, quantification of the differential stability of the two complexes (see high throughput method in (32)) will be a reliable way of assigning the most correct restriction element. It is yet unknown whether promiscuously binding peptides will actually be equally stable when complexed with different HLA molecules. This raises another point of complexity, that stability of pHLA complexes may be a better predictor of immunogenicity than binding affinity (32-36). It is, therefore, important to compare the stability of pHLA-A *versus* pHLA-B complexes, as this information might prove invaluable to confirm the potential higher densities of pHLA-B complexes at the cell surface, and its implication for the dominance of $CD8^+$ T-cell responses.

Infection: adding to the complexity

The ratio of expression levels of pHLA-B *versus* pHLA-A at the cell surface in 'steady state' is still not clear. Viral infection adds yet another level of complexity. In fact, viruses have been shown to interfere with HLA presentation by diverse mechanisms, which include hijacking HLAs from the cell surface and targeting them for destruction, and interfering with TAP-transport and peptide loading (37, 38). The net ratio of pathogen-derived pHLA complexes at the cell surface will depend on the balance of stimuli that support their upregulation (e.g.: cytokines) and those that promote downregulation as a part of specific pathogen evasion strategies. Expression of HLA-B molecules has been shown to respond more strongly to cytokines like IFNγ and IFNβ and TNFα than HLA-A, as measured by a more pronounced increase in mRNA levels

(29). Furthermore, the human immunodeficiency virus-1 (HIV-1) viral protein Nef, which down-regulates HLA class I surface expression, has been shown to have a more pronounced effect on HLA-A alleles compared to HLA-B (39). For the Epstein-Barr virus (EBV), the viral protein BILF1 also seems to differentially decrease the expression of distinct HLA alleles (40), while the human cytomegalovirus (HCMV) US2 may down-regulate all HLA-A molecules but only some HLA-B alleles (41). Finally, HLA-B molecules (and C molecules), more than HLA-A, are known ligands for NK cells. (42). As a consequence, pathogens in general are caught between 'a rock and a hard place': they down-regulate HLA-B molecules and trigger NK cells, or if they do not, $CD8^+$ T cells will detect them via recognition of pHLA-B complexes. A less pronounced down-regulation of HLA-B, as compared to HLA-A, may be a compromise to ensure that persistent viruses partially evade both NK and $CD8^+$ T cells. The abovementioned data suggest that HLA-B alleles may be more refractory to down-regulation, and more susceptible to 'infection-induced' cytokines. If this is the case, it might allow for a higher ratio of pHLA-B/pHLA-A complexes upon infection, a scenario that clearly needs further investigation.

Thus, the impact of a pathogen's evasion strategy may relate to the particular HLA associations with resistance/susceptibility to (progression of) infection and may partly explain the immunodominance of HLA-B-restricted T-cell responses.

In conclusion, despite the fact that a lot is known on antigen processing and presentation, there is still much to explore about the relative surface expression of HLA molecules under normal (uninfected) circumstances, and the changes that occur upon (viral) infection, such as the impact of pathogen evasion mechanisms. This information is necessary for interpreting peptidome diversity and abundance in health and changes upon (viral) disease. As the probability of epitope presentation correlates strongly with protein abundance, analysis of the peptidome will be most informative when sampled at different time points after infection (26, 43), adapted to the pathogen's replicative cycle. Although very demanding, epitope presentation should be put into context of the time-span of infection. Finally, most of the peptidome data has been obtained from B-cell lines, given the high numbers of cells needed for such analyses. B cells constitutively express immunoproteasomes, but non-immune cells will express a mix of constitutive and immunoproteasomes upon infection, with different cleavage motifs that affect the processed repertoire (44-46). Although the peptides presented by B-cell lines might resemble the ones presented by dendritic cells, the antigen-presenting cells of excellence, as they also constitutively express immunoproteasomes, it would be interesting to analyze the net differences in host peptidome in uninfected *versus* infected

non-immune cells, especially the targets of viral replication *in vivo*. This way, more parallels can be drawn between what $CD8^+$ T cells see on infected, target-organ cells, in comparison to antigen-presenting cells.

Do HLA-B molecules target more conserved epitopes?

Some viruses, including HIV and HCV, have fast replication kinetics, and an error-prone replication of the genetic material due to lack of proofreading of the viral polymerase. These highly mutable viruses are, thus, present in the host as a quasi-species, i.e.: as a population of viral mutants (47). When mutations in epitopes occur, and the epitopes can no longer be presented to or be recognized by cognate T-cell receptors (TCRs), the respective T cells become ineffective and redundant. This is one of the mechanisms exploited by persistent viruses to evade the immune system, and it has been widely exemplified in the literature. Not all regions of the virus proteomes are as easily mutable, causing some epitopes to be more conserved than others. Indeed, some mutations cause a high fitness cost on viral replication and, therefore, either revert in the absence of the original immune pressure upon transmission to HLA-mismatched hosts (48-55), or become fixed in the viral quasispecies upon accumulation of compensatory mutations (56-61). These regions of high vulnerability are, theoretically at least, the ones that should be targeted by the immune system in order to cripple rapidly evolving (i.e.: mutating) viruses.

In chapter 3, we have shown that HLA-B-restricted $CD8^+$ T-cell responses target more conserved regions of HIV-1 than HLA-A-restricted responses. In addition, mutations in HLA-B regions revert more often than mutations in HLA-A regions when the immune pressure drops in HLA-mismatched recipients (after transmission), underlining the selective pressure exerted by cognate responses triggered by pHLA-B complexes (chapter 3). In fact, HLA-B molecules, and especially the ones associated with delayed disease progression (HLA-B27 and B57) were even shown to drive adaptation of HIV to the host immune response (62).

This preferential targeting of conserved residues by HLA-B may also occur for another rapidly evolving virus that often remains persistent in the human host, the hepatitis C virus (HCV). We showed that HLA-B27 and B57, two HLA alleles associated with efficient clearance of HCV-infection, also have a preference to present peptides from the most conserved HCV proteins, core and NS5b (Rao et al, unpublished). It remains unclear whether in HCV infection HLA-B molecules in general have a stronger preference to present conserved viral peptides when compared to HLA-A molecules, or whether this feature is unique to the HLA-B molecules conferring best protection. Still, mutations in HLA-B-targeted HCV epitopes that revert during a stage of higher

immune 'tolerance' (i.e.: pregnancy) have been recently reported (63). This recent study on HCV evolution during and after pregnancy highlights how mutations in constrained regions impact virion production, and how their appearance and reversion correlates to the reactivity of autologous CTL lines to wild-type and variant epitopes. The epitopes undergoing mutation and reversion were restricted mainly by the HLA-B alleles B*5101, B*0801 and B*1501. However, additional mutations described in this study were present in regions where no CTL response has thus-far been found, and therefore, more studies are necessary to conclude that also during HCV infection, the epitopes presented by HLA-B molecules are under stronger selection pressure than those presented by HLA-A.

The reports above suggest that the preferential targeting of constrained regions of rapidly evolving viruses by HLA-B might be a general difference between HLA-A and HLA-B molecules. Although we cannot pinpoint the underlying reason, less easily mutable amino acids are more often present in the binding motifs of HLA-B molecules. However, additional data from HCV and other highly mutating persistent viruses is needed to determine if the targeting of more constrained regions by HLA-B molecules is a general feature that distinguishes them from HLA-A molecules. By targeting more constrained regions in the proteome of persistent, rapidly evolving viruses, HLA-B-restricted responses may be the ones better preserved and hence be detected as immunodominant.

The other side of the coin: the CD8$^+$ T cell populations recognizing pathogen-derived peptide-HLA class I complexes

As discussed above, several features of the epitope repertoire presented by MHC molecules at the cell surface can influence the immunodominance hierarchy of the reactive CD8$^+$ T cell responses. Whether or not immunodominance is directly related to T-cell precursor frequencies has been a motive of controversy (64-69). It goes without saying, however, that the presence of antigen-specific T cells in the T-cell repertoire is a key constraint for any epitope to be recognized. In principle, the thymus will have selected for the T cells with TCRs that do not react excessively to self antigens, but will recognize non-self, pathogen-derived pHLA-I complexes with enough affinity to set the signaling cascade in motion, resulting in proliferation and effector function of specific T cells. CD8$^+$ T-cell responses are responsible for killing virus-infected cells, limiting viral replication and decreasing dissemination, and thereby avoiding, or delaying, disease progression. The main goal of studying CD8$^+$ T cells is, thus, to establish the features of a T-cell response that allow fast elimination of infected cells and swift control of viral

replication. This knowledge can then be applied for the design of vaccines to diseases where antibodies able to neutralize the infectivity of the virus in the host are quickly surpassed by viral mutation (HIV-1, Influenza, and others). The main goal of a T-cell based vaccine is, then, to provide these high quality T cells a head start.

The T-cell receptor (TCR) repertoire of antigen-specific CD8⁺ T cells

Both quantitative and qualitative aspects of T-cell responses have been analyzed as determinants of control of virus infection. Among the qualitative features explored are the abilities of antigen-specific T cells to proliferate, to suppress viral replication and to respond to limited amounts of antigen (functional sensitivity/ functional avidity). All these characteristics reflect the response potential of the T-cell clones that make up a specific responding T-cell population. However, it is not known what defines an efficient immune response in terms of TCR repertoire, and conflicting views have been reported. While some studies emphasize the need for a broad repertoire (70-73), others highlight the impact of selection of particular TCR specificities (74, 75), including cross-reactive clonotypes (76-79). In contrast, other reports failed to find a correlation between TCR repertoire features (diversity, public clonotype usage) and disease progression/outcome (80-82).

The apparent conflicting results might stem from differences in study design, namely on the study of particular antigen-specific CD8⁺ T-cell populations, directed toward conserved *versus* mutating epitopes, upon different time spans after (relative) control of the virus or imbalance between viral replication and immune pressure.

We have characterized the TCR repertoire of antigen-specific CD8⁺ T cells recognizing frequently targeted epitopes from i) persistent DNA viruses that are predominantly kept in latency in healthy individuals (EBV/HCMV - chapter 5), ii) a persistent retrovirus with a high mutation rate (HIV-1 – chapter 6), and iii) an RNA virus causing acute infections generally cleared by the host's immune system (influenza A – chapter 7).

The CD8⁺ T-cell repertoire to EBV and HCMV has been suggested to be stable over years (83-85), and this might hold true for other latent, stable-genome, highly prevalent human viruses. We looked for evidence of differences in general T-cell repertoire breadth for pHLA-B *versus* pHLA-A complexes of frequently targeted, curated epitopes of EBV and HCMV (IEDB, (86)), focusing on common HLA alleles in the Caucasian population. Both pHLA-A and pHLA-B complexes seem to induce T-cell responses with similar levels of diversity (chapter 5). In essence, the HLA locus does not seem to have an inherent ability to elicit more or less diverse responses, nor does that apply for

particular HLA alleles (see B8-RL8 and B8-FL9, chapter 5). Surprisingly, some epitopes (e.g.: B8-RL8) are typically recognized by a wide T-cell repertoire while others (e.g.: A11-AK10) are typically recognized by a narrow TCR repertoire, even across different individuals. Also in mice, influenza-specific epitopes have been consistently associated with either narrow or diverse antigen-specific T-cell repertoires (87-89). It remains unknown why certain peptides tend to trigger a T-cell response of higher or lower diversity across individuals. One possibility is that within the population, naive T-cell precursor frequencies per antigen are highly comparable. Alanio et al. (90) have reported that, in humans, the precursor frequencies of 6 antigen-specific naive $CD8^+$ T-cell populations differed across epitopes but were similar among seronegative donors. Another factor that may influence the level of diversity of different T-cell responses is the sharing of $CD8^+$ T-cell responses between different antigens of distinct pathogens – so-called heterologous immunity (91-93). It has been recently shown that a cross-reactive T-cell response exists to a 9-mer epitope conserved in some α and γ-herpesviruses (7-8 amino acids of the 9-mer epitope are conserved in EBV, HSV, VZV) (94). Although the latter study did not analyze the TCR repertoire of the cross-reactive cells, their results suggest that the additive effect of having encountered different herpesviruses increases the probability to find a cross-reactive response. It has been described that, when T-cell responses to an epitope of EBV and one of influenza were cross-reactive, the breadth of clonotypic usage in the cross-reactive T-cell population was greater than that of the non-crossreactive T-cell population (95). It is tempting to speculate that more diverse repertoires might include a higher fraction of cross-reactive T-cells, initially triggered by childhood pathogens like VZV, especially for epitopes within conserved proteins of viruses of the same family. Conversely, heterologous immunity can also narrow the responding TCR repertoire (96). Yet, important factors related to heterologous immunity in humans are not always known, and hard to account for, including the event and severity of primary infection (childhood *versus* adulthood infection), the timeframe of viral persistence/latency and the frequency of viral reactivation. Moreover, the diversity of the TCR repertoire may be significantly affected by '3D-similarity' of the pathogen-derived pHLA complex with an alloantigen complex, as is the case for the EBV-derived epitope B*08-FL9 when compared to a self-peptide presented by B*44 (97-99). In our study, the diversity of the T-cell response against B*08-FL9 was very low for all individuals (chapter 5). All T-cell repertoires were significantly dominated by a TCRβ clonotype corresponding to, or differing maximally in one CDR3 amino acid from the public TCRβ clonotype (98). The only individual having a higher number of clonotypes recognizing B*08-FL9 indeed also expressed B*44, yet the main TCRβ contributor was still but one amino acid away from the

alloreactive clonotype. It remains a possibility that the B*08-FL9 complex is similar to other self-peptide-HLA complexes, highlighting the difficulty in understanding the causes of narrow and diverse antigen-specific TCR repertoires. Finally, the higher or lower TCR diversity of a given antigen-specific T-cell response may be due to the fact that some TCRs are produced more readily than others during VDJ recombination, e.g.: when they require few nucleotide additions (100-102).

The finding that there tends to be no big overlap in TRBV families between peptides sharing the same HLA restriction emphasizes the fundamental role of the presented peptide in shaping the diversity of the T-cell repertoire elicited by a pHLA-I complex. This finding seems hard to reconcile with the 'germline model' of TCR recognition, that proposes that each variable segment of the αβTCR (namely the germline-encoded CDR1 and 2) will interact with particular amino acids (interaction 'codons') in the MHC helices (103-105). Based on this model, we would have expected to find more overlap between the TCRs recognizing peptides in the context of the same HLA molecule. However, the lack of overlap can also have been motivated by differences in the length of analyzed epitopes restricted by the same HLA. As has been proposed by Wooldridge (106), TCRs might have a unique "peptide recognition signature," one of the influencing components being a peptide length preference.

We observed no correlation between the magnitude of T-cell responses against EBV/HCMV antigens and the respective breadth of TCR usage (chapter 5), even though this relation has been described for HIV (107). This difference between studies might arise from fundamental differences between these two groups of pathogens, namely in the frequency of antigen presentation and recognition by cognate T cells and the degree of within-host evolution of the pathogen's proteome: while EBV and HCMV are mainly kept latent in healthy donors (with no or few proteins expressed), HIV-1 maintains at least a low-level of replication and this event is highly error-prone (108).

The evolution of the HIV-1 proteome upon persistence in the host involves the accumulation of non-synonymous mutations, namely within T-cell epitopes. If the epitope is still presented (i.e.: binding is not abrogated), changes in the pHLA-I landscape might influence the diversity of the antigen-specific TCR repertoire. To fully understand the potential impact of viral epitope sequence variation in the T-cell response that recognizes it, one has to go beyond a cross-sectional study (chapter 6). While the diversity of a particular T-cell response is not necessarily higher for epitopes displayed as a higher number of sequence variants, changes in the clonotypic composition of a T-cell response were more evident when the epitope quasispecies evolved over time (chapter 6). We show that considerable changes in the magnitude of antigen-specific $CD8^+$ T-cell responses were often accompanied by deflation and

inflation of clonotypes and paralleled by mutations in the restricting epitope. Still, as often remarked by other studies, it is not possible to discern the chronology of these events.

Whereas there seemed to be a higher clonotypic turnover at higher viral load, as expected, there were still shifts in clonotypic composition when the viral burden was low. In addition, evolution of different antigen-specific T-cell responses within one patient in the same timeframe did not follow the same pattern in terms of changes in magnitude and diversity, or in the epitope quasispecies that was recognized. Consequently, care should be taken when drawing conclusions about the virus-immune system (im)balance based on results from single antigen-specific T-cell responses. Other factors will obviously play a role in determining the breadth of the antigen-specific T-cell repertoire, such as the T-cell precursor frequency, the degree of cross-reactivity of the cognate T cells to the original epitope and its variants, and the balance between triggering of new clonotypes and loss of senescent clonotypes (109, 110).

In acute infections like influenza A, internal proteins are the major targets of the $CD8^+$ T-cell response, and show far less mutability than surface proteins, which are under great selection pressure as targets of humoral immune responses (111). Within-host evolution of $CD8^+$ T-cell epitopes from internal proteins is also likely limited by the relatively short time span of viral replication. One of the epitopes that is highly constrained, and might not even mutate in situations of persistent replication (112), is the immunodominant A2-$M1_{58-66}$ epitope. Notwithstanding, we show that the TCR repertoire against this epitope can show significant changes in its composition upon re-infection (chapter 7). We observed that a new, non-TRBV19 clonotype took over after re-infection, while the dominantly contributing clonotype for the A2-$M1_{58-66}$ $CD8^+$ T-cell response prior to the infection episode was no longer detected after re-infection, suggesting that it might have failed to expand. This study shows that, in recurrent acute infections, re-exposure to a conserved epitope can nonetheless elicit a change in the responding TCR repertoire. Cross-reactivity between this epitope and an HLA-A2-restricted EBV epitope has been shown (113), but did not underlie the source of the new clonotype, as we were unable to detect co-staining with the two specific tetramers.

The common thread through the observations of the abovementioned studies and ours is that the antigen-specific $CD8^+$ T-cell repertoire should possess enough diversity/plasticity to allow expansion of other T-cell clonotypes if the original epitope mutates (chapter 6), or if the main T-cell clonotypes to the original epitope fail to expand (chapter 7), whether due to senescence or because of other (inclusive stochastic) reasons. Nonetheless, the reasons why a given antigen-specific $CD8^+$ T-cell response is diverse or not remains thus-far unknown.

Functional sensitivity of antigen-specific CD8⁺ T cells

Each antigen-specific T-cell population consists of a set of clones, each having their unique qualitative features. Among these features is the functional sensitivity (or functional avidity): the ability to respond to decreasing concentrations of cognate antigen. The functional sensitivity of a T-cell population is, thus, the result of the cumulative sensitivity of its clonotypes.

In animal models, a higher functional sensitivity of $CD8^+$ T cells has been shown to associate with better control of viral replication (72, 114-116), immunodominance and better proliferation capacity (117), and a higher efficiency of killing infected cells *in vitro* (118, 119), yet inducing immune escape (120). In humans, high functional sensitivity has been connected with better clearance of HCV and recognition of viral variants (121), as well as with slow progression to AIDS, a better suppressor ability and larger T-cell cross-reactivity (110, 122-125). On the other side, the T cells with the highest functional sensitivity may get more exhausted (126, 127) and thus be depleted over time (128), and be more likely to induce immune escape (82). However, other studies have not found a correlation between the functional sensitivity of $CD8^+$ T cells and suppression of HIV replication (129), with progressors exhibiting uncontrolled replication despite the presence of high avidity T cells (130). There is, again, inconsistency on whether the presence of high avidity T cells is beneficial for viral control, or whether HLA-A or HLA-B-restricted $CD8^+$ T-cell responses are superior in functional sensitivity (16, 122, 124, 127).

Making use of known, frequently targeted EBV and HCMV peptides, we analyzed the functional sensitivity of the responding T cells and, in contrast to Harari et al, (127) found no significant difference between T-cell responses restricted by HLA-B *versus* HLA-A alleles (Fig. 2A). In fact, and as noticed in our extensive repertoire analysis (chapter 5), some peptides also tended to elicit T-cell responses of similar functional avidity in different donors (Fig. 2B, e.g.: A*02-NV9, B*08-RL8). This similarity might be more obvious for latent viruses, as the epitope sequence will probably undergo little evolution, but should be taken into account as a factor when analyzing responses to the same antigens in a group of individuals. This analysis of functional sensitivity in healthy individuals should be complemented by information on the epitope sequence of the virus. Even though latent DNA viruses are stable and very conserved, we cannot exclude that the infecting and latently persisting virus strain may have been different between individuals, especially for the epitopes showing greatest variability/spread. Nevertheless, the antigen(s) chosen in a study as representative of the antiviral immune response might influence whether or not the impact of a functional feature is detected. A high avidity response can, thus, be the result of either a polyclonal TCR repertoire (A*02-YV9, Figs. 2B light blue, and 2C first pie chart) or a oligoclonal or monoclonal one

(B*07-RL9 Figs. 2B darker blue, and 2C third pie chart), as can a low avidity response (Figs. 2B and 2C, remainder).

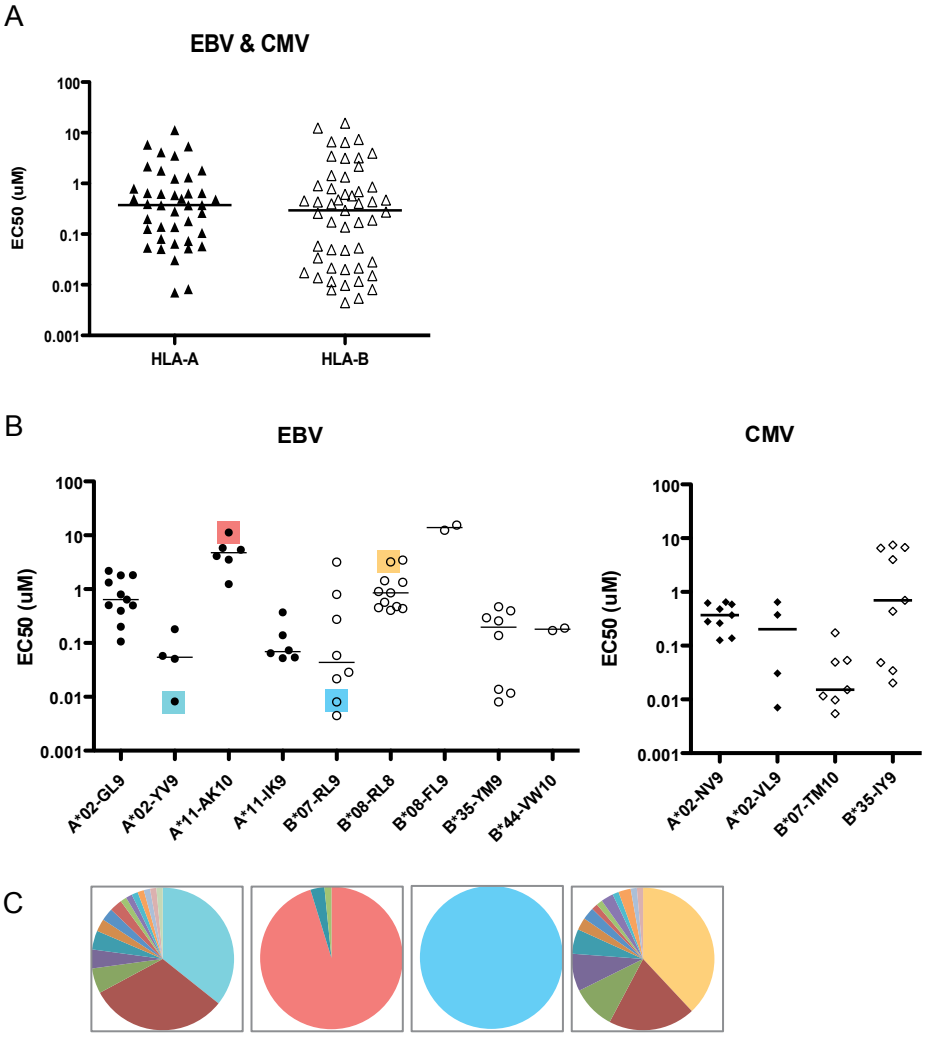


Figure 2. High functional sensitivity of an antigen-specific CD8⁺ T cell response can be achieved by a highly polyclonal or a monoclonal TCR repertoire, and vice-versa.
PBMC of healthy donors were stimulated with decreasing concentrations of known, frequently targeted EBV and HCMV peptides, and the decline in the response was used for the determination of the functional sensitivity of the response (i.e.: the concentration of peptide that elicits 50% of the maximal response (EC_{50}), as measured by IFNγ ELIspot). Functional sensitivity of antigen-specific T-cell responses was categorized per (A) HLA-locus restriction (HLA-A *versus* HLA-B), and per (B) cognate peptide. Two of the highest and two of the lowest functional sensitivity responses were highlighted in panel (B), and the composition of the TCR repertoire (each colored slice is a TCRβ clonotype) of the corresponding $CD8^+$ T-cell responses is shown in panel (C). The color attributed to the dominant clonotype in the response displays correspondence to the EC_{50} of the total antigen-specific $CD8^+$ T-cell pool.

In conclusion, the functional features of a $CD8^+$ T-cell response (i.e.: not only functional sensitivity, but also polyfunctionality and others) are not independent, but should be analyzed in the context of the particular antigen, the clonotypic diversity of the respective response (131), and the type of infection (persistent with (low-level) replication, latent or acute).

The level of TCR diversity of a given response is of particular importance in the case of viruses that persist and replicate, albeit at a low level, in the host.

It has been shown for HIV-1 that there is mainly one founder virus establishing infection. The viral set-point will most likely depend on the efficacy of the immunodominant acute-phase $CD8^+$ T-cell responses (132), before mutations start to accumulate. Parallel, longitudinal analysis during this initial window might be needed to study the 'basal' level of functional sensitivity and TCR diversity of antigen-specific $CD8^+$ T-cell responses, and whether these responses have similar qualitative traits across individuals. Progressively, these features will change more or less, depending on, among others, the level of antigen presentation, and the degree of mutation occurring in the epitope. T-cell responses to the most constrained epitopes and the ones exhibiting the highest functional sensitivity will probably be most dramatically affected by the loss of certain clonotypes that have become exhausted (122). A large number of clonotypes in an antigen-specific T-cell response will allow for the replacement of T cells that i) get lost by exhaustion, or ii) in the case of mutating epitopes, are 'functionally' lost by lack of recognition of the epitope variant.

However, it is hard to define what is the 'basal TCR diversity' of each particular antigen-specific T-cell response, especially when not based on 'laboratory-controlled' conditions such as animal models. The initial response to a given antigen depends on the recruitment of naive T cells, but it will potentially also involve the recruitment of T cells that cross-react with other pathogens encountered earlier in life. Naive T cells have been 'educated' in the thymus by a set of pHLA complexes that differ from those presented in the periphery (at least during positive selection, given the different cleavage motif of the thymoproteasome (133, 134)) and may react to pHLA complexes diverging in the bound epitope (i.e.: T-cell clones that react to epitopes of different pathogens). In fact, it has recently been shown that healthy adults may carry memory $CD4^+$ T cells specific for viruses they have not previously encountered (135). It seems crucial to understand how influential the pHLA 3D-landscape is in terms of epitope sequence similarity (same amino acids at TCR contact positions): how it influences thymic 'education', and how it determines that a T-cell response to a foreign antigen is elicited. An additional challenge is to understand the role of co-receptors and co-

stimulatory and inhibitory molecules in determining whether a cell that sees a similar 'pHLA landscape' will be activated and respond.

Experimental set-up and its influence on the study outcome

In an effort for uniformity and elimination of confounders, specific-pathogen free (SPF) mice are often used as models for particular infections. Of note, the MHC polymorphism in mice is significantly lower than that in humans. In these mice models, it was possible to reproduce immunodominance hierarchies (magnitude, targeting) of primary antigen-specific T-cell responses (influenza, LCMV) (87, 136-138), as well as their degrees of TCR diversity. For example, in the mouse model of influenza there are two epitopes systematically found as co-dominant in primary infection, with the cognate TCR repertoire of one being consistently diverse and private, while the other is public and narrow (87-89). Although the insights of these studies are invaluable, and have for example unveiled the impact of naïve precursor frequencies, recruitment and clonal expansion for immunodominance (139, 140), these mouse populations are, beside 'pathogen-free', so inbred that the nuances of MHC polymorphism are less detectable. The complexity of the immune system of the human population, 'outbred', with a very polymorphic HLA-I complex, and incompletely documented immunological histories, makes it hard to find reproducible correlates. In addition, if one would extrapolate the conclusions on chronic or recurrent acute infections from mice to men, one must take into account that, while the naive T-cell pool diversity in adult mice will continuously change by thymic output of new specificities, the naive T-cell pool in adult humans is mainly maintained by peripheral T-cell division (141). The impact of consecutive (and heterologous) infections in mice and men is thus potentially different.

The impact of the HLA polymorphism on $CD8^+$ T-cell responses, and specifically on their pattern of immunodominance, is already measurable at the level of antigen presentation. There may be a hierarchy in the HLA alleles as to the presentation of the broadest peptide repertoire: while HLA-B*07 seems to present the highest amount of peptides when compared with B*27, A*03 and A*01 (25), it may present a smaller set when compared to A*02 (142). The HLA-allele presenting the most diverse set of peptides is expected to have the highest probability of triggering an immunodominant response. As a consequence, the HLA coverage of the population chosen for a study can affect the outcome of the analysis on the most frequently targeted/highest magnitude T-cell responses.

Furthermore, if some epitopes have an intrinsic tendency to elicit responses with more/less clonotypic diversity, or more/less sensitivity to antigen, inconsistencies in the literature on functional sensitivity and TCR diversity could derive from the

particular epitopes under analysis. This may be more visible when studying responses to viruses like EBV and HCMV in healthy individuals, where they remain predominantly in a latent phase (with no or few expressed viral proteins) and where epitopes are under little pressure to mutate. In viruses like HCV and HIV-1, this impact becomes even more complicated upon accumulation of mutations during (low-level) replication. Here, even within one individual, responses follow different directions (chapter 6), and drawing conclusions from one particular epitope-specific response on the progression of the persistent viral disease is particularly tricky.

Finally, and also implicitly referred to above, the choice of the phase of infection in persistent viruses may condition the interpretation of the immunodominance data. We and others have frequently chosen to analyze epitopes that elicit immunodominant T-cell responses in the relatively late phase of chronic infection. In this light it is important to realize that immunodominance hierarchies change upon persistence of the pathogen, as exemplified in animal models (e.g.: (143-145). In addition, immunodominant responses in chronic infection have been established after bringing the pathogen into a relative balance by the immune response. In order to obtain better insights into immunodominance, it is, thus, of importance to also look into the *establishment* of dominance hierarchies in the acute phase, whenever possible.

Conclusion and perspectives

In this thesis, we aimed to assess whether the frequently reported immunodominance of HLA-B-restricted T-cell responses could stem from differences in the peptide repertoire presented by HLA-B molecules, relative to HLA-A. We found that, in general, peptides bind HLA-A molecules with higher affinity than HLA-B molecules. While the 'HLA/population-wide' peptide binder repertoire seems to be more diverse for HLA-A molecules, analyses on the naturally processed peptide repertoire suggest that HLA-B molecules present a more diverse set of peptides at the cell surface. It thus remains unclear whether the diversity of the bound peptide repertoire contributes to the immunodominance of HLA-B-restricted responses. In addition, HLA-A and HLA-B molecules display a similar level of promiscuity of binding. Our studies do suggest that HLA-B molecules target more conserved regions of viral proteomes, at least in HIV-1 and HCV infection. It remains to be studied whether this preference is also true for other pathogens. We did not find any correlation between the magnitude of a T-cell response and its TCR diversity, nor did we find a consistent difference between HLA-A *versus* HLA-B-restricted T-cell responses in terms of TCR diversity (EBV/HCMV). In

fact, the breadth of a T-cell response seems to be determined by the nature of the peptide, and not by the restricting HLA molecule. Nonetheless, in HIV-1 infection, substantial changes in the height of antigen-specific $CD8^+$ T-cell responses are frequently paralleled by deflation and inflation of clonotypes and by mutations in the restricting epitope. We found that even in situations of viral persistence at low-level replication (HIV-1), or upon re-infection (influenza A), the TCR repertoire of the antigen-specific T-cell response can change significantly, even if mutation of the epitope sequence is not observed or expected. Understanding why certain antigen-specific T-cell responses reach a higher level of diversity than others will require more insights into the patterns of epitope recognition occurring already at the level of thymic education, and how these affect (cross)reactivity. A major step toward this goal has recently been achieved in a complex study that scrutinized mice and human MHC class II ligand libraries for binding and discovery of patterns among peptides recognized by different TCRs (146). Similar studies for HLA class I-restricted T-cell responses will be invaluable for understanding the rules of engagement in TCR recognition of pHLA-I complexes.

References

1. Parham P, and Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science.* 1996;272(5258):67-74.

2. McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, and Watkins DI. A uniquely high level of recombination at the HLA-B locus. *Proceedings of the National Academy of Sciences of the United States of America.* 1994;91(13):5893-7.

3. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature.* 2004;432(7018):769-75.

4. Carrington M, and O'Brien SJ. The influence of HLA genotype on AIDS. *Annual review of medicine.* 2003;54:535-51.

5. Chuang WC, Sarkodie F, Brown CJ, Owusu-Ofori S, Brown J, Li C, Navarrete C, Klenerman P, and Allain JP. Protective effect of HLA-B57 on HCV genotype 2 infection in a West African population. *Journal of medical virology.* 2007;79(6):724-33.

6. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science.* 2007;317(5840):944-7.

7. Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, Feeney ME, Addo MM, Lichterfeld M, Lahaie MP, et al. Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes. *Nature immunology.* 2006;7(2):173-8.

8. Goulder PJ, and Watkins DI. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature reviews Immunology.* 2008;8(8):619-30.

9. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, and Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. *Nature.* 1991;352(6336):595-600.

10. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Ripke S, Brumme CJ, Pulit SL, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* 2010;330(6010):1551-7.

11. Kim AY, Kuntzen T, Timm J, Nolan BE, Baca MA, Reyor LL, Berical AC, Feller AJ, Johnson KL, Schulze zur Wiesch J, et al. Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. *Gastroenterology.* 2011;140(2):686-96 e1.

12. Leslie A, Matthews PC, Listgarten J, Carlson JM, Kadie C, Ndung'u T, Brander C, Coovadia H, Walker BD, Heckerman D, et al. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *Journal of virology.* 2010;84(19):9879-88.

13. Martin MP, and Carrington M. Immunogenetics of viral infections. *Current opinion in immunology.* 2005;17(5):510-6.

14. McKiernan SM, Hagan R, Curry M, McDonald GS, Kelly A, Nolan N, Walsh A, Hegarty J, Lawlor E, and Kelleher D. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology.* 2004;40(1):108-14.

15. Neumann-Haefelin C, McKiernan S, Ward S, Viazov S, Spangenberg HC, Killinger T, Baumert TF, Nazarova N, Sheridan I, Pybus O, et al. Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution. *Hepatology.* 2006;43(3):563-72.

16. Bihl F, Frahm N, Di Giammarino L, Sidney J, John M, Yusim K, Woodberry T, Sango K, Hewitt HS, Henry L, et al. Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *Journal of immunology.* 2006;176(7):4094-101.

17. Altfeld M, Addo MM, Rosenberg ES, Hecht FM, Lee PK, Vogel M, Yu XG, Draenert R, Johnston MN, Strick D, et al. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *Aids.* 2003;17(18):2581-91.

18. Lewinsohn DA, Winata E, Swarbrick GM, Tanner KE, Cook MS, Null MD, Cansler ME, Sette A, Sidney J, and Lewinsohn DM. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS pathogens.* 2007;3(9):1240-9.

19. Weiskopf D, Yauch LE, Angelo MA, John DV, Greenbaum JA, Sidney J, Kolla RV, De Silva AD, de Silva AM, Grey H, et al. Insights into HLA-restricted T cell responses in a novel mouse model of dengue virus infection point toward new implications for vaccine design. *Journal of immunology.* 2011;187(8):4268-79.

20. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, and Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *Journal of immunology.* 2013;191(12):5831-9.

21. Calis JJ, Sanchez-Perez GF, and Kesmir C. MHC class I molecules exploit the low G+C content of pathogen genomes for enhanced presentation. *European journal of immunology.* 2010;40(10):2699-709.

22. Granados DP, Yahyaoui W, Laumont CM, Daouda T, Muratore-Schroeder TL, Cote C, Laverdure JP, Lemieux S, Thibault P, and Perreault C. MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements. *Blood.* 2012;119(26):e181-91.

23. Granados DP, Sriranganadane D, Daouda T, Zieger A, Laumont CM, Caron-Lizotte O, Boucher G, Hardy MP, Gendron P, Cote C, et al. Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. *Nature communications.* 2014;5:3600.

24. Johnson KL, Ovsyannikova IG, Mason CJ, Bergen HR, 3rd, and Poland GA. Discovery of naturally processed and HLA-presented class I peptides from vaccinia virus infection using mass spectrometry for vaccine development. *Vaccine.* 2009;28(1):38-47.

25. Mommen GP, Frese CK, Meiring HD, van Gaans-van den Brink J, de Jong AP, van Els CA, and Heck AJ. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proceedings of the National Academy of Sciences of the United States of America.* 2014;111(12):4507-12.

26. Croft NP, Smith SA, Wong YC, Tan CT, Dudek NL, Flesch IE, Lin LC, Tscharke DC, and Purcell AW. Kinetics of antigen expression and epitope presentation during virus infection. *PLoS pathogens.* 2013;9(1):e1003129.

27. Crotzer VL, Christian RE, Brooks JM, Shabanowitz J, Settlage RE, Marto JA, White FM, Rickinson AB, Hunt DF, and Engelhard VH. Immunodominance among EBV-derived epitopes restricted by HLA-B27 does not correlate with epitope abundance in EBV-transformed B-lymphoblastoid cell lines. *Journal of immunology.* 2000;164(12):6120-9.

28. Greene JM, Wiseman RW, Lank SM, Bimber BN, Karl JA, Burwitz BJ, Lhost JJ, Hawkins OE, Kunstman KJ, Broman KW, et al. Differential MHC class I expression in distinct leukocyte subsets. *BMC immunology.* 2011;12:39.

29. Johnson DR. Locus-specific constitutive and cytokine-induced HLA class I gene expression. *Journal of immunology.* 2003;170(4):1894-902.

30. Weinzierl AO, Lemmel C, Schoor O, Muller M, Kruger T, Wernet D, Hennenlotter J, Stenzl A, Klingel K, Rammensee HG, et al. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Molecular & cellular proteomics : MCP.* 2007;6(1):102-13.

31. Neisig A, Wubbolts R, Zang X, Melief C, and Neefjes J. Allele-specific differences in the interaction of MHC class I molecules with transporters associated with antigen processing. *Journal of immunology.* 1996;156(9):3196-206.

32. Harndahl M, Rasmussen M, Roder G, Dalgaard Pedersen I, Sorensen M, Nielsen M, and Buus S. Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *European journal of immunology.* 2012;42(6):1405-16.

33. Brooks JM, Colbert RA, Mear JP, Leese AM, and Rickinson AB. HLA-B27 subtype polymorphism and CTL epitope choice: studies with EBV peptides link immunogenicity with stability of the B27:peptide complex. *Journal of immunology.* 1998;161(10):5252-9.

34. Pogue RR, Eron J, Frelinger JA, and Matsui M. Amino-terminal alteration of the HLA-A*0201-restricted human immunodeficiency virus pol peptide increases complex stability and in vitro immunogenicity. *Proceedings of the National Academy of Sciences of the United States of America.* 1995;92(18):8166-70.

35. Spierings E, Gras S, Reiser JB, Mommaas B, Almekinders M, Kester MG, Chouquet A, Le Gorrec M, Drijfhout JW, Ossendorp F, et al. Steric hindrance and fast dissociation explain the lack of immunogenicity of the minor histocompatibility HA-1Arg Null allele. *Journal of immunology.* 2009;182(8):4809-16.

36. van der Burg SH, Visseren MJ, Brandt RM, Kast WM, and Melief CJ. Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *Journal of immunology.* 1996;156(9):3308-14.

37. Hansen TH, and Bouvier M. MHC class I antigen presentation: learning from viral evasion strategies. *Nature reviews Immunology.* 2009;9(7):503-13.

38. Horst D, Verweij MC, Davison AJ, Ressing ME, and Wiertz EJ. Viral evasion of T cell immunity: ancient mechanisms offering new applications. *Current opinion in immunology.* 2011;23(1):96-103.

39. Rajapaksa US, Li D, Peng YC, McMichael AJ, Dong T, and Xu XN. HLA-B may be more protective against HIV-1 than HLA-A because it resists negative regulatory factor (Nef) mediated down-regulation. *Proceedings of the National Academy of Sciences of the United States of America.* 2012;109(33):13353-8.

40. Griffin BD, Gram AM, Mulder A, Van Leeuwen D, Claas FH, Wang F, Ressing ME, and Wiertz E. EBV BILF1 evolved to downregulate cell surface display of a wide range of HLA class I molecules through their cytoplasmic tail. *Journal of immunology.* 2013;190(4):1672-84.

41. Barel MT, Ressing M, Pizzato N, van Leeuwen D, Le Bouteiller P, Lenfant F, and Wiertz EJ. Human cytomegalovirus-encoded US2 differentially affects surface expression of MHC class I locus products and targets membrane-bound, but not soluble HLA-G1 for degradation. *Journal of immunology.* 2003;171(12):6757-65.

42. Cella M, Longo A, Ferrara GB, Strominger JL, and Colonna M. NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80. *The Journal of experimental medicine.* 1994;180(4):1235-42.

43. Hoof I, van Baarle D, Hildebrand WH, and Kesmir C. Proteome sampling by the HLA class I antigen processing pathway. *PLoS computational biology.* 2012;8(5):e1002517.

44. Kincaid EZ, Che JW, York I, Escobar H, Reyes-Vargas E, Delgado JC, Welsh RM, Karow ML, Murphy AJ, Valenzuela DM, et al. Mice completely lacking immunoproteasomes show major changes in antigen presentation. *Nature immunology.* 2012;13(2):129-35.

45. Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, Laplace C, Zwinderman A, Dick TP, Muller J, et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *The Journal of experimental medicine.* 2001;194(1):1-12.

46. Zanker D, Waithman J, Yewdell JW, and Chen W. Mixed proteasomes function to increase viral peptide diversity and broaden antiviral CD8+ T cell responses. *Journal of immunology.* 2013;191(1):52-9.

47. Lauring AS, and Andino R. Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens.* 2010;6(7):e1001005.

48. Crawford H, Lumm W, Leslie A, Schaefer M, Boeras D, Prado JG, Tang J, Farmer P, Ndung'u T, Lakhi S, et al. Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. *The Journal of experimental medicine.* 2009;206(4):909-21.

49. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, Derdeyn CA, Tang J, Kaslow RA, Bansal A, et al. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *The Journal of experimental medicine.* 2008;205(5):1009-17.

50. Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, Schneidewind A, Power KA, Toth I, Frahm N, 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. *Journal of virology.* 2009;83(4):1845-55.

51. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, Treurnicht F, de Rosa DA, Hide W, Karim SA, et al. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. *PLoS pathogens.* 2008;4(3):e1000033.

52. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine.* 2004;10(3):282-9.

53. Davenport MP, Loh L, Petravic J, and Kent SJ. Rates of HIV immune escape and reversion: implications for vaccination. *Trends in microbiology.* 2008;16(12):561-6.

54. Kuntzen T, Timm J, Berical A, Lewis-Ximenez LL, Jones A, Nolan B, Schulze zur Wiesch J, Li B, Schneidewind A, Kim AY, et al. Viral sequence evolution in acute hepatitis C virus infection. *Journal of virology.* 2007;81(21):11658-68.

55. Goulder PJ, Brander C, Tang Y, Tremblay C, Colbert RA, Addo MM, Rosenberg ES, Nguyen T, Allen R, Trocha A, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature.* 2001;412(6844):334-8.

56. Gijsbers EF, van Nuenen AC, Schuitemaker H, and Kootstra NA. Gag sequence variation in a human immunodeficiency virus type 1 transmission cluster influences viral replication fitness. *The Journal of general virology.* 2013;94(Pt 2):354-9.

57. Gijsbers EF, Feenstra KA, van Nuenen AC, Navis M, Heringa J, Schuitemaker H, and Kootstra NA. HIV-1 replication fitness of HLA-B*57/58:01 CTL escape variants is restored by the accumulation of compensatory mutations in gag. *PloS one.* 2013;8(12):e81235.

58. Schneidewind A, Brumme ZL, Brumme CJ, Power KA, Reyor LL, O'Sullivan K, Gladden A, Hempel U, Kuntzen T, Wang YE, et al. Transmission and long-term stability of compensated CD8 escape mutations. *Journal of virology.* 2009;83(8):3993-7.

59. Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, Reddy S, van der Stok M, Mncube Z, Brander C, Rousseau C, et al. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *Journal of virology.* 2007;81(15):8346-51.

60. Ruhl M, Chhatwal P, Strathmann H, Kuntzen T, Bankwitz D, Skibbe K, Walker A, Heinemann FM, Horn PA, Allen TM, et al. Escape from a dominant HLA-B*15-restricted CD8+ T cell response against hepatitis C virus requires compensatory mutations outside the epitope. *Journal of virology.* 2012;86(2):991-1000.

61. Neumann-Haefelin C, Oniangue-Ndza C, Kuntzen T, Schmidt J, Nitschke K, Sidney J, Caillet-Saguy C, Binder M, Kersting N, Kemper MW, et al. Human leukocyte antigen B27 selects for rare escape mutations that significantly impair hepatitis C virus replication and require compensatory mutations. *Hepatology.* 2011;54(4):1157-66.

62. Schellens IM, Navis M, van Deutekom HW, Boeser-Nunnink B, Berkhout B, Kootstra N, Miedema F, Kesmir C, Schuitemaker H, van Baarle D, et al. Loss of HIV-1-derived cytotoxic T lymphocyte epitopes restricted by protective HLA-B alleles during the HIV-1 epidemic. *Aids.* 2011;25(14):1691-700.

63. Honegger JR, Kim S, Price AA, Kohout JA, McKnight KL, Prasad MR, Lemon SM, Grakoui A, and Walker CM. Loss of immune escape mutations during persistent HCV infection in pregnancy enhances replication of vertically transmitted viruses. *Nature medicine.* 2013;19(11):1529-33.

64. De Boer RJ, Homann D, and Perelson AS. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *Journal of immunology.* 2003;171(8):3928-35.

65. Iglesias MC, Briceno O, Gostick E, Moris A, Meaudre C, Price DA, Ungeheuer MN, Saez-Cirion A, Mallone R, and Appay V. Immunodominance of HLA-B27-restricted HIV KK10-specific CD8(+) T-cells is not related to naive precursor frequency. *Immunology letters.* 2013;149(1-2):119-22.

66. Jenkins MK, and Moon JJ. The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *Journal of immunology.* 2012;188(9):4135-40.

67. Moon JJ, Chu HH, Pepper M, McSorley SJ, Jameson SC, Kedl RM, and Jenkins MK. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity.* 2007;27(2):203-13.

68. Obar JJ, Khanna KM, and Lefrancois L. Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection. *Immunity.* 2008;28(6):859-69.

69. Schmidt J, Neumann-Haefelin C, Altay T, Gostick E, Price DA, Lohmann V, Blum HE, and Thimme R. Immunodominance of HLA-A2-restricted hepatitis C virus-specific CD8+ T cell responses is linked to naive-precursor frequency. *Journal of virology.* 2011;85(10):5232-6.

70. Price DA, West SM, Betts MR, Ruff LE, Brenchley JM, Ambrozak DR, Edghill-Smith Y, Kuroda MJ, Bogdan D, Kunstman K, et al. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. *Immunity.* 2004;21(6):793-803.

71. Meyer-Olson D, Shoukry NH, Brady KW, Kim H, Olson DP, Hartman K, Shintani AK, Walker CM, and Kalams SA. Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape. *The Journal of experimental medicine.* 2004;200(3):307-19.

72. Messaoudi I, Guevara Patino JA, Dyall R, LeMaoult J, and Nikolich-Zugich J. Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science.* 2002;298(5599):1797-800.

73. Wang GC, Dash P, McCullers JA, Doherty PC, and Thomas PG. T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Science translational medicine.* 2012;4(128):128ra42.

74. Dong T, Stewart-Jones G, Chen N, Easterbrook P, Xu X, Papagno L, Appay V, Weekes M, Conlon C, Spina C, et al. HIV-specific cytotoxic T cells from long-term survivors select a unique T cell receptor. *The Journal of experimental medicine.* 2004;200(12):1547-57.

75. Price DA, Asher TE, Wilson NA, Nason MC, Brenchley JM, Metzler IS, Venturi V, Gostick E, Chattopadhyay PK, Roederer M, et al. Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. *The Journal of experimental medicine.* 2009;206(4):923-36.

76. Simons BC, Vancompernolle SE, Smith RM, Wei J, Barnett L, Lorey SL, Meyer-Olson D, and Kalams SA. Despite biased TRBV gene usage against a dominant HLA B57-restricted epitope, TCR diversity can provide recognition of circulating epitope variants. *Journal of immunology.* 2008;181(7):5137-46.

77. van Bockel DJ, Price DA, Munier ML, Venturi V, Asher TE, Ladell K, Greenaway HY, Zaunders J, Douek DC, Cooper DA, et al. Persistent survival of prevalent clonotypes within an immunodominant HIV gag-specific CD8+ T cell response. *Journal of immunology.* 2011;186(1):359-71.

78. Chen H, Ndhlovu ZM, Liu D, Porter LC, Fang JW, Darko S, Brockman MA, Miura T, Brumme ZL, Schneidewind A, et al. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nature immunology.* 2012;13(7):691-700.

79. Ladell K, Hashimoto M, Iglesias MC, Wilmann PG, McLaren JE, Gras S, Chikata T, Kuse N, Fastenackels S, Gostick E, et al. A molecular basis for the control of preimmune escape variants by HIV-specific CD8+ T cells. *Immunity.* 2013;38(3):425-36.

80. Miles JJ, Thammanichanond D, Moneer S, Nivarthi UK, Kjer-Nielsen L, Tracy SL, Aitken CK, Brennan RM, Zeng W, Marquart L, et al. Antigen-driven patterns of TCR bias are shared across diverse outcomes of human hepatitis C virus infection. *Journal of immunology.* 2011;186(2):901-12.

81. Mendoza D, Royce C, Ruff LE, Ambrozak DR, Quigley MF, Dang T, Venturi V, Price DA, Douek DC, Migueles SA, et al. HLA B*5701-positive long-term nonprogressors/elite controllers are not distinguished from progressors by the clonal composition of HIV-specific CD8+ T cells. *Journal of virology.* 2012;86(7):4014-8.

82. Iglesias MC, Almeida JR, Fastenackels S, van Bockel DJ, Hashimoto M, Venturi V, Gostick E, Urrutia A, Wooldridge L, Clement M, et al. Escape from highly effective public CD8+ T-cell clonotypes by HIV. *Blood.* 2011;118(8):2138-49.

83. Iancu EM, Corthesy P, Baumgaertner P, Devevre E, Voelter V, Romero P, Speiser DE, and Rufer N. Clonotype selection and composition of human CD8 T cells specific for persistent herpes viruses varies with differentiation but is stable over time. *Journal of immunology.* 2009;183(1):319-31.

84. Klarenbeek PL, Remmerswaal EB, ten Berge IJ, Doorenspleet ME, van Schaik BD, Esveldt RE, Koch SD, ten Brinke A, van Kampen AH, Bemelman FJ, et al. Deep sequencing of antiviral T-cell responses to HCMV and EBV in humans reveals a stable repertoire that is maintained for many years. *PLoS pathogens.* 2012;8(9):e1002889.

85. Miles JJ, Silins SL, Brooks AG, Davis JE, Misko I, and Burrows SR. T-cell grit: large clonal expansions of virus-specific CD8+ T cells can dominate in the peripheral circulation for at least 18 years. *Blood.* 2005;106(13):4412-3.

86. Hislop AD, Taylor GS, Sauce D, and Rickinson AB. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. *Annual review of immunology.* 2007;25:587-617.

87. Kedzierska K, La Gruta NL, Turner SJ, and Doherty PC. Establishment and recall of CD8+ T-cell memory in a model of localized transient infection. *Immunological reviews.* 2006;211:133-45.

88. Kedzierska K, Turner SJ, and Doherty PC. Conserved T cell receptor usage in primary and recall responses to an immunodominant influenza virus nucleoprotein epitope. *Proceedings of the National Academy of Sciences of the United States of America.* 2004;101(14):4942-7.

89. Turner SJ, Kedzierska K, Komodromou H, La Gruta NL, Dunstone MA, Webb AI, Webby R, Walden H, Xie W, McCluskey J, et al. Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific CD8+ T cell populations. *Nature immunology.* 2005;6(4):382-9.

90. Alanio C, Lemaitre F, Law HK, Hasan M, and Albert ML. Enumeration of human antigen-specific naive CD8+ T cells reveals conserved precursor frequencies. *Blood.* 2010;115(18):3718-25.

91. Sharma S, and Thomas PG. The two faces of heterologous immunity: protection or immunopathology. *Journal of leukocyte biology.* 2014;95(3):405-16.

92. Selin LK, Brehm MA, Naumov YN, Cornberg M, Kim SK, Clute SC, and Welsh RM. Memory of mice and men: CD8+ T-cell cross-reactivity and heterologous immunity. *Immunological reviews.* 2006;211:164-81.

93. Welsh RM, Che JW, Brehm MA, and Selin LK. Heterologous immunity between viruses. *Immunological reviews.* 2010;235(1):244-66.

94. Chiu C, McCausland M, Sidney J, Duh FM, Rouphael N, Mehta A, Mulligan M, Carrington M, Wieland A, Sullivan NL, et al. Broadly reactive human CD8 T cells that recognize an epitope conserved between VZV, HSV and EBV. *PLoS pathogens.* 2014;10(3):e1004008.

95. Clute SC, Naumov YN, Watkin LB, Aslan N, Sullivan JL, Thorley-Lawson DA, Luzuriaga K, Welsh RM, Puzone R, Celada F, et al. Broad cross-reactive TCR repertoires recognizing dissimilar Epstein-Barr and influenza A virus epitopes. *Journal of immunology.* 2010;185(11):6753-64.

96. Cornberg M, Chen AT, Wilkinson LA, Brehm MA, Kim SK, Calcagno C, Ghersi D, Puzone R, Celada F, Welsh RM, et al. Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. *The Journal of clinical investigation.* 2006;116(5):1443-56.

97. Burrows SR, Khanna R, Burrows JM, and Moss DJ. An alloresponse in humans is dominated by cytotoxic T lymphocytes (CTL) cross-reactive with a single Epstein-Barr virus CTL epitope: implications for graft-versus-host disease. *The Journal of experimental medicine.* 1994;179(4):1155-61.

98. Burrows SR, Silins SL, Moss DJ, Khanna R, Misko IS, and Argaet VP. T cell receptor repertoire for a viral epitope in humans is diversified by tolerance to a background major histocompatibility complex antigen. *The Journal of experimental medicine.* 1995;182(6):1703-15.

99. Macdonald WA, Chen Z, Gras S, Archbold JK, Tynan FE, Clements CS, Bharadwaj M, Kjer-Nielsen L, Saunders PM, Wilce MC, et al. T cell allorecognition via molecular mimicry. *Immunity.* 2009;31(6):897-908.

100. Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR, Carlson CS, and Warren EH. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Science translational medicine.* 2010;2(47):47ra64.

101. Venturi V, Price DA, Douek DC, and Davenport MP. The molecular basis for public T-cell responses? *Nature reviews Immunology.* 2008;8(3):231-8.

102. Venturi V, Quigley MF, Greenaway HY, Ng PC, Ende ZS, McIntosh T, Asher TE, Almeida JR, Levy S, Price DA, et al. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *Journal of immunology.* 2011;186(7):4285-94.

103. Garcia KC. Reconciling views on T cell receptor germline bias for MHC. *Trends in immunology.* 2012;33(9):429-36.

104. Garcia KC, Adams JJ, Feng D, and Ely LK. The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nature immunology.* 2009;10(2):143-7.

105. Yin L, Scott-Browne J, Kappler JW, Gapin L, and Marrack P. T cells and their eons-old obsession with MHC. *Immunological reviews.* 2012;250(1):49-60.

106. Wooldridge L. Individual MHCI-Restricted T-Cell Receptors are Characterized by a Unique Peptide Recognition Signature. *Frontiers in immunology.* 2013;4:199.

107. Balamurugan A, Ng HL, and Yang OO. Rapid T cell receptor delineation reveals clonal expansion limitation of the magnitude of the HIV-1-specific CD8+ T cell response. *Journal of immunology.* 2010;185(10):5935-42.

108. Roberts JD, Bebenek K, and Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science.* 1988;242(4882):1171-3.

109. Appay V, Douek DC, and Price DA. CD8+ T cell efficacy in vaccination and disease. *Nature medicine.* 2008;14(6):623-8.

110. Davenport MP, Price DA, and McMichael AJ. The T cell repertoire in infection and vaccination: implications for control of persistent viruses. *Current opinion in immunology.* 2007;19(3):294-300.

111. Li CK, Rappuoli R, and Xu XN. Correlates of protection against influenza infection in humans--on the path to a universal vaccine? *Current opinion in immunology.* 2013;25(4):470-6.

112. Valkenburg SA, Quinones-Parra S, Gras S, Komadina N, McVernon J, Wang Z, Halim H, Iannello P, Cole C, Laurie K, et al. Acute emergence and reversion of influenza A virus quasispecies within CD8+ T cell antigenic peptides. *Nature communications.* 2013;4:2663.

113. Clute SC, Watkin LB, Cornberg M, Naumov YN, Sullivan JL, Luzuriaga K, Welsh RM, and Selin LK. Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. *The Journal of clinical investigation.* 2005;115(12):3602-12.

114. Alexander-Miller MA, Leggatt GR, and Berzofsky JA. Selective expansion of high- or low-avidity cytotoxic T lymphocytes and efficacy for adoptive immunotherapy. *Proceedings of the National Academy of Sciences of the United States of America.* 1996;93(9):4102-7.

115. Belyakov IM, Kuznetsov VA, Kelsall B, Klinman D, Moniuszko M, Lemon M, Markham PD, Pal R, Clements JD, Lewis MG, et al. Impact of vaccine-induced mucosal high-avidity CD8+ CTLs in delay of AIDS viral dissemination from mucosa. *Blood.* 2006;107(8):3258-64.

116. Derby M, Alexander-Miller M, Tse R, and Berzofsky J. High-avidity CTL exploit two complementary mechanisms to provide better protection against viral infection than low-avidity CTL. *Journal of immunology.* 2001;166(3):1690-7.

117. Dzutsev AH, Belyakov IM, Isakov DV, Margulies DH, and Berzofsky JA. Avidity of CD8 T cells sharpens immunodominance. *International immunology.* 2007;19(4):497-507.

118. Kitazono T, Okazaki T, Araya N, Yamano Y, Yamada Y, Nakamura T, Tanaka Y, Inoue M, and Ozaki S. Advantage of higher-avidity CTL specific for Tax against human T-lymphotropic virus-1 infected cells and tumors. *Cellular immunology.* 2011;272(1):11-7.

119. Kattan T, MacNamara A, Rowan AG, Nose H, Mosley AJ, Tanaka Y, Taylor GP, Asquith B, and Bangham CR. The avidity and lytic efficiency of the CTL response to HTLV-1. *Journal of immunology.* 2009;182(9):5723-9.

120. O'Connor DH, Allen TM, Vogel TU, Jing P, DeSouza IP, Dodds E, Dunphy EJ, Melsaether C, Mothe B, Yamamoto H, et al. Acute phase cytotoxic T lymphocyte escape is a hallmark of simian immunodeficiency virus infection. *Nature medicine.* 2002;8(5):493-9.

121. Yerly D, Heckerman D, Allen TM, Chisholm JV, 3rd, Faircloth K, Linde CH, Frahm N, Timm J, Pichler WJ, Cerny A, et al. Increased cytotoxic T-lymphocyte epitope variant cross-recognition and functional avidity are associated with hepatitis C virus clearance. *Journal of virology.* 2008;82(6):3147-53.

122. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, Bornstein E, Asher TE, Samri A, Schnuriger A, Theodorou I, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of experimental medicine.* 2007;204(10):2473-85.

123. Almeida JR, Sauce D, Price DA, Papagno L, Shin SY, Moris A, Larsen M, Pancino G, Douek DC, Autran B, et al. Antigen sensitivity is a major determinant of CD8+ T-cell polyfunctionality and HIV-suppressive activity. *Blood.* 2009;113(25):6351-60.

124. Berger CT, Frahm N, Price DA, Mothe B, Ghebremichael M, Hartman KL, Henry LM, Brenchley JM, Ruff LE, Venturi V, et al. High-functional-avidity cytotoxic T lymphocyte responses to HLA-B-restricted Gag-derived epitopes associated with relative HIV control. *Journal of virology.* 2011;85(18):9334-45.

125. Mothe B, Llano A, Ibarrondo J, Zamarreno J, Schiaulini M, Miranda C, Ruiz-Riol M, Berger CT, Herrero MJ, Palou E, et al. CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. *PloS one.* 2012;7(1):e29717.

126. Conrad JA, Ramalingam RK, Smith RM, Barnett L, Lorey SL, Wei J, Simons BC, Sadagopal S, Meyer-Olson D, and Kalams SA. Dominant clonotypes within HIV-specific T cell responses are programmed death-1high and CD127low and display reduced variant cross-reactivity. *Journal of immunology.* 2011;186(12):6871-85.

127. Harari A, Cellerai C, Bellutti Enders F, Kostler J, Codarri L, Tapia G, Boyman O, Castro E, Gaudieri S, James I, et al. Skewed association of polyfunctional antigen-specific CD8 T cell populations with HLA-B genotype. *Proceedings of the National Academy of Sciences of the United States of America.* 2007;104(41):16233-8.

128. Lichterfeld M, Yu XG, Mui SK, Williams KL, Trocha A, Brockman MA, Allgaier RL, Waring MT, Koibuchi T, Johnston MN, et al. Selective depletion of high-avidity human immunodeficiency virus type 1 (HIV-1)-specific CD8+ T cells after early HIV-1 infection. *Journal of virology.* 2007;81(8):4199-214.

129. Chen H, Piechocka-Trocha A, Miura T, Brockman MA, Julg BD, Baker BM, Rothchild AC, Block BL, Schneidewind A, Koibuchi T, et al. Differential neutralization of human immunodeficiency virus (HIV) replication in autologous CD4 T cells by HIV-specific cytotoxic T lymphocytes. *Journal of virology.* 2009;83(7):3138-49.

130. Draenert R, Verrill CL, Tang Y, Allen TM, Wurcel AG, Boczanowski M, Lechner A, Kim AY, Suscovich T, Brown NV, et al. Persistent recognition of autologous virus by high-avidity CD8 T cells in chronic, progressive human immunodeficiency virus type 1 infection. *Journal of virology.* 2004;78(2):630-41.

131. Moffat JM, Handel A, Doherty PC, Turner SJ, Thomas PG, and La Gruta NL. Influenza epitope-specific CD8+ T cell avidity, but not cytokine polyfunctionality, can be determined by TCRbeta clonotype. *Journal of immunology.* 2010;185(11):6850-6.

132. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, Keele BF, Learn GH, Turnbull EL, Salazar MG, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *The Journal of experimental medicine.* 2009;206(6):1253-72.

133. Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, and Tanaka K. Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science.* 2007;316(5829):1349-53.

134. Klein L, Kyewski B, Allen PM, and Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nature reviews Immunology.* 2014;14(6):377-91.

135. Su LF, Kidd BA, Han A, Kotzin JJ, and Davis MM. Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity.* 2013;38(2):373-83.

136. Chen W, Bennink JR, Morton PA, and Yewdell JW. Mice deficient in perforin, CD4+ T cells, or CD28-mediated signaling maintain the typical immunodominance hierarchies of CD8+ T-cell responses to influenza virus. *Journal of virology.* 2002;76(20):10332-7.

137. Gallimore A, Dumrese T, Hengartner H, Zinkernagel RM, and Rammensee HG. Protective immunity does not correlate with the hierarchy of virus-specific cytotoxic T cell responses to naturally processed peptides. *The Journal of experimental medicine.* 1998;187(10):1647-57.

138. van der Most RG, Sette A, Oseroff C, Alexander J, Murali-Krishna K, Lau LL, Southwood S, Sidney J, Chesnut RW, Matloubian M, et al. Analysis of cytotoxic T cell responses to dominant and subdominant epitopes during acute and chronic lymphocytic choriomeningitis virus infection. *Journal of immunology.* 1996;157(12):5543-54.

139. La Gruta NL, Rothwell WT, Cukalac T, Swan NG, Valkenburg SA, Kedzierska K, Thomas PG, Doherty PC, and Turner SJ. Primary CTL response magnitude in mice is determined by the extent of naive T cell recruitment and subsequent clonal expansion. *The Journal of clinical investigation.* 2010;120(6):1885-94.

140. van Heijst JW, Gerlach C, Swart E, Sie D, Nunes-Alves C, Kerkhoven RM, Arens R, Correia-Neves M, Schepers K, and Schumacher TN. Recruitment of antigen-specific CD8+ T cells in response to infection is markedly efficient. *Science.* 2009;325(5945):1265-9.

141. den Braber I, Mugwagwa T, Vrisekoop N, Westera L, Mogling R, de Boer AB, Willems N, Schrijver EH, Spierenburg G, Gaiser K, et al. Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. *Immunity.* 2012;36(2):288-97.

142. Hassan C, Kester MG, de Ru AH, Hombrink P, Drijfhout JW, Nijveen H, Leunissen JA, Heemskerk MH, Falkenburg JH, and van Veelen PA. The human leukocyte antigen-presented ligandome of B lymphocytes. *Molecular & cellular proteomics : MCP.* 2013;12(7):1829-43.

143. Wherry EJ, Blattman JN, Murali-Krishna K, van der Most R, and Ahmed R. Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *Journal of virology.* 2003;77(8):4911-27.

144. Crowe SR, Turner SJ, Miller SC, Roberts AD, Rappolo RA, Doherty PC, Ely KH, and Woodland DL. Differential antigen presentation regulates the changing patterns of CD8+ T cell immunodominance in primary and secondary influenza virus infections. *The Journal of experimental medicine*. 2003;198(3):399-410.

145. Chen W, Pang K, Masterman KA, Kennedy G, Basta S, Dimopoulos N, Hornung F, Smyth M, Bennink JR, and Yewdell JW. Reversal in the immunodominance hierarchy in secondary CD8+ T cell responses to influenza A virus: roles for cross-presentation and lysis-independent immunodomination. *Journal of immunology*. 2004;173(8):5021-7.

146. Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, Ozkan E, Davis MM, Wucherpfennig KW, and Garcia KC. Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell*. 2014;157(5):1073-87.

A

Appendix

Summary

The immune system

We are frequently exposed to agents that can cause disease, such as bacteria and viruses. Yet, and fortunately, we are not permanently sick. Our immune system, with its many fronts of defense, protects us from several of the 'intruders' – pathogens -, either by avoiding their entrance, or by impeding or controlling their dissemination in the body. It does so by recognizing pathogens' structures as different from the body's *self* 'building materials': the pathogen is foreign, *non-self*. The cells from the innate arm of the immune system distinguish patterned structures common to groups of pathogens (e.g.: the cell membrane of bacteria). In contrast, the adaptive immune system, even though it takes time to be working full force, is more specific and potent, and most importantly, it carries the ability to remember the particular pathogen it once reacted to, i.e. it develops (immunological) memory. This way, if its cells encounter the pathogen again, the reaction will be faster, even more potent, and potentially stop the (progression of the) infection at a point when no, or few symptoms, are evident. The cells of the adaptive immune system, the lymphocytes, owe their specificity to the receptors they have at their surface. These receptors allow the recognition of distinct fractions of the pathogens, and trigger the cell to act. A particular set of lymphocytes is responsible for killing the cells of the body that have been infected by intracellular pathogens (viruses, some bacteria): the $CD8^+$ T cells or cytotoxic *killer* T lymphocytes (CTLs).

Recognition of *non-self* by $CD8^+$ T cells

The cells of the body maintain their integrity and perform their functions during their lifespan by producing and renewing the structures they are made of. The major building blocks of the cells - *proteins* - are, thus, constantly produced, degraded and destroyed (into smaller fragments called *peptides*) inside a cell. A fraction of peptides of the cells' own proteins (*self* peptides) is shown at the cell surface bound to structures called HLA class-I molecules (proteins), the most abundant ones being HLA-A and HLA-B molecules. In general, $CD8^+$ T cells have been 'educated' to see the complexes of HLA with the self-peptide and not react to them so that the body is spared from attack by its own defense cells; when this safe-mechanism is not working properly, auto-

immune diseases ensue. When a virus infects a cell, it takes hold of the building machinery of the cell to make its own progeny. However, also the viral proteins will be degraded, and consequently a fraction of (*non-self*) viral peptides will bind to the HLA and head to the cell surface, flagging danger. $CD8^+$ T cells will recognize some of these *non-self* peptide-HLA complexes and become activated to respond, destroying the infected cell. Peptides, or *antigens*, that are able to bind to HLA molecules are frequently named *ligands*, and those that additionally are able to activate an immune response are called *epitopes*.

Peptides are made of smaller units, named amino acids. For the peptides to bind to the HLA, they have to be compatible: their amino acids must be able to bind to the amino acids of the HLA protein. Each individual has 2 types of HLA-A and 2 types of HLA-B proteins, inherited from each of the parents. As there are thousands of different HLA (-A or –B) proteins in the human population, several combinations of HLA are possible in one person, and they can, coincidentally, be the same. (For example, one person can have two HLA-A*02 proteins, one HLA-B*07 and one HLA-B*44 protein: this means that both parents had HLA-A*02 in their genes, while they differed on the HLA-B that they passed on to their child). Given the thousands of different HLA proteins in the human population, each with their own specific binding requirements, it is no wonder that individuals react differently to the very same viral infection. Above all, the HLA combination of an individual (HLA *background*) will impact the efficiency of the immune response that ensues.

Each individual will, thus, mount a fundamentally different immune response against a particular pathogen, composed of a group of T cells with distinct specificities i.e.: they recognize different pathogen-derived, peptide-HLA complexes. Still, certain T-cell specificities are found more often in the blood of infected individuals, and often to higher numbers. These T-cell responses, of higher *frequency* in the infected population, and/or of higher *magnitude* in the individual, are called *immunodominant*.

Both HLA-A and HLA-B are similar in (3D)-structure and present peptides at the cell surface. However, they are different in some aspects: 1) there are more 'variants' of HLA-B proteins in the human population; 2) individuals with particular HLA-B proteins seem to be more resistant, or more susceptible, to certain infections; and 3) several studies have reported more *immunodominant* responses of $CD8^+$ T cells to peptide-HLA-B complexes.

In this thesis, we investigated if variables intrinsic to the peptide set presented to $CD8^+$ T cells (peptide *repertoire*) could contribute to the occurrence of *immunodominant* T-cell responses (chapters 2 to 4). In addition (chapters 5 to 7), we looked into features of the T-cell response itself that may impact immunodominance.

Peptide repertoire presented by HLA-A versus HLA-B

In chapters 2 to 4 of this thesis, we analyzed whether there would be differences in the peptides that HLA-A and HLA-B molecules present that could contribute to the findings stated above. In **Chapter 2** we studied whether HLA-B proteins are able to present more peptides (a more diverse *ligand* repertoire), and hence call more $CD8^+$ T cells into action, and whether such peptides bind more strongly to the HLA-B, so the T cells could see the peptide more often and/or connect stronger with it. Previous research has led to the development of computer programs that can predict which peptides will bind to which HLA protein. These programs are based on the systematic assessment of patterns in the amino acid sequence of peptides (*binding motifs*) that have been identified in lab experiments to bind with enough strength to certain HLA molecules. We have made use of sets of pathogen-derived $CD8^+$ T-cell epitopes that have been discovered experimentally, reported and collected in a database (Immune Epitope DataBase, IEDB), and of peptides that we predicted as binders using computer predictor programs - *in silico* predicted peptides. In addition, we predicted which peptides from the human proteins (*self*) would be binders. In contrast to our expectations, HLA-B proteins bound fewer peptides, and with lower strength, than HLA-A molecules. Also, HLA-A molecules had the ability to bind more *self*-peptides. Thus, the immunodominance of T-cell responses reacting to peptides presented by HLA-B, as part of the total response against pathogens, cannot be explained by the presentation of a more diverse set of peptides, nor by a greater binding strength to these molecules.

$CD8^+$ T cells will detect infected cells presenting pathogen-derived peptide-HLA complexes that 'fit' the requirements of their particular receptor (T-cell receptor, TCR). These specific $CD8^+$ T cells can, consequently, be 'blind-sided' and not react if there are changes in that complex, for example if it is no longer presented at the cell surface (because of changes – *mutations* – in the peptide that do not allow it to use the HLA as a scaffold), or it is no longer recognized (because mutations in the amino acids of the peptide make the complex look different). This occurs when pathogens, while 'propagating' (producing progeny, *replicating*), make mistakes in building up their proteins that are missed by the 'quality control'. While some of these errors might ensure that the pathogen *replicates* better because, for instance, the infected cell is not detected and killed by the $CD8^+$ T cells, many others lead to reduced fitness of the pathogen, i.e.: it no longer functions properly, so ultimately, if it does not mutate back – *revert* -, its progeny/lineage will 'die out'. The best-case scenarios for the infected individual are: i) that the errors do not occur (the peptide is *conserved*), and thus the

$CD8^+$ T cells keep recognizing the infected cells and killing them; or, if errors occur, ii) the complex is still formed and recognized by the T cells (the mutations do not (totally) hamper binding to HLA nor recognition of infected cells), or iii) the changes in the protein affect its function or structure so much that the lineage of the mutated pathogen will die out. In other words, it is beneficial for a T-cell response if the peptide it targets is conserved and *constrained*, i.e. cannot mutate without affecting the pathogen's 'survival'. In **Chapter 3**, we analyzed the proteins of the human immunodeficiency virus (HIV-1), and mapped the amino acids that are part of known epitopes *restricted* by HLA-A, HLA-B or both (or HLA-C), based on the database HIV-LANL (collects verified HIV-1-derived epitopes). In addition, we used HIV-1 protein sequences found in patients collected in the same database to determine to which level each amino acid in each protein has been found to mutate. When combining this information, we were able to detect that amino acids within HLA-B epitopes are more conserved than those within HLA-A epitopes. This might mean that the amino acids that bind HLA-B molecules mutate less or revert more, because mutations at those positions are too 'harmful' for the virus. In our additional analysis of published data from another group, we found indications that mutations in HLA-B epitopes might in fact revert more.

Published reports and our analysis in **Chapter 4** looked into the binding promiscuity of HLA class-I epitopes, again experimentally verified (and collected in IEDB) and *in silico*-predicted peptides. A peptide is called promiscuous if it can bind at least two HLA molecules. If HLA-B molecules would tend to share more peptides (have a more promiscuous binding set), more individuals in the population could have T-cell responses to such peptides, hence contributing to the more frequent detection – *immunodominance* - of HLA-B-restricted $CD8^+$ T cell responses. However, we did not find a significant and consistent difference between the levels of promiscuity of the peptide sets of HLA-B versus HLA-A molecules. In fact, there even seem to exist peptides with the ability to bind both HLA-A and HLA-B molecules.

Circulating antiviral $CD8^+$ T-cell responses

When recognizing their antigen (peptide complexed to the HLA class I molecule), $CD8^+$ T cells become activated: they undergo a series of events that lead to their division into daughter cells that carry the same TCR (*proliferation* or *clonal expansion*), and to changes in their functioning (*differentiation*) that allows them to, directly or indirectly, kill infected cells that present the same (or similar) antigen at their surface via HLA. These

pre-set (TCR specificity) and new abilities (*killing* upgrades) could potentially be related to *immunodominance*.

Each specific $CD8^+$ T cell (and its progeny) has its own T-cell receptor (TCR), with which it recognizes a (limited number of) peptide-HLA complex(es). In turn, there is normally a group of T cells, more or less diverse, that recognizes a given peptide-HLA complex. But as the set of peptide-HLA complexes exposed by infected cells differs between individuals (as they often differ in HLA *background*), so does the set of T cells that recognize that set of pathogen-derived epitopes. The diversity of the T-cell population (i.e. the T-cell repertoire diversity) is, thus, a complex but important feature that will impact the reaction to the infection.

In chapters 5 to 7, we analyzed some features of the T cells that recognize epitopes from different viruses. We studied viruses that i) cause lifelong infection but are typically relatively harmless and kept under control in healthy individuals (EBV, HCMV – Chapter 5); ii) cause lifelong infection that is not usually kept under control and has life-threatening repercussions (HIV-1 – Chapter 6); and iii) cause acute infection and are eliminated from the body, but can re-enter and re-infect it (influenza A – Chapter 7).

In **Chapter 5,** we looked into variables that could influence how diverse a T-cell response to a given pathogen-derived peptide-HLA complex is. We used EBV and HCMV as model viruses, as at least half of the human population is thought to carry these viruses, and $CD8^+$ T-cell responses directed to them can be found in the circulating blood. In particular, we isolated cells that recognized certain virus-specific peptide-HLA complexes from the blood of healthy individuals, and assessed how diverse (how many different TCRs) these populations are. In this study, we found that it was neither the infecting virus, nor the specific presenting HLA group (HLA-A versus HLA-B) or molecule (e.g.: HLA-A*02 versus A*11), but rather the presented peptide that seemed to guide whether more or less TCRs reacted to it, i.e. whether the T-cell repertoire was narrower, or more diverse. The diversity of the responding T-cell population was not related to whether the particular response was immunodominant.

While EBV and HCMV are considered relatively stable, HIV-1 has been shown to mutate its proteins very frequently. This is one of the reasons why the virus is so hard to control, and till now impossible to eliminate (with one hopeful exception). When the peptides mutate at certain amino acid positions, at least some of the $CD8^+$ T cells may fail to detect them and fail to kill the infected cells. In **Chapter 6**, we show that, in patients infected with HIV-1 who are not under antiviral medication, larger changes in the total size of an antigen-specific T-cell population (its *magnitude*) frequently occur together with alterations in its composition (changes in the TCR repertoire) and with

mutations in the antigen/peptide itself. This general picture, not without exceptions, illustrates that both virus and the immune system challenge each other in a struggle for *survival*. However, it is hard to draw generalized conclusions on how effective the immune system is acting as a whole at any given time point during infection, because the available T-cell responses specific for distinct peptides may evolve differently, or remain stable, and may change without parallel changes in the amount of circulating virus. In **Chapter 7,** we made use of blood samples collected from an individual before and after a re-infection with the flu virus and compared features of an immunodominant T-cell response. Upon new flu infection, $CD8^+$ T cells that had previously recognized a conserved influenza epitope, and hence were part of the previously mentioned *immunological memory*, were joined by new cells (new TCRs) called into action, which then persisted in the 'new' *immunological memory* formed. These cells seemed to become more versatile in function after encountering the flu epitope again, but these changes were temporary, in contrast to the stable alterations in TCR repertoire.

Main Conclusions

The immune responses in general, and the $CD8^+$ T cells and their targets (peptides) herein studied in particular, are set upon extremely dynamic and complex interactions. These interactions take place not only between the different types of cooperating immune cells and the infected cells they detect, but also between the immune cells and the pathogens themselves. The imprint of the interactions between $CD8^+$ T cells and pathogens is potentially noticeable in T-cell repertoire changes and in a landscape of mutations in the pathogen, respectively.

In conclusion, we have shown that the reported immunodominance of HLA-B-restricted T-cell responses stems neither from a greater or stronger-binding peptide repertoire when compared to HLA-A (chapter 2), nor from a difference in the peptide-binding promiscuity of the two types of HLA molecules (chapter 4), nor from differences in the diversity of the responding T-cell population (chapter 5). We provided some indications that, at least for HIV-1, HLA-B molecules may target regions of the viral proteins that are less likely to mutate without negatively impacting the virus. Consequently, those responses are probably more likely detectable in infected individuals (chapter 3). In addition, the peptide that is being presented by the HLA molecule has a dominant role in favoring either narrower or more diverse responding T-cell repertoires (chapter 5). There is a clear interplay between T cells and the epitopes

they recognize, and while we cannot prove causality, both seem to drive the other to evolve. When a given epitope mutates (chapter 6), or when the pre-existing cells do not proliferate (enough) when re-encountering this original trigger (chapter 7), the immune system, if not yet exhausted, will have the flexibility to bring new, or less used, specific cells (with new TCRs) into action.

It is of great interest to understand the targets (epitopes) of $CD8^+$ T cells, and how they affect or (re)direct the type of T-cell response that is built up, and saved as memory. This will have a direct influence in the field of T-cell vaccine research, when considering the antigens that should be incorporated in the design of effective vaccines with low 'collateral damage'.

Nederlandse Samenvatting

Het immuunsysteem

We worden voortdurend blootgesteld aan stoffen of agentia die ziektes veroorzaken, zoals bacteriën en virussen. Toch zijn wij gelukkig niet constant ziek. Ons immuunsysteem, met zijn vele verdedigingsmechanismen, beschermt ons tegen een aantal van de 'indringers' - ziekteverwekkers -, hetzij door te vermijden dat ze het lichaam binnendringen, of door hun verspreiding in het lichaam te belemmeren of te beperken. Dit gebeurt door het herkennen van de ziekteverwekkers als 'anders dan de lichaamseigen bouwblokken': de ziekteverwekker is vreemd, '*niet-eigen*'. De cellen van het aangeboren immuunsysteem kunnen structuren onderscheiden die groepen van ziekteverwekkers gemeen hebben (b.v.: het celmembraan van bacteriën). Het adaptieve ('verworven') immuunsysteem daarentegen, al kost het wat tijd eer het op volle kracht werkt, is veel specifieker en potenter. Het is vooral belangrijk dat het verworven immuunsysteem het vermogen bezit om specifieke ziekteverwekkers te onthouden zodra ze ertegen gereageerd heeft, het ontwikkelt een (immunologisch) geheugen. Hierdoor zal de reactie tegen de ziekteverwekker de volgende keer sneller en krachtiger zijn waardoor de ontwikkeling van een infectie tegengegaan kan worden op een moment waarop er nog weinig of geen symptomen zijn. De cellen van het verworven immuunsysteem, de lymfocyten, danken hun specificiteit aan de receptoren die zich aan de buitenkant van die cellen bevinden. Deze receptoren zorgen voor een gerichte herkenning van bepaalde delen van de ziekteverwekkers en zetten zo de cel aan tot actie. Een bepaalde groep van lymfocyten is verantwoordelijk voor het doden van lichaamseigen cellen die geïnfecteerd zijn door intracellulaire ziekteverwekkers (virussen, sommige bacteriën): de $CD8^+$ T-cellen of de cytotoxische *killer* T-lymfocyten (CTLs).

Herkenning van *lichaamsvreemd* door $CD8^+$ T-cellen

Onze lichaamscellen behouden hun integriteit en voeren hun taken uit gedurende hun levensduur door de onderdelen waaruit ze zijn opgebouwd te produceren en te vernieuwen. De belangrijkste van deze bouwblokken – *eiwitten* - worden dus continu gevormd en afgebroken (in kleinere delen, *peptiden* genaamd) binnenin een cel. Een deel van de peptiden van de celeigen eiwitten wordt aan het celoppervlak gepresenteerd

verbonden aan structuren die HLA type I moleculen (eiwitten) genoemd worden; de meest voorkomende onder hen zijn de HLA-A en de HLA-B eiwitten. De $CD8^+$ T-cellen hebben de 'instructie' meegekregen om de complexen gevormd van de HLA eiwitten met celeigen peptiden op te merken, maar er niet op te reageren, zodat het lichaam niet door zijn eigen verdedigingscellen wordt aangevallen; Als dit toch gebeurt, spreekt men van auto-immuun ziektes. Wanneer een virus de cel infecteert, zal het het bouwend vermogen van de cel overnemen om zichzelf te kunnen vermenigvuldigen. Ook de virale eiwitten zullen dan afgebroken worden. Als gevolg daarvan zullen een deel van de (*niet-eigen*) virale peptiden binden aan die uithangborden en zo het gevaar uitdragen. $CD8^+$ T-cellen kunnen die peptide-HLA complexen herkennen en op die manier geactiveerd worden en de geïnfecteerde cel vernietigen. Peptiden, of *antigenen*, die aan HLA eiwitten kunnen binden worden l*iganden* genoemd. De liganden die vervolgens een immuunreactie kunnen uitlokken worden *epitopen* genoemd.

Peptiden zijn opgebouwd uit nog kleinere stukjes, aminozuren genaamd. Om te kunnen binden aan de HLA eiwitten moeten de aminozuren van de peptide kunnen binden aan de aminozuren van het HLA-eiwit. Elk van ons heeft 2 types HLA-A en 2 types HLA-B eiwitten, geërfd van onze ouders. Aangezien er duizenden verschillende HLA (-A of –B) eiwitten bestaan in de menselijke populatie, zijn er meerdere combinaties mogelijk in 1 persoon en de combinatie kan bij wijze van toeval ook identiek zijn. (Bijvoorbeeld, iemand kan 2 HLA-A*02 eiwitten hebben, 1 HLA-B*07 en 1 HLA-B*44 eiwit: dit betekent dat beide ouders HLA-A*02 eiwitten in hun genen hadden, terwijl ze verschilden op het vlak van de HLA-B eiwitten die ze aan de volgende generatie meegaven). Als je daar dan nog eens bijneemt dat elk van de HLA eiwitten specifieke bindingsmechanismen heeft, is het niet verwonderlijk dat mensen individueel verschillende reacties hebben op eenzelfde virale infectie. De individuele HLA-combinatie (HLA *achtergrond*) van een persoon bepaalt de efficiëntie van zijn immuunreacties.

Elk van ons zal dus een fundamenteel verschillende immuunreactie vertonen tegen een zekere ziekteverwekker, samengesteld uit een groep van T-cellen met onderscheiden kenmerken i.e.: ze herkennen verschillende ziekteverwekker-afkomstige peptide-HLA complexen. Bepaalde T-cel kenmerken worden vaker in het bloed van geïnfecteerde individuen aangetroffen en regelmatig ook in grotere omvang. De T-celreacties, met een hogere frequentie in de geïnfecteerde populatie en/of een grotere omvang in het individu, worden *immunodominant* genoemd.

HLA-A en HLA-B lijken op elkaar wat hun (3D)-structuur betreft en beide presenteren peptiden op het celoppervlak. Ze verschillen echter in een aantal opzichten: 1) er zijn meer HLA-B varianten in de menselijke populatie; 2) mensen met bepaalde

HLA-B eiwitten lijken resistenter tegen, of juist vatbaarder voor, sommige infecties; en 3) verschillende studies beschrijven dat $CD8^+$ T-cel reacties tegen peptide-HLA-B complexen meer *immunodominant* zijn.

In deze thesis hebben we onderzocht of er variabelen bestaan, inherent aan de peptiden die aan de $CD8^+$ T-cellen gepresenteerd worden (peptiderepertoire), die bijdragen aan de *immunodominantie* van T-celreacties (hoofdstukken 2 tot 4). Daarnaast (hoofdstukken 5 tot 7), hebben we kenmerken van de T-celreacties zelf bestudeerd, en hun eventuele invloed op de *immunodominantie*.

Peptiden gepresenteerd door HLA-A versus HLA-B

In hoofdstuk 2, 3 en 4 van dit proefschrift hebben we onderzocht of er verschillen zijn in de peptiden die door HLA-A en HLA-B eiwitten gepresenteerd worden die de bovenstaande verschillen mogelijk kunnen verklaren. In **Hoofdstuk 2** onderzochten we of HLA-B eiwitten meer verschillende peptiden presenteren (een meer divers *ligand* repertoire), en op die manier meer $CD8^+$ T-cellen activeren. Ook onderzochten we of peptiden sterker binden aan HLA-B eiwitten, zodat de T-cellen deze peptiden vaker zouden kunnen opmerken en/of er sterker aan zouden kunnen binden. Eerder onderzoek heeft geleid tot de ontwikkeling van computerprogramma's die voorspellen welke peptide op welk HLA eiwit kan binden. Deze programma's zijn gebaseerd op patronen in de aminozuur volgorde van peptiden (*binding motief*) die in laboratorium experimenten voldoende sterk bonden aan bepaalde HLA eiwitten. We maakten gebruik van zowel bekende $CD8^+$ T-cel epitopen van verschillende pathogenen (die experimenteel ontdekt zijn, en beschreven staan in een databank (Immune Epitope Database, IEDB)), en van peptiden waarvan wij voorspelden dat ze zouden binden. De voorspellingen deden we met behulp van zo'n computerprogramma – *in silico* voorspelde peptiden. Verder voorspelden we welke lichaams*eigen* peptiden zouden binden. In tegenstelling tot onze verwachtingen vonden we dat HLA-B eiwitten minder verschillende peptiden kunnen binden, en minder sterk, dan HLA-A eiwitten. HLA-A eiwitten bonden ook nog eens meer *eigen* peptiden. Dus de immunodominantie van T-cel reacties op door HLA-B gepresenteerde peptiden kon niet worden verklaard door een grotere diversiteit aan peptiden die gepresenteerd worden, noch door een sterkere binding aan HLA eiwitten.

$CD8^+$ T-cellen kunnen geïnfecteerde cellen herkennen aan de hand van de peptide-HLA complexen die exact passen op hun receptoren (T-cel receptoren, TCR). Deze specificiteit is tevens ook hun zwakke schakel. De $CD8^+$ T-cellen zullen niet reageren als

het complex niet meer aan het celoppervlak gepresenteerd wordt (omdat veranderingen – *mutaties* - in een peptide ervoor zorgen dat het de HLA eiwitten niet meer als uithangbord kan gebruiken), of als het complex niet meer herkend wordt (omdat de vorm van het complex veranderd is door mutaties in de aminozuren van een peptide). Deze veranderingen treden op wanneer ziekteverwekkers fouten maken in de opbouw van hun eiwitten terwijl ze zich vermenigvuldigen (reproduceren, *repliceren*), die niet door de 'kwaliteitscontrole´ worden opgemerkt. Sommige van deze fouten kunnen ervoor zorgen dat de ziekteverwekker zich beter kan vermenigvuldigen, omdat bijvoorbeeld de geïnfecteerde cel niet meer wordt opgemerkt en vernietigd door de $CD8^+$ T-cellen. Veel van die fouten zullen echter leiden tot een verzwakking van de ziekteverwekker: het zal niet meer normaal functioneren en zijn nakomelingen zullen uiteindelijk, als de mutatie niet wordt teruggedraaid, uitsterven. Het beste scenario voor een geïnfecteerd individu is dat er zich geen fouten voordoen (de peptide blijft precies zoals hij was) en de $CD8^+$ T-cel gewoon zijn werk kan doen. Als er toch fouten optreden dan is het het beste als het peptide-HLA complex ondanks de fout nog steeds gevormd en herkend wordt door de T-cellen, of dat de fout het functioneren van de ziekteverwekker dermate in het gedrang brengt dat zijn 'gemuteerde' nakomelingen uitsterven. Met andere woorden, een T-cel reactie heeft er baat bij als de peptide die zijn doelwit is niet verandert, en '*constrained*' is, i.e. niet kan muteren zonder de 'overlevingskans' van de ziekteverwekker te ondermijnen. In **Hoofdstuk 3** analyseerden we de eiwitten van het humaan immunodeficiëntievirus (HIV-1), en brachten we de aminozuren in kaart die deel uitmaken van bekende epitopen die gepresenteerd worden ('*restricted*') door HLA-A, HLA-B of beiden (of HLA-C), gebaseerd op de HIV-LANL database (een database waarin geverifieerde HIV-1 epitopen verzameld worden). Daarnaast gebruikten we HIV-1 eiwitsequenties, aangetroffen in patiënten, uit diezelfde database, om na te gaan in welke mate elk van de aminozuren in elk eiwit muteert. Deze informatie combinerend, vonden we dat de aminozuren in HLA-B epitopen beter bewaard blijven dan die in HLA-A epitopen. Dat zou kunnen betekenen dat de aminozuren van peptiden die HLA-B-eiwitten binden ofwel minder vaak muteren, ofwel vaker de mutatie terugdraaien, omdat mutaties op die posities te 'nadelig' zijn voor het virus. In onze bijkomende analyse van gepubliceerde data van een andere groep, stelden we indicaties vast dat mutaties in HLA-B epitopen mogelijks vaker worden teruggedraaid.

In **Hoofdstuk 4** onderzochten we de binding promiscuïteit van HLA type I epitopen, zowel van experimenteel geverifieerde (verzameld in IEDB) en van *in-silico*-voorspelde peptiden. Een peptide is promiscu als hij tenminste 2 verschillende HLA-eiwitten kan binden. Als HLA-B eiwitten meer peptiden gemeen zouden hebben (een

promiscuere binding set), dan zouden meer individuen in de populatie T-cel reacties kunnen hebben tegen dergelijke peptiden, en zo bijdragen aan een hogere frequentie – *immunodominantie* - van HLA-B-gepresenteerde $CD8^+$ T-cel-reacties. Desondanks hebben we geen significant en consistent verschil gevonden in de mate van promicuïteit van peptidereeksen van HLA-B versus HLA-A eiwitten. Er blijken zelfs peptiden te bestaan die zowel HLA-A als HLA-B eiwitten binden.

Circulerende $CD8^+$ T-cel reacties op virussen

Wanneer $CD8^+$ T-cellen hun antigeen (het complex van peptide en HLA type I eiwit) herkennen, worden ze geactiveerd. Ze ondergaan een aantal gebeurtenissen die leiden tot dochtercellen die dezelfde TCR dragen (*proliferatie* of snelle toename; *klonale expansie*) en tot veranderingen in hun functioneren (*differentiatie*). Dit zorgt ervoor dat ze, zowel direct als indirect, geïnfecteerde cellen die dezelfde (of gelijkwaardige) antigenen op hun celoppervlak via HLA presenteren, kunnen vernietigen. Deze voorgeprogrammeerde (TCR specificiteit) en nieuwe bekwaamheden (*killing* upgrade) zouden in verband kunnen staan met *immunodominantie.*

Elke specifieke $CD8^+$ T-cel (en zijn nageslacht) heeft zijn eigen T-cel receptor (TCR), waarmee hij een (beperkt aantal) peptide-HLA-complex(en) herkent. Daartegenover staat dat een gegeven peptide-HLA-complex door een aantal T-cellen (meer of minder divers) herkend kan worden. Aangezien de set van peptide-HLA complexen die door de geïnfecteerde cellen gepresenteerd worden per individu kan verschillen (omdat mensen verschillen in hun HLA-*achtergrond*), is ook de groep van T-cellen die voor de herkenning van de epitopen van ziekteverwekkers zorgt verschillend. De diversiteit van de T-cel populatie (i.e. de T-cel repertoire diversiteit) is een ingewikkelde, maar belangrijke eigenschap met een grote impact op de reactie van de gastheer op een infectie.

In Hoofdstukken 5, 6 en 7, onderzochten we een aantal eigenschappen van T-cellen die epitopen herkennen van verschillende virussen. We bestudeerden virussen die i) een levenslange infectie veroorzaken, maar relatief onschadelijk zijn en onder controle gehouden worden in gezonde individuen (EBV, HCMV – Hoofdstuk 5); ii) een levenslange infectie veroorzaken, maar niet onder controle gehouden kunnen worden en levensbedreigende gevolgen hebben (HIV-1 – Hoofdstuk 6); en iii) die acute infecties veroorzaken en uit het lichaam geëlimineerd worden, maar het lichaam opnieuw kunnen binnendringen en voor een her-infectie kunnen zorgen (influenza A – Hoofdstuk 7).

In **Hoofdstuk 5**, onderzochten we een aantal variabelen die een invloed zouden kunnen hebben op hoe divers een T-cel reactie op een peptide-HLA complex is. We gebruikten EBV en HCMV als modelvirussen, omdat het merendeel van de populatie drager is van deze virussen, en omdat $CD8^+$ T-cel reacties tegen deze virussen in het bloed teruggevonden kunnen worden. Uit het bloed van gezonde donoren hebben we cellen geïsoleerd die bepaalde virusspecifieke peptide-HLA complexen konden herkennen. Vervolgens hebben we onderzocht hoe divers (hoeveel verschillende TCRs) deze populaties zijn. In deze studie ontdekten we dat niet zozeer het virus, de specifieke HLA groep (HLA-A versus HLA-B) of HLA-eiwit (bv.: HLA-A*02 versus HLA-A*11), bepaalde of er meer of minder TCRs gingen reageren i.e. of het T-cel repertoire divers of niet divers genoemd kon worden, maar juist de gepresenteerde peptide zelf. De diversiteit van de reagerende T-cel populatie bleek niet in verband te staan met de immunodominantie van de T-cel reactie.

EBV en HCMV worden gezien als relatief stabiele virussen. Van HIV-1 is daarentegen bekend dat er heel vaak mutaties optreden. Dat is één van de redenen waarom het zo moeilijk is dit virus onder controle te houden en het vooralsnog onmogelijk is om het te elimineren (op een enkele hoopgevende uitzondering na). Wanneer een peptiden muteren op bepaalde aminozuurposities, zal op z'n minst een deel van de $CD8^+$ T-cellen de geïnfecteerde cellen niet meer kunnen herkennen en vernietigen. In **Hoofdstuk 6** laten we zien dat in HIV-1 geïnfecteerde patiënten die geen antivirale medicatie namen, grote veranderingen in de totale omvang van een antigeenspecifieke T-cel populatie (de *magnitude*) vaak gelijk optreden met veranderingen in de samenstelling (veranderingen in het TCR-repertoire) en met mutaties in het antigeen/peptide. Dit illustreert dat zowel virus als het immunsysteem elkaar uit kunnen dagen in deze overlevingsstrijd. Het is echter moeilijk om een algemeen beeld op te maken van hoe effectief het immuunsysteem in zijn geheel is op een gegeven moment gedurende een infectie, want de voorhanden zijnde T-celreacties tegen bepaalde peptiden kunnen afwijkend evolueren, of stabiel blijven, en kunnen veranderen zonder parallelle veranderingen in de hoeveelheid van circulerend virus. In **Hoofdstuk 7** maakten we gebruik van bloedstalen, genomen bij een individu voor en na her-infectie met griep, en vergeleken we eigenschappen van een immunodominante T-celreactie. $CD8^+$ T-cellen die eerder een stabiel influenza epitoop hadden herkend, en dus deel uitmaakten van het eerder vermelde *immunologische geheugen*, kregen na blootstelling aan een nieuwe griepinfectie de hulp van nieuwe cellen (nieuwe TCRs). Samen vormden al deze TCRs vervolgens het nieuwe *immunologisch geheugen*. Deze cellen leken veelzijdiger te worden in hun functies na een nieuwe blootstelling aan de

griep epitoop, maar deze veranderingen waren slechts tijdelijk, in tegenstelling tot de veranderingen in het TCR-repertoire.

Algemene conclusies

Immuunreacties in het algemeen, en de interactie tussen $CD8^+$ T-cellen en hun doelwitten (peptiden) in het bijzonder, zijn erg dynamisch en complex. Deze interacties vinden niet alleen plaats tussen de verschillende type afweercellen en de geïnfecteerde cellen die ze herkennen, maar ook tussen de afweercellen en de ziekteverwekkers zelf. Gevolgen van de interactie tussen $CD8^+$ T-cellen en ziekteverwekkers zijn mogelijk terug te zien in respectievelijk veranderingen in het T-cel repertoire, en in een landschap van mutaties in de ziekteverwekkers.

Tot slot hebben we aangetoond dat de beschreven immunodominantie van HLA-B-gecontroleerde T-cel reacties niet voortkomt uit een groter of beter bindend peptiderepertoire in vergelijking met HLA-A (hoofdstuk 2), noch uit een verschil in peptidebindende promiscuïteit van beide types HLA-eiwitten (hoofdstuk 4), noch uit een verschil in diversiteit van de reagerende T-cel populatie (hoofdstuk 5). We verschaften aanwijzingen dat, tenminste voor HIV-1, HLA-B eiwitten vooral delen van de virale eiwitten presenteren die minder vaak muteren zonder negatieve gevolgen voor het virus zelf. Als gevolg daarvan worden de T-cel reacties tegen deze peptiden mogelijk vaker gevonden in geïnfecteerde individuen (hoofdstuk 3). Daarnaast speelt de door een HLA-eiwit gepresenteerde peptide zelf een dominante rol in het bepalen van de voorkeur voor een meer of minder divers T-cel repertoire (hoofdstuk 5). Er is een duidelijke interactie tussen T-cellen en de epitopen die ze herkennen, en ondanks dat we geen oorzakelijk verband kunnen bewijzen, lijken beiden afzonderlijk de ander aan te zetten tot evolueren. Wanneer een epitoop muteert (hoofdstuk 6), of wanneer de bestaande T cellen niet (genoeg) vermenigvuldigen bij hernieuwde blootstelling aan een initiële trigger (hoofdstuk 7), zal het immuunsysteem de flexibiliteit vertonen om nieuwe, of minder gebruikte, specifieke cellen (met nieuwe TCRs) te mobiliseren (op voorwaarde dat het nog niet is uitgeput).

Het is van groot belang om de doelwitten (epitopen) van $CD8^+$ T-cellen beter te begrijpen, en hoe zij de soort T-cel reactie die wordt opgebouwd en als geheugen wordt opgeslagen, beïnvloeden of (om)leiden. Kennis hierover zal een directe invloed hebben op T-cel vaccin onderzoek, omdat we dan beter kunnen bepalen welke antigenen moeten worden opgenomen in de ontwikkeling van een effectief vaccin met weinig 'collateral damage'.

Acknowledgements

The years I spent devoting myself to this PhD project have been of learning and changing, not only professionally but also personally. Here, I would like to express my sincere gratitude to those who 'walked' beside me.

Debbie, thank you for giving me the opportunity of not only doing an internship, but also of joining your group afterwards as a PhD student. I really appreciated the guidance, the fruitful meetings and pragmatism, the support and availability. Frank, thank you for being my promotor. The meetings we had together were no doubt stimulating.

Can, José, you have been there for me professionally and personally. I am grateful for your teachings, critical input, support, and especially for believing.

I thank the members of the reading committee for assessing my thesis.

Rao, Dan, I really appreciate all the work we've done together, the discussions, the exchange of thoughts and ideas. Rao, thanks for all the help, even when already in a new stage of your career! Dan, thank you for everything, it was great to have worked together toward a few papers!

Group Debbie, thank you all! I have crossed paths with a lot of you, including Hilde, Ingrid, Nening, Sanne, Esther, Bart, Thijs, Joop, Justin, Soeradj, and the older 'crew' Jolanda, Charlotte, Anne, Floor, Leonie, Ronald… and the students, Evelien, Karlijn, Marietta, and especially Stephanie (as my part-time student! ☺)… Thank you all for the *gezelligheid,* the enthusiasm, support and fruitful work-discussions. Ingrid, first 'teacher', very good friend, thank you for all the work and non-work related talks, discussions, advice and support! Nen, Sanne, I value all the hard work, long days elispot-ing, I am grateful for you being there for all the lab questions and beyond work. Hilde, always love our time together… and the CDs in the HIV lab were top! ☺

The department of Immunology, now Laboratory of Translational Immunology, was a nice place to work at: a big thank you to all colleagues! Prof. Erik Hack, I am grateful for having stayed longer at the department during the period of the FCT grant.

I hereby would also like to acknowledge the Portuguese Foundation, *Fundação para a Ciência e Tecnologia*, for granting me a one-year fellowship to finish some work.

This PhD project was done together with a very inspiring, critical and thought-provoking group: the Theoretical Biology/Bioinformatics group. Prof. Hogeweg, Rob, Can, I really appreciated my period at the department, and am very grateful for having experienced

working there! A big thank you to the people I met at the department, including Tendai, Jorg, Boris, Christian, Henk-Jan, Ilka… I definitely enjoyed the time together!

David, I really appreciate having spent some time in your lab. Thank you for the opportunity. Together with Kristin, Kath, Kelly and John, my stay and our discussions and correspondence, in and out of the lab, were helpful, productive and very enjoyable. The contribution of you all to the work described in this thesis was fundamental.

Yvonne and Saskia, you were always there to help me with everything that needed to be arranged: thank you!

Gerrit and Koos, I appreciated the hours spent together behind the Aria, not only because I could always count on you to explain all ins-and-outs and sort lots of samples, but also because of our interesting talks on the most diverse topics. Sigrid, I am grateful that you were available and helpful, and surely enjoyed our talks in the HIV-lab.

Kristin Denzer, you were very important in these years I spent in the Netherlands. In the beginning, as a teacher and a support to me as a newly arrived foreign student. Afterwards, as a good friend, someone with whom I could share my *saudade e amor por Portugal. Obrigada!*

My longing for, the *saudade* of Portugal, and my love of science, have been translated beautifully in the cover drawing by Filipa Loureiro (Doctus Design, Portugal). Thank you, Filipa, for drawing it for me from a few ideas jotted down in an email! ☺

The friends I made, made my life in the Netherlands far richer. Thank you for making me feel loved and wanted, I cherish every moment we have together! The Waalstraat girls, Anita, Mara, Taam, Nienke and Sophie, you made my coming to my room, coming home. I learned a lot from you, had a lot of fun, and 'began' understanding the Dutch (people and language)! ☺ To the game/movie-night 'gang', Ingrid, Floor, Margreet, Jolanda, it was great to share moments together and relax with a game or a feel-good movie! To the Salsa 'crew', especially Yvette and Albert, we met 'swinging', and life is much nicer when we swing and enjoy each other's company on the dance floor and out! Zsuzsi, Silvia and Robbert, Anna, you have also walked beside me for long, and I am grateful for all the laughs, the advice, and the support. Alina, Ramiro and Kevin, my 'borrowed nephews', you made me feel loved and also taught me it's not how much time we are together, but the quality of it. Marieke and Anita, great friends, thank you for listening, supporting, helping, being there no matter what. Anita, your house was a fantastic home to write the last part of this thesis, thank you for it, for being yourself, and giving yourself whole and transparently.

Carien Kalmeijer, thank you for reminding me to look for pleasure in the little things! I made it this far…

Carina, although we haven't met that often it is always great quality time. I am thankful to have met you, and will continue enjoying every talk and every short reunion we have! Together with Edite, Sabrina, the Monteiro family and Hakim, we make cheese fondue- and new year-evenings memorable!

Sabrina, Hilde, I really appreciate that you will stand beside me as paranimfs while I defend my thesis… you have certainly done that throughout these years. Your amazing friendship, optimism and support have helped me get here…big hug!

As Fernando Pessoa once wrote, *A minha patria é a língua Portuguesa* – (My homeland is the Portuguese language). Therefore, hereon I'll continue mostly in Portuguese, as I find myself best in it.

Edite, o teu optimismo, sorriso à vida mesmo quando ela não nos dá a melhor mão, é inspirador… obrigada, amiga! Sandra, o teu apoio nesta última fase foi muito importante: obrigada! Conhecemo-nos no 'agora', mas o nosso entendimento e amizade já o ultrapassa. Ao meu grupo de amigos de Viseu, e da Faculdade, obrigada por estarem presentes e me terem apoiado ao longo destes anos. Especialmente as meninas da 'roda-dentada', porque os momentos e as conversas, mesmo sendo espaçados, os trago todos no coração… Obrigada pelo apoio! Murta, minha mão na tua, sempre.

E, por último, mas sempre, sempre, agradeço à minha família. À de sangue, e não só (não é, Neia e Dani? ☺)! Pelo apoio e conselhos, por acreditarem, por me acompanharem mesmo a 2000 km, e sobretudo pelo amor incondicional! Pode haver longe, mas não há distância… Neia, obrigada por me mostrar que com pouco se tem tanto! Dani, minha mana de coração e comadre, porque me fazes rir com as tuas histórias, estás presente, e desde pequenina já tanto me ensinaste! Filipe, obrigada pela ajuda, e sobretudo pela dedicação e humildade! Madrinha, por despertares em mim, criança, o amor pela ciência embrulhado nas primeiras garrafas de cultura de células cheias de detergente, obrigada! Avós, perdi-vos quase todos nestes anos, mas nunca deixei de vos encontrar nos ensinamentos que nem sempre consigo pôr em prática, mas que (sempre) me guiam...estão sempre no meu pensamento. Pais, porque nunca deixaram de acreditar, porque me apoiaram nas decisões, e me tentaram compreender nas dúvidas: obrigada! Mãe, és a minha rocha e a minha estrela-guia, a minha força e a minha bússola. Emprestaste-me a tua Fé nos momentos de crise e desilusão. Admiro-te. Cheguei aqui, graças a ti. Cris, mana, minha inspiração, que continuemos a construir

castelos de lutas e sonhos, Juntas! Agradeço muito a ajuda e apoio nos últimos 'parágrafos' desta tese. Admiro-te. Cheguei aqui, graças a ti. Diederik, meu companheiro, melhor amigo, não há palavras que descrevam os vales e montanhas que atravessámos juntos, mas fizemo-lo juntos, e saímos mais unidos. Amo-te, hoje e sempre, porque no meio dos redemoinhos nunca deixaste de me Ver, estiveste sempre aqui e acreditaste. E por último, à minha vida, o meu Tiago, porque em ti encontrei o fôlego para os últimos passos desta etapa e a alegria para um recomeçar. Na voz de Mariza…*(excerpt of a Fado song, Miguel Gameiro/Mariza)*

Eu sei
Que a vida tem pressa
Que tudo aconteça
Sem que a gente peça
Eu sei

Eu sei
Que o tempo não pára
O tempo é coisa rara
E a gente só repara
Quando ele já passou

Não sei se andei depressa demais
Mas sei, que algum sorriso eu perdi
Vou pedir ao tempo que me dê mais tempo
Para olhar para ti
De agora em diante, não serei distante
Eu vou estar aqui

I know
Life is in a hurry
For everything to happen
Without us asking for it
I know

I know
That time does not stop
Time is precious
And we only notice it
Once it has gone by

I don't know if I've walked too fast
But I do know that I've lost a smile
I'll ask time to give me some more time
Just to look at you
From now on, I will not be distant
I will be here.

Curriculum vitae

Ana Costa was born on the 5th of November 1979, in Viseu, Portugal. She attended the Alves Martins secondary school (Viseu, Portugal) branch Science and Nature, and graduated in 1997. She studied Veterinary Medicine at the Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa (Lisbon, Portugal) and graduated in 2003, upon completion of a curricular internship in Virology at the Departamento de Virologia Experimental, Laboratório Nacional de Investigação Veterinária in Lisboa (Experimental Virology service of the National Veterinary Investigation Laboratory). After a period of voluntary work at the same laboratory, she came to the Netherlands end 2004 to follow a master's program in Biomedical Sciences of Utrecht University: Immunity & Infection. She graduated *cum laude*. Her first research training (9 months) was done at the Department of Infectious Diseases and Immunology – Molecular Infectiology, Virology – of the Faculty of Veterinary Medicine of Utrecht University, under the supervision of dr. Raoul de Groot. She studied the impact of incorporation of codon-optimized genes (human codon usage) on the replication capacity of the murine hepatitis virus. The second research training (6 months), at the department of Immunology of the University Medical Center Utrecht, was part of a project on the role of cytotoxic T lymphocytes in HIV-1 infection. In her 6-week elective component, she learned some basics of advanced imaging techniques (e.g.: FRET) at the department of Cell Biology & Biochemistry of the Faculty of Veterinary Sciences, Utrecht University, under the supervision of dr. Richard Wubbolts.

In March 2007 Ana started her PhD at the department of Immunology, now Laboratory of Translational Immunology, at the University Medical Center Utrecht, under the supervision of dr. Debbie van Baarle (copromotor) and Prof. dr. Frank Miedema (promotor). The research was within the scope of the HiPo project *Role of HLA-B alleles in immunodominance and outcome of viral infections* (Utrecht University). She additionally obtained a one-year grant from the Fundação para a Ciência e Tecnologia for completion of her studies. The results of her PhD research are described in this thesis.

List of Publications

Rao X, Costa AI, van Baarle D, and Keşmir C. A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant $CD8^+$ T cell responses. *Journal of immunology*. 2009;182(3):1526-32.

Fontaine Costa AI*, Rao X*, Lechenadec E, van Baarle D, and Kesmir C. HLA-B molecules target more conserved regions of the HIV-1 proteome. *Aids.* 2010;24(2):211-5.

Rao X, Hoof I, Costa AI, van Baarle D, and Keşmir C. HLA class I allele promiscuity revisited. *Immunogenetics.* 2011;63(11):691-701.

Koning D, Costa AI, Hoof I, Miles JJ, Nanlohy NM, Ladell K, Matthews KK, Venturi V, Schellens IM, Borghans JA, et al. $CD8^+$ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex. *Journal of immunology*. 2013;190(3):931-9.

Koning D, Costa AI, Hasrat R, Grady BP, Spijkers S, Nanlohy N, Keşmir C, and van Baarle D. In vitro expansion of antigen-specific $CD8^+$ T cells distorts the T-cell repertoire. *Journal of immunological methods.* 2014;405:199-203.

Costa AI*, Koning D*, Ladell K, McLaren JE, Grady BPX, Schellens IMM, van Ham P, Nijhuis M, Borghans JAM, Keşmir C, Price DA, van Baarle D. Complex TCR repertoire dynamics underlie the $CD8^+$ T-cell response to HIV-1. *Submitted for publication*.

* These authors contributed equally