A detailed comparison of peptides presented by different HLA class I loci: an *in silico* approach

Xiangyu Rao

Cover: Zhejiang University Publisher
Printed by: CPI Wohrmann Print Service, Zutphen
ISBN: 978-90-8570-989-3

A detailed comparison of peptides presented by different HLA class I loci: an *in silico* approach

A detailed comparison of peptides presented by different HLA class I loci: an *in silico* approach

(with a summary in English)

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 dinsdag 20 maart 2012 des middags te 12.45 uur

door

Xiangyu Rao

geboren op 21 Mei 1980 te Hangzhou, China

Promotor: Prof. dr. R.J.de Boer

Co-promotor: Dr. C. Keşmir

Content

Chapter1. General Introduction

The host genes that are most often associated with disease outcome belong to the Major Histocompatibility Complex (MHC) gene family, which in humans is called the Human Leukocyte Antigen (HLA) (reviewed in (1)). The HLA complex is located on chromosome 6 (The MHC sequence consortium (1999), quoted in (2)). The HLA class I gene complex 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 (Figure 1.1, reviewed in (3–6)) 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 2-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 via 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), 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 (8).

Pathogens, and viruses in particular, have evolved mechanisms to survive the counter-attack of the immune system by interfering with the antigen presentation pathway (9) 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 (10, 11). 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.

1.1 Human Leukocyte Antigen region

In humans the HLA region codes for class I, class II, and class III subgroup genes. A simplified representation of this region is shown in Figure 1.2. 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).

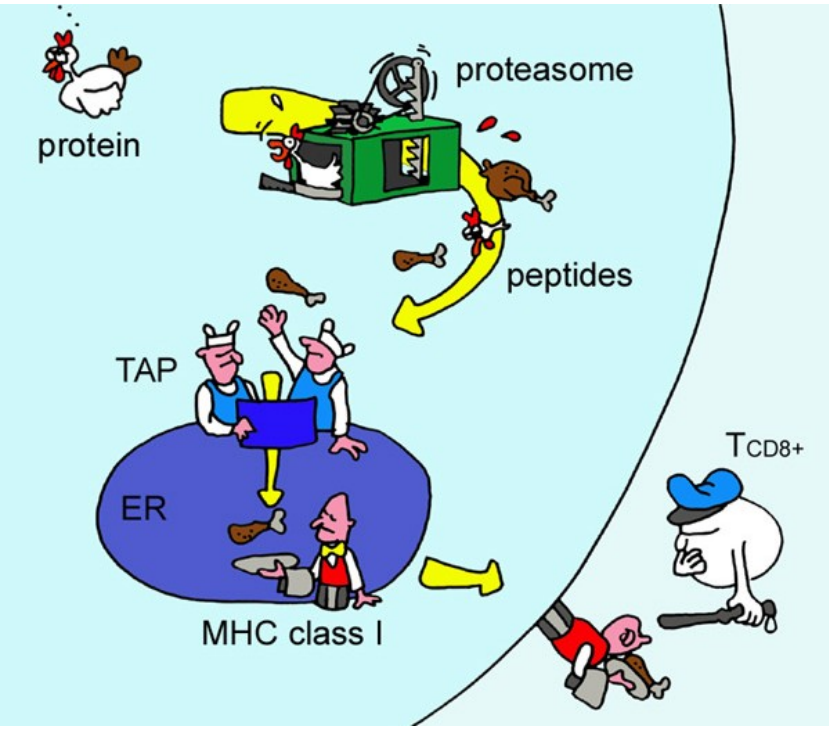

Figure 1.1 The MHC class I antigen-presentation pathway. By *Eric Reits*, 2003 (12)
MHC class I antigen presentation: the basics. Intracellular proteins are degraded by the proteasome into peptides. The transporter for antigen processing (TAP) then translocates peptides into the lumen of the endoplasmic reticulum (ER). Newly synthesised MHC class I molecules require peptide binding to be released from the ER and be transported to the plasma membrane, where the peptide is presented to the immune system.

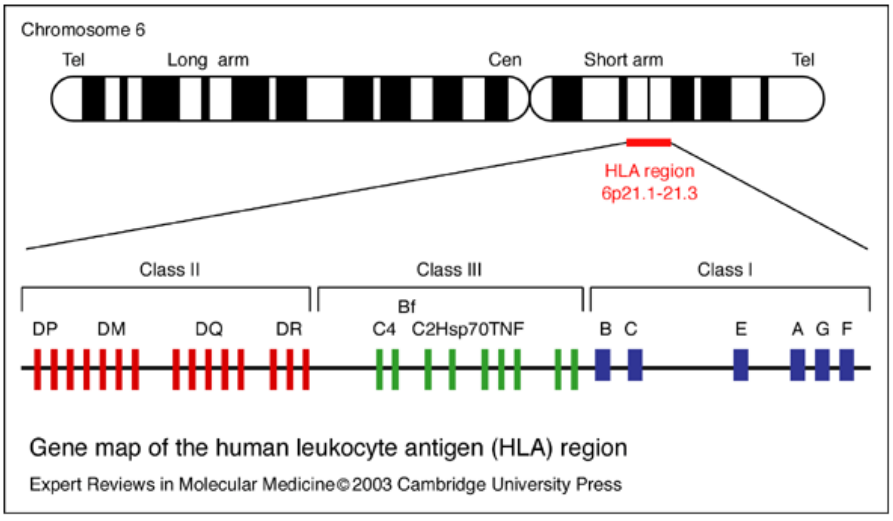

Figure 1.2 Gene map of the HLA region (in *Expert Reviews in Molecular Medicine*, 2003 (11)).
Schematic representation of the HLA complex on chromosome 6. Bars mark the localization of the major genes. Reproduced from Mehra N.K. and Kaur G. (2003) MHC based vaccination approaches: progress and perspectives. *Expert Reviews in Molecular Medicine* 5(07): 1–17, with permission from Cambridge University Press.

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 1.3.

This high polymorphism of the three major class I HLA loci had been studied extensively (reviewed in (1, 13)). 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 ((14, 15), 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 (16). 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, 13, 17)). The **over-dominant selection theory**, or heterozygote advantage, relies on the fact that HLA genes are co-dominant, which means that an individual having 6 distinct alleles will be able to present a larger variety of peptides, and elicit a broader CTL response, than a subject having less alleles. Heterozygosity ensures a superior immune surveillance. This was first reported over thirty years ago in H-2 heterozygous mice (18). The reported associations of homozygosity at the class I HLA loci with rapid progression to AIDS (19, 20) supports the importance of heterozygote advantage

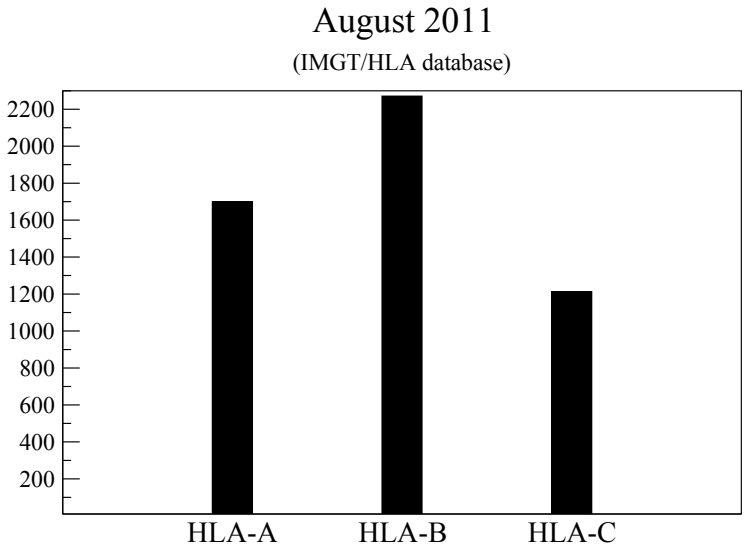

Figure 1.3 Distribution of the current number of described alleles over the three major HLA class I loci.
Number of described alleles over the HLA-A, B and C loci (n=1698, 2271, and 1213, respectively), as reported in the IMGT database (http://www.ebi.ac.uk/imgt/hla/stats.html) in August 2011.

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 (21). In addition, a more recent study by Leslie and coworkers (22) did not find a homozygosity disadvantage during 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 (23). It has actually been reported that HIV adapts to the most frequent alleles in humans (23, 24). 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. 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 (25). 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 elegantly illustrated in a recent study reporting the adaptation of HIV-1 to less common allles that were thus-far associated with a low low relative hazard (26). 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 (27). 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 an allele associated with protection from severe malaria, was observed to increase in prevalence in those regions where malaria incidence is high. (28). 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* (29). 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 (30).

1.2 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 (31). Most likely, HLA-B and –C were duplicated after the divergence between apes and Old World monkeys reviewed in (15, 32)). Curiously, even though HLA-B seems to be younger than HLA-A, it has a higher degree of polymorphism (Figure 1.3). Actually, Mungall and colleagues have reported that HLA–B is the most polymorphic locus in the whole human genome ((2)), and is definitely the gene that is most rapidly evolving and diversifying in the MHC complex (33). For example, studies of isolated Amerindian populations have shown that “new” recombinant B alleles (in relation to the alleles of the founding populations) exist in larger variety than “new” A and C alleles ((29, 34), reviewed in (17)).This is probably due to the fact that the HLA-B locus possess an exceptionally high level of

recombination and therefore evolves differently from HLA-A, which mainly evolves via point mutations (35). 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 HLA-B locus (36). 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 (16). Indeed, HLA-B alleles have been associated with opposing disease outcomes, namely susceptibility to, or protection from, parasitic (28) and viral infections, the most striking cases being described for HIV-1.

Several studies have highlighted that particular HLA-B alleles are protective (e.g.: HLA-B27; B57; B*5801), whereas others are related to a swift progression towards AIDS (e.g.: B*35, B*5301, B*5802) (reviewed in (37–39)). HLA-B alleles influence the establishment of viral set-points, and are also most strongly associated with variation in absolute CD4 counts and, consequently, in rate of progression to AIDS (33). 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, and most of these were located in HLA-B molecules (40). 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. B57 (41). 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 (26). 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 acquired immunodeficiency syndrome (AIDS) upon infection with SIV^{cpz}, the "parental" virus of HIV-1, was suggested to have gone through such a selective sweep by an ancestral HIV-like retrovirus (42). Interestingly, HLA alleles associated with low viral load in HIV long-term non-progressors (LNTPs) target identical epitopes (43), and share binding motifs with the chimpanzee MHC (44, 45). 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 (46)).

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 ((47, 48) (49) and reviewed in (50)). 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 A or C alleles (e.g., A*02 (and Cw*08) with protection, and B*5401 as a detrimental allele in HTLV infection (51)), as well

as examples of viruses in which the HLA-A alleles take the main role (e.g., EBV (52)(53)and Dengue virus (54–56)) have been reported.

1.3 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) an epiope 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 (57, 58)).

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 (33). Bihl and co-workers confirmed this finding, and added that the HLA-B responses were of higher magnitude (58). In fact, an older study had already shown the superiority of HLA-B57-restricted responses in acute infection, them being broader and stronger than those restricted by all other co-expressed class I alleles combined (59). This was concomitant with a lower incidence of symptomatic acute infection, suggestive of the protective effect of the elicited responses (59).

Upon HCV infection, spontaneous viral clearance is associated with the expression of HLA-B27. 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 (47, 60), thus highlighting the pressure induced by this dominant response. Finally, mutations in epitopes restricted by B*07, B*25, B*37 (61), and B*08 (with reversion upon transmission to B8- individuals) (62, 63) stand as other indications of a strong B-allele-specific selection pressure. It is notable that for certain HLA 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, 65).

For Influenza infection, there is also no consensus. Although a dominant HLA-A2 restricted matrix M1.58-specific response has been reported in very early studies (66, 67), and frequently ever since, B35 and B27-restricted immune responses against Influenza A, and B8 in Influenza B have also been described to be immunodominant (68). In the latter study, only in particular cases, co-expression of A2 in influenza A-infected individuals changed the hierarchy of the responses (68). Additionally, in a previous study by the same group, influenza A-infected people carrying the alleles A1, A2, A3, B8, B27, and B35 in different combinations were analysed for specific HLA-restricted responses. There seems to exist a tendency for A2 or B27-restricted responses to be immunodominant, while B8 was superior in the absence of B27 (69). These studies illustrate how co-expressed genes from the MHC family (e.g., HLA class I and II alleles,

TAP variants) influence the hierarchy of immunodominant responses, which varies not only with the virus targeted, but also with the combination of HLA class I molecules that are co-expressed in the host (68, 69).

In conclusion, most data suggest that HLA-B alleles are major determinants of the effectivety 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 (70). However, the mechanisms underlying the dominance of these responses remain elusive.

1.4 Our approach : a comparative analysis of experimental and predicted CTL epitopes

The above summary of the literature demonstrates the facts and contradictions around the role of HLA-B responses in the literature. To obtain a more consistent picture, we performed large scale comparisons of HLA-A, and –B restricted epitopes in several studies presented in this thesis. Whenever possible, we included multiple alleles and multiple pathogens in our comparisons. Our approach was to first analyse data collected on verified CTL epitopes. Although we downloaded, filtered, and analysed experimental verified epitopes from several public databases, unavoidable biases remain present in such data sets. For example, certain dominant HLA alleles like A*02 are more frequently studied than others, which potentially increases its epitope repertoire in the public databases, while several other rare HLA class I molecules may have insufficient data for analysis. Moreover, most experimental epitopes come from extensive studied human viruses like HIV-1, and are hence not necessarily appropriate for drawing general conclusion on the role of HLA-B restricted responses in infectious diseases. To eliminate these biases, we have chosen to supplement the analysis based on experimental epitopes with *in silico* predictions of peptide-HLA binding affinities. The accuracy of *in silico* predictors of antigen presentation has been increasing over time because more experimental binding data has become available. We selected and applied *in silico* predictions for the main steps of antigen presentation pathway: MHC binding, TAP binding probability and proteasome cleavage. Here I briefly introduce the two main algorithms used in this thesis.

Stabilized Matrix Method (SMM)

SMM is a matrix-based algorithm, which identifies the functional relationship between a peptide sequence and its binding affinity to a particular HLA molecule. The relationship between sequence and affinity can be roughly approximated by the independent binding assumption, i.e., amino acids at different positions of a peptide contribute independently to the overall binding affinity of the peptide (71). For each HLA molecule a specific scoring matrix quantifying residue contributions to binding was generated using the experimental data. The final matrix can be used to predict an IC50 value for novel peptides binding the HLA molecule in question. It is not trivial to define the true set of binders for a specific MHC molecule based on these IC50 predictions. Throughout this thesis we applied two strategies: first, as suggested by

Ruppert (72), an IC50 threshold of 500 nm was used to distinguish binding peptides from nonbinding peptides, and second, we ranked all the peptides according to their predicted binding affinity and assumed that the top 1-2% of the peptides are potential binders. Besides predicting HLA binding affinity, SMM was also applied to predict the TAP binding probability, and proteasomal cleavage sites (73).
The SMM algorithm is especially attractive to use when only little experimental data is available (74)(75).

Artificial Neural Networks (ANN)

An artificial neural network (ANN) is a computational model that has been applied for solving pattern recognition problems in biological sequence analysis. An ANN is ideally suited to account for the correlations where the binding affinity of a given amino acid at one position is influenced by amino acids at other positions in the peptide. For example, two adjacent amino acids may compete for the space in the binding grove of an MHC molecule. The simple architecture of a "feed-forward multilayer network" (as often used in biological applications of ANNs) is composed of a layer of input nodes, one or several hidden layers, and an output layer. Each layer is fully connected by weighted edges with the next layer. In this thesis, two state-of-the-art ANN-based predictors have been employed for HLA-peptide binding prediction: NetMHC (76) and NetMHCpan (77). Both methods assign an IC50 value to the candidate peptide as its predicted binding affinity to a certain HLA molecule. The ANN-based predictor NetChop was used for predicting proteasomal cleavage or the combined presentation likelihood.

1.5 Structure of the thesis

Although HLA-A, HLA-B and HLA–C molecules are structurally very similar, the HLA-B locus has a special status: i) it has the largest diversity, ii) HLA-B molecules are most strongly associated with resistance or susceptibility to infectious disease, and iii) T-cell responses restricted through HLA-B alleles are often immunodominant. In order to explore why HLA-B displays these characteristic features, we perform an extensive comparison of HLA-A and HLA–B peptide repertoires using both experimental and *in silico* predicted data. We focus on the following questions:

1. Do HLA-B molecules present a broader repertoire of peptides, and thereby decrease the chance of pathogen evasion? We address this question in Chapter 2 by contrasting features of epitope-presentation on HLA-A or HLA-B molecules based upon the curated pathogen-derived epitope data retrieved from the Immune Epitope Database and Analysis Resource, and upon *in silico* predicted epitopes.
2. Do HLA-B molecules target more conserved/constraint parts of the pathogen proteins and thus make pathogen evasion less likely? To address this issue, we specify the regions (proteins) and their conservation levels targeted by different HLA molecules from well-studied pathogens, like HIV (chapter 3) and HCV (chapter 4), based on the known experimental data from LANL database.
3. Does the ligand promiscuity of HLA-A and HLA–B molecules differ? Does this have an influence in the disease outcome (chapter 5)?

4. Do the binding motifs of the HLA-A and HLA–B molecules in a HLA haplotype influence the prevalence of that haplotype in the population (chapter 6)?

Chapter 2. A Comparative Study Of HLA Binding Affinity and Ligand Diversity: Implications for Generating Immunodominant CD8+ T Cell Reponses

Xiangyu Rao [*], Ana Isabel C.A.Fontaine Costa[†], Debbie van Baarle [†], Can Keşmir [2*‡]

*Dept Theoretical Biology/Bioinformatics, Utrecht University
†Dept Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht
‡Academic Biomedical Centre, Utrecht University

The Journal of Immunology 2009, 182:1526-1532

2.1 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. In addition, 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 antigen presentation level, we have analysed the diversity and binding affinity of their peptide repertoire, by making use of curated pathogen-derived epitope data retrieved from the Immune Epitope Database (IEDB), and *in silico* predicted epitopes.

In contrast to our expectations, HLA-B alleles were found to have a less diverse peptide repertoire, which points towards 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.

2.2 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 (78) is most likely pathogen-driven (79–83).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 (84). Indeed, it seems to be the most rapidly evolving locus1 and it has an exceptional rate of recombination (85), 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 (82). Furthermore, HLA-B-restricted pathogen-derived epitopes seem to be more frequently targeted by $CD8^+$ T cells than those presented on HLA-A within a given host. This has been shown for viral pathogens such as HIV, EBV, Influenza and CMV (86–89), and for an intracellular bacterial pathogen, *Mycobacterium tuberculosis* (90). In addition, HLA-B dictated T cell responses seem to be of higher magnitude (87, 90, 91), and have been described to influence infectious disease course and outcome. In HIV infection, a specific group of B alleles (e.g. B*3502, 3503, 3504, B*5301, B*5802) has been associated with fast progression towards AIDS (86, 92, 93), 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 (94, 95)). 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 (86). Also, in Hepatitis C virus (HCV) infection, HLA-B27 seems to be the allele most significantly associated with spontaneous viral clearance (96). Furthermore, the association of B53 with protection from severe malaria has most likely contributed to its high frequencies in West Africa (97). 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 (94, 98) .

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 (99, 100). Intrinsic differences in the amount/diversity and binding affinity 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, present less peptides than their HLA-A homologs, and that 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 effects of antigen presentation on immunodominance.

2.3 Materials and Methods

2.3.1 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% (101), all epitopes with less 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.

2.3.2 Genomic data

Sequences of the human proteome as a resource of self peptides, 13 common bacteria and 17 common virus proteomes (Table. S2.1) as resources of pathogenic (non-self) peptides were downloaded from the **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 which 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: i) Protein shuffling: Shuffling residues within a protein by keeping the frequency of amino acids and the length of the protein intact. ii) Generating artificial proteomes: Artificial proteomes were generated by using an uniform amino acid distribution (i.e frequency of each amino acid is 0.05), without changing the size of original proteome. All these sequence modifications were performed by PERL programming.

2.3.3 HLA-peptide binding predictions

There are at the moment several HLA-peptide binding predictors, such as polynomial method (**PM**) (102), artificial neural networks (**ANN**) (103), **BIMAS** (104), a classification tree (**CART**) (105), additive method (106) and Stabilized Matrix Method (**SMM**) (71). The performance of these algorithms to identify new epitopes has recently been benchmarked on experimental data (107). Ingeneral **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 an measured binding affinity is as good as the

correlation between measurements from different laboratories (107) . Another demonstration of how powerful these predictions are is a study by Schellens et. al (84), who identified 18 new CTL epitopes out of a set of 22 predicted CTL epitopes using ANN 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 most broad HLA coverage at the time this project was initialized. For each HLA molecule a specific **SMM** matrix was generated using the experimental binding data. The details of this algorithm are given in (71). We downloaded **SMM** prediction software and training datasets from **IEDB** (http://mhcbindingpredictions.immuneepitope.org/), and generated scoring matrixes 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 are generated because of training data limitation.

As suggested by Ruppert J *et al* (72) an IC50 threshold of 500nM was used to distinguish binding peptides from non-binding peptides.

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

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

2.3.4 Shannon entropy calculation and binding motif visualization

The Shannon entropy 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 same frequency ($-20\frac{1}{20}\log_2(1/20) \approx 4.3$) and the minimum Shannon entropy is zero if a position is fully conserved. The binding motifs of HLA molecules were visualized by **sequence logo technique** (108) in this study (see Figure caption of Figure. S2.5 for an explanation).

2.4 Result

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

In order to explore the differences in the epitope repertoire of HLA molecules, publicly available experimental epitope data was analyzed. At the moment IEDB (http://www.immuneepitope.org) is the

most complete database for experimentally verified epitopes. All epitopes which can bind to HLA-A or HLA-B alleles were retrieved from IEDB (downloads were made in Oct 2006). In average the database contains 216 unique epitopes (SD $\pm$ 447) for each HLA-A allele, and 37 unique epitopes (SD $\pm$ 70) for each HLA-B allele. After excluding all HLA-A*02 epitopes, which are considerably abundant in the database, the average number of unique epitopes per HLA-A allele dropped to 150 (SD $\pm$ 319), however, it is still significantly higher than that of HLA-B alleles. Thus the diversity of HLA-A binding epitopes is significantly larger than that of HLA-B. After excluding alleles having less than 20 epitopes, we calculated the epitope length distribution for 19 HLA-A and 16 HLA-B alleles (Figure. 2.1). 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) (Figure. 2.1). For the rest of the paper this data set is used and referred as "experimental binding data".

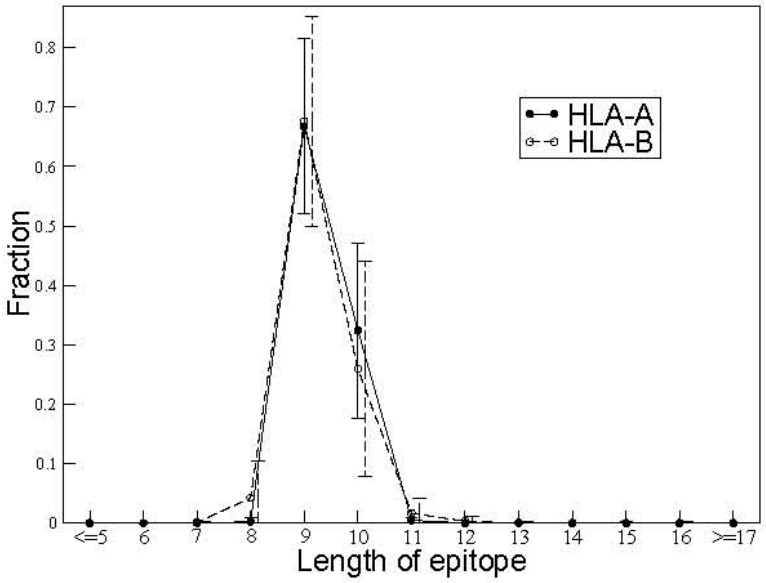

Figure 2.1 Length distribution of experimentally verified HLA-A and HLA-B epitopes.
11878 unique experimental epitopes for 19 HLA-A alleles and 2353 unique experimental epitopes for 16 HLA-B alleles were downloaded from IEDB (http://www.immuneepitope.org), and were classified according to the length. The lines show the average fraction for HLA-A (solid) and HLA-B (dashed) epitopes, while error bars represent the standard deviation.

The experimental data might have some biases, e.g. due to frequency 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, 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 9mer-HLA pair an IC50 value indicating the binding affinity. The lower IC50 value, implies the higher binding affinity of the 9-mer to a certain HLA molecule. As suggested earlier (72), an IC50 threshold of 500nM was used to distinguish binding 9-mers from non-binding 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. Figure. 2.2 shows that on average the binding fraction of HLA-A alleles is about 5%, which is significantly higher than the average binding fraction of HLA-B alleles, 2%, in the human proteome, (Figure. 2.2, $p<0.001$, Mann-Whitney test). Similar results were obtained for viral 9-mers and bacterial 9-mers (Figure. 2.2).Changing the IC50 threshold (using 50nM and 5000nM instead of 500nM) or excluding two alleles, that are either too generic (A*3002, with a binding fraction higher than 20%) or too specific (B*0801, which does not have a predicted binder in many viral proteomes), did not change our results (result not shown). Thus, HLA-A alleles seem to bind 2 fold more peptides than 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 Figure. S2.1).

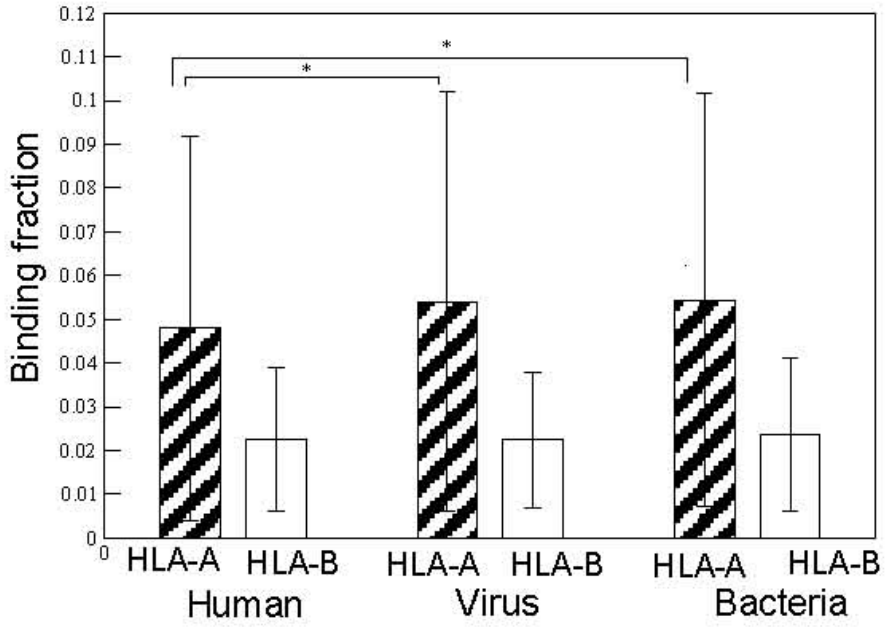

Figure 2.2 Average binding fraction of HLA-A and HLA-B alleles in human, viral and bacterial proteomes.
All possible 9-mers from human, virus and bacteria proteomes which can bind to certain HLA-A or HLA-B molecules were predicted by **SMM** matrices as described in Material and Methods. The average binding fractions, defined as the average fraction of 9-mers that were predicted to bind an HLA molecule with an IC50 value lower than 500 nM in a proteome, are shown for HLA-A (shaded 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 test). Moreover, HLA-A molecules bind significantly less self (human) 9-mers than non-self (viral or bacterial) 9-mers. ($p<0.001$, Mann-Whitney test, indicated with *).

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 two fold higher than that of HLA-B alleles (Figure. S2.2). 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 less human 9-mers than pathogen-derived 9-mers ($p<0.001$ Mann-Whitney test), which was not the case for HLA-B molecules (Figure. 2.2). 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 (Figure. S2.3) may create a preference for presenting non-self.

Experimentally verified epitope-binding data of a given HLA molecule are used to construct a **SMM** scoring matrix that can predict new epitopes of this HLA molecule (109). The number of experimental epitopes used to construct SMM predictors, i.e. 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 (Figure S 2.4). 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 HLA-A molecules.

2.4.2 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 (110).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 IC50 value associated were retrieved from **IEDB** database, and are shown in Figure 2.3. *In vitro* measured binding affinities of HLA-A epitopes are significantly higher than that of HLA-B epitopes ($p< 0.001$, Mann-Whitney test).

We then analyzed the *in silico*-predicted binding affinities from the human, viral and bacterial proteomes. To exclude a potential bias introduced by the two-fold higher amount of predicted HLA-A epitopes as compared to 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 (Figure. 2.3). The results were confirmed by repeating the same analysis for the top 10 epitopes of each HLA molecule (results not shown).

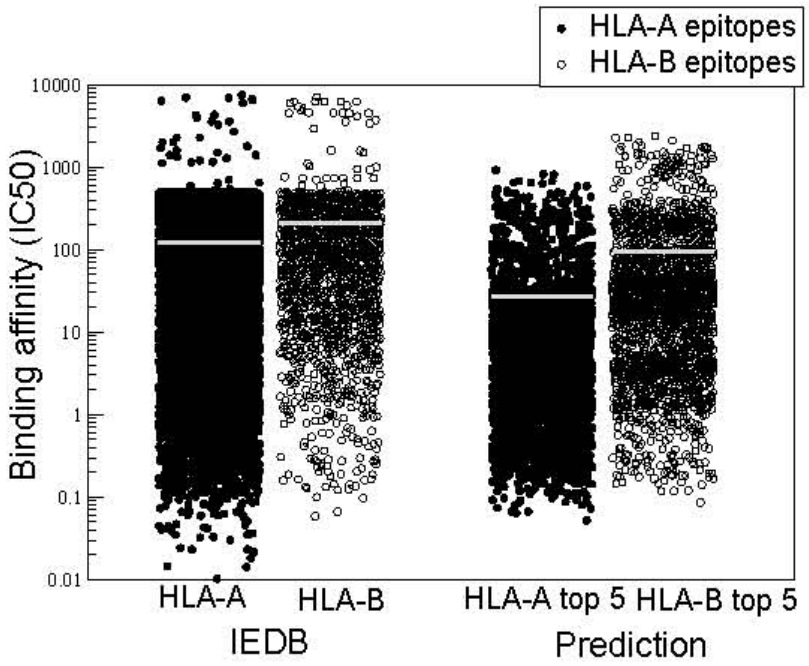

Figure 2.3 Binding affinities of HLA-A epitopes are significantly higher than HLA-B epitopes.
Left: Binding affinity (IC50 value) for HLA-A and HLA-B epitopes measured in *vitro* (data downloaded from IEDB). Right: Binding affinity of top 5 predicted epitopes in human, virus and bacteria proteomes for 19 HLA-A and 15 HLA-B molecules. The grey lines show the average. In all cases the binding affinities of HLA-A molecules were significantly higher ($p<0.001$, Mann-Whitney test) than HLA-B molecules.

Although the predicted binding affinities show the same tendency as the experimental data (Figure 2.3), one can argue that the predictions are not true validations of the experimental data, because they are not independent, i.e., 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 choose three HLA-A and three HLA-B alleles randomly from all 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 non-overlapping. 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 HLA-A alleles. One can, thus, 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 Figure. 2.4 we plotted the binding fraction versus the binding affinity of 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 HLA-B epitopes, and a significant positive correlation was found between the binding fraction and the binding affinity of HLA molecules.

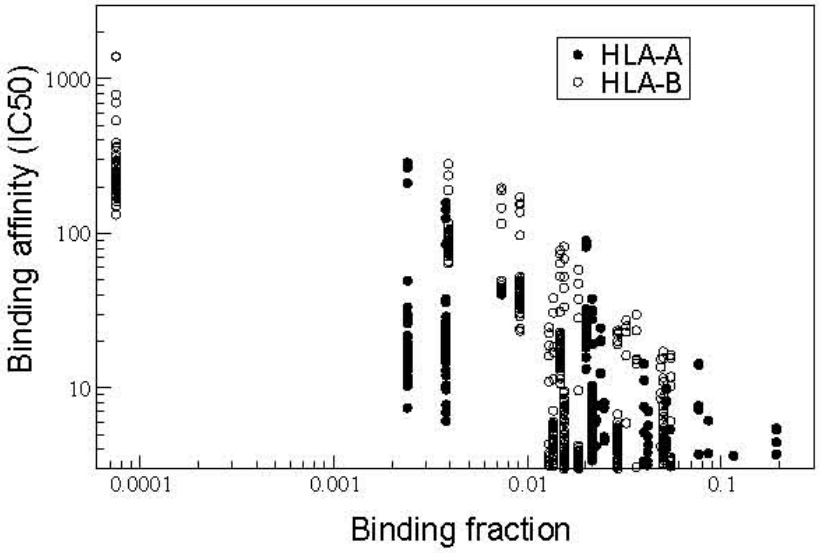

Figure 2.4 Correlation between binding fraction and binding affinity.
A significant correlation (cor=-0.25, p<0.001, Mann-Whitney test) was found between predicted binding fraction of 19 HLA-A and 15 HLA-B molecules and predicted binding affinity of their top 5 epitopes in bacterial proteomes. Filled circles depict HLA-A epitopes and empty circles depict HLA-B epitopes. Similar results were found in human and viral proteomes (human: cor=-0.32, p<0.001; virus: cor=-0.31, p<0.001).

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

Sequence patterns in cohort of epitopes specific for a certain HLA molecule define its binding motif (111). Some positions of HLA binding motif (e.g. P 2, P 3, P 5, P 9) which are very important for epitope binding were defined as anchor positions (112–115).

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 ariability at each position in terms of the Shannon entropy explained in Materials and Methods. The larger Shannon entropy indicates higher level of variation in a position. Only the sequence conservation in P 2 and P 9 were significantly different between HLA-A and HLA-B epitopes (p<0.001, Mann-Whitney test, Figure. 2.5). Interestingly, P 2 is the most conserved position for HLA-B alleles, while for HLA-A alleles P 9 has the lowest Shannon entropy (Figure. 2.5).

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 which have a similar level of conservation in P 2 and P 9, and contains both HLA-A and HLA-B molecules. Group 2 contains only HLA-B molecules, defined by a markedly conserved P 2. Group 3 encloses only HLA-A molecules, and these share a conservative P 9 in their binding motifs. Figure. S2.5 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. P 2 and P 9), where P 2 is more important for HLA-B alleles and P 9 is more important for HLA-A alleles.

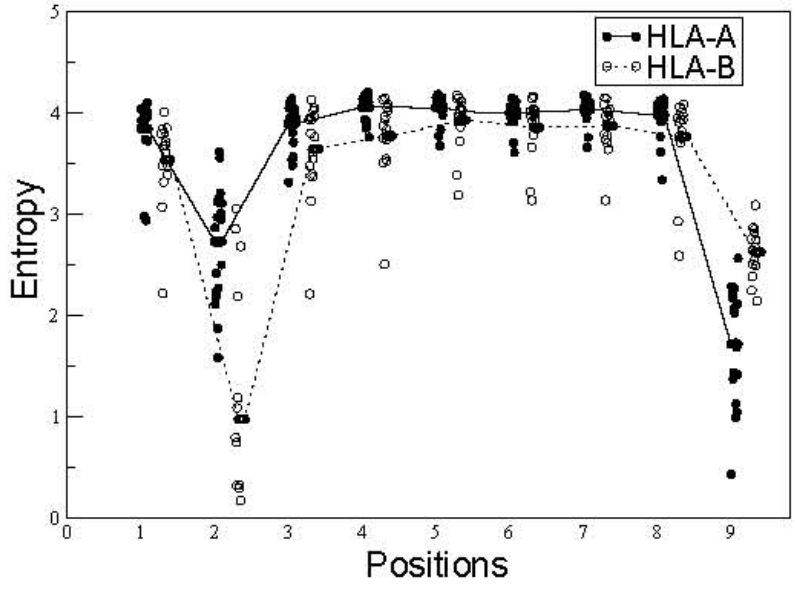

Figure 2.5 Variability in experimentally verified HLA-A and HLA-B epitopes.
The entropy of HLA-A (filled circles) and HLA-B (empty circles) epitopes were calculated based on 7238 unique epitopes (length 9) of 19 HLA-A alleles and 1585 unique epitopes (length 9) of 16 HLA-B alleles downloaded from **IEDB** database. Only in P2 and P9 the variability of HLA-A and –B epitopes is significantly different (both cases $p<0.001$, Mann-Whitney test). Shannon entropy for P2 of HLA-B epitopes is significantly lower ($p=0.017$, Mann-Whitney test) than that of the best conserved position of HLA-A, the P2.

2.5 Discussion

Immune responses restricted by HLA-B alleles to several pathogens have been previously shown to be immunodominant (86–88, 90, 91). In addition, particular HLA-B alleles seem to be associated with either protection or susceptibility to infectious diseases, see reviews (94–97). It has been suggested that this association of HLA-B molecules with a favourable 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 (96, 98). 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 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

these have been historically less studied, thus the data is more scarce and, consequently, no good predictors are currently available.

Surprisingly, we show that HLA-B alleles present a less diverse set of epitopes, approximately five-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 (Figure. 2.2). The familiar notion that 9-mers are the most preferred MHC class I ligands (116, 117) can be confirmed in this analysis, and holds true for both loci (Figure. 2.1). 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 (Figure.2.2), and is lost when the subtle differences among amino acid frequencies are excluded (see S2 and S3).

High HLA peptide binding affinity is not strictly correlated with immunodominance (118). Moreover, Assarsson *et al* showed that dominant epitopes might have lower affinities than subdominant epitopes (110). 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 (Figure 2.3). 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 favour of complexes with longer cell surface half-lives enclosing epitopes of higher binding affinity (119–122). In fact, thus far only B alleles have been shown to associate poorly with this chaperone, among which B27 and B*4405 (122–124). 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 (122), and the complex has been proposed to act by disrupting peptide-MHC class I complex formation until high affinity peptides are bound (121).

Interestingly, binding fraction and binding affinity seem to be positively correlated (Figure. 2.4). HLA-A alleles seem to have a more permissive binding motif, where the most conserved residue is the C-terminal (Figure. 2.5, S2.5). 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 P 2 (Figure. 2.5, S2.5). In accordance, the information content for P 2 of HLA-B epitopes is significantly higher ($p=0.017$, Mann-Whitney 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 (125). 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 favour 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 less 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 is 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 here 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, though unexpected and counterintuitive, call for novel hypotheses to explain why HLA-B alleles are associated with immunodominance.

First, if epitopes bind indeed 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 PD-1 (126). As these high avidity T cells will be preferentially deleted (127–129), the "surviving" HLA-B-restricted responses would become dominant. In contrast, Almeida and coworkers (130) have found HLA-B T-cell responses to be of higher avidity during HIV infection. Still, the highest avidity response tested, though 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, i.e. 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 (118, 131). 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 (132). 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.

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. Though 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.

2.6 Acknowledgements

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

2.7 Supplemental materials

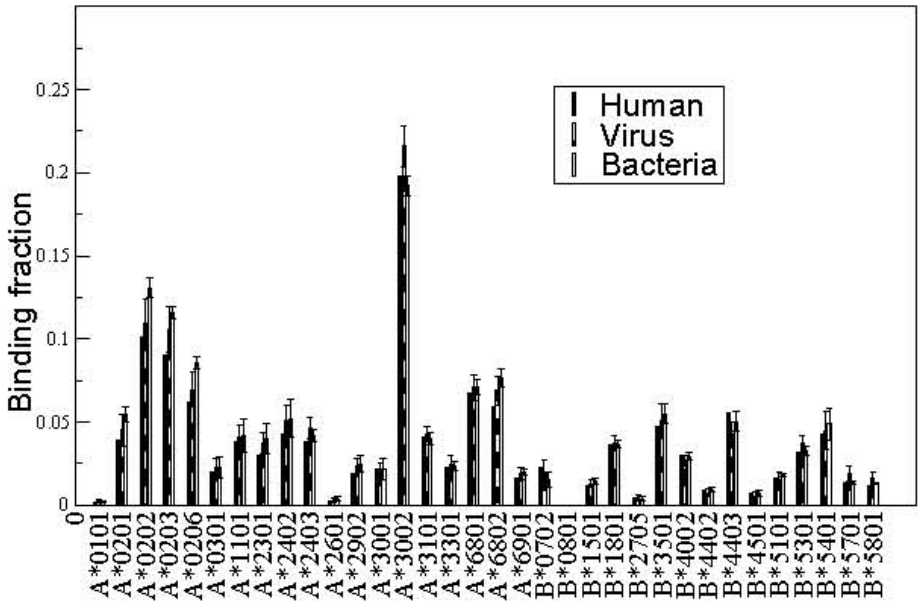

Figure S 2.1 Binding fractions of individual HLA alleles in human, viral and bacterial proteomes. Average binding fractions, predicted by SMM method, for individual alleles were displayed for the different proteomes (human: filled bars; virus: shaded bars; bacteria: empty bars).

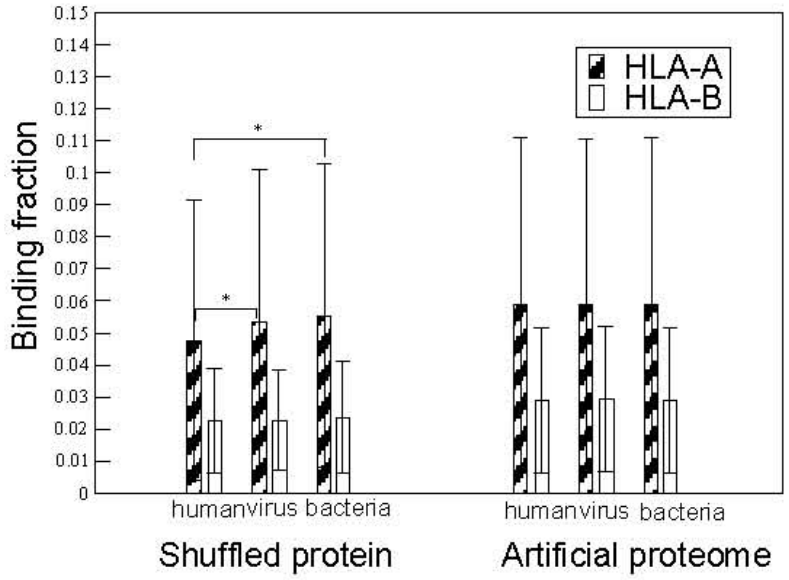

Figure S 2.2 Average binding fraction of HLA-A and HLA-B alleles in shuffled and artificial proteomes.
Left: To test whether HLA-A and -B binding is sensitive to sequence patterns accruing in natural proteins, every protein in human, viral and bacterial proteomes was shuffled and average binding fraction 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.

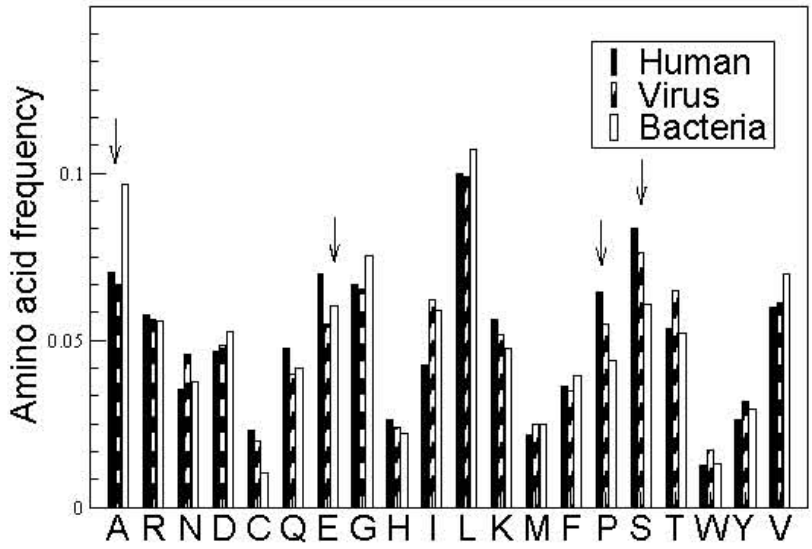

Figure S 2.3 Amino acid frequencies in human, viral and bacterial proteomes.
Filled bars indicate the frequencies of 20 amino acids in human proteome, shaded bars present the amino acids frequencies in a combined set of 17 different virus proteomes and similarly empty bars depict the amino acids frequencies in a combined set of 13 bacterial proteomes (see Table. S2.1 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 too some extent abundant in the human proteome.

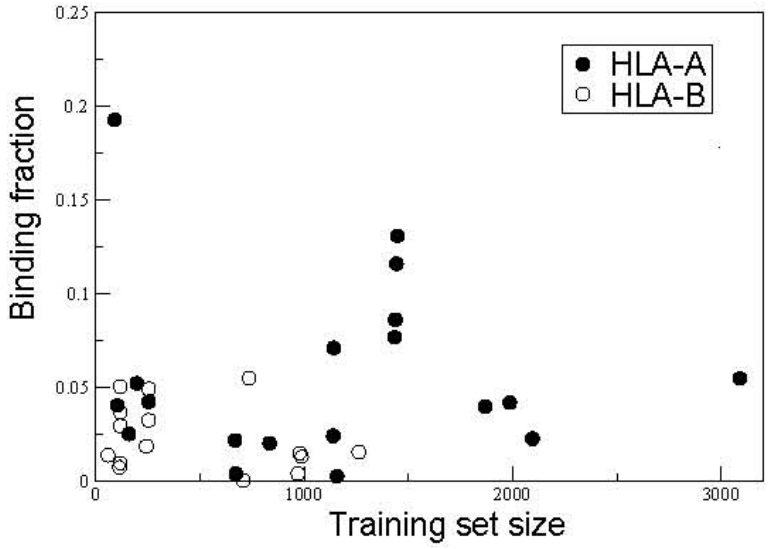

Figure S 2.4 Lack of Correlation between predicted binding fraction and training set size for HLA-A and B.
The number of 9-mers used to train HLA-A and HLA-B predictors (in Peters et al, 2006) is plotted against the average binding fraction for the respective HLA alleles for bacterial proteomes. Similar 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.

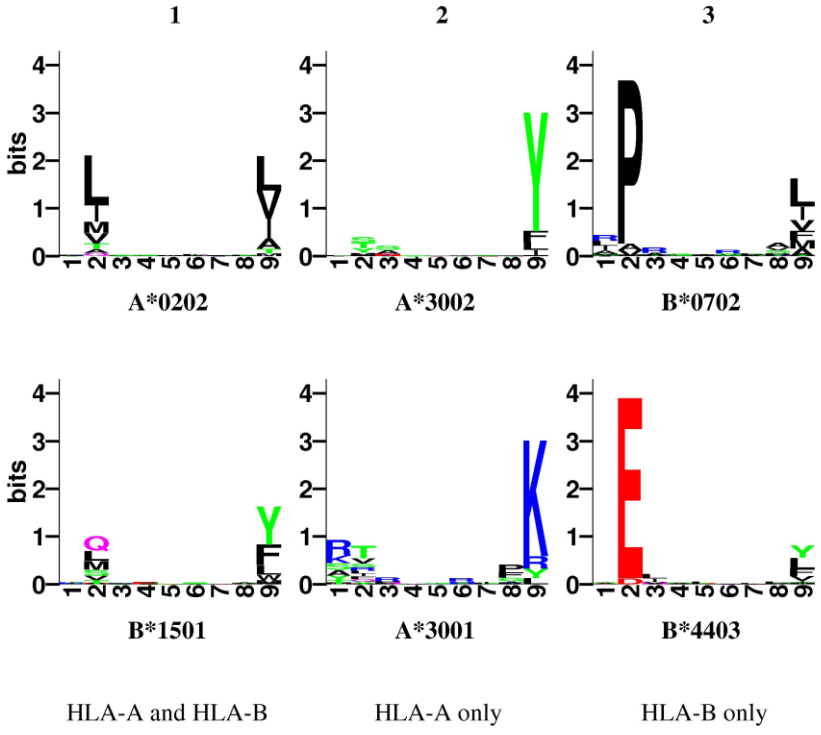

Figure S 2.5 Examples of HLA-A and HLA-B binding motifs.
The binding motifs of HLA molecules were visualized by sequence logo technique. In the logo plot, the height of the each bar gives the Shannon information content at that position which can be used as a measure of amino acid conservation. The higher Shannon information content presents 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 sub groups by their binding motifs. Left panel: Two examples of HLA molecules whose binding motif has similar conservation at P 2 and P 9. Middle panel: Two examples of HLA molecules whose binding motif is most conserved in P 9. This group contains only HLA-A molecules. Right panel: Two examples of HLA molecules whose binding motif is most conserved in P 2. This group contains only HLA-B molecules.

Table S2.1. List of viral and bacterial proteomes which were 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

Chapter 3.HLA-B molecules target more conserved regions of the HIV-1 proteome

Fontaine Costa AI[1#], Rao X[2#], LeChenadec E[2], van Baarle D[1,4], Keşmir C[2,3]

[#] These authors contributed equally to this work.

[1]Dept Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht
[2]Dept Theoretical Biology/Bioinformatics, Utrecht University,
[3]The Academic Biomedical Centre, Utrecht University, The Netherlands
[4] Dept Internal Medicine and Infectious Diseases, University Medical Center Utrecht, Heidelberglaan 100, The Netherlands

AIDS 2010, 24:211-215

3.1 Abstract

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

Objective: We sought to identify if the association of HLA-B alleles with rate of progression towards 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 *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.

Conclusions: 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 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.

3.2 Introduction

Cytotoxic T lymphocytes (CTL) are believed to have a central role in controlling HIV-1 infection (reviewed in (133)). T cells responding to HLA-B-restricted epitopes seem to be immunodominant (134), and, through not yet understood reasons, have a major impact on progression towards AIDS (reviewed in (133)). 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.

3.3 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 1981-1985 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 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).

3.4 Results and Discussion

3.4.1 Analysis of HIV-1 clade B MHC 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, -B or -C epitope (positions A, B, or C), or both an A and B epitope (X), or empty (E), using

publicly available CTL epitope lists at the LANL Immunology database (http://www.hiv.lanl.gov/content/immunology, details of this coding schema are explained in the legend of Table 3.1).

Table 3.1 summarizes the fraction of residues comprised in HLA-A, -B, and -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, (135–142) and also highly immunogenic, even in different ethnicities (143). 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 (144, 145).

In the total proteome, the fraction of unique A and B positions is equivalent (Table 3.1, 23.5 *vs* 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 to Nef, have been previously shown to dominate the total HIV-specific response, in both breadth and magnitude (146). Tat is also preferentially targeted by HLA-B alleles (16.8%, *vs* 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 and three-fold higher than B positions, respectively), which are among the most variable proteins in the proteome (147).

3.4.2 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. 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 Yusim and coworkers (147), who found an inverse correlation 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 ($p<0.025$).

Table 3.1

Protein	Length (aa)	% A	% B	% C	% X (A and B)	% 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+RNase H	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, -B or -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: i) the best defined CTL epitope list (epitopes with defined optimal length and HLA-restriction, verified by several independent research groups); and ii) 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, though 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 was 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%, though 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, Tat), while grey highlights the ones with a B/A fraction of ≤ 0.5 (Rev, gp160). Numbers in parantheses 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).

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. 3.1) 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 and -B targeted regions is not significantly different.

3.4.3 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 and colleagues (148). 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 Wang et al, data 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 can not be used as HLA-A or -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 ($p=0.002$). This 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 and colleagues have illustrated that mutations at conserved sites revert more rapidly (149), 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, B57) become fixed in the population (Schellens et al. manuscript submitted) and correlate with the prevalence of the corresponding HLA (150). 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 become 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) (151).

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 (152), 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.

Chapter 4. Protective HLA Molecules Determine Infection Outcome in Hepatitis C Virus Infection by Preferential Presentation of Peptides From Conserved Viral Proteins

Xiangyu Rao **,** Ilka Hoof, Debbie van Baarle †*, Can Keşmir

Dept Theoretical Biology/Bioinformatics, Utrecht University,
†Dept Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht
*Dept Internal Medicine and Infectious Diseases, UMCU.

Manuscript in submitted

4.1 Abstract

Hepatitis C virus infections affect worldwide more than 170 million people. While the majority of these individuals are chronically infected, some clear the infection rapidly. Host factors seem to play a key role in the clearance of HCV, among them the human leukocyte antigen (HLA) class I molecules. Particular HLA molecules, e.g. B*27 and B*57, are associated with viral clearance. Here we estimated the sequence variability of HCV proteins and analyzed the HLA-restriction of both experimentally verified and *in silico* predicted HCV epitopes. We demonstrate that HLA molecules associated with HCV clearance preferentially present epitopes from conserved HCV proteins, especially from NS5B. In contrast, none of the known susceptible HLA molecules, i.e., HLA molecules associated with HCV persistence, have a similar preference.

Conclusion: Our analysis suggests that by targeting the most constrained --and thereby conserved-- parts of the HCV genome, “protective” HLA molecules reduce the potential of HCV to escape the cytotoxic T cell (CTL) response of the infected host.

4.2 Introduction

Hepatitis C virus (HCV) affects more than 170 million people worldwide (153). Approximately 25% of the infected individuals spontaneously resolve the infection, whereas most develop a persistent viral infection, which may lead to liver cirrhosis, liver failure, or hepatocellular carcinoma (154). Although the determinants of HCV clearance, persistence, and disease outcome are not well known, it has been shown in numerous studies that the magnitude, diversity, and quality of T cell responses play an essential role in controlling the infection and determine the outcome of HCV infection (155–158). Especially, efficient (CD4+) T cell responses against nonstructural proteins might be important for viral control (159, 160). The importance of cellular immune responses is also supported by the finding that certain human leukocyte antigen (HLA) class I molecules are associated with differential HCV infection outcome. Numerous HCV-HLA association studies have been performed covering different ethnic groups and

different HCV genotypes. Kuniholm et al. (161) have recently reviewed these studies and compiled a list of strong associations based on the level of evidence given in the literature. Interestingly, two allele groups, B*27 and B*57, that have previously been associated with slow disease progression in HIV-1 positive individuals, are also associated with viral clearance in HCV infection.

The HCV genome encodes a polyprotein precursor of about 3000 amino acid residues, which is processed into several mature structural proteins (CORE, E1, E2) and nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B, P7) (162). A possible mechanism behind the observed associations between certain HLA class I molecules and HCV infection outcome could be the differential presentation of HCV proteins by certain HLA molecules, as was shown to be the case for HLA associations with HIV-1 (163) and HTLV-1 (164). To test this hypothesis for HCV, we analyzed the sequence variability of HCV proteins and the HCV epitope repertoires of several HLA allotypes that have been associated with viral clearance or persistence. We employed clinical data from public databases and used *in silico* predictions of HLA-peptide binding to define these HLA epitope repertoires. We found that protective HLA molecules show a preference to present epitopes from conserved HCV proteins (e.g. Core and NS5B), while non-protective HLA molecules preferentially target HCV proteins that are significantly less conserved (e.g. NS5A). Taken together, our analysis suggests a relationship between the protective potential of an HLA molecule and the degree of sequence conservation of the HCV protein targeted by that HLA molecule.

4.3 Material and Methods

4.3.1 Experimentally verified HCV T cell epitopes

All experimentally verified HCV $CD8^+$ T cell epitopes restricted by either HLA-A, -B or -C molecules were downloaded (May 2010) from two public databases: 1) the Los Alamos HCV immunology database (http://hcv.lanl.gov/content/immuno/immuno-main.html, note that maintenance of this database stopped in 2007) (165) and the Immune Epitope Database and Analysis Resource (IEDB, http://www.immuneepitope.org/) (166). HCV is classified into six phylogenetically distinct genotypes (167). As HCV genotype 1 is the most dominant strain worldwide and is well studied (168), the epitopes with genotype 1 annotation were directly included in our analysis. For the epitopes without a genotype annotation, we included only those for which the exact sequence occurred in at least one of the HCV genotype 1 sequences in the Los Alamos HCV sequence database (http://hcv.lanl.gov/components/sequence/HCV/search/searchi.html) (168). To make sure that all epitopes we considered are from the HCV strains infecting humans, the epitopes identified using HLA transgenic mammalian cells were excluded. This procedure resulted in 256 experimentally verified HCV genotype 1 T cell epitopes with known HLA restriction. To calculate the distribution of CTL epitopes over the HCV proteins for each HLA allele, we limited our analysis to 187 unique experimental epitopes, where we consider an epitope "unique" if it differs in at least three positions from all other CTL epitopes restricted by the same HLA in the data set. Only five epitopes were restricted by HLA-C molecules. This set of curated HCV epitopes is available upon request.

4.3.2 Sequence Conservation

Pre-aligned HCV protein sequences (CORE, E1, E2, NS2, NS3, NS4A, NS4B, NS5A, NS5B, and P7 sequences) dated 2008 or older were downloaded from the Los Alamos HCV sequence database (downloads were made in June 2010). All alignment files were manually modified to exclude the protein sequences that are too short (shorter than 20 amino acids), using the Jalview multiple sequence alignment editor (version 2.6.1) (169). The alignments contain only the "oldest" sequence per patient based on their genbank submission date and cover sequences from all HCV genotypes (genotype1-6). To estimate the sequence conservation, the Shannon entropy (170) was calculated for each position of each protein, i.e., $E(i) = -\sum_{L=1}^{20} q_i^L \log_2 q_i^L$,where *E(i)* indicates the Shannon entropy at position *i*, and q_i^L is the observed frequency of a certain amino acid *L* at position *i* in the alignment. The Shannon entropy reaches its maximum value of 4.3 if all amino acids are equally probable in that position, and is 0 if the position is fully conserved. Different numbers of sequences were available for different HCV proteins, ranging from 466 sequences for NS5B to 918 sequences for NS3. To prevent a bias due to different numbers of sequences in the alignments, we randomly selected 466 sequences from all proteins (i.e., the minimum number of sequences we had per protein) 100 times and calculated the corresponding entropy values.

4.3.3 HLA-peptide binding predictions

We applied two different algorithms to predict epitopes from the HCV reference strain (accession number AAA45534.1). The artificial neural network based predictor NetMHC 3.2 (171) was used to predict all 9-mer or 10-mer HCV epitopes restricted by common HLA-A or HLA-B alleles. This predictor assigns to each peptide-HLA pair an IC50 value, which can be used as a predicted binding affinity. Using a widely accepted IC50 value of 500 nM as a threshold to distinguish binders from non-binders generates different repertoire sizes for different HLA molecules (data not shown). We therefore repeated our analysis by predicting the top 1% HCV epitopes with the highest HLA binding affinities (i.e., 30 epitopes) for each HLA allele, thereby excluding potential bias introduced by different repertoire sizes.

As NetMHC 3.2 does not provide ligand predictions for HLA-C loci, we predicted the epitopes of the three HLA-C molecules in Table 4.1 and Supporting Table 4.1 (Cw*05, Cw*01, and Cw*04) using NetMHCpan 2.4 (172) which has a larger allele coverage.

The Stabilized Matrix Method (SMM) was applied to predict proteasomal cleavage (173), which is the main selection step during peptide generation for antigen presentation. Using a threshold of 1.2, we were able to simulate the specificity of the proteasome, i.e., about one fourth of all possible HCV peptides were predicted to be processed (174) .

4.3.4 HCV proteome coding

In order to distinguish epitope-covered regions from epitope-free regions of the HCV proteome, each residue of the HCV consensus sequence was defined either as "epitope-covered" if it is part of a reported epitope, or as "epitope-free" if none of the epitopes in our experimental HCV epitope data set (see above) overlapped with the residue.

4.3.5 Statistical Analysis

All statistical analysis was performed using the R environment (http://www.r-project.org/).

4.4 Results

4.4.1 Classification of HCV proteins by their sequence variability

The relative variability of HCV proteins was assessed by estimating the average Shannon entropy of all positions for each protein using pre-aligned amino acid sequences from all genotypes (genotypes 1-6) available from the Los Alamos HCV database. Each of these alignments includes only one sequence per patient. In order to allow for direct comparison between the proteins, we corrected for the number of sequences per protein alignment (see Material and Methods). We observed that HCV proteins vary in variability (see Figure 4.1, in which the HCV proteins are ranked with respect to their median entropy). CORE is the least variable HCV protein, although the levels of conservation of CORE, NS5B and NS3 are indistinguishable ($p>0.1$, Mann-Whitney test), and E1 is the most variable. The three most conserved proteins, CORE, NS5B and NS3, are significantly less variable than the other proteins ($p<0.0001$, Mann-Whitney test by pooling the entropy values for CORE, NS5B, and NS3 into one group and the other proteins into another group). The degree of conservation is probably reflecting the structural flexibility of these proteins: The structural protein CORE, for example, constitutes the capsid of the virus particle and affects several cellular processes (175). For most of the HCV proteins, the functional and structural constraints remain to be identified.

In a previous study, we have shown that HIV CD8+ T cell epitopes are located in more conserved regions of the HIV proteome (176). To see if this is the case also for HCV CD8+ T cell epitopes, we mapped experimentally verified HCV epitopes to the proteome sequence of the HCV reference strain (accession number AAA45534.1) and compared the level of sequence conservation between epitope covered residues and non-epitope residues (see Material and Methods for details). The epitope data was obtained from the LANL HCV immune database and the IEDB epitope repository, resulting in 187 unique $CD8^+$ T cell epitopes. These epitopes are partially overlapping and cover in total 1173 residues of the HCV proteome. The level of conservation was similar in HCV positions within CTL epitopes (median entropy 0.09) and those positions that never overlap with epitopes (median entropy 0.11). Note that this result is based on a limited epitope data set. Far more epitopes restricted by HLA-A alleles (n=124) have been identified, compared to epitopes restricted by HLA-B (n=57) and HLA-C alleles (n=5). Future identification of a

larger number of CTL epitopes, especially HLA-B and -C epitopes, will provide a better estimate of the epitope coverage of the HCV proteome and may therefore alter our current result.

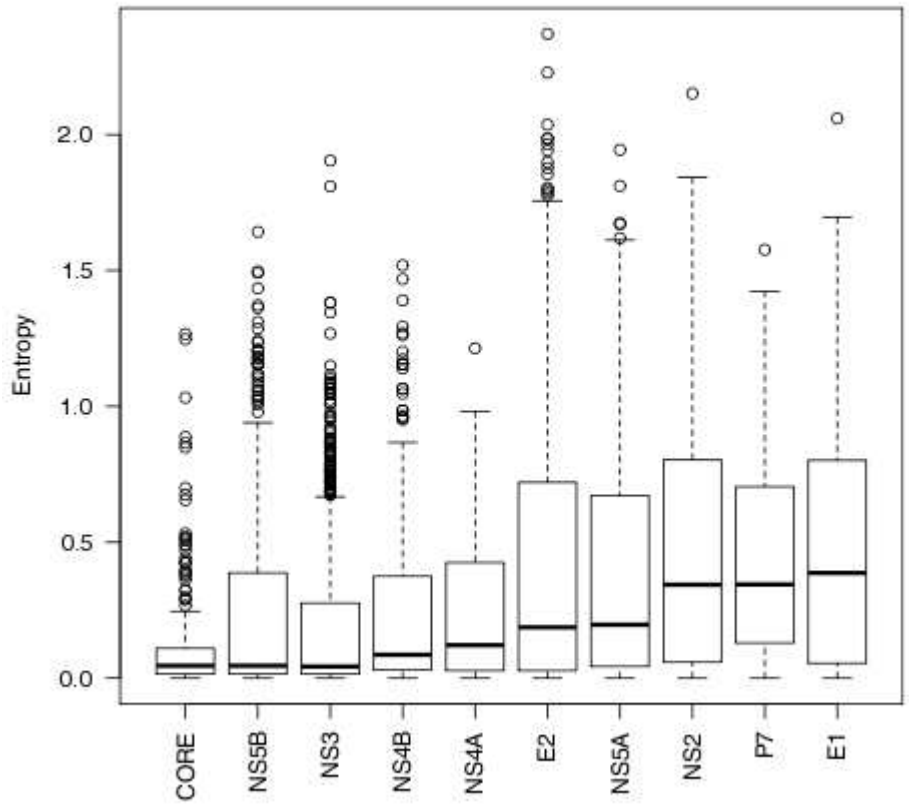

Figure 4.1. Core, NS5B, and NS3 are the most conserved HCV proteins.
Protein variability was estimated using Shannon entropy scores calculated for each position in a protein alignment for each of the HCV proteins (see Material and Methods). The median entropies of all positions in the alignment of each of the 10 HCV proteins are shown (the horizontal line indicates the median, the box indicates the 25- and 75 percentile, and the whiskers indicate 1.5 times interquartile range). Each protein alignment contains the same number of sequences (n=466 being the lowest number of sequences per protein available). For the alignments where more sequences were available, a subset of sequences was chosen randomly. This process was repeated 100 times, each time resulting in very similar entropies (none of the runs produced significantly different entropies than the one shown in the figure). The figure shows the entropies calculated for one batch of alignments.

4.4.2 HCV-HLA associations and analysis of known HCV epitopes

Since HCV is known to escape CTL pressure by continuous evolution (177) , CTL responses to structurally and functionally constrained parts of the HCV genome are likely to be most beneficial. Differences in the tendency of HLA molecules to present conserved parts of HCV may therefore underlie the HLA-related differential HCV infection outcome. To test this hypothesis, we studied the HLA-restriction of all known CTL epitopes from the HCV LANL and IEDB databases (see Methods). Recently, Kuniholm et al. (161) conducted a literature review of HCV-HLA associations. Using stringent selection criteria, the authors identified five HLA class I allele groups (B*18, B*27, B*57, Cw*01, and Cw*04) that are strongly associated with either HCV clearance or persistence (Table 4.1). Of this list of

HLA allele groups, CTL epitopes have only been reported for B*27 and B*57. To extend the list of HLA groups for this part of our analysis, we included HLA class I associations reported by additional studies (see Supporting Table 4.1). Given this extended list, CTL epitopes have been identified only for five of eight HLA-B groups and none of the HLA-C allele groups. Based on this small data set, we observed that six out of ten experimentally verified HCV $CD8^+$ T cell epitopes restricted by protective HLA allele groups (B*57 and B*27) are located in the HCV protein NS5B (Figure 4.2), which is one of the most conserved HCV proteins (Figure 4.1). The remaining four epitopes are distributed over four other HCV proteins (Figure 4.2 and Supporting Table 4.2). In other words, the epitopes presented by protective HLA allele groups are significantly enriched in NS5B ($p<0.01$, permutation test, Figure 4.2B and 4.2C). More epitopes are reported for susceptible HLA allele groups (n=27, see Supporting Table 4.2), and the distribution of these HCV epitopes does not differ from the expected distribution ($p>0.1$ in all tests).

Table 4.1: HLA class I alleles with strong associations with HCV viremia as reported by Kuniholm et al. (178).

Allele	Reference	HCV RNA Status	N_E [a]	Population	Epitope-enriched proteins [b]
B*57	(178)(179)(180)(181)	Clearance	4	Mixed	NS5B
B*27	(64) (182) (183)	Clearance	6	Irish, German	CORE, NS5B
B*18	(64)	Persistence	-	Irish	NS5A
Cw*01	(178)(179)(64)	Clearance	-	Cau	
Cw*04	(179)(184)	Persistence	-	Cau, Irish	

[a] Stands for the number of CTL epitopes identified so far.
[b]This analysis is based on in silico predictions, as there are too few CTL epitopes identified so far. We looked at the protein distribution of top 30 predicted HCV epitopes per HLA allele and compared it with the expected number of epitopes based on the protein length. Enrichment indicates a significantly larger number of predicted epitopes ($p<0.05$ permutation test) than expected. Result of B*27 and B*18 is based on the 9mer epitope prediction while result of B*57 is derived from 10mer epitope prediction.

4.4.3 Protective HLA class I molecules preferentially present peptides from conserved HCV proteins

The analysis based on experimentally verified epitopes suggests that protective HLA allele groups, like B*27 and B*57, prefer to present epitopes from the conserved HCV protein NS5B. However, the power of this analysis is limited because for many alleles no CTL epitopes have been reported yet. We therefore investigated differences between protective and susceptible HLA alleles by analyzing the predicted HCV peptide repertoires (and, thus, potential HCV epitopes) of the HLA alleles listed in Table 4.1 (see Methods for details). This analysis confirmed the preference of protective allele groups to present epitopes from the conserved HCV protein NS5B: 12 out of the 30 top-ranking HCV epitopes for HLA-B*2705 stem from NS5B (40%), which is a significantly higher number than expected ($p=0.02$, permutation test, Figure 4.3). Similarly, HLA-B*5701 has a preference to present NS5B ($p=0.04$, permutation test, Figure 4.3). For HLA-Cw*0101, which is also associated with clearance of HCV (161, 185, 186), the presentation of NS3, another conserved protein, is enriched, though not significantly. The non-protective alleles, on the other hand, target different HCV proteins. For example, B*1801 is predicted to present a significantly higher number of epitopes from NS5A ($p<0.05$, permutation test, Figure 4.3). The predicted preferences of other

HLA alleles are summarized in Supporting Table 4.1. Omitting the filtering of the predicted HLA ligands for their probability to be generated by the proteasome (see Material and Methods) did not change these results (data not shown).

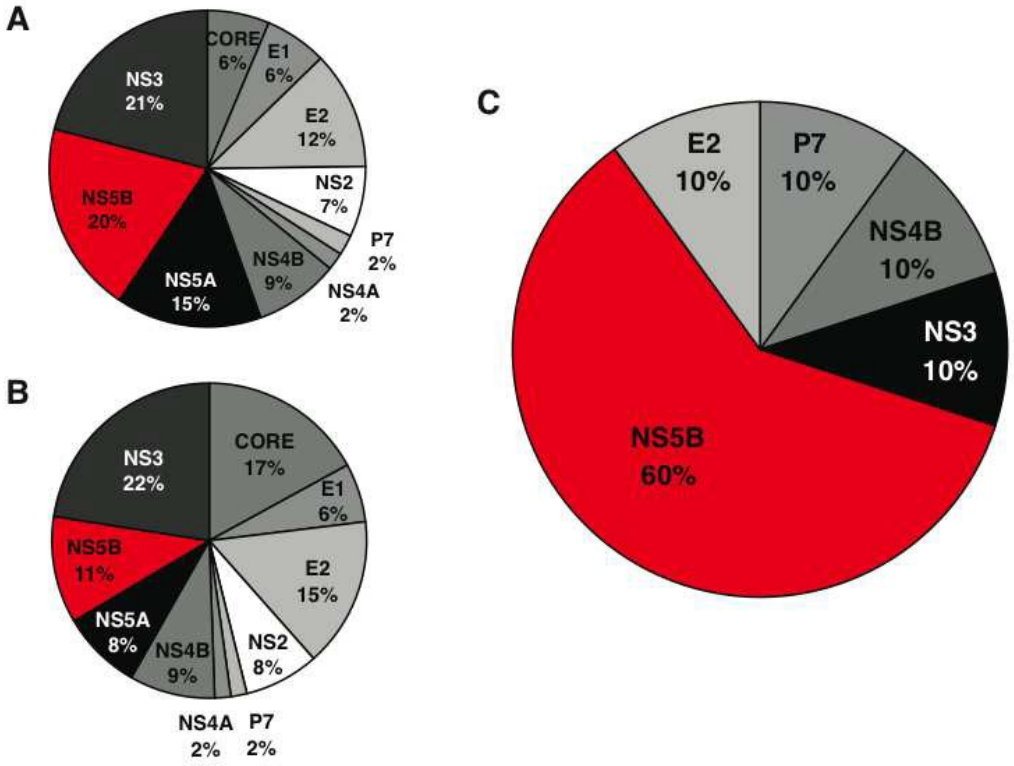

Figure 4.2 NS5B is enriched in experimentally verified epitopes restricted by protective HLA allele groups (HLA-B*27, HLA-B*57).
A. The normalized length of each protein in the complete HCV proteome. **B.** The distribution of all unique epitopes reported in the IEDB and LANL databases. **C.** The distribution of unique experimentally verified epitopes restricted by the protective alleles. This distribution is significantly different from what is expected based on the distributions given in **A** and **B** ($p<0.01$ permutation test).

In an alternative approach to study preferential targeting of HCV proteins by protective and susceptible HLA alleles, we followed the strategy taken in previous studies on HIV-1 (187) and HTLV-1 infection (164). We ranked the predicted HCV epitopes by their predicted HLA binding affinity within all HCV proteins and then plotted the ranks of the two best-binding peptides derived from the conserved HCV proteins CORE, NS5B and NS3 among all other HCV epitopes (Figure 4.4). Also this analysis shows that the three HLA alleles associated with HCV clearance have a significantly stronger preference for peptides from the conserved proteins CORE, NS5B and NS3 than the susceptible alleles B*1801 and Cw*0401 (Mann-Whitney test, $p=0.02$, Figure 4.4A). At the single protein level, this preference is clear for NS5B ($p=0.04$, Figure 4.4B), while for CORE and NS3 the difference between the protective and susceptible alleles is not significant (Supporting Table 4.3). The susceptible HLA molecules have a tendency to present peptides from NS5A (Mann-Whitney test, $p=0.09$, Figure 4.4B, Supporting Table 4.3). However, the differences at the protein level disappear when we correct for multiple testing of significance using the

false discovery rate (FDR) procedure (188). Nevertheless, the results are rather robust: removing any of the protective or susceptible alleles or taking more epitopes into account (up to 10 per protein) does not change these results (data not shown). Taken together, the *in silico* analysis as well as our analysis based on a limited number of known HCV CD8+ T cell epitopes suggest that HLA molecules associated with clearance of HCV viremia tend to target conserved proteins of HCV (especially NS5B), while HLA molecules enriched among patients with chronic HCV infection target other (and less conserved) HCV proteins.

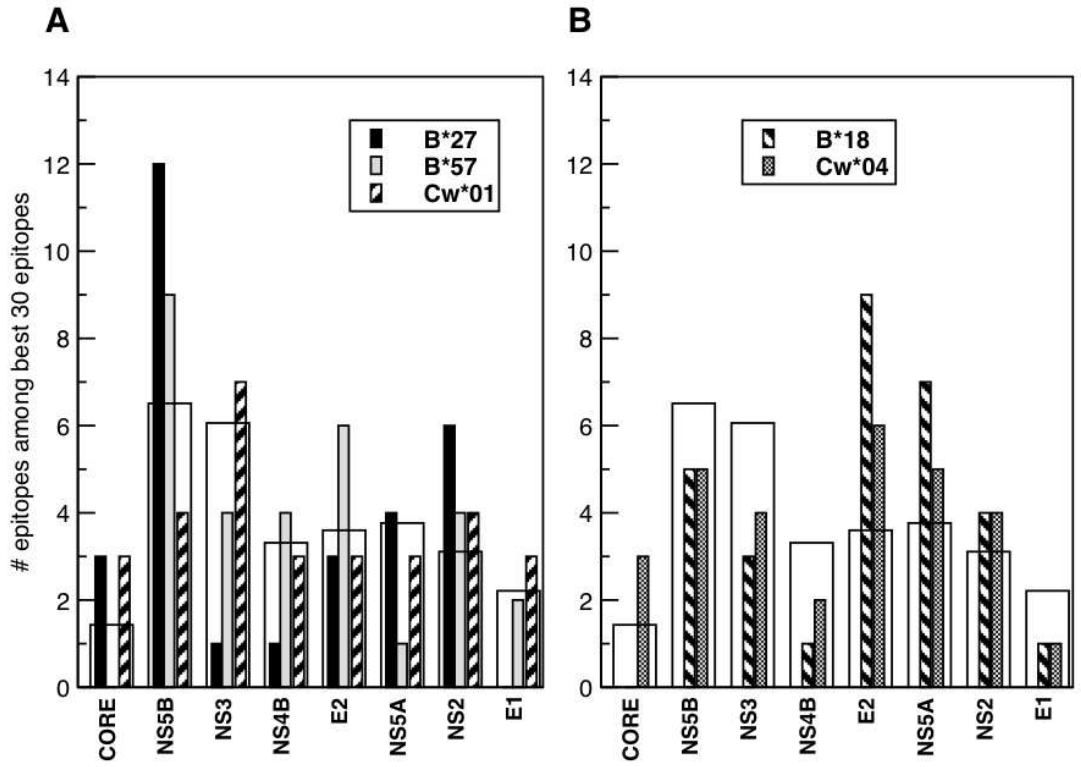

Figure 4.3 NS5B is enriched in predicted epitopes presented by HLA-B*27 and B*57.
The bar plot shows the distribution of the top 30 predicted epitopes over all HCV proteins for (A) protective alleles, and (B) susceptible alleles. Open bars show the expected number of epitopes based on the number of peptides that are predicted to be cleaved properly in every HCV protein (i.e., the more peptides of length 9 or 10 are produced by the proteasome from a specific protein, the more epitopes are expected to be found). Alternatively, the expected number of epitopes can be calculated using the protein length (i.e., the longer the protein, the more epitopes are expected to be found). The latter approach produces very similar values as the ones reported here (results not shown). The allele names are given in the legend and HCV proteins are sorted with respect to their Shannon entropies given in Figure 4.1.

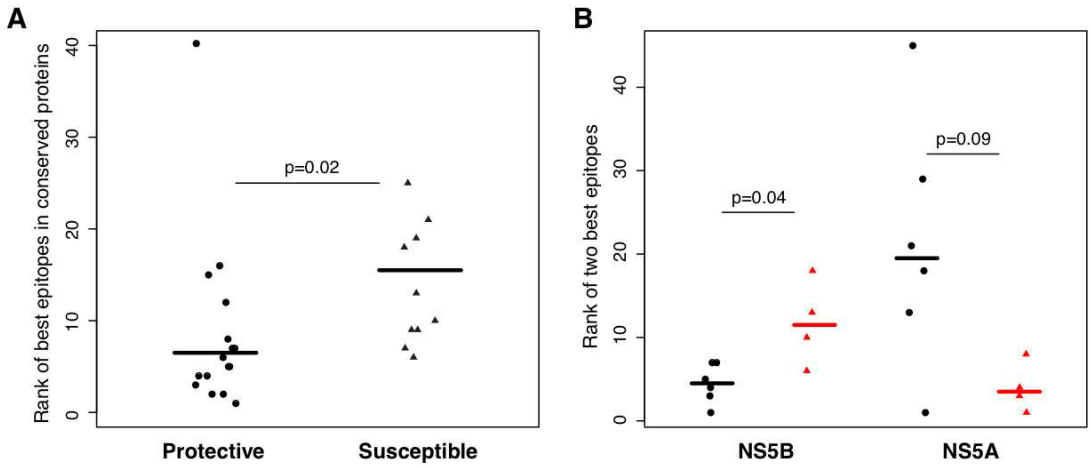

Figure 4.4 HLA alleles associated with HCV clearance preferentially present peptides from conserved proteins.
(**A**) The relative rank of epitopes from the three most conserved HCV proteins, Core, NS5B and NS3, are plotted for the protective (B*5701, B*2705 and Cw*0101, black circles) and susceptible (B*1801, Cw*0401, red triangles) HLA molecules. The median of each group is indicated by a horizontal line. The relative rank is calculated with respect to all predicted epitopes of the HCV proteome. (**B**) The preferential presentation of peptides from NS5B and NS5A by the protective (circle) and susceptible (triangle) HLA molecules, respectively. For each molecule the rank of the two best peptides from either NS5 protein is plotted.

4.5 Discussion

Several studies have shown that both $CD4^+$ and $CD8^+$ T cell responses play an important role in controlling HCV replication during infection (155–158) , and certain HLA alleles have been significantly associated with differential infection outcome (161). The underlying mechanism behind these associations remains unclear. In case of HIV-1 and HTLV-1, it was shown that the immune system can effectively control infectious diseases by targeting certain pathogen-derived proteins (164, 187). Similarly, other studies found that T cell epitopes are not equally distributed over the genome of HIV-1 and *M. tuberculosis*, but that they cluster together in conserved regions (176, 189, 190). We hypothesized that targeting conserved regions within the HCV proteome by HLA alleles may be a potential mechanism to control viral replication during the acute phase of HCV infection, which may then further determine disease outcome.

To test this hypothesis, we first determined the differences in sequence conservation between HCV proteins. In line with previous studies (191), we found that CORE, NS3 and NS5B are the three most conserved proteins, while the other proteins show a significantly higher level of variability. Next, we performed a literature study to compile a list of protective and susceptible HLA alleles. This turned out to be rather difficult due to limited data and differences between ethnic groups. Using the five strong HLA class I/HCV associations defined by Kuniholm et al. (161) as a basis for our analysis, we found that both experimentally verified and *in silico* predicted $CD8^+$ T cell epitopes restricted by protective HLA molecules are enriched within conserved HCV proteins, especially in NS5B. This result does not fully explain the differences between protective and susceptible alleles. When we analyzed the distribution of HCV CTL epitopes presented by susceptible alleles, we did not observe a (strong) bias towards variable HCV proteins. The most striking difference was seen for the presentation of NS5A, which tends to be

enriched in CTL epitopes restricted by susceptible HLA class I molecules compared to the protective ones. These observations suggest that the main mechanism behind the association between HCV persistence and HLA molecules may be the lack of sufficient peptide presentation from conserved HCV proteins (especially NS5B) rather than a preference for presenting peptides from more variable HCV proteins like NS2 and E2.

The analysis concerning the comparison of epitope repertoires of protective and susceptible alleles is sensitive to the precise definition of susceptible HLA alleles. For example, including B*1801 as the only susceptible allele drastically increases the differences in targeting NS5A and NS5B by the two groups of HLA molecules, while adding B*0801 as another susceptible allele diminishes this difference. The latter would be in agreement with a previous study reporting that B*08 is associated with clearance of HCV infection (192). Obviously, additional large cohort studies are needed to better define HLA associations with HCV disease outcome. Still, we believe that our study is presenting the first clues on how different HLA molecules might shape HCV infection outcome.

So far there is no direct evidence that targeting more conserved HCV proteins by $CD8^+$ T cells in a host helps controlling HCV replication. However, it has been widely argued that $CD8^+$ T cell responses specific towards conserved regions of HIV-1 (especially Gag p24) are associated with slow disease progression (187, 193, 194). As in the case of HIV-1, the most conserved HCV proteins are also the most functionally and structurally constrained ones, which makes it more difficult for the pathogen to escape from host immune control in these proteins than in other regions. In line with this, it has been shown that a point mutation within a B*27 restricted dominant T cell epitope in NS5B entails a large fitness cost (195), which is probably largely due to the important role of NS5B in viral replication. Another conserved HCV protein, CORE, forms the capsid of the virus particle, a role similar to HIV-Gag. Taken together, these data (in part) shed light on why immune responses targeted to conserved proteins of HCV might result in a favorable infection outcome.

4.6 Supplemental materials

Table S 4.1 Additional HLA class I alleles with strong associations with HCV viremia from the literature.

Allele	Reference	HCV RNA Status	N_E [a]	Population	Epitope-enriched proteins[b]
B*15	(65)	Persistence	-	West India	
B*55	(65)	Persistence	2	West India	
B*08	(180)(64)	Persistence	6	Mixed	NS4B
B*35	(196)	Persistence	15	Korean	
B*46	(196)	Persistence	-	Korean	
A*2301	(179)	Persistence	1	Cau	P7
A*29	(197)	Persistence	3	Egyptian	P7
Cw*05	(198)	Clearance	-	Cau	

[a] Stands for the number of CTL epitopes identified so far.
[b]This analysis is based on in silico predictions, as there are too few CTL epitopes identified so far. We looked at the protein distribution of top 30 predicted HCV epitopes per HLA allele and compared it with the expected number of epitopes based on the protein length. Enrichment indicates a significantly larger number of predicted epitopes ($p<0.05$ permutation test) than expected.

Table S 4.2 Experimentally verified CTL epitopes restricted by the susceptible/protective HLA class I alleles listed in Table 1 and S1.

Epitope	Protein	Position[a]	HLA allele[b]
LPGCSFSIF	CORE	169_177	B35
PQRKTKRNTNR	CORE	7_17	B8
CPNSSIVY	E1	16_23	B35
GNASRCWVAM	E1	42_51	B35
NASRCWVAM	E1	43_51	B35
ASRCWVAM	E1	44_51	B35
VPASQVCGPVY	E2	114_124	B35
NTRPPLGNWF	**E2**	**158_167**	**B57**
YISWCLWWL	NS2	29_37	A23
LMALTLSPYYKRY	NS2	17_29	A29
MALTLSPY	NS2	18_25	A29
STDATSILGI	NS3	294_303	B35
HSNIEEVAL	NS3	333_341	B35
LVRLKPTL	NS3	585_592	B8
ELAAKLVGL	NS3	376_384	B8
HSKKKCDEV	NS3	369_377	B8
HPITKYIMACMSADL	NS3	613_627	B8
YSTYGKFLAD	NS3	267_276	B35
YGKAIPLEVI	NS3	350_359	B35
GRGKPGIYRF	**NS3**	**466_475**	**B27**
TPAETTVRL	NS3	505_513	B35
EVTLTHPITKYIMTCMSA	NS3	608_625	B8
IPDREVLY	NS4A	38_45	B35
LTTSQTLLF	**NS4B**	**90_98**	**B57**

LPYIEQGMML	NS4B	4_13	B35
PCEPEPDVAVL	NS5A	189_199	B35
EPEPDVAVL	NS5A	191_199	B35
GRAAICGKY	**NS5B**	**516_524**	**B27**
KSKKTPMGF	**NS5B**	**209_217**	**B57**
QPEKGGRKPA	NS5B	148_157	B55
KGGRKPARLIVFPDL	**NS5B**	**151_165**	**B27**
ARHTPVNSW	**NS5B**	**400_408**	**B27**
ARMILMTHF	**NS5B**	**421_429**	**B27**
SPGEINRVAA	NS5B	478_487	B55
LGVPPLRAWR	**NS5B**	**492_501**	**B57**
GRWVPGAAY	**P7**	**34_42**	**B27**
FYGMWPLLL	P7	44_52	A29

[a] The position is with respect to the HCV consensus strain, H77.
[b] The epitopes restricted by the protective alleles are highlighted in bold.

Table S4.3 Comparison of the strength of binding of alleles associated with HCV clearance (Protective) and HCV persistence (Detrimental) to each of the HCV proteins. Since P7 is a very short protein, we analyzed top ranking epitopes from P7 and NS2 together. Similarly, epitopes within NS4A and NS4B were pooled together.

Protein	P-value	Group with strongest binding
Core	0.17	Protective
NS5B	0.04	Protective
NS4	0.52	-
NS5A	0.09	Detrimental
NS3	0.61	-
E2	0.28	Detrimental
E1	0.91	-
NS2+P7	0.67	-

Chapter 5. HLA class I allele promiscuity revisited

Rao, X[1]., Hoof, I[1]., Costa, A[2]., Van Baarle, D.[2], and Keşmir, C[1].

[1] Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, The Netherlands
[2]Department of Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
Immunogenetics ,2011

5.1 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 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.

5.2 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 (199–201). 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 (202–208).

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 (209–213), 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" (214–220). Few other reports showed promiscuity across supertypes or even loci; these findings were considered as "exceptions" (221, 222). 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.* (223) 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.* (224) 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.* (217) 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 (224) and Table 5.3 of this manuscript), 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 (225)) 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 (226–228), 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.

5.3 Materials and Methods

5.3.1 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 IC50 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.

5.3.2 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).

5.3.3 Promiscuity at supertype level

Throughout our analysis, we followed allele-supertype associations defined by (208) 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 (224).

5.3.4 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 (229, 230). 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 S5.4A, S5.4B). The peptide-MHC binding predictions for majority of these molecules are of high quality (228).

HLA class I ligand prediction

To have an as large as possible population coverage, we used NetMHCpan (228) to predict peptide-HLA binding affinity. NetMHCpan assigns to each peptide-HLA pair a predicted IC50 value, indicative of the predicted binding affinity. An IC50 threshold of 500 nM was used to discriminate HLA binding ligands from nonbinding peptide. 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 (227, 231, 232). 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 Figure. 5.1). 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 IC50 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 (233). To avoid this, we repeated the analysis by defining the top-ranking 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 TAP transport efficiency and proteasomal cleavage, which are the two main steps of antigen processing (234). Applying an alternative predictor of antigen processing, NetChop (235, 236), did not affect our results.

5.3.5 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 S5.1) 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 S5.3 to test the dependence of our results on the initial set of viral proteomes.

5.4 Results

5.4.1 HLA class I binding shows a high degree of promiscuity

To our knowledge Frahm *et al.* were the first to study HLA class I binding promiscuity systematically (223). 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 (223). Using the pan-specific MHC class I binding predictor NetMHCpan (226, 228) 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 Frahm *et al*. data is significantly lower than it is 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 (237–239).

To test the HLA class I binding promiscuity in an independent data set, we analyzed HLA class I binding data from the IEDB database (225) (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 5.1). Using a more stringent criterion, e.g., when including only the peptides which were tested on eight or 10 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 5.1). 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 6 HLA alleles, 82 (60%) elicited responses in the context of two or more HLA molecules.

Table 5.1 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/ (229)). We predicted potential 9-mer binders to these HLA molecules within common viral proteomes (n=17, see Table S5.1) using NetMHCpan (226, 228). This prediction method was demonstrated to be the best one in a recent large benchmark performance test (240). The most frequent HLA molecules (top 20 for A- and B-locus, respectively, listed in Table S5.4) 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 (Figure. 5.1). As expected, using a more stringent threshold to define the peptide binding (e.g., a predicted IC50 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 (227) (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 5.1 and Figure. 5.1 (results not shown). In order to evaluate whether antigen processing has an impact on ligand promiscuity, we then added TAP and proteasomal cleavage predictions (234–236) to our MHC binding predictions. The level of

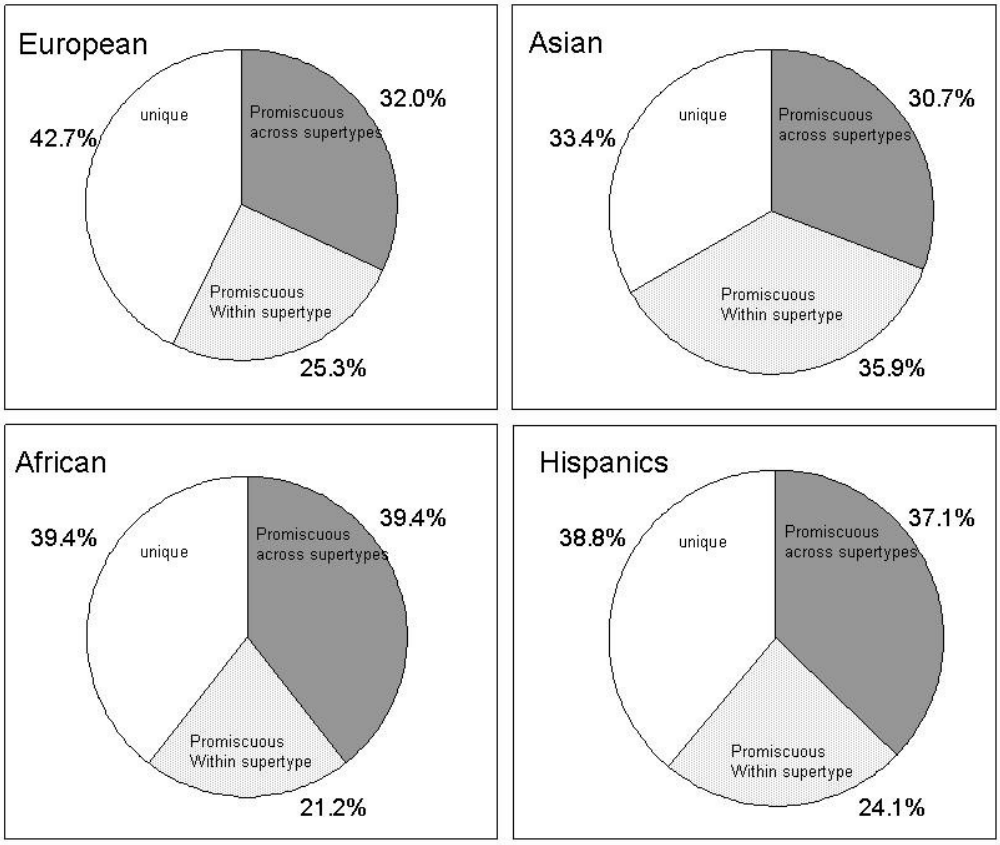

Figure 5.1 Distribution of predicted HLA class I ligands of viral origin.
All predicted ligands of the 20 most frequent HLA-A and HLA-B molecules in US subpopulations of a certain ethnic background (European, African, Asian and Hispanic) were classified into three categories: unique ligands (exclusively 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).

promiscuity remained very similar (data not shown), implying that (predicted) antigen processing does not significantly influence the ligand sharing among HLA class I molecules. 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 have exceptionally many peptides that exhibit promiscuous HLA binding (Table 5.2). The HLA-B44 supertype, analyzed by Hillen *et al*. (224) 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 5.2). A similar result was obtained in *in silico* analysis where we estimate the HLA peptide binding promiscuity on the population level (see Table S5.2).

Table 5.2 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 are exclusively bind to the HLA molecules outside of the supertype.

5.4.2 Comparison of different experimental approaches for HLA peptide binding promiscuity

Hillen *et al*. (224) 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 5.3). 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 (224). The binding promiscuity of ligands from different allele varies between 18% to none, with an average of 10% (Table 5.3). 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 (223) ii) eluted peptides (224) 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 5.3). 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 5.3). 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 (224). 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 5.3). T cell response data extracted from IEDB show a similar trend: out of 14 T cell epitopes

Table 5.3 HLA-B44 alleles used in Hillen et al., 2008 and HLA-B44 epitopes tested in Frahm et al., 2007.

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.

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 above results 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 5.2, 57% within supertype).

5.4.3HLA-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., (241, 242)). 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., (219, 242–244). 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 tested, 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, Figure. S5.1).

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 Figure. S5.2, p=0.34, Mann-Whitney U test). We repeated this analysis using a different criteria to define "binders" (i.e., using the top 1-2% percentile, see Materials and Methods) and by extending our viral data set (Table S5.3), 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.

5.4.4 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 (245). 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 (243, 246). 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. In line with the studies demonstrating that HLA-B alleles show the strongest association with disease outcome in HIV-1 infection (22, 193, 247), we found a strong negative correlation between the fraction of uniquely presented epitopes of HLA-B molecules and median viral loads reported by (248) (Figure. 5.2A, r=-0.57, p=0.02, Spearman correlation test) or the relative hazard (RH) reported by (249) (Figure. 5.2B, 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 S5.5), implying that

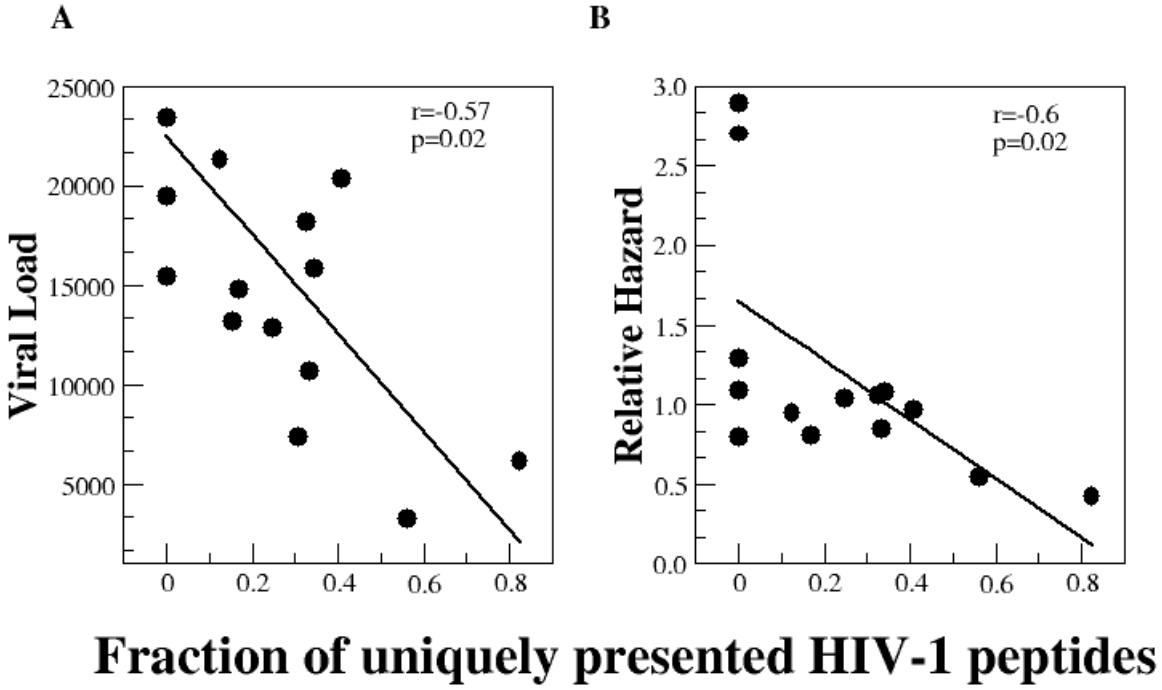

Figure 5.2 Correlation between the predicted fraction of unique ligands for HLA-B molecules and mean set point viral load associated with the same molecule or the relative hazard (RH).
The fraction of uniquely presented HIV-1 peptides for an allele was calculated by comparing the predicted HIV-1 peptides 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 S5.4) in the Caucasian population. The predictions were performed with NetMHCpan (228) 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 (248) and (B) RH taken from (249) 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 S5.5.

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 (22), the fraction of uniquely presented epitopes no longer significantly correlated with the median viral loads (results not shown).

5.5 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 (202–208). 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 (250, 251). These findings were challenged by Hillen *et al*. who showed that even within a supertype, eluted HLA ligands can show as little as 3% promiscuity (224).

We have taken another approach to estimate HLA peptide binding promiscuity by using *in vitro* binding measurements reported in the IEDB database (225). 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 5.1, Figure. 5.1). 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 5.2), considering that around 100,000 MHC molecules are expected to be on the cell surface at a time (252). 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 repertoire 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 (253–255). 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 Figure. 5.2). 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 (256). The heterozygous advantage was suggested to generate a more effective immune response and therefore resulted in better control of HIV-1 infection (257). 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.

5.6 Acknowlegments

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

5.7 Supplemental Materials

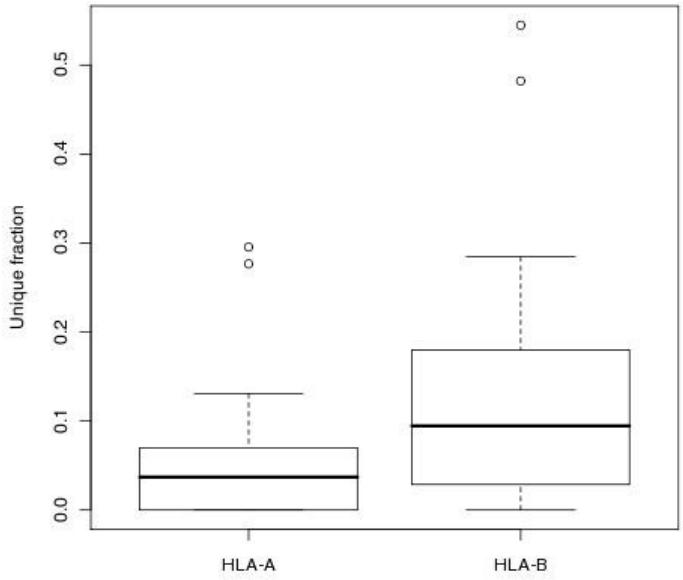

Figure S5.1 Promiscuity of HLA-A and HLA-B ligands based on IEDB MHC binding data.

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).

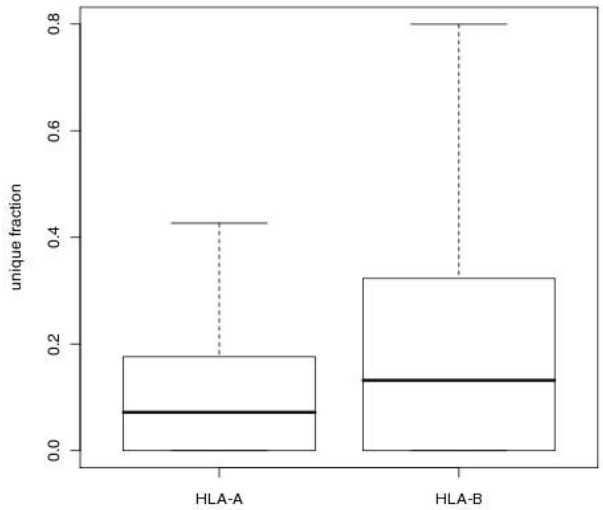

Figure S5.2 Promiscuity of predicted HLA-A and HLA-B ligands based on NetMHCpan predictions.
All possible 9-mer viral ligands from viral proteomes (listed in Table S5.1 and S5.3) were predicted for the 20 most frequent HLA-A and HLA-B alleles in USA subpopulation of European ethnic background. The fraction 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 S5.1 List of viral proteomes used as the resource of NetMHCpan predicted ligands.

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 /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 S5.2 Ligand promiscuity of HLA supertypes for different ethnic groups based on NetMHCpan prediction data. We obtained very similar results with NetMHC3.2 predictor.

	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 S 5.3. List of human viral proteomes used in the extension of the viral set.

HUMAN VIRUSES	
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 S 5.4 A 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 S 5.4 B 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
B0702	0.140	1	0.073	2	0.026	15	0.055	4
B0801	0.125	2	0.038	9			0.045	6
B4402	0.090	3	0.021	17			0.033	9
B1501	0.067	4	0.030	13	0.035	11	0.029	10
B3501	0.057	5	0.065	3	0.043	5	0.064	1
B4001	0.056	6			0.080	1		
B4403	0.050	7	0.054	6	0.042	6	0.061	2
B1801	0.046	8	0.036	10			0.040	8
B5101	0.045	9	0.022	16	0.063	2	0.058	3
B5701	0.038	10			0.021	18		
B2705	0.033	11					0.017	18
B1402	0.031	12	0.022	15			0.041	7
B1302	0.026	13			0.023	17		
B3801	0.022	14					0.019	16
B5501	0.017	15						
B3503	0.016	16			0.024	16		
B3701	0.013	17						
B4901	0.013	18	0.028	14			0.024	12
B3502	0.011	19						
B4002	0.010	20			0.031	14	0.049	5
B5301			0.112	1				
B1503			0.062	4			0.016	20
B4201			0.055	5				
B5802			0.041	8				
B5801			0.035	11	0.058	4		
B5703			0.034	12				
B8101			0.020	18				
B1516			0.017	19				
B5201			0.014	20	0.037	8	0.027	11
B4501			0.045	7				
B4601					0.061	3		
B4006					0.037	7		
B3802					0.037	9		
B1502					0.036	10		
B5401					0.033	12		
B1301					0.033	13		
B4801					0.020	19	0.021	14
B0705					0.020	20		
B3905							0.023	13
B3906							0.020	15
B3512							0.019	17
B3517							0.016	19

Table S 5.5 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.

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. 5.2 of the main text.

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

Chapter 6. Do the binding motifs of HLA molecules influence the haplotype frequencies?

Rao,X., De Boer, R., Debbie van Baarle[a], Maiers, M[b], and Kesmir, C.

Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, The Netherlands
[a]Immunology Department, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands
[b]National Marrow Donor Program, Minneapolis, Minnesota

Manuscript in preparation

6.1 Abstract

Different HLA haplotypes (specific sets of HLA-A, -B, -DR locus alleles inherited together from a parent) are observed in different frequencies in human populations. Some haplotypes, like HLA-A1-B8, are very common, reaching upto 10% in caucasian population, while others remain very low. Numerous studies identified associations between HLA haplotypes and diseases, and the differences in the haplotype frequencies can be in part explained due to these associations: stronger the association with a severe (autoimmune) disease, lower the HLA haplotype frequency. Additionally, the peptide repertoire of HLA molecules composing an haplotype might influence the frequency of an haplotype. For example, it could be advantegous to have MHC molecules with totally non-overlapping binding specificities in a haplotype, as individuals having such an haplotype would have most diverse set of peptides presented on the cell surface. To test this hypothesis, we estimate the total ligand repetoire of HLA class I haplotypes (HLA-A-B) using *in silico* predictions and compare the size of this repertoire to the HLA haplotype frequencies reported in NMDP. We find that the most common HLA-A-B haplotypes contain HLA-A and –B genes with completely non-overlapping peptide repertories, i.e., with distinct binding motifs. This result suggests that on the haplotype level there might be a selection pressure, though not so strong, that favors the haplotypes having HLA alleles with distinct binding motifs and thereby the largest possible presented peptide repertoires.

6.2 Introduction

The human leukocyte antigen (HLA) genes are the most polymorphic coding loci known in humans. HLA gene cluster is located on the major histocompatibility complex (MHC) on the chromosome 6 and contains over 200 genes. Two groups of loci, coding for MHC class I and class II genes, are the most poymorphic

ones and dictates T cell responses. It is generally believed that this variability is maintained by a balancing selection as individuals being heterozygous in HLA class I and II loci seem to have a better outcome in infections diseases (see e.g. for HIV-1 (258)). In line with this, it has been demonstrated that in several species, and up to a certain extent in humans, females prefer to mate with males having dissimilar HLA to their own, increasing the chance of their offspring to survive infectious diseases (259). However, we argued that the heterozygous advantage on its own is not enough to maintain such a large polymorphism and that co-evolution with pathogens might play a major role as well (260).

On the functional level, HLA polymorphism seems to be much more limited: it was well established that MHC class II molecules have largely overlapping peptide repertoires (see e.g., (261)). In the last few years we and others showed this to be true for class I alleles as well (262)(263)(264)(265). The promiscuos peptide binding is not limited to within a locus, but it extends to other MHC class I loci (263)(265). This property has important consequences: An heterozygous individual carrying two different HLA molecules in a locus with largely overlapping peptide repertoires should nevertheless be considered functionally as an homozygous individual, diminishing the heterozgous advantage.

Frequent recombinations in HLA region results in novel HLA haplotypes continuously (266). However, the number of haplotypes found in human populations is far lower than the number of alleles observed (267), suggesting that not all the HLA haplotypes have the same chance of being established in human populations. Indeed, different haplotypes occur with very different frequencies in different ethnic groups/subpopulations (www.allelefrequencies.net, www.nmdp.org). It is likely that the ``usefullness'' of HLA molecules play a role in determining the haplotype frequencies: haplotypes carrying HLA molecules that are protective for certain diseases are expected to be in higher frequencies in the endemic areas. Indeed, several HLA haplotype disease associations are established, e.g., autoimmuno disease (268), squamous cell cervical cancer (269) and recurrent aphthous stomatitis (270). Along these lines, one can speculate that it might be advantegous to have MHC molecules with totally non-overlapping binding specificities in a haplotype, because such haplotypes would give an individual larger epitope repertoires than haplotypes containing HLA molecules with similar binding motifs. To test this possibility, we here study the peptide repertoires of HLA-A and HLA-B molecules that are estimated to be in an HLA-A-B haplotype (i.e., the ones with a strong linkage disequlibrium) using an *in silico* approach. The peptide-HLA predictions for the other loci (HLA-C, HLA-DR etc.) are of lower quality than HLA-A and –B predictions (77), and therefore are left out of this analysis. In four major ethnic groups (European, African, Asian and Spanic) HLA-A-B haplotypes with no overlapping peptide repertories have in general higher frequencies than overlapping haplotypes. Especially, the most common haplotypes seem to be composed of HLA molecules with non-overlaping peptide repertoires. These results suggest a possible selection pressure, besides many others, to maintain HLA haplotypes with functionally distinct class I molecules in high frequencies in human populations.

6.3 Materials and Methods

6.3.1 NMDP data

All estimated HLA-A-B haplotype frequencies were downloaded from the National Marrow Donor Program (NMDP) which was established to develop and maintain a registry of HLA-typed volunteer unrelated donors for patients in need of a hematopoietic stem cell transplant (http://bioinformatics.nmdp.org/). Four predominant US ethnic and racial groups were included in this data set: European Americans, African Americans, Asians and Hispanic (271). Haplotype frequencies were estimated separately for each ethnic group using an implementation of the expectation maximization (EM) algorithm (272)(273).

Linkage disequilibrium, *D*, between two alleles in each haplotype was expressed as the difference between the observed and expected haplotype frequency:

$D = f_{AB} - f_A f_B$, where f_{AB} is the observed (estimated) haplotype frequency, f_A is the allele frequency of the HLA-A molecule in the haplotype and f_B is the allele frequency of the HLA-B molecule in this haplotype. D is easy to calculate, but has the disadvantage of depending on the frequency of the alleles. In order to overcome this drawback, the normalized measure, D', was calculated as:

$D' = \frac{D}{D_{\max}}$, where $D_{\max}$ is the lesser of f_{AB} or $(1-f_A)(1-f_B)$ when D<0 and is the lesser of $f_A(1-f_B)$ or $f_B(1-f_A)$ when *D>0*. The advantage of this measure of disequilibrium is that it has a range of -1 to 1, regardless of the allelic frequencies in the sample. $|D'|=1$ indicates complete LD and $|D'|=0$ corresponds to total absence of LD.

Linkage disequilibrium statistics were calculated for each haplotype using by M.Maiers to identify the haplotypes that have a significantly positive D'.

6.3.2 HLA ligand prediction

To have an as large as possible population coverage, we used NetMHCpan (77) to predict peptide-HLA binding affinity. NetMHCpan assigns to each peptide-HLA pair a predicted IC50 value, indicative of the predicted binding affinity. To assess whether a peptide binds to an HLA molecule depends on the choice of binding threshold, and recently it has been discussed what the optimal threshold is (274). If one assumes that all HLA molecules use a fixed threshold, the default threshold of 500 nM (275)(276) can be used, otherwise a 5000 nM threshold can be used to allow for the comparison of more weakly binding peptides. However, using a fixed threshold of 500 nM IC50 to define predicted binders may result in differences in predicted repertoire sizes between HLA molecules, which in turn may introduce a bias into our analysis. To avoid this, we defined the top-ranking 1% of the peptides as candidate binders for each HLA molecule,

thereby ensuring the same ligand repertoire size for each HLA molecule. To ensure consistency of our results with respect to parameters, we repeated every analysis with the fixed threshold of 500 nM. Results presented are derived using the scaled threshold and similar for other thresholds, unless mentioned otherwise.

6.3.3 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 S6.1) as the source of potential HLA ligands. We used the HLA binding predictor to screen all possible unique virus-derived 9-mer peptides.

6.3.4 Peptide repertoire overlap

We define the peptide repertoire overlap between two HLA alleles in the same HLA-A-B haplotype, F_p, as the fraction of overlap ligands between these two HLA class I molecules among all of their ligands: $F_p = \frac{A \bigcap B}{A \bigcup B}$, where A and B are the ligand numbers of HLA-A and -B molecule, respectively. $F_p = 1$ implies that the HLA-A and HLA-B molecules belonging to the same haplotype have the same epitope repertories while $F_p = 0$ indicates completely different epitope repertoires for two HLA molecules.

6.4 Results and Discussion

6.4.1 Determining HLA-A-B Haplotypes

The ``true'' HLA haplotypes can be determined either by molecular haplotyping or family-based segregation studies (277)(278)(279). However, both approaches are expensive and laborous, and therefore, statistical methods are used to infer haplotypes from datasets covering large population of individuals with known HLA genotypes (280). Several methods are proposed to infer HLA haplotypes from genotype data, and in recent studies the performance of two most commonly used approaches, Expectation-Maximization algorithm based (implemented in Arlequin V3.0) and Bayesian algorithm based (implemented in PHASE V2.1.1) are compared (281)(282). Unfortunately, neither of the methods could infer all of the known haplotypes: incorrect haplotypes are estimated in more than 30 % of the cases (281)(282). However, once the sample size increases, the power of these statistical methods are expected to increase tremendously.

The national marrow donor program (NMDP) provides, to our knowledge, the largest repository of HLA-typed donors, where use of statistical methods becomes more reliable (283): for A-B haplotype, the total chromosome counts (2N) for major four ethnic groups exceeds 2000 in this data set. In NMDP webpage, bioinformatics.nmdp.org, the high-resolution allele and haplotype frequencies (estimated by EM method) are available (271). Focusing on HLA-A-B haplotypes, the most common haplotypes found in

US population (seperated into four main ethnic groups) are summarized in Table 6.1 (adopted from bioinformatics.nmdp.org, Dec 2007 version). Alternatively, www.allelefrequencies.net provides allele frequencies established in smaller, but probably better defined studies (284).

Table 6.1 Occurences of the three most common HLA-A-B frequency ranked haplotypes in four major ethnic groups in US (adopted from bioinformatics.nmdp.org). EUR: Caucasian, AFA: African, API: Asian, and HIS: Hispanic. F stands for population frequency in precentages.

		EUR		AFA		API		HIS	
HLA-A	HLA-B	F (%)	Rank	F (%)	Rank	F (%)	Rank	F (%)	Rank
0101	0801	9.55	1	1.50	6	0.41	46	2.21	2
0201	4402	5.70	3	1.33	9	0.17	130	1.94	4
0201	4501	0.05	200	1.66	3	-	-	0.23	105
0201	5101	2.00	9	0.61	26	0.91	25	2.20	3
0207	4601	-	-	-	-	3.34	2	-	-
0301	0702	6.01	2	1.73	2	0.26	82	1.92	5
2902	4403	2.38	7	1.08	15	0.03	433	2.54	1
3001	4201	-	-	2.96	1	-	-	0.40	50
3303	4403	-	-	0.09	261	2.94	3	0.16	156
3303	5801	0.08	162	0.28	88	4.53	1	0.10	230

For European American sampling, 660 possible haplotypes are reported in NMDP database, however, many of those are bound to be false positives. To decrease the amount of false positive haplotypes in our analysis, we apply a rather strict criteria for considering a predicted haplotype a ``true'' haplotype: we demand to have a positive LD value significantly different than zero ($p<0.01$, see materials and methods). This criteria results in 60 haplotypes for the european Americans. These 60 haplotypes are estimated to cover 58% of the population (see Table 6.2), i.e., current statistical methods and data sets (even the large repositories like NMDP) are rather limited in providing HLA haplotype diversity within a population. For other ethnical groups, the number of reliable haplotypes drops to 30-40 per ethnic group, eventhough the number of possible haplotypes remain the same or higher (ses Table 6.2 and bioinformatics.nmdp.org). The population coverage in non-european groups are lower, down to 30-40% of their respective populations, possible due to the lower number of individuals with known HLA-typing. All together the number of unique haplotypes that could be detected in either of the ethnical groups adds upto 120 (see Table S6.1).

6.4.2 Peptide repertoire of an haplotype

Having identified the HLA-A-B haplotypes for the US population, we next estimated the overlaps between peptide repertoires of HLA-A and –B molecules that belong to the same haplotype. To get a non

Table 6.2 The numbers of different haplotypes with a significantly positive LD in four major US ethnic groups. Population coverage in percentages are given within parenthesis.

Ethnicity	Haplotypes (%)	Non-overlapping (%)	Overlapping (%)
EUR	60(57.7)	19(28.1)	41(29.6)
API	34(35.4)	10(15.4)	24(20.0)
HIS	43(33.5)	9(9.7)	34(23.8)
AFA	43(33.7)	15(12.9)	28(20.8)

biased estimate of how distinct or similar the peptide repertoire of two alleles are, we used an *in silico* approach and we predicted the peptide repertoire of all HLA-A and HLA-B alleles part of the identified haplotypes using the genomes of common viruses (see Table S6.2) and art-of-the-state HLA-peptide binding predictors (285) (see materials and methods). These predictors assign to each peptide-HLA pair an IC50 value, which can be used as the predicted binding affinity. Using a widely accepted IC50 value of 500 nM as a threshold to distinguish binders from non-binders generates variable predicted repertoire sizes for different HLA molecules (data not shown). Besides using this fixed IC50 threshold of 500 nM, we repeated our analysis by defining the peptide repertoire as being the top 1% epitopes with the highest HLA binding affinities for each HLA allele, thereby minimizing the potential bias introduced by different repertoire sizes. Unless mentioned otherwise, our results remained the same with both definitions of predicted peptide repertoires (more details are given in Methods section).

The unique haplotypes listed in Table S6.1 have on average a low peptide overlap (1.1%, with a standard deviation of 1.8%, and median being 0.44%). We and Frahm et.al. 2007 have previously reported a high cross loci peptide promiscuity, ranging around 23-44% (265) (263). However, in those studies the peptide promiscuity was estimated at the population level, i.e., these numbers reflect the fraction of the peptide repertoire of an HLA-A molecule that is predicted to be presented by at least one HLA-B molecule in the population (and vice versa). Within an individual having at most 2 different HLA-A and HLA-B molecules, however, extensive peptide binding promiscuity across loci is not likely to be realized, as the peptide repertoires of alleles belonging to the same haplotype hardly overlap (the distribution of the overlaps in all unique haplotypes is given in Figure 6.1). The degree of overlap changes somewhat between different haplotypes. First of all, out of 120 haplotypes 36 (30%) have non-overlapping peptide repertoires, at least for the common viruses that we tested (see Table S6.1). Only for five haplotypes the repertoire overlap is higher than 5%, HLA-A0101-B1517 being with a 11.8% highest, see Figure 6.1. HLA-A0101-B1517 is a rare haplotype: only among the Asian Americans it is estimated as an HLA-A-B haplotype with population frequency of 0.4%. To compare these results with ``random'' haplotypes, we shuffled HLA-A and HLA-B molecules in 120 ``true'' haplotypes and calculated the peptide repertoire overlap in the shuffled (artificial) haplotypes. Though the HLA molecules in random haplotypes can have overlapping peptide repertoires upto 28% (results not shown), the distribution of the overlaps is not significantly different than the distribution given in Figure 6.1 (Kolmogorov-Smirnov test, p=0.4). This result suggests that having almost distinct binding preferences is a generic property of HLA-A and –B in pairs in general, and not a unique property of HLA molecules having strong linkage disequlibrium.

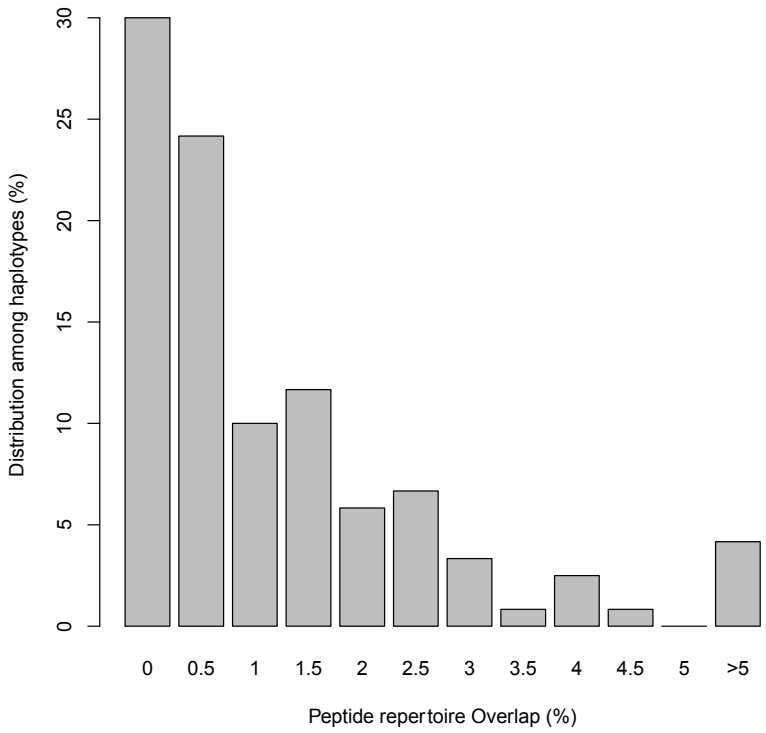

Figure 6.1 The distribution of the presented peptide overlap (given in percentages, see Methods) by HLA-A and HLA-B molecules belonging to the same haplotype (the analysis is done for the 120 unique haplotypes we could identify in the US population).

6.4.3 Does the presented peptide overlap effect HLA haplotype frequencies?

Next, we focused on the most common HLA-A-B haplotypes, since if non-overlapping peptide repertoires give a selective advantage, then the most common HLA-A-B haplotypes should clearly be the ones that have distinct binding specificities. Among european Americans, three most common haplotypes (A0101-B0801, A0301-B0701, and A0201-B4402, see Table 6.1) have all non-overlapping peptide repertoires (with both approaches of defining predicted epitopes, see Methods). This is significantly different than expected (p=0.027 in Chi-square test). In other words, it would be rather unexpected to find the three most common haplotypes containing HLA molecules with distinct binding motifs given that only 30% of the haplotypes have this property. Extending this analysis to other ethnic groups, we find that the most common haplotypes in US population (10 unique if one takes first three most common haplotypes in four ethnic groups, see Table 6.1) show the same trend to be enriched in containing HLA molecules with distinct peptide repertoires (p=0.061 in Chi-square test).

Finally, to test whether non-overlapping haplotypes are beneficial and therefore would occur in higher frequencies than non-overlapping ones, we compared the population frequency of overlapping and non-overlapping haplotypes. Every ethnic group contained different number of overlapping and non-overlapping haplotypes (see Table 6.2). In all ethnicities, the mean frequency of haplotypes with non-overlapping A-B pairs were higher than the mean frequency of haplotypes with overlapping alleles.

Especially among European Americans and Asians, the average frequency of non-overlapping HLA-A-B haplotypes were two fold higher than the overlapping ones (1.5% vs 0.7% among European background and 1.54% vs 0.83% among Asian background). The 19 non-overlapping haplotypes found in European ethnic group had almost the same total frequency as the other 41 overlapping haplotypes (see Table 6.2). However, in none of the ethnical groups the difference between the frequency of overlapping and non-overlapping haplotypes was significant. When we pooled the data, we found that non-overlapping haplotypes have in general higher frequencies (on average1.5%) than HLA-A-B haplotypes that have overlapping (average of 0.83%) peptide repertoires ($p=0.079$, Mann-Whitney test, where every haplotype is assigned the maximum frequency found in different ethnic groups). However, the frequency of an haplotype has only a weak negative correlation with the degree of peptide overlap among HLA-A and HLA-B molecules ($r=-.1$, $p=0.2$ spearman rank correlation).

6.5 Conclusion

Using high resolution data available in NMDP database on haplotype frequencies, we here tested whether or not having HLA molecules with distinct binding motifs have an influence on the frequency of HLA-A-B haplotypes. First of all, we found that in all the haplotypes we could identify in US population, the peptide repertoire overlap between HLA-A and HLA-B molecules is very low per haplotype. This suggests that the extensive peptide binding promiscuity among HLA class I molecules, found by us and the others, diminishes at the haplotype level. Second, the most common HLA-A-B haplotypes almost entirely contain HLA molecules with totally distinct peptide binding motifs. In general such HLA-A-B haplotypes composed of molecules with distinct binding motifs occur twice as frequent in US population as the other haplotypes. Taken together, these results suggest the complementarity of binding motifs in an HLA haplotype might be advantegous. Though we realize that the frequency of an HLA haplotype determined by complex interactions among several factors, like birth weight (286), that vary between haploytpes.

6.6 Acknowledgments

This study was financially supported by the University of Utrecht through a High Potential grant. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

6.7 Supplemental Materials

Table S 6.1 List of all reliable haplotypes (with an LD positive and significantly different than zero, $p<0.01$) different of four ethnic groups from NMDP database. O500 and Otop corresponds to the overlaps estimated using

two different approaches to define the peptide repertoire of an HLA molecule (see Methods). The last four columns give the population frequencies in European, African, Asian, and Hispanic Americans, respectively.

Haplotype	O_{500}	O_{top}	EUR	AFA	API	HIS
A0101_B0801	0	0	0.095	0.015	0.0041	0.022
A0101_B1517	0.049	0.12	-	-	0.0039	-
A0101_B3701	0	0	0.0082	0.0021	0.0087	-
A0101_B5701	0.034	0.053	0.021	0.0026	0.014	0.0074
A0102_B4901	0	0.00088	-	0.0029	-	-
A0201_B1501	0.077	0.026	0.037	0.0047	-	-
A0201_B1511	0.027	0.017	-	-	0.0038	-
A0201_B1515	0.031	0.025	-	-	-	0.005
A0201_B2705	0.003	0.0044	0.015	-	-	-
A0201_B3512	0.0004	0.014	-	-	-	0.011
A0201_B4001	0.0062	0.0053	0.028	0.005	-	-
A0201_B4402	0	0	0.057	0.013	-	0.019
A0201_B4501	0.0021	0	-	0.017	-	-
A0201_B5101	0.0012	0.012	0.011	0.0061	-	0.022
A0201_B5401	0.0022	0.011	-	-	0.0079	-
A0202_B1516	0.1	0.0071	-	-	-	0.0024
A0202_B4101	0.0028	0.00088	0.00057	-	-	-
A0202_B5301	0.0035	0.00088	-	0.01	-	-
A0203_B3802	0	0.028	-	-	0.012	-
A0205_B4901	0.0046	0.022	0.0014	-	-	0.0029
A0205_B5001	0.021	0.017	0.0036	-	0.0028	-
A0205_B5801	0.021	0.0035	0.001	0.0039	-	0.0025
A0206_B2705	0.0028	0.0062	0.00068	-	-	-
A0206_B3905	0.024	0.061	-	-	-	0.0064
A0206_B4002	0.053	0.038	-	-	-	0.0085
A0206_B4801	0.022	0.086	0.00044	-	-	0.0052
A0206_B5101	0.0026	0.019	-	-	0.0094	-
A0206_B5901	0.015	0.034	-	-	0.0035	-
A0207_B4601	0	0.022	-	-	0.033	-
A0211_B4006	0.023	0.0071	-	-	0.0047	-
A0301_B0702	0	0	0.06	0.017	-	0.019
A0301_B1402	0	0	0.0072	-	-	-
A0301_B3501	0.014	0.0071	0.011	0.012	0.0049	0.011
A0301_B3503	0	0.0026	0.0031	-	-	-
A0301_B4701	0.0025	0.00088	0.0021	-	-	-
A1101_B1301	0	0	-	-	0.012	-
A1101_B1502	0.034	0.0088	-	-	0.011	-
A1101_B2705	0.003	0.0035	-	-	-	0.0058
A1101_B3501	0.021	0.0044	0.017	0.0036	-	0.0071
A1101_B3901	0	0	-	-	0.0065	-
A1101_B4001	0	0	-	-	0.024	-
A1101_B5101	0	0	0.0043	-	-	-
A1101_B5201	0	0	0.0028	-	-	-
A1101_B5501	0	0	0.0058	0.0021	-	-
A1102_B2704	0.0015	0.0044	-	-	0.0034	-
A2301_B1503	0.034	0.035	-	0.015	-	0.003
A2301_B4403	0.0074	0.0097	0.0086	-	-	0.0097
A2301_B4501	0.0056	0.0035	-	0.0097	-	0.0033
A2301_B4901	0.0068	0.011	0.0037	-	-	0.0038
A2402_B1501	0.0096	0.022	0.0095	-	-	-
A2402_B1507	0.011	0.018	0.00063	-	-	-
A2402_B3502	0	0.0035	0.0049	-	-	0.0062

A2402_B3508	0.00081	0.0018	0.0012	-	-	-
A2402_B3906	0.0029	0.0053	0.0027	0.0015	-	0.012
A2402_B4002	0.011	0.012	-	-	0.012	0.012
A2402_B4005	0.0097	0.012	-	-	-	0.0034
A2402_B4801	0.0038	0.015	-	-	0.0092	-
A2402_B5401	0	0	-	-	0.012	-
A2402_B5501	0	0.00088	0.0032	-	-	-
A2403_B1801	0.011	0.0035	-	-	0.0011	-
A2407_B3505	0.0013	0.0044	-	-	0.0064	-
A2417_B1502	0.0049	0.023	-	-	0.0031	-
A2501_B1801	0.0072	0.0071	0.011	0.0029	-	0.0048
A2501_B3901	0.0035	0.0044	0.00089	-	-	-
A2601_B0801	0.012	0.012	-	0.0031	0.0052	-
A2601_B1401	0	0.0053	0.00094	-	-	-
A2601_B2705	0	0	0.0021	-	-	-
A2601_B3801	0	0.0035	0.011	-	-	0.0085
A2601_B4501	0	0	0.00082	-	-	-
A2601_B5501	0	0.012	0.0016	-	-	-
A2901_B0705	0.0013	0.00088	0.001	-	0.012	-
A2902_B4403	0.0044	0.0018	0.024	0.011	-	0.025
A2902_B4404	0.0047	0.0018	0.00057	-	-	-
A2902_B4501	0.00067	0.00088	0.0021	-	-	0.0035
A2902_B4901	0	0.0018	-	0.0036	-	-
A3001_B1302	0.00035	0.0035	0.011	-	0.019	0.0053
A3001_B4201	0.0024	0.021	-	0.021	-	0.004
A3001_B4202	0.014	0.023	-	0.0058	-	-
A3002_B1402	0	0	-	0.0044	-	-
A3002_B1801	0.0065	0.0053	0.0058	-	-	0.01
A3002_B5703	0.0041	0.012	-	0.0053	-	-
A3004_B1401	0	0.0026	-	-	0.0028	-
A3101_B2705	0.0068	0.0044	0.0018	-	-	-
A3101_B4001	0	0	0.0088	0.0027	-	-
A3101_B5101	0	0	0.0025	-	0.0072	-
A3101_B5102	0.00044	0	-	-	-	0.0028
A3201_B1401	0	0.021	0.0038	-	-	-
A3201_B4002	0.034	0.011	0.0025	-	-	-
A3201_B4402	0	0.0018	0.0041	-	-	-
A3201_B8101	0.0005	0.024	-	0.0022	-	-
A3301_B1402	0	0	0.0093	0.003	-	0.015
A3301_B7801	0.001	0	-	0.0049	-	-
A3303_B1516	0.00035	0	-	0.0055	-	-
A3303_B4403	0	0	-	-	0.029	-
A3303_B5301	0	0	-	0.013	-	-
A3303_B5801	0	0	0.00076	-	0.045	-
A3401_B1521	0.061	0.071	-	-	0.0011	-
A3401_B1535	0.031	0.045	-	-	0.0036	-
A3402_B3501	0.033	0.016	-	0.0056	-	-
A3402_B4403	0	0	-	0.0078	-	-
A3601_B5301	0.028	0.035	-	0.015	-	0.002
A6601_B4102	0	0	0.0019	-	-	0.0025
A6601_B5802	0	0.013	-	0.0043	-	-
A6602_B5801	0.017	0.011	-	0.0037	-	-
A6801_B3503	0	0	0.0023	-	-	-
A6801_B4001	0	0	0.0036	-	-	-
A6801_B4002	0.0003	0	-	-	-	0.0078

A6801_B44020	0	0.0056	-	-	-
A6801_B48010	0	-	-	-	0.0058
A6801_B51010	0	0.0028	-	-	-
A6801_B52010	0	-	-	0.0041	-
A6801_B58020	0.00088	-	0.012	-	-
A6802_B14020	0.013	0.0057	-	-	0.0056
A6802_B15100.007	0.013	-	0.012	-	0.0025
A6802_B53010.02	0.014	0.002	0.014	-	0.0052
A6803_B39050.00063	0.00088	-	-	-	0.0066
A6901_B55010.012	0.028	0.00064	-	-	0.0025
A7401_B15030.0019	0	-	0.016	-	0.0025
A7401_B57030.0022	0	-	0.0051	-	-
A8001_B18010.0063	0.0062	-	0.0021	-	-

Table S6.2. List of human viral proteomes used in the extension of the viral set.

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 /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

Chapter 7. Conclusions and outlook

7.1 General conclusions

During the last decade there have been several studies suggesting a distinct role for HLA-B restricted CTL responses in infectious diseases (see the introduction chapter for an extensive list of references). We, and others in the field, hypothesized that a likely mechanism to explain why HLA-B alleles shape the outcome of infectious disease might lie in their tendency to evoke immunodominant T cell responses. The focus of this thesis work is an extensive comparison of HLA-A and HLA-B antigen presentation in order to discover potential mechanisms which may explain why HLA-B restricted T cell responses tend to be immunodominant.

In chapter 2 we analyzed the epitope diversity and binding affinity of the molecules from HLA-A and HLA-B loci. Unexpectedly, the amount of epitopes presented by HLA-A molecules is about 2-fold higher than that of HLA-B molecules, and the binding affinities of HLA-A epitopes are significantly higher than those of HLA-B epitopes. Alternatively, targeting more constrained regions of a pathogen genome might be favourable for the outcome of an infection (e.g., a slow rate of disease progression or low viral load during infection). Therefore, we compared HIV-1 CTL epitopes restricted by HLA-A and HLA-B molecules in Chapter 3, and found that conserved HIV proteins, like Gag-p24, contain more HLA-B epitopes than HLA-A epitopes. In addition, the residues targeted by HLA-B alleles in the HIV-1 proteome were found to be significantly more conserved than the ones targeted by HLA-A molecules. In chapter 4 a similar analysis was done on HCV CTL epitopes by analyzing the variability of different HCV proteins and the distribution of CTL epitopes restricted by protective and detrimental HLA alleles. Similar to our results on HIV-1, we found that HLA molecules associated with HCV clearance preferentially present epitopes from the conserved proteins, like NS5B and Core.

Unexpectedly, it was recently shown that HLA-peptide binding is highly promiscuous (263), i.e., a large fraction of epitopes binds at least two HLA molecules. Our results suggest that less promiscuous peptide presentation might contribute to viral control during HIV-1 infection (chapter 5). In order to test whether the dominant role of HLA-B restricted T cell responses is due to promiscuous binding of HLA-B restricted epitopes, we compared their binding promiscuity to that of HLA-A restricted epitopes in chapter 5. To this end, we studied the HLA ligand promiscuity based on both experimental epitopes from IEDB and *in silico* predicted epitopes. In line with the high HLA ligand promiscuity observed by Frahm et al (263) we found that above 60% of all HLA class I ligands are promiscuous, but we failed to find any consistent difference in the peptide binding promiscuity of HLA-A and HLA–B molecules During this analysis, we found that there is also a considerable promiscuity between the two loci: significant fraction of HLA-A ligands can be presented by at least one HLA-B molecule in the population. To test whether this across loci promiscuity has an effect on the frequencies of HLA class I haplotypes, we estimated the total ligand repertoires for all HLA class I haplotypes (HLA-A-B) by *in silico* prediction and found that

the most common HLA-A-B haplotypes are enriched in HLA-A and HLA–B pairs with distinct peptide binding motifs (chapter 6).

The chapters of this thesis present only the work that had been published or is under consideration for publication. But our efforts were not limited to the results presented here. Much more analysis was done in our group to discover possible characteristics of HLA-B responses explaining their generation of immunodominant responses. We shortly describe these additional analyses below.

We can predict the specificity of the proteasome and TAP using bioinformatics tools developed by us and by others (287). Thus, given a pathogen genome, we can predict the peptide pool in the ER that originates from this pathogen. We have previously shown that proteasomal cleavage and TAP translocation efficiency have been co-evolving with HLA binding motifs (288, 289). In other words, human MHC molecules are more likely to bind the peptides generated in the MHC class I pathway than random peptides. We looked in detail if this preference for processed peptides change with MHC class I loci (using the allele set of Chapter 2). Using TAP binding predictor suggested by Peters et al (290), we found that almost all predicted HLA-A binders would be transported by TAP (on average 99%), while slightly fewer, on average 95%, of all predicted HLA-B binders have a good TAP affinity. However, this difference is not significant (Mann Whitney U test per allele, $p=0.19$, results not shown). Conversely, the proteasome specificity seems to have co-evolved more with HLA-B molecules. On average only half of predicted HLA-A binders are predicted to be generated by the proteasome, and for predicted HLA-B ligands this average goes up to 66% (using SMM matrices for predicting proteasome specificity as suggested by (73). This difference was approaching statistical significant ($p=0.08$ Mann Whitney U test). Given the fact that the proteasome is not the only protease in acting in generation of the MHC presented peptides (291), the physiological role of this small difference between HLA-A and –B molecules remains doubtful. Taken together, these analysis do not provide strong evidence for an enhanced supply of ligands for HLA-B molecules compared to their HLA-A counterparts.

Initially, we hypothesized that an enhanced presentation of non-self compared to self might explain why HLA-B alleles are more correlated with protection and immunodominance in anti-viral infections. However, in chapter 2 we report that HLA-A, but not HLA-B molecules, bind significantly less self (human) 9-mers than non-self (viral or bacterial) 9-mers. ($p<0.001$, Mann-Whitney test, Figure.2.2). The underlying reason for this difference was discovered by colleagues in our group: Calis et al showed that the G+C content of pathogenic intracellular bacteria and viruses is significantly lower than that of non-pathogenic bacteria ($p<0.01$), or that of human coding sequences ($p<0.01$) (292). This implies that a low G+C content seems to be a universal signature of organisms with a pathogenic lifestyle (293)(294). A significant majority of HLA-A alleles prefers to present the peptides derived from organisms with a low G+C content (292). Thus HLA-A, but not HLA-B molecules, seems to have evolved to make use of this pathogen-specific signature to shape the $CD8^+$ T cells responses.

Another possibility to explain the observed immunodominant CTL responses restricted to HLA-B molecules is their level of expression. If HLA-B genes were expressed more than other HLA class I genes, either constitutively or during infection, the antigen presentation in the context of HLA-B molecules would be enhanced, which in turn could result in immunodominant T cell responses. Early studies measured the expression of HLA genes using locus specific PCRs. Johnson, 2003 demonstrated that HLA-B locus responds most strongly to the stimulation with inflammatory cytokines like TNF-alpha, IFN-beta and IFN-gamma (295). However, constitutive expression of different HLA loci depends heavily on the cell types (296), which makes it difficult to have a clear conclusion. Unfortunately, more recent high-throughput techniques like microarray platforms are not designed to have probe sets that are specific for different HLA locus. Therefore, any analysis comparing the expression level of HLA-B molecules with other locus using standard microarray platforms can only remain indicative, and hardly become conclusive. We performed such a preliminary analysis and found some evidence for elevated levels of HLA-B expression. In short, we analyzed the transcriptomics data provided by Su et.al, 2004 (297), which contains 79 tissue samples from healthy donors (HLA typing unknown) using Affymetrix platform. This data consists of two probe sets annotated as HLA-A11, and three probe sets annotated as HLA-B71, -B39, and B2711. We filtered these probe sets so that only the probes that i) are locus specific, i.e. having only perfect matches either in HLA-A or HLA-B alleles, and ii) have a good population coverage, i.e. having a perfect hit with at least half of the alleles within one locus, are included in the further analysis. This resulted in 4-8 probes per the above mentioned probe sets. Using this limited set of "specific" probes, we found that the expression of HLA-A and HLA-B molecules are very similar when we pool all tissues (n=79) together. If we look at immune tissues only (n=22), HLA-B expression is significantly higher than HLA-A expression ($p=0.003$, Mann-Whitney test). However, Greene et.al, 2011 recently sequenced HLA class I region from two subjects and found that transcription of MHC class I genes was consistent across the leukocyte subsets studied with only small differences detected (298). More specific measurements of HLA class I gene expression will be needed to draw a clear conclusion on the effect of differential expression of HLA molecules on the CTL immunodominance.

In conclusion, despite these efforts we identified hardly any characteristics of antigen presentation by HLA-B molecules that were significantly “better” than HLA-A molecules, and therefore can explain their generation of immunodominant CTL responses. On the contrary, HLA-A molecules seems to outperform HLA-B molecules in many aspects of the antigen presentation: HLA-A molecules present significantly more epitopes with higher affinity compared to their HLA-B counterparts, and the presentation of nonself is enhanced in the context of HLA-A molecules. Some of these results should not come to us as a big surprise: HLA binding affinity does not strictly correlate with immunodominance (118), and certain sets of subdominant epitopes have even higher binding affinity than dominant epitopes (110). Furthermore, Crotzer and colleagues reported an immunodominant EBV epitope, which is least abundant on the cell surface, indicating that immunodominance does not need to correlate with epitope abundance (299). Thus, the targeting of more conserved regions in the pathogen proteomes, as we have shown to be the case for

HIV and HCV, is the one property of HLA-B antigen presentation that currently stands out as a crucial factor for controlling the viral replication during an infectious disease.

7.2 Outlook

Every research year spent during my thesis work brought me closer to believe that factors other than antigen presentation should be responsible for the protective and immunodominant nature of HLA-B-restricted T cell responses. More data is accumulating to support this view. For example, after peptide stimulation HLA-B57-restricted HIV-specific T cells have the highest functional capacity (i.e., IFN-gamma production) due to their high TCR affinity for the HLA-B57-peptide complex (300). Several other studies suggested an important role for high avidity CTL responses. However, all these studies have been based on single epitope/allele combinations and/or used T cell cultures after in *vitro* expansion, and therefore it had been difficult to make a general conclusion. Very recently, Berger et al. (301) used CTL directly *ex vivo* in a cohort of more than 100 HIV-1 infected individuals, and showed that HLA-B restricted Gag specific high avidity CTL populations are linked to good control of viral prelication. Berger et al (301) cite our results presented in Chapter 3 to explain their findings: "*The HLA-B-restricted HIV epitope repertoire is thought to be enriched for epitopes with increased sequence conservation, which might limit viral escape from CTL-mediated immune pressure and select for a narrower TCR repertoire of high avidity* (302)*. Such selective presentation of more conserved epitopes on HLA-B alleles could result from the need for specific anchor residues for B alleles that require amino acids that pose stringent functional and structural constraints on the viral protein and thus complicate viral escape*." This is an interesting hypothesis and can indeed be tested more in depth using a bioinformatics approach.

HLA class I molecules have more functions than just their essential role in antigen presentation pathway, and also function as the ligands of immunoglobulin-like receptors (KIR) regulating the NK cell function. NK cells are important components of the immune response against viral infections, and are key mediators of innate immunity by killing aberrant target cells, or by secreting cytokines, without prior exposure to the target. Various allelic combinations of HLA-B alleles and KIR are associated with infectious diseases: KIR3DS1 provides significant protection during clearance of HCV infection in the presence of HLA-B Bw4 (303). The KIR3DS1/HLA-B Bw4-80I was shown to strongly inhibit HIV replication, and is significantly associated with lower viral load and slower progression to AIDS (304, 305). Presence of both the inhibitory receptor KIR3DL1 and the B*57 supertype carrying the Bw4-80I epitope also has a significant protective effect on progression to AIDS (306). Finally, Carrington and colleagues found that the presence of activating receptor KIR3DS1 in combination with the absence of specific HLA allele groups, like C2 and/or Bw4 which are the ligands for the inhibitory KIR2DL1 and KIR3DL1 respectively, was associated with increased risk of developing cervical neoplasia (307). This suggested that the vigorous immune response against chronic HPV infection is mediated partly by KIR-HLA interactions, and if not tightly regulated, may contribute to HPV pathogenesis in the cervix. Collectively these data

suggest that the seemingly important role of HLA-B locus in infectious diseases need not involve acquired immune responses (i.e. CD8+ T cell responses) but can also be due to innate immunity (i.e. NK cells).

Bibliography

1. Cooke, G. S., and A. V. Hill. 2001. Genetics of susceptibility to human infectious disease. *Nat. Rev. Genet.* 2: 967-977.

2. Mungall, A. J., S. A. Palmer, S. K. Sims, C. A. Edwards, J. L. Ashurst, L. Wilming, M. C. Jones, R. Horton, S. E. Hunt, C. E. Scott, J. G. R. Gilbert, M. E. Clamp, G. Bethel, S. Milne, R. Ainscough, J. P. Almeida, K. D. Ambrose, T. D. Andrews, R. I. S. Ashwell, A. K. Babbage, C. L. Bagguley, J. Bailey, R. Banerjee, D. J. Barker, K. F. Barlow, K. Bates, D. M. Beare, H. Beasley, O. Beasley, C. P. Bird, S. Blakey, S. Bray-Allen, J. Brook, A. J. Brown, J. Y. Brown, D. C. Burford, W. Burrill, J. Burton, C. Carder, N. P. Carter, J. C. Chapman, S. Y. Clark, G. Clark, C. M. Clee, S. Clegg, V. Cobley, R. E. Collier, J. E. Collins, L. K. Colman, N. R. Corby, G. J. Coville, K. M. Culley, P. Dhami, J. Davies, M. Dunn, M. E. Earthrowl, A. E. Ellington, K. A. Evans, L. Faulkner, M. D. Francis, A. Frankish, J. Frankland, L. French, P. Garner, J. Garnett, M. J. R. Ghori, L. M. Gilby, C. J. Gillson, R. J. Glithero, D. V. Grafham, M. Grant, S. Gribble, C. Griffiths, M. Griffiths, R. Hall, K. S. Halls, S. Hammond, J. L. Harley, E. A. Hart, P. D. Heath, R. Heathcott, S. J. Holmes, P. J. Howden, K. L. Howe, G. R. Howell, E. Huckle, S. J. Humphray, M. D. Humphries, A. R. Hunt, C. M. Johnson, A. A. Joy, M. Kay, S. J. Keenan, A. M. Kimberley, A. King, G. K. Laird, C. Langford, S. Lawlor, D. A. Leongamornlert, M. Leversha, C. R. Lloyd, D. M. Lloyd, J. E. Loveland, J. Lovell, S. Martin, M. Mashreghi-Mohammadi, G. L. Maslen, L. Matthews, O. T. McCann, S. J. McLaren, K. McLay, A. McMurray, M. J. F. Moore, J. C. Mullikin, D. Niblett, T. Nickerson, K. L. Novik, K. Oliver, E. K. Overton-Larty, A. Parker, R. Patel, A. V. Pearce, A. I. Peck, B. Phillimore, S. Phillips, R. W. Plumb, K. M. Porter, Y. Ramsey, S. A. Ranby, C. M. Rice, M. T. Ross, S. M. Searle, H. K. Sehra, E. Sheridan, C. D. Skuce, S. Smith, M. Smith, L. Spraggon, S. L. Squares, C. A. Steward, N. Sycamore, G. Tamlyn-Hall, J. Tester, A. J. Theaker, D. W. Thomas, A. Thorpe, A. Tracey, A. Tromans, B. Tubby, M. Wall, J. M. Wallis, A. P. West, S. S. White, S. L. Whitehead, H. Whittaker, A. Wild, D. J. Willey, T. E. Wilmer, J. M. Wood, P. W. Wray, J. C. Wyatt, L. Young, R. M. Younger, D. R. Bentley, A. Coulson, R. Durbin, T. Hubbard, J. E. Sulston, I. Dunham, J. Rogers, and S. Beck. 2003. The DNA sequence and analysis of human chromosome 6. *Nature* 425: 805-811.

3. Purcell, A. W., and T. Elliott. 2008. Molecular machinations of the MHC-I peptide loading complex. *Curr. Opin. Immunol.* 20: 75-81.

4. Peaper, D. R., and P. Cresswell. 2008. Regulation of MHC class I assembly and peptide binding. *Annu. Rev. Cell Dev. Biol.* 24: 343-368.

5. Jensen, P. E. 2007. Recent advances in antigen processing and presentation. *Nat. Immunol.* 8: 1041-1048.

6. Groothuis, T. A. M., A. C. Griekspoor, J. J. Neijssen, C. A. Herberts, and J. J. Neefjes. 2005. MHC class I alleles and their exploration of the antigen-processing machinery. *Immunol. Rev.* 207: 60-76.

7. Burroughs, N. J., R. J. de Boer, and C. Keşmir. 2004. Discriminating self from nonself with short peptides from large proteomes. *Immunogenetics* 56: 311-320.

8. Sijts, E. J. A. M., and P. M. Kloetzel. 2011. The role of the proteasome in the generation of MHC class I ligands and immune responses. *Cell. Mol. Life Sci.* 68: 1491-1502.

9. Horst, D., M. C. Verweij, A. J. Davison, M. E. Ressing, and E. J. H. J. Wiertz. 2011. Viral evasion of T cell immunity: ancient mechanisms offering new applications. *Curr. Opin. Immunol.* 23: 96-103.

10. Yokomaku, Y., H. Miura, H. Tomiyama, A. Kawana-Tachikawa, M. Takiguchi, A. Kojima, Y. Nagai, A. Iwamoto, Z. Matsuda, and K. Ariyoshi. 2004. 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.* 78: 1324-1332.

11. Kwun, H. J., S. R. da Silva, I. M. Shah, N. Blake, P. S. Moore, and Y. Chang. 2007. 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.* 81: 8225-8235.

12. Yewdell, J. W., E. Reits, and J. Neefjes. 2003. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat Rev Immunol* 3: 952-961.

13. Jeffery, K. J., and C. R. Bangham. 2000. Do infectious diseases drive MHC diversity? *Microbes Infect.* 2: 1335-1341.

14. Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.

15. Lienert, K., and P. Parham. 1996. Evolution of MHC class I genes in higher primates. *Immunol. Cell Biol.* 74: 349-356.
16. Prugnolle, F., A. Manica, M. Charpentier, J. F. Guégan, V. Guernier, and F. Balloux. 2005. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15: 1022-1027.
17. Parham, P., and T. Ohta. 1996. Population biology of antigen presentation by MHC class I molecules. *Science* 272: 67-74.
18. Doherty, P. C., and R. M. Zinkernagel. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256: 50-52.
19. 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.
20. Tang, J., C. Costello, I. P. Keet, C. Rivers, S. Leblanc, E. Karita, S. Allen, and R. A. Kaslow. 1999. HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection. *AIDS Res. Hum. Retroviruses* 15: 317-324.
21. De Boer, R. J., J. A. M. Borghans, M. van Boven, C. Keşmir, and F. J. Weissing. 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* 55: 725-731.
22. Leslie, A., P. C. Matthews, J. Listgarten, J. M. Carlson, C. Kadie, T. Ndung'u, C. Brander, H. Coovadia, B. D. Walker, D. Heckerman, and P. J. R. Goulder. 2010. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *J. Virol.* 84: 9879-9888.
23. Trachtenberg, E., B. Korber, C. Sollars, T. B. Kepler, P. T. Hraber, E. Hayes, R. Funkhouser, M. Fugate, J. Theiler, Y. S. Hsu, K. Kunstman, S. Wu, J. Phair, H. Erlich, and S. Wolinsky. 2003. Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* 9: 928-935.
24. Moore, C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. A. Mallal. 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296: 1439-1443.
25. Frahm, N., P. Kiepiela, S. Adams, C. H. Linde, H. S. Hewitt, K. Sango, M. E. Feeney, M. M. Addo, M. Lichterfeld, M. P. Lahaie, E. Pae, A. G. Wurcel, T. Roach, M. A. St John, M. Altfeld, F. M. Marincola, C. Moore, S. Mallal, M. Carrington, D. Heckerman, T. M. Allen, J. I. Mullins, B. T. Korber, P. J. R. Goulder, B. D. Walker, and C. Brander. 2006. Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes. *Nat. Immunol.* 7: 173-178.
26. Schellens, I. M. M., M. Navis, H. W. M. van Deutekom, B. Boeser-Nunnink, B. Berkhout, N. Kootstra, F. Miedema, C. Keşmir, H. Schuitemaker, D. van Baarle, and J. A. M. Borghans. 2011. Loss of HIV-1-derived cytotoxic T lymphocyte epitopes restricted by protective HLA-B alleles during the HIV-1 epidemic. *AIDS (London, England)* 25: 1691-1700.
27. de Groot, N. G., N. Otting, R. Argüello, D. I. Watkins, G. G. Doxiadis, J. A. Madrigal, and R. E. Bontrop. 2000. Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. *Immunogenetics* 51: 398-409.
28. 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.
29. Watkins, D. I., S. N. McAdam, X. Liu, C. R. Strang, E. L. Milford, C. G. Levine, T. L. Garber, A. L. Dogon, C. I. Lord, and S. H. Ghim. 1992. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* 357: 329-333.
30. Borghans, J. A. M., J. B. Beltman, and R. J. De Boer. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55: 732-739.
31. Otting, N., C. M. C. Heijmans, R. C. Noort, N. G. de Groot, G. G. M. Doxiadis, J. J. van Rood, D. I. Watkins, and R. E. Bontrop. 2005. Unparalleled complexity of the MHC class I region in rhesus macaques. *Proc. Natl. Acad. Sci. U.S.A.* 102: 1626-1631.
32. Fukami-Kobayashi, K., T. Shiina, T. Anzai, K. Sano, M. Yamazaki, H. Inoko, and Y. Tateno. 2005. Genomic evolution of MHC class I region in primates. *Proc. Natl. Acad. Sci. U.S.A.* 102: 9230-9234.
33. Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger, C. Day, P. Klenerman, J. Mullins, B. Korber, H. M. Coovadia, B. D. Walker, and P. J. R. Goulder. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432: 769-775.

34. Belich, M. P., J. A. Madrigal, W. H. Hildebrand, J. Zemmour, R. C. Williams, R. Luz, M. L. Petzl-Erler, and P. Parham. 1992. Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature* 357: 326-329.
35. McAdam, S. N., J. E. Boyson, X. Liu, T. L. Garber, A. L. Hughes, R. E. Bontrop, and D. I. Watkins. 1994. A uniquely high level of recombination at the HLA-B locus. *Proc. Natl. Acad. Sci. U.S.A.* 91: 5893-5897.
36. Barber, L. D., L. Percival, K. L. Arnett, J. E. Gumperz, L. Chen, and P. Parham. 1997. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *J. Immunol.* 158: 1660-1669.
37. Carrington, M., and S. J. O'Brien. 2003. The influence of HLA genotype on AIDS. *Annu. Rev. Med.* 54: 535-551.
38. Martin, M. P., and M. Carrington. 2005. Immunogenetics of viral infections. *Curr. Opin. Immunol.* 17: 510-516.
39. Goulder, P. J. R., and D. I. Watkins. 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat. Rev. Immunol.* 8: 619-630.
40. Pereyra, F., X. Jia, P. J. McLaren, A. Telenti, P. I. W. de Bakker, B. D. Walker, S. Ripke, C. J. Brumme, S. L. Pulit, M. Carrington, C. M. Kadie, J. M. Carlson, D. Heckerman, R. R. Graham, R. M. Plenge, S. G. Deeks, L. Gianniny, G. Crawford, J. Sullivan, E. Gonzalez, L. Davies, A. Camargo, J. M. Moore, N. Beattie, S. Gupta, A. Crenshaw, N. P. Burtt, C. Guiducci, N. Gupta, X. Gao, Y. Qi, Y. Yuki, A. Piechocka-Trocha, E. Cutrell, R. Rosenberg, K. L. Moss, P. Lemay, J. O'Leary, T. Schaefer, P. Verma, I. Toth, B. Block, B. Baker, A. Rothchild, J. Lian, J. Proudfoot, D. M. L. Alvino, S. Vine, M. M. Addo, T. M. Allen, M. Altfeld, M. R. Henn, S. Le Gall, H. Streeck, D. W. Haas, D. R. Kuritzkes, G. K. Robbins, R. W. Shafer, R. M. Gulick, C. M. Shikuma, R. Haubrich, S. Riddler, P. E. Sax, E. S. Daar, H. J. Ribaudo, B. Agan, S. Agarwal, R. L. Ahern, B. L. Allen, S. Altidor, E. L. Altschuler, S. Ambardar, K. Anastos, B. Anderson, V. Anderson, U. Andrady, D. Antoniskis, D. Bangsberg, D. Barbaro, W. Barrie, J. Bartczak, S. Barton, P. Basden, N. Basgoz, S. Bazner, N. C. Bellos, A. M. Benson, J. Berger, N. F. Bernard, A. M. Bernard, C. Birch, S. J. Bodner, R. K. Bolan, E. T. Boudreaux, M. Bradley, J. F. Braun, J. E. Brndjar, S. J. Brown, K. Brown, S. T. Brown, J. Burack, L. M. Bush, V. Cafaro, O. Campbell, J. Campbell, R. H. Carlson, J. K. Carmichael, K. K. Casey, C. Cavacuiti, G. Celestin, S. T. Chambers, N. Chez, L. M. Chirch, P. J. Cimoch, D. Cohen, L. E. Cohn, B. Conway, D. A. Cooper, B. Cornelson, D. T. Cox, M. V. Cristofano, G. Cuchural Jr, J. L. Czartoski, J. M. Dahman, J. S. Daly, B. T. Davis, K. Davis, S. M. Davod, E. DeJesus, C. A. Dietz, E. Dunham, M. E. Dunn, T. B. Ellerin, J. J. Eron, J. J. W. Fangman, C. E. Farel, H. Ferlazzo, S. Fidler, A. Fleenor-Ford, R. Frankel, K. A. Freedberg, N. K. French, J. D. Fuchs, J. D. Fuller, J. Gaberman, J. E. Gallant, R. T. Gandhi, E. Garcia, D. Garmon, J. C. Gathe Jr, C. R. Gaultier, W. Gebre, F. D. Gilman, I. Gilson, P. A. Goepfert, M. S. Gottlieb, C. Goulston, R. K. Groger, T. D. Gurley, S. Haber, R. Hardwicke, W. D. Hardy, P. R. Harrigan, T. N. Hawkins, S. Heath, F. M. Hecht, W. K. Henry, M. Hladek, R. P. Hoffman, J. M. Horton, R. K. Hsu, G. D. Huhn, P. Hunt, M. J. Hupert, M. L. Illeman, H. Jaeger, R. M. Jellinger, M. John, J. A. Johnson, K. L. Johnson, H. Johnson, K. Johnson, J. Joly, W. C. Jordan, C. A. Kauffman, H. Khanlou, R. K. Killian, A. Y. Kim, D. D. Kim, C. A. Kinder, J. T. Kirchner, L. Kogelman, E. M. Kojic, P. T. Korthuis, W. Kurisu, D. S. Kwon, M. LaMar, H. Lampiris, M. Lanzafame, M. M. Lederman, D. M. Lee, J. M. L. Lee, M. J. Lee, E. T. Y. Lee, J. Lemoine, J. A. Levy, J. M. Llibre, M. A. Liguori, S. J. Little, A. Y. Liu, A. J. Lopez, M. R. Loutfy, D. Loy, D. Y. Mohammed, A. Man, M. K. Mansour, V. C. Marconi, M. Markowitz, R. Marques, J. N. Martin, H. L. Martin Jr, K. H. Mayer, M. J. McElrath, T. A. McGhee, B. H. McGovern, K. McGowan, D. McIntyre, G. X. Mcleod, P. Menezes, G. Mesa, C. E. Metroka, D. Meyer-Olson, A. O. Miller, K. Montgomery, K. C. Mounzer, E. H. Nagami, I. Nagin, R. G. Nahass, M. O. Nelson, C. Nielsen, D. L. Norene, D. H. O'Connor, B. O. Ojikutu, J. Okulicz, O. O. Oladehin, E. C. Oldfield 3rd, S. A. Olender, M. Ostrowski, W. F. Owen Jr, E. Pae, J. Parsonnet, A. M. Pavlatos, A. M. Perlmutter, M. N. Pierce, J. M. Pincus, L. Pisani, L. J. Price, L. Proia, R. C. Prokesch, H. C. Pujet, M. Ramgopal, A. Rathod, M. Rausch, J. Ravishankar, F. S. Rhame, C. S. Richards, D. D. Richman, B. Rodes, M. Rodriguez, R. C. Rose 3rd, E. S. Rosenberg, D. Rosenthal, P. E. Ross, D. S. Rubin, E. Rumbaugh, L. Saenz, M. R. Salvaggio, W. C. Sanchez, V. M. Sanjana, S. Santiago, W. Schmidt, H. Schuitemaker, P. M. Sestak, P. Shalit, W. Shay, V. N. Shirvani, V. I. Silebi, J. M. Sizemore Jr, P. R. Skolnik, M. Sokol-Anderson, J. M. Sosman, P. Stabile, J. T. Stapleton, S. Starrett, F. Stein, H.-J. Stellbrink, F. L. Sterman, V. E. Stone, D. R. Stone, G. Tambussi, R. A. Taplitz, E. M. Tedaldi, A. Telenti, W. Theisen, R. Torres, L. Tosiello, C. Tremblay, M. A. Tribble, P. D. Trinh, A. Tsao, P. Ueda, A. Vaccaro, E.

Valadas, T. J. Vanig, I. Vecino, V. M. Vega, W. Veikley, B. H. Wade, C. Walworth, C. Wanidworanun, D. J. Ward, D. A. Warner, R. D. Weber, D. Webster, S. Weis, D. A. Wheeler, D. J. White, E. Wilkins, A. Winston, C. G. Wlodaver, A. van't Wout, D. P. Wright, O. O. Yang, D. L. Yurdin, B. W. Zabukovic, K. C. Zachary, B. Zeeman, and M. Zhao. 2010. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330: 1551-1557.

41. Matthews, P. C., A. J. Leslie, A. Katzourakis, H. Crawford, R. Payne, A. Prendergast, K. Power, A. D. Kelleher, P. Klenerman, J. Carlson, D. Heckerman, T. Ndung'u, B. D. Walker, T. M. Allen, O. G. Pybus, and P. J. R. Goulder. 2009. HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering. *J. Virol.* 83: 4605-4615.

42. de Groot, N. G., N. Otting, G. G. M. Doxiadis, S. S. Balla-Jhagjhoorsingh, J. L. Heeney, J. J. van Rood, P. Gagneux, and R. E. Bontrop. 2002. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 99: 11748-11753.

43. Balla-Jhagjhoorsingh, S. S., G. Koopman, P. Mooij, T. G. Haaksma, V. J. Teeuwsen, R. E. Bontrop, and J. L. Heeney. 1999. Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. *J. Immunol.* 162: 2308-2314.

44. Hoof, I., C. Keşmir, O. Lund, and M. Nielsen. 2008. Humans with chimpanzee-like major histocompatibility complex-specificities control HIV-1 infection. *AIDS* 22: 1299-1303.

45. de Groot, N. G., C. M. C. Heijmans, Y. M. Zoet, A. H. de Ru, F. A. Verreck, P. A. van Veelen, J. W. Drijfhout, G. G. M. Doxiadis, E. J. Remarque, I. I. N. Doxiadis, J. J. van Rood, F. Koning, and R. E. Bontrop. 2010. 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.* 107: 15175-15180.

46. Bontrop, R. E., and D. I. Watkins. 2005. MHC polymorphism: AIDS susceptibility in non-human primates. *Trends Immunol.* 26: 227-233.

47. McKiernan, S. M., R. Hagan, M. Curry, G. S. A. McDonald, A. Kelly, N. Nolan, A. Walsh, J. Hegarty, E. Lawlor, and D. Kelleher. 2004. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* 40: 108-114.

48. Chuang, W. C.-M., F. Sarkodie, C. J. Brown, S. Owusu-Ofori, J. Brown, C. Li, C. Navarrete, P. Klenerman, and J.-P. Allain. 2007. Protective effect of HLA-B57 on HCV genotype 2 infection in a West African population. *J. Med. Virol.* 79: 724-733.

49. Kuniholm, M. H., A. Kovacs, X. Gao, X. Xue, D. Marti, C. L. Thio, M. G. Peters, N. A. Terrault, R. M. Greenblatt, J. J. Goedert, M. H. Cohen, H. Minkoff, S. J. Gange, K. Anastos, M. Fazzari, T. G. Harris, M. A. Young, H. D. Strickler, and M. Carrington. 2010. Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology* 51: 1514-1522.

50. Neumann-Haefelin, C., and R. Thimme. 2007. Impact of the genetic restriction of virus-specific T-cell responses in hepatitis C virus infection. *Genes Immun.* 8: 181-192.

51. Jeffery, K. J., A. A. Siddiqui, M. Bunce, A. L. Lloyd, A. M. Vine, A. D. Witkover, S. Izumo, K. Usuku, K. I. Welsh, M. Osame, and C. R. Bangham. 2000. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. *J. Immunol.* 165: 7278-7284.

52. McAulay, K. A., C. D. Higgins, K. F. Macsween, A. Lake, R. F. Jarrett, F. L. Robertson, H. Williams, and D. H. Crawford. 2007. HLA class I polymorphisms are associated with development of infectious mononucleosis upon primary EBV infection. *J. Clin. Invest.* 117: 3042-3048.

53. Niens, M., R. F. Jarrett, B. Hepkema, I. M. Nolte, A. Diepstra, M. Platteel, N. Kouprie, C. P. Delury, A. Gallagher, L. Visser, S. Poppema, G. J. te Meerman, and A. van den Berg. 2007. HLA-A*02 is associated with a reduced risk and HLA-A*01 with an increased risk of developing EBV+ Hodgkin lymphoma. *Blood* 110: 3310-3315.

54. Loke, H., D. B. Bethell, C. X. Phuong, M. Dung, J. Schneider, N. J. White, N. P. Day, J. Farrar, and A. V. Hill. 2001. Strong HLA class I--restricted T cell responses in dengue hemorrhagic fever: a double-edged sword? *J. Infect. Dis.* 184: 1369-1373.

55. Nguyen, T. P. L., M. Kikuchi, T. Q. H. Vu, Q. H. Do, T. T. Tran, D. T. Vo, M. T. Ha, V. T. Vo, T. P. N. Cao, V. D. Tran, T. Oyama, K. Morita, M. Yasunami, and K. Hirayama. 2008. 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* 2: e304.

56. Stephens, H. A. F., R. Klaythong, M. Sirikong, D. W. Vaughn, S. Green, S. Kalayanarooj, T. P. Endy, D. H. Libraty, A. Nisalak, B. L. Innis, A. L. Rothman, F. A. Ennis, and D. Chandanayingyong. 2002. 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* 60: 309-318.
57. Edwards, B. H., A. Bansal, S. Sabbaj, J. Bakari, M. J. Mulligan, and P. A. Goepfert. 2002. 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.* 76: 2298-2305.
58. Ogg, G. S., X. Jin, S. Bonhoeffer, P. R. Dunbar, M. A. Nowak, S. Monard, J. P. Segal, Y. Cao, S. L. Rowland-Jones, V. Cerundolo, A. Hurley, M. Markowitz, D. D. Ho, D. F. Nixon, and A. J. McMichael. 1998. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* 279: 2103-2106.
59. Altfeld, M., M. M. Addo, E. S. Rosenberg, F. M. Hecht, P. K. Lee, M. Vogel, X. G. Yu, R. Draenert, M. N. Johnston, D. Strick, T. M. Allen, M. E. Feeney, J. O. Kahn, R. P. Sekaly, J. A. Levy, J. K. Rockstroh, P. J. Goulder, and B. D. Walker. 2003. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* 17: 2581-2591.
60. Neumann-Haefelin, C., S. McKiernan, S. Ward, S. Viazov, H. C. Spangenberg, T. Killinger, T. F. Baumert, N. Nazarova, I. Sheridan, O. Pybus, F. von Weizsäcker, M. Roggendorf, D. Kelleher, P. Klenerman, 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.
61. Ray, S. C., L. Fanning, X.-H. Wang, D. M. Netski, E. Kenny-Walsh, and D. L. Thomas. 2005. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J. Exp. Med.* 201: 1753-1759.
62. Timm, J., G. M. Lauer, D. G. Kavanagh, I. Sheridan, A. Y. Kim, M. Lucas, T. Pillay, K. Ouchi, L. L. Reyor, J. Schulze zur Wiesch, R. T. Gandhi, R. T. Chung, N. Bhardwaj, P. Klenerman, B. D. Walker, and T. M. Allen. 2004. CD8 epitope escape and reversion in acute HCV infection. *J. Exp. Med.* 200: 1593-1604.
63. Gaudieri, S., A. Rauch, L. P. Park, E. Freitas, S. Herrmann, G. Jeffrey, W. Cheng, K. Pfafferott, K. Naidoo, R. Chapman, M. Battegay, R. Weber, A. Telenti, H. Furrer, I. James, M. Lucas, and S. A. Mallal. 2006. Evidence of viral adaptation to HLA class I-restricted immune pressure in chronic hepatitis C virus infection. *J. Virol.* 80: 11094-11104.
64. McKiernan, S. M., R. Hagan, M. Curry, G. S. A. McDonald, A. Kelly, N. Nolan, A. Walsh, J. Hegarty, E. Lawlor, and D. Kelleher. 2004. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* 40: 108-114.
65. Tripathy, A. S., U. Shankarkumar, M. S. Chadha, K. Ghosh, and V. A. Arankalle. 2009. Association of HLA alleles with hepatitis C infection in Maharashtra, western India. *Indian J. Med. Res* 130: 550-555.
66. Bednarek, M. A., S. Y. Sauma, M. C. Gammon, G. Porter, S. Tamhankar, A. R. Williamson, and H. J. Zweerink. 1991. The minimum peptide epitope from the influenza virus matrix protein. Extra and intracellular loading of HLA-A2. *J. Immunol.* 147: 4047-4053.
67. Morrison, J., J. Elvin, F. Latron, F. Gotch, R. Moots, J. L. Strominger, and A. McMichael. 1992. 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.* 22: 903-907.
68. Boon, A. C. M., G. De Mutsert, R. A. M. Fouchier, K. Sintnicolaas, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan. 2004. Preferential HLA usage in the influenza virus-specific CTL response. *J. Immunol.* 172: 4435-4443.
69. Boon, A. C. M., G. de Mutsert, Y. M. F. Graus, R. A. M. Fouchier, K. Sintnicolaas, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan. 2002. 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.* 76: 582-590.
70. Lewinsohn, D. A., E. Winata, G. M. Swarbrick, K. E. Tanner, M. S. Cook, M. D. Null, M. E. Cansler, A. Sette, J. Sidney, and D. M. Lewinsohn. 2007. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS Pathog.* 3: 1240-1249.
71. Peters, B., W. Tong, J. Sidney, A. Sette, and Z. Weng. 2003. Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. *Bioinformatics* 19: 1765-1772.
72. Ruppert, J., J. Sidney, E. Celis, R. T. Kubo, H. M. Grey, and A. Sette. 1993. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* 74: 929-937.
73. Tenzer, S., B. Peters, S. Bulik, O. Schoor, C. Lemmel, M. M. Schatz, P. M. Kloetzel, H. G. Rammensee, H. Schild, and H. G. Holzh\ütter. 2005. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences* 62: 1025–1037.

74. Borrás-Cuesta, F., J. Golvano, M. García-Granero, P. Sarobe, J. Riezu-Boj, E. Huarte, and J. Lasarte. 2000. Specific and general HLA-DR binding motifs: comparison of algorithms. *Hum. Immunol.* 61: 266-278.

75. Brusic, V., C. Schönbach, M. Takiguchi, V. Ciesielski, and L. C. Harrison. 1997. Application of genetic search in derivation of matrix models of peptide binding to MHC molecules. *Proc Int Conf Intell Syst Mol Biol* 5: 75-83.

76. Lundegaard, C., K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen. 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: W509-512.

77. Hoof, I., B. Peters, J. Sidney, L. E. Pedersen, A. Sette, O. Lund, S. Buus, and M. Nielsen. 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61: 1–13.

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

79. Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.

80. Vogel, T. U., D. T. Evans, J. A. Urvater, D. H. O'Connor, A. L. Hughes, and D. I. Watkins. 1999. Major histocompatibility complex class I genes in primates: co-evolution with pathogens. *Immunol. Rev.* 167: 327-337.

81. Cooke, G. S., and A. V. Hill. 2001. Genetics of susceptibility to human infectious disease. *Nat. Rev. Genet.* 2: 967-977.

82. Prugnolle, F., A. Manica, M. Charpentier, J. F. Guégan, V. Guernier, and F. Balloux. 2005. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15: 1022-1027.

83. Borghans, J. A. M., J. B. Beltman, and R. J. De Boer. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55: 732-739.

84. Schellens, I. M. M., C. Keşmir, F. Miedema, D. van Baarle, and J. A. M. Borghans. 2008. An unanticipated lack of consensus cytotoxic T lymphocyte epitopes in HIV-1 databases: the contribution of prediction programs. *AIDS* 22: 33-37.

85. McAdam, S. N., J. E. Boyson, X. Liu, T. L. Garber, A. L. Hughes, R. E. Bontrop, and D. I. Watkins. 1994. A uniquely high level of recombination at the HLA-B locus. *Proc. Natl. Acad. Sci. U.S.A.* 91: 5893-5897.

86. Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger, C. Day, P. Klenerman, J. Mullins, B. Korber, H. M. Coovadia, B. D. Walker, and P. J. R. Goulder. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432: 769-775.

87. Bihl, F., N. Frahm, L. Di Giammarino, J. Sidney, M. John, K. Yusim, T. Woodberry, K. Sango, H. S. Hewitt, L. Henry, C. H. Linde, J. V. Chisholm 3rd, 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.

88. Boon, A. C. M., G. De Mutsert, R. A. M. Fouchier, K. Sintnicolaas, A. D. M. E. Osterhaus, and G. F. Rimmelzwaan. 2004. Preferential HLA usage in the influenza virus-specific CTL response. *J. Immunol.* 172: 4435-4443.

89. Lacey, S. F., M. C. Villacres, C. La Rosa, Z. Wang, J. Longmate, J. Martinez, J. C. Brewer, S. Mekhoubad, R. Maas, J. M. Leedom, S. J. Forman, J. A. Zaia, and D. J. Diamond. 2003. 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. *Hum. Immunol.* 64: 440-452.

90. Lewinsohn, D. A., E. Winata, G. M. Swarbrick, K. E. Tanner, M. S. Cook, M. D. Null, M. E. Cansler, A. Sette, J. Sidney, and D. M. Lewinsohn. 2007. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS Pathog.* 3: 1240-1249.

91. Weichold, F. F., S. Mueller, C. Kortsik, W. E. Hitzler, M. J. Wulf, D. M. Hone, J. C. Sadoff, and M. J. Maeurer. 2007. Impact of MHC class I alleles on the M. tuberculosis antigen-specific CD8+ T-cell response in patients with pulmonary tuberculosis. *Genes Immun.* 8: 334-343.

92. 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.
93. Gao, X., G. W. Nelson, P. Karacki, M. P. Martin, J. Phair, R. Kaslow, J. J. Goedert, S. Buchbinder, K. Hoots, D. Vlahov, S. J. O'Brien, and M. Carrington. 2001. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N. Engl. J. Med.* 344: 1668-1675.
94. Martin, M. P., and M. Carrington. 2005. Immunogenetics of viral infections. *Curr. Opin. Immunol.* 17: 510-516.
95. Carrington, M., and S. J. O'Brien. 2003. The influence of HLA genotype on AIDS. *Annu. Rev. Med.* 54: 535-551.
96. Neumann-Haefelin, C., S. McKiernan, S. Ward, S. Viazov, H. C. Spangenberg, T. Killinger, T. F. Baumert, N. Nazarova, I. Sheridan, O. Pybus, F. von Weizsäcker, M. Roggendorf, D. Kelleher, P. Klenerman, 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.
97. 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.
98. Khakoo, S. I., and M. Carrington. 2006. KIR and disease: a model system or system of models? *Immunol. Rev.* 214: 186-201.
99. Yewdell, J. W., and J. R. Bennink. 1999. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu. Rev. Immunol.* 17: 51-88.
100. Yewdell, J. W. 2006. Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. *Immunity* 25: 533-543.
101. Burroughs, N. J., R. J. de Boer, and C. Keşmir. 2004. Discriminating self from nonself with short peptides from large proteomes. *Immunogenetics* 56: 311-320.
102. Rammensee, H., J. Bachmann, N. P. Emmerich, O. A. Bachor, and S. Stevanović. 1999. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50: 213-219.
103. Gulukota, K., J. Sidney, A. Sette, and C. DeLisi. 1997. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol.* 267: 1258-1267.
104. Parker, K. C., M. A. Bednarek, and J. E. Coligan. 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* 152: 163-175.
105. Segal, M. R., M. P. Cummings, and A. E. Hubbard. 2001. Relating amino acid sequence to phenotype: analysis of peptide-binding data. *Biometrics* 57: 632-642.
106. Doytchinova, I. A., M. J. Blythe, and D. R. Flower. 2002. Additive method for the prediction of protein-peptide binding affinity. Application to the MHC class I molecule HLA-A*0201. *J. Proteome Res.* 1: 263-272.
107. Peters, B., H.-H. Bui, S. Frankild, M. Nielson, C. Lundegaard, E. Kostem, D. Basch, K. Lamberth, M. Harndahl, W. Fleri, S. S. Wilson, J. Sidney, O. Lund, S. Buus, and A. Sette. 2006. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput. Biol.* 2: e65.
108. Schneider, T. D., and R. M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18: 6097-6100.
109. Peters, B., and A. Sette. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6: 132.
110. Assarsson, E., J. Sidney, C. Oseroff, V. Pasquetto, H.-H. Bui, N. Frahm, C. Brander, B. Peters, H. Grey, and A. Sette. 2007. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J. Immunol.* 178: 7890-7901.
111. Rammensee, H. G., T. Friede, and S. Stevanoviíc. 1995. MHC ligands and peptide motifs: first listing. *Immunogenetics* 41: 178-228.
112. Falk, K., O. Rötzschke, B. Grahovac, D. Schendel, S. Stevanović, G. Jung, and H. G. Rammensee. 1993. Peptide motifs of HLA-B35 and -B37 molecules. *Immunogenetics* 38: 161-162.
113. Takamiya, Y., C. Schönbach, K. Nokihara, M. Yamaguchi, S. Ferrone, K. Kano, K. Egawa, and M. Takiguchi. 1994. HLA-B*3501-peptide interactions: role of anchor residues of peptides in their binding to HLA-B*3501 molecules. *Int. Immunol.* 6: 255-261.
114. DiBrino, M., K. C. Parker, D. H. Margulies, J. Shiloach, R. V. Turner, W. E. Biddison, and J. E. Coligan. 1994. The HLA-B14 peptide binding site can accommodate peptides with different

combinations of anchor residues. *J. Biol. Chem.* 269: 32426-32434.
115. Lamas, J. R., A. Paradela, F. Roncal, and J. A. López de Castro. 1999. Modulation at multiple anchor positions of the peptide specificity of HLA-B27 subtypes differentially associated with ankylosing spondylitis. *Arthritis Rheum.* 42: 1975-1985.
116. Falk, K., O. Rötzschke, S. Stevanović, G. Jung, and H.-G. Rammensee. 2006. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. 1991. *J. Immunol.* 177: 2741-2747.
117. Yewdell, J. W., E. Reits, and J. Neefjes. 2003. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* 3: 952-961.
118. Kotturi, M. F., I. Scott, T. Wolfe, B. Peters, J. Sidney, H. Cheroutre, M. G. von Herrath, M. J. Buchmeier, H. Grey, and A. Sette. 2008. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. *J. Immunol.* 181: 2124-2133.
119. Howarth, M., A. Williams, A. B. Tolstrup, and T. Elliott. 2004. Tapasin enhances MHC class I peptide presentation according to peptide half-life. *Proc. Natl. Acad. Sci. U.S.A.* 101: 11737-11742.
120. Kienast, A., M. Preuss, M. Winkler, and T. P. Dick. 2007. Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. *Nat. Immunol.* 8: 864-872.
121. Wearsch, P. A., and P. Cresswell. 2007. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat. Immunol.* 8: 873-881.
122. Purcell, A. W., and T. Elliott. 2008. Molecular machinations of the MHC-I peptide loading complex. *Curr. Opin. Immunol.* 20: 75-81.
123. Turnquist, H. R., H. J. Thomas, K. R. Prilliman, C. T. Lutz, W. H. Hildebrand, and J. C. Solheim. 2000. HLA-B polymorphism affects interactions with multiple endoplasmic reticulum proteins. *Eur. J. Immunol.* 30: 3021-3028.
124. Groothuis, T. A. M., A. C. Griekspoor, J. J. Neijssen, C. A. Herberts, and J. J. Neefjes. 2005. MHC class I alleles and their exploration of the antigen-processing machinery. *Immunol. Rev.* 207: 60-76.
125. Kesmir, C., V. van Noort, R. J. de Boer, and P. Hogeweg. 2003. Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome. *Immunogenetics* 55: 437-449.
126. Harari, A., C. Cellerai, F. B. Enders, J. Köstler, 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.
127. Van Parijs, L., and A. K. Abbas. 1998. Homeostasis and self-tolerance in the immune system: turning lymphocytes off. *Science* 280: 243-248.
128. Molldrem, J. J., P. P. Lee, S. Kant, E. Wieder, W. Jiang, S. Lu, C. Wang, and M. M. Davis. 2003. Chronic myelogenous leukemia shapes host immunity by selective deletion of high-avidity leukemia-specific T cells. *J. Clin. Invest.* 111: 639-647.
129. Lichterfeld, M., X. G. Yu, S. K. Mui, K. L. Williams, A. Trocha, M. A. Brockman, R. L. Allgaier, M. T. Waring, T. Koibuchi, M. N. Johnston, D. Cohen, T. M. Allen, E. S. Rosenberg, B. D. Walker, and M. Altfeld. 2007. Selective depletion of high-avidity human immunodeficiency virus type 1 (HIV-1)-specific CD8+ T cells after early HIV-1 infection. *J. Virol.* 81: 4199-4214.
130. Almeida, J. R., D. A. Price, L. Papagno, Z. A. Arkoub, D. Sauce, E. Bornstein, T. E. Asher, A. Samri, A. Schnuriger, I. Theodorou, D. Costagliola, C. Rouzioux, H. Agut, A.-G. Marcelin, D. Douek, B. Autran, and V. Appay. 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.
131. De Boer, R. J., D. Homann, and A. S. Perelson. 2003. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *J. Immunol.* 171: 3928-3935.
132. Obar, J. J., K. M. Khanna, and L. Lefrançois. 2008. Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection. *Immunity* 28: 859-869.
133. Goulder, P. J. R., and D. I. Watkins. 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat. Rev. Immunol.* 8: 619-630.
134. Bihl, F., N. Frahm, L. Di Giammarino, J. Sidney, M. John, K. Yusim, T. Woodberry, K. Sango, H. S. Hewitt, L. Henry, C. H. Linde, J. V. Chisholm 3rd, 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.

135. Borghans, J. A. M., A. Mølgaard, R. J. de Boer, and C. Keşmir. 2007. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PLoS ONE* 2: e920.

136. Chopera, D. R., Z. Woodman, K. Mlisana, M. Mlotshwa, D. P. Martin, C. Seoighe, F. Treurnicht, D. A. de Rosa, W. Hide, S. A. Karim, C. M. Gray, and C. Williamson. 2008. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. *PLoS Pathog.* 4: e1000033.

137. Edwards, B. H., A. Bansal, S. Sabbaj, J. Bakari, M. J. Mulligan, and P. A. Goepfert. 2002. 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.* 76: 2298-2305.

138. Goepfert, P. A., W. Lumm, P. Farmer, P. Matthews, A. Prendergast, J. M. Carlson, C. A. Derdeyn, J. Tang, R. A. Kaslow, A. Bansal, K. Yusim, D. Heckerman, J. Mulenga, S. Allen, P. J. R. Goulder, and E. Hunter. 2008. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *J. Exp. Med.* 205: 1009-1017.

139. Kiepiela, P., K. Ngumbela, C. Thobakgale, D. Ramduth, I. Honeyborne, E. Moodley, S. Reddy, C. de Pierres, Z. Mncube, N. Mkhwanazi, K. Bishop, M. van der Stok, K. Nair, N. Khan, H. Crawford, R. Payne, A. Leslie, J. Prado, A. Prendergast, J. Frater, N. McCarthy, C. Brander, G. H. Learn, D. Nickle, C. Rousseau, H. Coovadia, J. I. Mullins, D. Heckerman, B. D. Walker, and P. Goulder. 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat. Med.* 13: 46-53.

140. Novitsky, V., P. Gilbert, T. Peter, M. F. McLane, S. Gaolekwe, N. Rybak, I. Thior, T. Ndung'u, R. Marlink, T. H. Lee, and M. Essex. 2003. Association between virus-specific T-cell responses and plasma viral load in human immunodeficiency virus type 1 subtype C infection. *J. Virol.* 77: 882-890.

141. Rolland, M., D. Heckerman, W. Deng, C. M. Rousseau, H. Coovadia, K. Bishop, P. J. R. Goulder, B. D. Walker, C. Brander, and J. I. Mullins. 2008. Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads. *PLoS ONE* 3: e1424.

142. Zuñiga, R., A. Lucchetti, P. Galvan, S. Sanchez, C. Sanchez, A. Hernandez, H. Sanchez, N. Frahm, C. H. Linde, H. S. Hewitt, W. Hildebrand, M. Altfeld, T. M. Allen, B. D. Walker, B. T. Korber, T. Leitner, J. Sanchez, and C. Brander. 2006. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *J. Virol.* 80: 3122-3125.

143. Frahm, N., B. T. Korber, C. M. Adams, J. J. Szinger, R. Draenert, M. M. Addo, M. E. Feeney, K. Yusim, K. Sango, N. V. Brown, D. SenGupta, A. Piechocka-Trocha, T. Simonis, F. M. Marincola, A. G. Wurcel, D. R. Stone, C. J. Russell, P. Adolf, D. Cohen, T. Roach, A. StJohn, A. Khatri, K. Davis, J. Mullins, P. J. R. Goulder, B. D. Walker, and C. Brander. 2004. Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. *J. Virol.* 78: 2187-2200.

144. Altfeld, M., M. M. Addo, R. Shankarappa, P. K. Lee, T. M. Allen, X. G. Yu, A. Rathod, J. Harlow, K. O'Sullivan, M. N. Johnston, P. J. R. Goulder, J. I. Mullins, E. S. Rosenberg, C. Brander, B. Korber, and B. D. Walker. 2003. Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences. *J. Virol.* 77: 7330-7340.

145. Betts, M. R., D. R. Ambrozak, D. C. Douek, S. Bonhoeffer, J. M. Brenchley, J. P. Casazza, R. A. Koup, and L. J. Picker. 2001. Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection. *J. Virol.* 75: 11983-11991.

146. Addo, M. M., X. G. Yu, A. Rathod, D. Cohen, R. L. Eldridge, D. Strick, M. N. Johnston, C. Corcoran, A. G. Wurcel, C. A. Fitzpatrick, M. E. Feeney, W. R. Rodriguez, N. Basgoz, R. Draenert, D. R. Stone, C. Brander, P. J. R. Goulder, E. S. Rosenberg, M. Altfeld, and B. D. Walker. 2003. Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* 77: 2081-2092.

147. Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber. 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human

immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol.* 76: 8757-8768.
148. Wang, Y. E., B. Li, J. M. Carlson, H. Streeck, A. D. Gladden, R. Goodman, A. Schneidewind, K. A. Power, I. Toth, N. Frahm, G. Alter, C. Brander, M. Carrington, B. D. Walker, M. Altfeld, D. Heckerman, and T. M. Allen. 2009. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *J. Virol.* 83: 1845-1855.
149. Li, B., A. D. Gladden, M. Altfeld, J. M. Kaldor, D. A. Cooper, A. D. Kelleher, and T. M. Allen. 2007. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J. Virol.* 81: 193-201.
150. Kawashima, Y., K. Pfafferott, J. Frater, P. Matthews, R. Payne, M. Addo, H. Gatanaga, M. Fujiwara, A. Hachiya, H. Koizumi, N. Kuse, S. Oka, A. Duda, A. Prendergast, H. Crawford, A. Leslie, Z. Brumme, C. Brumme, T. Allen, C. Brander, R. Kaslow, J. Tang, E. Hunter, S. Allen, J. Mulenga, S. Branch, T. Roach, M. John, S. Mallal, A. Ogwu, R. Shapiro, J. G. Prado, S. Fidler, J. Weber, O. G. Pybus, P. Klenerman, T. Ndung'u, R. Phillips, D. Heckerman, P. R. Harrigan, B. D. Walker, M. Takiguchi, and P. Goulder. 2009. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* 458: 641-645.
151. Rapin, N., I. Hoof, O. Lund, and M. Nielsen. 2008. MHC motif viewer. *Immunogenetics* 60: 759-765.
152. Schellens, I. M. M., C. Keşmir, F. Miedema, D. van Baarle, and J. A. M. Borghans. 2008. An unanticipated lack of consensus cytotoxic T lymphocyte epitopes in HIV-1 databases: the contribution of prediction programs. *AIDS* 22: 33-37.
153. Lavanchy, D. 2009. The global burden of hepatitis C. *Liver Int.* 29 Suppl 1: 74-81.
154. Marcellin, P. 1999. Hepatitis C: the clinical spectrum of the disease. *J. Hepatol* 31 Suppl 1: 9-16.
155. Thimme, R., D. Oldach, K. M. Chang, C. Steiger, S. C. Ray, and F. V. Chisari. 2001. Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J. Exp. Med* 194: 1395-1406.
156. Grüner, N. H., T. J. Gerlach, M. C. Jung, H. M. Diepolder, C. A. Schirren, W. W. Schraut, R. Hoffmann, R. Zachoval, T. Santantonio, M. Cucchiarini, A. Cerny, and G. R. Pape. 2000. Association of hepatitis C virus-specific CD8+ T cells with viral clearance in acute hepatitis C. *J. Infect. Dis* 181: 1528-1536.
157. Lechner, F., D. K. Wong, P. R. Dunbar, R. Chapman, R. T. Chung, P. Dohrenwend, G. Robbins, R. Phillips, P. Klenerman, and B. D. Walker. 2000. Analysis of successful immune responses in persons infected with hepatitis C virus. *J. Exp. Med* 191: 1499-1512.
158. Cooper, S., A. L. Erickson, E. J. Adams, J. Kansopon, A. J. Weiner, D. Y. Chien, M. Houghton, P. Parham, and C. M. Walker. 1999. Analysis of a successful immune response against hepatitis C virus. *Immunity* 10: 439-449.
159. van den Berg, C. H. S. B., T. A. Ruys, N. M. Nanlohy, S. E. Geerlings, J. T. van der Meer, J.-W. Mulder, J. A. Lange, and D. van Baarle. 2009. Comprehensive longitudinal analysis of hepatitis C virus (HCV)-specific T cell responses during acute HCV infection in the presence of existing HIV-1 infection. *J. Viral Hepat.* 16: 239-248.
160. Ruys, T. A., N. M. Nanlohy, C. H. S. B. van den Berg, E. Hassink, M. Beld, T. van de Laar, S. Bruisten, F. Wit, A. Krol, M. Prins, J. Lange, and D. van Baarle. 2008. HCV-specific T-cell responses in injecting drug users: evidence for previous exposure to HCV and a role for CD4+ T cells focussing on nonstructural proteins in viral clearance. *J. Viral Hepat.* 15: 409-420.
161. Kuniholm, M. H., A. Kovacs, X. Gao, X. Xue, D. Marti, C. L. Thio, M. G. Peters, N. A. Terrault, R. M. Greenblatt, J. J. Goedert, M. H. Cohen, H. Minkoff, S. J. Gange, K. Anastos, M. Fazzari, T. G. Harris, M. A. Young, H. D. Strickler, and M. Carrington. 2010. Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology* 51: 1514-1522.
162. Shimotohno, K., Y. Tanji, Y. Hirowatari, Y. Komoda, N. Kato, and M. Hijikata. 1995. Processing of the hepatitis C virus precursor protein. *J. Hepatol* 22: 87-92.
163. Chuang, W. C.-M., F. Sarkodie, C. J. Brown, S. Owusu-Ofori, J. Brown, C. Li, C. Navarrete, P. Klenerman, and J.-P. Allain. 2007. Protective effect of HLA-B57 on HCV genotype 2 infection in a West African population. *J. Med. Virol* 79: 724-733.
164. Macnamara, A., A. Rowan, S. Hilburn, U. Kadolsky, H. Fujiwara, K. Suemori, M. Yasukawa, G. Taylor, C. R. M. Bangham, and B. Asquith. 2010. HLA class I binding of HBZ determines outcome in HTLV-1 infection. *PLoS Pathog* 6.

165. Yusim, K., R. Richardson, N. Tao, A. Dalwani, A. Agrawal, J. Szinger, R. Funkhouser, B. Korber, and C. Kuiken. 2005. Los alamos hepatitis C immunology database. *Appl. Bioinformatics* 4: 217-225.
166. Vita, R., L. Zarebski, J. A. Greenbaum, H. Emami, I. Hoof, N. Salimi, R. Damle, A. Sette, and B. Peters. 2010. The immune epitope database 2.0. *Nucleic Acids Res.* 38: D854-862.
167. Simmonds, P., J. Bukh, C. Combet, G. Deléage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspé, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J.-M. Pawlotsky, F. Penin, E. Sablon, T. Shin-I, L. J. Stuyver, H.-J. Thiel, S. Viazov, A. J. Weiner, and A. Widell. 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 42: 962-973.
168. Kuiken, C., K. Yusim, L. Boykin, and R. Richardson. 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics* 21: 379-384.
169. Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191.
170. Shannon, C. E. 1948. *A mathematical theory of communication*,. American Telephone and Telegraph Company.
171. Lundegaard, C., K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen. 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: W509-512.
172. Hoof, I., B. Peters, J. Sidney, L. E. Pedersen, A. Sette, O. Lund, S. Buus, and M. Nielsen. 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61: 1–13.
173. Tenzer, S., B. Peters, S. Bulik, O. Schoor, C. Lemmel, M. M. Schatz, P. M. Kloetzel, H. G. Rammensee, H. Schild, and H. G. Holzh\ütter. 2005. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences* 62: 1025–1037.
174. Toes, R. E., A. K. Nussbaum, S. Degermann, M. Schirle, N. P. Emmerich, M. Kraft, C. Laplace, A. Zwinderman, T. P. Dick, J. Müller, B. Schönfisch, C. Schmid, H. J. Fehling, S. Stevanovic, H. G. Rammensee, and H. Schild. 2001. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J. Exp. Med.* 194: 1-12.
175. McLauchlan, J. 2000. Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J. Viral Hepat* 7: 2-14.
176. Fontaine Costa, A. I., X. Rao, E. Lechenadec, D. van Baarle, and C. Keşmir. 2010. HLA-B molecules target more conserved regions of the HIV-1 proteome. *AIDS* 24: 211-215.
177. Guglietta, S., A. R. Garbuglia, V. Pacciani, C. Scottà, M. P. Perrone, L. Laurenti, E. Spada, A. Mele, M. R. Capobianchi, G. Taliani, A. Folgori, A. Vitelli, L. Ruggeri, A. Nicosia, E. Piccolella, and P. Del Porto. 2005. Positive selection of cytotoxic T lymphocyte escape variants during acute hepatitis C virus infection. *Eur. J. Immunol* 35: 2627-2637.
178. Kuniholm, M. H., A. Kovacs, X. Gao, X. Xue, D. Marti, C. L. Thio, M. G. Peters, N. A. Terrault, R. M. Greenblatt, J. J. Goedert, M. H. Cohen, H. Minkoff, S. J. Gange, K. Anastos, M. Fazzari, T. G. Harris, M. A. Young, H. D. Strickler, and M. Carrington. 2010. Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology* 51: 1514-1522.
179. Thio, C. L., X. Gao, J. J. Goedert, D. Vlahov, K. E. Nelson, M. W. Hilgartner, S. J. O'Brien, P. Karacki, J. Astemborski, M. Carrington, and D. L. Thomas. 2002. HLA-Cw*04 and hepatitis C virus persistence. *J. Virol* 76: 4792-4797.
180. Kim, A. Y., T. Kuntzen, J. Timm, B. E. Nolan, M. A. Baca, L. L. Reyor, A. C. Berical, A. J. Feller, K. L. Johnson, J. S. Z. Wiesch, G. K. Robbins, R. T. Chung, B. D. Walker, M. Carrington, T. M. Allen, and G. M. Lauer. 2011. Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. *Gastroenterology* 140: 686-696.e1.
181. Chuang, W. C.-M., F. Sarkodie, C. J. Brown, S. Owusu-Ofori, J. Brown, C. Li, C. Navarrete, P. Klenerman, and J.-P. Allain. 2007. Protective effect of HLA-B57 on HCV genotype 2 infection in a West African population. *J. Med. Virol* 79: 724-733.
182. Neumann-Haefelin, C., J. Timm, J. Schmidt, N. Kersting, K. Fitzmaurice, C. Oniangue-Ndza, M. N. Kemper, I. Humphreys, S. McKiernan, D. Kelleher, V. Lohmann, P. Bowness, D. Huzly, H. R. Rosen, A. Y. Kim, G. M. Lauer, T. M. Allen, E. Barnes, M. Roggendorf, H. E. Blum, and R. Thimme. 2010. Protective effect of human leukocyte antigen B27 in hepatitis C virus infection requires the

presence of a genotype-specific immunodominant CD8+ T-cell epitope. *Hepatology* 51: 54-62.
183. Neumann-Haefelin, C., S. McKiernan, S. Ward, S. Viazov, H. C. Spangenberg, T. Killinger, T. F. Baumert, N. Nazarova, I. Sheridan, O. Pybus, F. von Weizsäcker, M. Roggendorf, D. Kelleher, P. Klenerman, 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.
184. Fanning, L. J., E. Kenny-Walsh, and F. Shanahan. 2004. Persistence of hepatitis C virus in a white population: associations with human leukocyte antigen class 1. *Hum. Immunol* 65: 745-751.
185. Thio, C. L., X. Gao, J. J. Goedert, D. Vlahov, K. E. Nelson, M. W. Hilgartner, S. J. O'Brien, P. Karacki, J. Astemborski, M. Carrington, and D. L. Thomas. 2002. HLA-Cw*04 and hepatitis C virus persistence. *J. Virol* 76: 4792-4797.
186. McKiernan, S. M., R. Hagan, M. Curry, G. S. A. McDonald, A. Kelly, N. Nolan, A. Walsh, J. Hegarty, E. Lawlor, and D. Kelleher. 2004. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* 40: 108-114.
187. Borghans, J. A. M., A. Mølgaard, R. J. de Boer, and C. Keşmir. 2007. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PLoS ONE* 2: e920.
188. Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57: 289-300.
189. Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber. 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol* 76: 8757-8768.
190. Comas, I., J. Chakravartti, P. M. Small, J. Galagan, S. Niemann, K. Kremer, J. D. Ernst, and S. Gagneux. 2010. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nat. Genet* 42: 498-503.
191. Li, H., A. L. Hughes, N. Bano, S. McArdle, S. Livingston, H. Deubner, B. J. McMahon, L. Townshend-Bulson, R. McMahan, H. R. Rosen, and D. R. Gretch. 2011. Genetic diversity of near genome-wide hepatitis C virus sequences during chronic infection: evidence for protein structural conservation over time. *PLoS ONE* 6: e19562.
192. Hraber, P., C. Kuiken, and K. Yusim. 2007. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. *Hepatology* 46: 1713-1721.
193. Kiepiela, P., K. Ngumbela, C. Thobakgale, D. Ramduth, I. Honeyborne, E. Moodley, S. Reddy, C. de Pierres, Z. Mncube, N. Mkhwanazi, K. Bishop, M. van der Stok, K. Nair, N. Khan, H. Crawford, R. Payne, A. Leslie, J. Prado, A. Prendergast, J. Frater, N. McCarthy, C. Brander, G. H. Learn, D. Nickle, C. Rousseau, H. Coovadia, J. I. Mullins, D. Heckerman, B. D. Walker, and P. Goulder. 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat. Med.* 13: 46-53.
194. Goulder, P. J. R., and D. I. Watkins. 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat. Rev. Immunol.* 8: 619-630.
195. Dazert, E., C. Neumann-Haefelin, S. Bressanelli, K. Fitzmaurice, J. Kort, J. Timm, S. McKiernan, D. Kelleher, N. Gruener, J. E. Tavis, H. R. Rosen, J. Shaw, P. Bowness, H. E. Blum, P. Klenerman, R. Bartenschlager, and R. Thimme. 2009. Loss of viral fitness and cross-recognition by CD8+ T cells limit HCV escape from a protective HLA-B27-restricted human immune response. *J. Clin. Invest* 119: 376-386.
196. Yoon, S. K., J. Y. Han, C.-W. Pyo, J. M. Yang, J. W. Jang, C. W. Kim, U. I. Chang, S. H. Bae, J. Y. Choi, K. W. Chung, H. S. Sun, H. B. Choi, and T.-G. Kim. 2005. Association between human leukocytes antigen alleles and chronic hepatitis C virus infection in the Korean population. *Liver Int* 25: 1122-1127.
197. Zekri, A.-R. N., H. A. El-Mahallawy, A. Hassan, N. H. A. El-Din, and A. M. Kamel. 2005. HLA alleles in Egyptian HCV genotype-4 carriers. *Egypt J Immunol* 12: 77-86.
198. Wang, J. H., X. Zheng, X. Ke, M. T. Dorak, J. Shen, B. Boodram, M. O'Gorman, K. Beaman, S. J. Cotler, R. Hershow, and L. Rong. 2009. Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. *Virol. J* 6: 46.
199. Bjorkman, P. J., and P. Parham. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem* 59: 253-288.
200. Parham, P., and T. Ohta. 1996. Population biology of antigen presentation by MHC class I molecules. *Science* 272: 67-74.

201. Marsh, S. G. E., E. D. Albert, W. F. Bodmer, R. E. Bontrop, B. Dupont, H. A. Erlich, M. Fernández-Viña, D. E. Geraghty, R. Holdsworth, C. K. Hurley, M. Lau, K. W. Lee, B. Mach, M. Maiers, W. R. Mayr, C. R. Müller, P. Parham, E. W. Petersdorf, T. Sasazuki, J. L. Strominger, A. Svejgaard, P. I. Terasaki, J. M. Tiercy, and J. Trowsdale. 2010. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens* 75: 291-455.

202. Sette, A., and J. Sidney. 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics.* 50: 201–212.

203. Lund, O., M. Nielsen, C. Kesmir, A. G. Petersen, C. Lundegaard, P. Worning, C. Sylvester-Hvid, K. Lamberth, G. Roder, S. Justesen, S. Buus, and S. Brunak. 2004. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics.* 55: 797–810.

204. Doytchinova, I. A., P. Guan, and D. R. Flower. 2004. Identifiying human MHC supertypes using bioinformatic methods. *J. Immunol* 172: 4314-4323.

205. Kangueane, P., M. K. Sakharkar, G. Rajaseger, S. Bolisetty, B. Sivasekari, B. Zhao, M. Ravichandran, P. Shapshak, and S. Subbiah. 2005. A framework to sub-type HLA supertypes. *Front. Biosci* 10: 879-886.

206. Reche, P. A., and E. L. Reinherz. 2007. Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. *Methods Mol. Biol* 409: 163-173.

207. Hertz, T., and C. Yanover. 2007. Identifying HLA supertypes by learning distance functions. *Bioinformatics* 23: e148-155.

208. Sidney, J., B. Peters, N. Frahm, C. Brander, and A. Sette. 2008. HLA class I supertypes: a revised and updated classification. *BMC Immunol* 9: 1.

209. Sidney, J., M. F. del Guercio, S. Southwood, V. H. Engelhard, E. Appella, H. G. Rammensee, K. Falk, O. Rötzschke, M. Takiguchi, and R. T. Kubo. 1995. Several HLA alleles share overlapping peptide specificities. *J. Immunol* 154: 247-259.

210. Barber, L. D., B. Gillece-Castro, L. Percival, X. Li, C. Clayberger, and P. Parham. 1995. Overlap in the repertoires of peptides bound in vivo by a group of related class I HLA-B allotypes. *Curr. Biol* 5: 179-190.

211. Doolan, D. L., S. L. Hoffman, S. Southwood, P. A. Wentworth, J. Sidney, R. W. Chesnut, E. Keogh, E. Appella, T. B. Nutman, A. A. Lal, D. M. Gordon, A. Oloo, and A. Sette. 1997. Degenerate cytotoxic T cell epitopes from P. falciparum restricted by multiple HLA-A and HLA-B supertype alleles. *Immunity* 7: 97-112.

212. Bertoni, R., J. Sidney, P. Fowler, R. W. Chesnut, F. V. Chisari, and A. Sette. 1997. Human histocompatibility leukocyte antigen-binding supermotifs predict broadly cross-reactive cytotoxic T lymphocyte responses in patients with acute hepatitis. *J. Clin. Invest* 100: 503-513.

213. Threlkeld, S. C., P. A. Wentworth, S. A. Kalams, B. M. Wilkes, D. J. Ruhl, E. Keogh, J. Sidney, S. Southwood, B. D. Walker, and A. Sette. 1997. Degenerate and promiscuous recognition by CTL of peptides presented by the MHC class I A3-like superfamily: implications for vaccine development. *J. Immunol* 159: 1648-1657.

214. Brusic, V., N. Petrovsky, G. Zhang, and V. B. Bajic. 2002. Prediction of promiscuous peptides that bind HLA class I molecules. *Immunol. Cell. Biol.* 80: 280–285.

215. Ueno, T., H. Tomiyama, and M. Takiguchi. 2002. Single T cell receptor-mediated recognition of an identical HIV-derived peptide presented by multiple HLA class I molecules. *J. Immunol* 169: 4961-4969.

216. Burrows, S. R., R. A. Elkington, J. J. Miles, K. J. Green, S. Walker, S. M. Haryana, D. J. Moss, H. Dunckley, J. M. Burrows, and R. Khanna. 2003. Promiscuous CTL recognition of viral epitopes on multiple human leukocyte antigens: biological validation of the proposed HLA A24 supertype. *J. Immunol* 171: 1407-1412.

217. Sidney, J., S. Southwood, V. Pasquetto, and A. Sette. 2003. Simultaneous prediction of binding capacity for multiple molecules of the HLA B44 supertype. *J. Immunol* 171: 5964-5974.

218. Takedatsu, H., S. Shichijo, K. Katagiri, H. Sawamizu, M. Sata, and K. Itoh. 2004. Identification of peptide vaccine candidates sharing among HLA-A3+, -A11+, -A31+, and -A33+ cancer patients. *Clin. Cancer Res* 10: 1112-1120.

219. Frahm, N., S. Adams, P. Kiepiela, C. H. Linde, H. S. Hewitt, M. Lichterfeld, K. Sango, N. V. Brown, E. Pae, A. G. Wurcel, M. Altfeld, M. E. Feeney, T. M. Allen, T. Roach, M. A. St John, E. S. Daar, E. Rosenberg, B. Korber, F. Marincola, B. D. Walker, P. J. R. Goulder, and C. Brander. 2005. HLA-B63 presents HLA-B57/B58-restricted cytotoxic T-lymphocyte epitopes and is associated with low human immunodeficiency virus load. *J. Virol* 79: 10218-10225.

220. Leslie, A., D. A. Price, P. Mkhize, K. Bishop, A. Rathod, C. Day, H. Crawford, I. Honeyborne, T. E. Asher, G. Luzzi, A. Edwards, C. M. Rousseau, C. M. Rosseau, J. I. Mullins, G. Tudor-Williams, V. Novelli, C. Brander, D. C. Douek, P. Kiepiela, B. D. Walker, and P. J. R. Goulder. 2006. Differential selection pressure exerted on HIV by CTL targeting identical epitopes but restricted by distinct HLA alleles from the same HLA supertype. *J. Immunol* 177: 4699-4708.

221. Sabbaj, S., A. Bansal, G. D. Ritter, C. Perkins, B. H. Edwards, E. Gough, J. Tang, J. J. Szinger, B. Korber, C. M. Wilson, R. A. Kaslow, M. J. Mulligan, and P. A. Goepfert. 2003. Cross-reactive CD8+ T cell epitopes identified in US adolescent minorities. *J. Acquir. Immune Defic. Syndr* 33: 426-438.

222. Masemola, A. M., T. N. Mashishi, G. Khoury, H. Bredell, M. Paximadis, T. Mathebula, D. Barkhan, A. Puren, E. Vardas, M. Colvin, L. Zijenah, D. Katzenstein, R. Musonda, S. Allen, N. Kumwenda, T. Taha, G. Gray, J. McIntyre, S. A. Karim, H. W. Sheppard, and C. M. Gray. 2004. Novel and promiscuous CTL epitopes in conserved regions of Gag targeted by individuals with early subtype C HIV type 1 infection from southern Africa. *J. Immunol* 173: 4607-4617.

223. Frahm, N., K. Yusim, T. J. Suscovich, S. Adams, J. Sidney, P. Hraber, H. S. Hewitt, C. H. Linde, D. G. Kavanagh, T. Woodberry, L. M. Henry, K. Faircloth, J. Listgarten, C. Kadie, N. Jojic, K. Sango, N. V. Brown, E. Pae, M. T. Zaman, F. Bihl, A. Khatri, M. John, S. Mallal, F. M. Marincola, B. D. Walker, A. Sette, D. Heckerman, B. T. Korber, and C. Brander. 2007. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur. J. Immunol* 37: 2419-2433.

224. Hillen, N., G. Mester, C. Lemmel, A. O. Weinzierl, M. Müller, D. Wernet, J. Hennenlotter, A. Stenzl, H.-G. Rammensee, and S. Stevanović. 2008. Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. *Eur. J. Immunol* 38: 2993-3003.

225. Vita, R., L. Zarebski, J. A. Greenbaum, H. Emami, I. Hoof, N. Salimi, R. Damle, A. Sette, and B. Peters. 2010. The immune epitope database 2.0. *Nucleic Acids Res* 38: D854-862.

226. Nielsen, M., C. Lundegaard, T. Blicher, K. Lamberth, M. Harndahl, S. Justesen, G. Roder, B. Peters, A. Sette, O. Lund, and S. Buus. 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.

227. Lundegaard, C., K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen. 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: W509-512.

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

229. Maiers, M., L. Gragert, and W. Klitz. 2007. High-resolution HLA alleles and haplotypes in the United States population. *Hum. Immunol* 68: 779-788.

230. Mori, M., P. G. Beatty, M. Graves, K. M. Boucher, and E. L. Milford. 1997. HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. *Transplantation* 64: 1017-1027.

231. Buus, S., S. L. Lauemoller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, and S. Brunak. 2003. Sensitive quantitative predictions of peptide-MHC binding by a "Query by Committee" artificial neural network approach. *Tissue Antigens* 62: 378–384.

232. Nielsen, M., C. Lundegaard, T. Blicher, B. Peters, A. Sette, S. Justesen, S. Buus, and O. Lund. 2008. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS. Comput. Biol.* 4: e1000107.

233. MacNamara, A., U. Kadolsky, C. R. Bangham, and B. Asquith. 2009. T-cell epitope prediction: rescaling can mask biological variation between MHC molecules. *PLoS. Comput. Biol.* 5: e1000327.

234. Tenzer, S., B. Peters, S. Bulik, O. Schoor, C. Lemmel, M. M. Schatz, P.-M. Kloetzel, H.-G. Rammensee, H. Schild, and H.-G. Holzhütter. 2005. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell. Mol. Life Sci* 62: 1025-1037.

235. Kesmir, C., A. K. Nussbaum, H. Schild, V. Detours, and S. Brunak. 2002. Prediction of proteasome cleavage motifs by neural networks. *Protein. Eng.* 15: 287–296.

236. Nielsen, M., C. Lundegaard, O. Lund, and C. Kesmir. 2005. The role of the proteasome in generating cytotoxic T cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics.* 57: 33-41.

237. Feltkamp, M. C., M. P. Vierboom, W. M. Kast, and C. J. Melief. 1994. Efficient MHC class

I-peptide binding is required but does not ensure MHC class I-restricted immunogenicity. *Mol. Immunol* 31: 1391-1401.
238. Sette, A., A. Vitiello, B. Reherman, P. Fowler, R. Nayersina, W. M. Kast, C. J. Melief, C. Oseroff, L. Yuan, J. Ruppert, J. Sidney, M. F. del Guercio, S. Southwood, R. T. Kubo, R. W. Chesnut, H. M. Grey, and F. V. Chisari. 1994. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol* 153: 5586-5592.
239. Fortier, M.-H., E. Caron, M.-P. Hardy, G. Voisin, S. Lemieux, C. Perreault, and P. Thibault. 2008. The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med* 205: 595-610.
240. Zhang, H., C. Lundegaard, and M. Nielsen. 2009. Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. *Bioinformatics* 25: 83-89.
241. Bihl, F., N. Frahm, L. Di Giammarino, J. Sidney, M. John, K. Yusim, T. Woodberry, K. Sango, H. S. Hewitt, L. Henry, C. H. Linde, J. V. Chisholm, 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.
242. Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger, C. Day, P. Klenerman, J. Mullins, 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.
243. Carrington, M., and S. J. O'Brien. 2003. The influence of HLA genotype on AIDS. *Annu. Rev. Med.* 54: 535–551.
244. Pereyra, F., X. Jia, P. J. McLaren, A. Telenti, P. I. W. de Bakker, B. D. Walker, S. Ripke, C. J. Brumme, S. L. Pulit, M. Carrington, C. M. Kadie, J. M. Carlson, D. Heckerman, R. R. Graham, R. M. Plenge, S. G. Deeks, L. Gianniny, G. Crawford, J. Sullivan, E. Gonzalez, L. Davies, A. Camargo, J. M. Moore, N. Beattie, S. Gupta, A. Crenshaw, N. P. Burtt, C. Guiducci, N. Gupta, X. Gao, Y. Qi, Y. Yuki, A. Piechocka-Trocha, E. Cutrell, R. Rosenberg, K. L. Moss, P. Lemay, J. O'Leary, T. Schaefer, P. Verma, I. Toth, B. Block, B. Baker, A. Rothchild, J. Lian, J. Proudfoot, D. M. L. Alvino, S. Vine, M. M. Addo, T. M. Allen, M. Altfeld, M. R. Henn, S. Le Gall, H. Streeck, D. W. Haas, D. R. Kuritzkes, G. K. Robbins, R. W. Shafer, R. M. Gulick, C. M. Shikuma, R. Haubrich, S. Riddler, P. E. Sax, E. S. Daar, H. J. Ribaudo, B. Agan, S. Agarwal, R. L. Ahern, B. L. Allen, S. Altidor, E. L. Altschuler, S. Ambardar, K. Anastos, B. Anderson, V. Anderson, U. Andrady, D. Antoniskis, D. Bangsberg, D. Barbaro, W. Barrie, J. Bartczak, S. Barton, P. Basden, N. Basgoz, S. Bazner, N. C. Bellos, A. M. Benson, J. Berger, N. F. Bernard, A. M. Bernard, C. Birch, S. J. Bodner, R. K. Bolan, E. T. Boudreaux, M. Bradley, J. F. Braun, J. E. Brndjar, S. J. Brown, K. Brown, S. T. Brown, J. Burack, L. M. Bush, V. Cafaro, O. Campbell, J. Campbell, R. H. Carlson, J. K. Carmichael, K. K. Casey, C. Cavacuiti, G. Celestin, S. T. Chambers, N. Chez, L. M. Chirch, P. J. Cimoch, D. Cohen, L. E. Cohn, B. Conway, D. A. Cooper, B. Cornelson, D. T. Cox, M. V. Cristofano, G. Cuchural, J. L. Czartoski, J. M. Dahman, J. S. Daly, B. T. Davis, K. Davis, S. M. Davod, E. DeJesus, C. A. Dietz, E. Dunham, M. E. Dunn, T. B. Ellerin, J. J. Eron, J. J. W. Fangman, C. E. Farel, H. Ferlazzo, S. Fidler, A. Fleenor-Ford, R. Frankel, K. A. Freedberg, N. K. French, J. D. Fuchs, J. D. Fuller, J. Gaberman, J. E. Gallant, R. T. Gandhi, E. Garcia, D. Garmon, J. C. Gathe, C. R. Gaultier, W. Gebre, F. D. Gilman, I. Gilson, P. A. Goepfert, M. S. Gottlieb, C. Goulston, R. K. Groger, T. D. Gurley, S. Haber, R. Hardwicke, W. D. Hardy, P. R. Harrigan, T. N. Hawkins, S. Heath, F. M. Hecht, W. K. Henry, M. Hladek, R. P. Hoffman, J. M. Horton, R. K. Hsu, G. D. Huhn, P. Hunt, M. J. Hupert, M. L. Illeman, H. Jaeger, R. M. Jellinger, M. John, J. A. Johnson, K. L. Johnson, H. Johnson, K. Johnson, J. Joly, W. C. Jordan, C. A. Kauffman, H. Khanlou, R. K. Killian, A. Y. Kim, D. D. Kim, C. A. Kinder, J. T. Kirchner, L. Kogelman, E. M. Kojic, P. T. Korthuis, W. Kurisu, D. S. Kwon, M. LaMar, H. Lampiris, M. Lanzafame, M. M. Lederman, D. M. Lee, J. M. L. Lee, M. J. Lee, E. T. Y. Lee, J. Lemoine, J. A. Levy, J. M. Llibre, M. A. Liguori, S. J. Little, A. Y. Liu, A. J. Lopez, M. R. Loutfy, D. Loy, D. Y. Mohammed, A. Man, M. K. Mansour, V. C. Marconi, M. Markowitz, R. Marques, J. N. Martin, H. L. Martin, K. H. Mayer, M. J. McElrath, T. A. McGhee, B. H. McGovern, K. McGowan, D. McIntyre, G. X. Mcleod, P. Menezes, G. Mesa, C. E. Metroka, D. Meyer-Olson, A. O. Miller, K. Montgomery, K. C. Mounzer, E. H. Nagami, I. Nagin, R. G. Nahass, M. O. Nelson, C. Nielsen, D. L. Norene, D. H. O'Connor, B. O. Ojikutu, J. Okulicz, O. O. Oladehin, E. C. Oldfield, S. A. Olender, M. Ostrowski, W. F. Owen, E. Pae, J. Parsonnet, A. M. Pavlatos, A. M. Perlmutter, M. N. Pierce, J. M. Pincus, L. Pisani, L. J. Price, L. Proia, R. C. Prokesch, H. C. Pujet, M. Ramgopal, A. Rathod, M. Rausch, J.

Ravishankar, F. S. Rhame, C. S. Richards, D. D. Richman, B. Rodes, M. Rodriguez, R. C. Rose, E. S. Rosenberg, D. Rosenthal, P. E. Ross, D. S. Rubin, E. Rumbaugh, L. Saenz, M. R. Salvaggio, W. C. Sanchez, V. M. Sanjana, S. Santiago, W. Schmidt, H. Schuitemaker, P. M. Sestak, P. Shalit, W. Shay, V. N. Shirvani, V. I. Silebi, J. M. Sizemore, P. R. Skolnik, M. Sokol-Anderson, J. M. Sosman, P. Stabile, J. T. Stapleton, S. Starrett, F. Stein, H.-J. Stellbrink, F. L. Sterman, V. E. Stone, D. R. Stone, G. Tambussi, R. A. Taplitz, E. M. Tedaldi, A. Telenti, W. Theisen, R. Torres, L. Tosiello, C. Tremblay, M. A. Tribble, P. D. Trinh, A. Tsao, P. Ueda, A. Vaccaro, E. Valadas, T. J. Vanig, I. Vecino, V. M. Vega, W. Veikley, B. H. Wade, C. Walworth, C. Wanidworanun, D. J. Ward, D. A. Warner, R. D. Weber, D. Webster, S. Weis, D. A. Wheeler, D. J. White, E. Wilkins, A. Winston, C. G. Wlodaver, A. van't Wout, D. P. Wright, O. O. Yang, D. L. Yurdin, B. W. Zabukovic, K. C. Zachary, B. Zeeman, and M. Zhao. 2010. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330: 1551-1557.

245. Borghans, J. A. M., J. B. Beltman, and R. J. De Boer. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics.* 55: 732–739.

246. Jeffery, K. J., A. A. Siddiqui, M. Bunce, A. L. Lloyd, A. M. Vine, A. D. Witkover, S. Izumo, K. Usuku, K. I. Welsh, M. Osame, and C. R. Bangham. 2000. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. *J. Immunol* 165: 7278-7284.

247. Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger, C. Day, P. Klenerman, J. Mullins, B. Korber, H. M. Coovadia, B. D. Walker, and P. J. R. Goulder. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432: 769-775.

248. Fellay, J., D. Ge, K. V. Shianna, S. Colombo, B. Ledergerber, E. T. Cirulli, T. J. Urban, K. Zhang, C. E. Gumbs, J. P. Smith, A. Castagna, A. Cozzi-Lepri, A. De Luca, P. Easterbrook, H. F. Günthard, S. Mallal, C. Mussini, J. Dalmau, J. Martinez-Picado, J. M. Miro, N. Obel, S. M. Wolinsky, J. J. Martinson, R. Detels, J. B. Margolick, L. P. Jacobson, P. Descombes, S. E. Antonarakis, J. S. Beckmann, S. J. O'Brien, N. L. Letvin, A. J. McMichael, B. F. Haynes, M. Carrington, S. Feng, A. Telenti, and D. B. Goldstein. 2009. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* 5: e1000791.

249. Gao, X., G. W. Nelson, P. Karacki, M. P. Martin, J. Phair, R. Kaslow, J. J. Goedert, S. Buchbinder, K. Hoots, D. Vlahov, S. J. O'Brien, and M. Carrington. 2001. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N. Engl. J. Med.* 344: 1668–1675.

250. Axelsson-Robertson, R., F. Weichold, D. Sizemore, M. Wulf, Y. A. W. Skeiky, J. Sadoff, and M. J. Maeurer. 2010. 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* 129: 496-505.

251. Nakagawa, M., K. H. Kim, T. M. Gillam, and A.-B. Moscicki. 2007. HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein. *J. Virol* 81: 1412-1423.

252. Yewdell, J. W., E. Reits, and J. Neefjes. 2003. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* 3: 952–961.

253. Sidney, J., S. Southwood, M. F. del Guercio, H. M. Grey, R. W. Chesnut, R. T. Kubo, and A. Sette. 1996. Specificity and degeneracy in peptide binding to HLA-B7-like class I molecules. *J. Immunol* 157: 3480-3490.

254. Sidney, J., S. Southwood, D. L. Mann, M. A. Fernandez-Vina, M. J. Newman, and A. Sette. 2001. Majority of peptides binding HLA-A*0201 with high affinity crossreact with other A2-supertype molecules. *Hum. Immunol* 62: 1200-1216.

255. Sette, A., J. Sidney, B. D. Livingston, J. L. Dzuris, C. Crimi, C. M. Walker, S. Southwood, E. J. Collins, and A. L. Hughes. 2003. Class I molecules with similar peptide-binding specificities are the result of both common ancestry and convergent evolution. *Immunogenetics.* 54: 830–841.

256. Dean, M., M. Carrington, and S. J. O'Brien. 2002. Balanced polymorphism selected by genetic versus infectious human disease. *Annu Rev Genomics Hum Genet* 3: 263-292.

257. 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.

258. 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.
259. Apanius, V., D. Penn, P. R. Slev, L. R. Ruff, and W. K. Potts. 1997. The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* 17: 179-224.
260. De Boer, R. J., J. A. M. Borghans, M. van Boven, C. Keşmir, and F. J. Weissing. 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* 55: 725-731.
261. O'Sullivan, D., T. Arrhenius, J. Sidney, M. F. Del Guercio, M. Albertson, M. Wall, C. Oseroff, S. Southwood, S. M. Colón, and F. C. Gaeta. 1991. On the interaction of promiscuous antigenic peptides with different DR alleles. Identification of common structural motifs. *J. Immunol.* 147: 2663-2669.
262. Axelsson-Robertson, R., F. Weichold, D. Sizemore, M. Wulf, Y. A. W. Skeiky, J. Sadoff, and M. J. Maeurer. 2010. 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* 129: 496-505.
263. Frahm, N., K. Yusim, T. J. Suscovich, S. Adams, J. Sidney, P. Hraber, H. S. Hewitt, C. H. Linde, D. G. Kavanagh, T. Woodberry, L. M. Henry, K. Faircloth, J. Listgarten, C. Kadie, N. Jojic, K. Sango, N. V. Brown, E. Pae, M. T. Zaman, F. Bihl, A. Khatri, M. John, S. Mallal, F. M. Marincola, B. D. Walker, A. Sette, D. Heckerman, B. T. Korber, and C. Brander. 2007. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur. J. Immunol.* 37: 2419-2433.
264. Nakagawa, M., K. H. Kim, T. M. Gillam, and A.-B. Moscicki. 2007. HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein. *J. Virol.* 81: 1412-1423.
265. Rao, X., I. Hoof, A. I. C. A. Fontaine Costa, D. van Baarle, and C. Keşmir. 2011. HLA class I allele promiscuity revisited. *Immunogenetics* .
266. Carrington, M. 1999. Recombination within the human MHC. *Immunol. Rev.* 167: 245-256.
267. Begovich, A. B., G. R. McClure, V. C. Suraj, R. C. Helmuth, N. Fildes, T. L. Bugawan, H. A. Erlich, and W. Klitz. 1992. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J. Immunol.* 148: 249-258.
268. Smith, W. P., Q. Vu, S. S. Li, J. A. Hansen, L. P. Zhao, and D. E. Geraghty. 2006. Toward understanding MHC disease associations: partial resequencing of 46 distinct HLA haplotypes. *Genomics* 87: 561-571.
269. Madeleine, M. M., L. G. Johnson, A. G. Smith, J. A. Hansen, B. B. Nisperos, S. Li, L.-P. Zhao, J. R. Daling, S. M. Schwartz, and D. A. Galloway. 2008. Comprehensive analysis of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 loci and squamous cell cervical cancer risk. *Cancer Res.* 68: 3532-3539.
270. Albanidou-Farmaki, E., A. Deligiannidis, A. K. Markopoulos, V. Katsares, K. Farmakis, and E. Parapanissiou. 2008. HLA haplotypes in recurrent aphthous stomatitis: a mode of inheritance? *Int. J. Immunogenet.* 35: 427-432.
271. Kollman, C., M. Maiers, L. Gragert, C. Müller, M. Setterholm, M. Oudshoorn, and C. K. Hurley. 2007. Estimation of HLA-A, -B, -DRB1 haplotype frequencies using mixed resolution data from a National Registry with selective retyping of volunteers. *Hum. Immunol.* 68: 950-958.
272. Long, J. C., R. C. Williams, and M. Urbanek. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am. J. Hum. Genet.* 56: 799-810.
273. Excoffier, L., and M. Slatkin. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12: 921-927.
274. MacNamara, A., U. Kadolsky, C. R. M. Bangham, and B. Asquith. 2009. T-cell epitope prediction: rescaling can mask biological variation between MHC molecules. *PLoS Comput. Biol.* 5: e1000327.
275. Buus, S., S. L. Lauem\oller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, and S. Brunak. 2003. Sensitive quantitative predictions of peptide-MHC binding by a "Query by Committee"artificial neural network approach. *Tissue Antigens* 62: 378–384.
276. Nielsen, M., C. Lundegaard, P. Worning, S. L. Lauem\oller, K. Lamberth, S. Buus, S. Brunak, and O. Lund. 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science* 12: 1007–1017.
277. Crawford, D. C., and D. A. Nickerson. 2005. Definition and clinical importance of haplotypes.

Annu. Rev. Med. **56: 303-320.**
278. Yan, H., N. Papadopoulos, G. Marra, C. Perrera, J. Jiricny, C. R. Boland, H. T. Lynch, R. B. Chadwick, A. de la Chapelle, K. Berg, J. R. Eshleman, W. Yuan, S. Markowitz, S. J. Laken, C. Lengauer, K. W. Kinzler, and B. Vogelstein. 2000. Conversion of diploidy to haploidy. *Nature* 403: 723-724.
279. Douglas, J. A., M. Boehnke, E. Gillanders, J. M. Trent, and S. B. Gruber. 2001. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat. Genet.* 28: 361-364.
280. Niu, T. 2004. Algorithms for inferring haplotypes. *Genet. Epidemiol.* 27: 334-347.
281. Bettencourt, B. F., M. R. Santos, R. N. Fialho, A. R. Couto, M. J. Peixoto, J. P. Pinheiro, H. Spínola, M. G. Mora, C. Santos, A. Brehm, and J. Bruges-Armas. 2008. Evaluation of two methods for computational HLA haplotypes inference using a real dataset. *BMC Bioinformatics* 9: 68.
282. Castelli, E. C., C. T. Mendes-Junior, L. C. Veiga-Castelli, N. F. Pereira, M. L. Petzl-Erler, and E. A. Donadi. 2010. Evaluation of computational methods for the reconstruction of HLA haplotypes. *Tissue Antigens* 76: 459-466.
283. Maiers, M., L. Gragert, and W. Klitz. 2007. High-resolution HLA alleles and haplotypes in the United States population. *Hum. Immunol.* 68: 779-788.
284. Middleton, D., L. Menchaca, H. Rood, and R. Komerofsky. 2003. New allele frequency database: http://www.allelefrequencies.net. *Tissue Antigens* 61: 403-407.
285. Nielsen, M., C. Lundegaard, T. Blicher, K. Lamberth, M. Harndahl, S. Justesen, G. R\oder, B. Peters, A. Sette, O. Lund, and others. 2007. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and-B locus protein of known sequence. *PLoS One* 2: 796.
286. Capittini, C., A. Pasi, P. Bergamaschi, C. Tinelli, A. De Silvestri, M. P. Mercati, C. Badulli, F. Garlaschelli, I. Sbarsi, M. Guarene, M. Martinetti, L. Salvaneschi, and M. Cuccia. 2009. HLA haplotypes and birth weight variation: is your future going to be light or heavy? *Tissue Antigens* 74: 156-163.
287. Lund, O. 2005. *Immunological bioinformatics*,. MIT Press.
288. Kesmir, C., V. van Noort, R. J. de Boer, and P. Hogeweg. 2003. Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome. *Immunogenetics* 55: 437-449.
289. Nielsen, M., C. Lundegaard, O. Lund, and C. Keşmir. 2005. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57: 33-41.
290. Peters, B., S. Bulik, R. Tampe, P. M. Van Endert, and H.-G. Holzhütter. 2003. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol.* 171: 1741-1749.
291. Shen, X. Z., S. Billet, C. Lin, D. Okwan-Duodu, X. Chen, A. E. Lukacher, and K. E. Bernstein. 2011. The carboxypeptidase ACE shapes the MHC class I peptide repertoire. *Nat. Immunol.* 12: 1078-1085.
292. Calis, J. J. A., G. F. Sanchez-Perez, and C. Keşmir. 2010. MHC class I molecules exploit the low G+C content of pathogen genomes for enhanced presentation. *Eur. J. Immunol.* 40: 2699-2709.
293. Rocha, E. P. C., and A. Danchin. 2002. Base composition bias might result from competition for metabolic resources. *Trends Genet.* 18: 291-294.
294. Bentley, S. D., and J. Parkhill. 2004. Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* 38: 771-792.
295. Johnson, D. R. 2003. Locus-specific constitutive and cytokine-induced HLA class I gene expression. *J. Immunol.* 170: 1894-1902.
296. Johnson, D. R. 2000. Differential expression of human major histocompatibility class I loci: HLA-A, -B, and -C. *Hum. Immunol.* 61: 389-396.
297. Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker, and J. B. Hogenesch. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* 101: 6062-6067.
298. Greene, J. M., R. W. Wiseman, S. M. Lank, B. N. Bimber, J. A. Karl, B. J. Burwitz, J. J. Lhost, O. E. Hawkins, K. J. Kunstman, K. W. Broman, S. M. Wolinsky, W. H. Hildebrand, and D. H. O'Connor. 2011. Differential MHC class I expression in distinct leukocyte subsets. *BMC Immunol.* 12: 39.
299. Crotzer, V. L., R. E. Christian, J. M. Brooks, J. Shabanowitz, R. E. Settlage, J. A. Marto, F. M.

White, A. B. Rickinson, D. F. Hunt, and V. H. Engelhard. 2000. Immunodominance among EBV-derived epitopes restricted by HLA-B27 does not correlate with epitope abundance in EBV-transformed B-lymphoblastoid cell lines. *J. Immunol.* 164: 6120-6129.

300. Jansen, C. A., S. Kostense, K. Vandenberghe, N. M. Nanlohy, I. M. De Cuyper, E. Piriou, E. H. Manting, F. Miedema, and D. van Baarle. 2005. High responsiveness of HLA-B57-restricted Gag-specific CD8+ T cells in vitro may contribute to the protective effect of HLA-B57 in HIV-infection. *Eur. J. Immunol.* 35: 150-158.

301. Berger, C. T., N. Frahm, D. A. Price, B. Mothe, M. Ghebremichael, K. L. Hartman, L. M. Henry, J. M. Brenchley, L. E. Ruff, V. Venturi, F. Pereyra, J. Sidney, A. Sette, D. C. Douek, B. D. Walker, D. E. Kaufmann, and C. Brander. 2011. High-functional-avidity cytotoxic T lymphocyte responses to HLA-B-restricted Gag-derived epitopes associated with relative HIV control. *J. Virol.* 85: 9334-9345.

302. Fontaine Costa, A. I., X. Rao, E. Lechenadec, D. van Baarle, and C. Keşmir. 2010. HLA-B molecules target more conserved regions of the HIV-1 proteome. *AIDS* 24: 211-215.

303. Khakoo, S. I., C. L. Thio, M. P. Martin, C. R. Brooks, X. Gao, J. Astemborski, J. Cheng, J. J. Goedert, D. Vlahov, M. Hilgartner, S. Cox, A.-M. Little, G. J. Alexander, M. E. Cramp, S. J. O'Brien, W. M. C. Rosenberg, D. L. Thomas, and M. Carrington. 2004. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 305: 872-874.

304. Qi, Y., M. P. Martin, X. Gao, L. Jacobson, J. J. Goedert, S. Buchbinder, G. D. Kirk, S. J. O'Brien, J. Trowsdale, and M. Carrington. 2006. KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. *PLoS Pathog.* 2: e79.

305. Alter, G., M. P. Martin, N. Teigen, W. H. Carr, T. J. Suscovich, A. Schneidewind, H. Streeck, M. Waring, A. Meier, C. Brander, J. D. Lifson, T. M. Allen, M. Carrington, and M. Altfeld. 2007. Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *J. Exp. Med.* 204: 3027-3036.

306. López-Vázquez, A., A. Miña-Blanco, J. Martínez-Borra, P. D. Njobvu, B. Suárez-Alvarez, M. A. Blanco-Gelaz, S. González, L. Rodrigo, and C. López-Larrea. 2005. Interaction between KIR3DL1 and HLA-B*57 supertype alleles influences the progression of HIV-1 infection in a Zambian population. *Hum. Immunol.* 66: 285-289.

307. Carrington, M., S. Wang, M. P. Martin, X. Gao, M. Schiffman, J. Cheng, R. Herrero, A. C. Rodriguez, R. Kurman, R. Mortel, P. Schwartz, A. Glass, and A. Hildesheim. 2005. Hierarchy of resistance to cervical neoplasia mediated by combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci. *J. Exp. Med.* 201: 1069-1075.

Summary

Human Leukocyte Antigen (HLA) class I is a group of genes located on human chromosome 6 which play a crucial role in initiating potentially protective immune responses, by presenting pathogen-derived peptides to CD8+ T cells and thus targeting infected cells for elimination. Compare to other HLA class I loci with high functional similarity, HLA-B molecules show some specific properties: 1) HLA-B alleles have been associated with opposing disease outcomes, namely susceptibility to, or protection from, parasitic and viral infection. 2) HLA-B elicits more dominant T cell responses with higher frequencies or superior magnitude.

In order to discover potential mechanisms which may explain why HLA-B restricted T cell responses tend to be immunodominant, the focus of my PHD project has been on an extensive comparison of HLA-A and HLA-B antigen presentation by analysing their ligands. Experimental verified epitopes were downloaded, filtered and analyzed from several public databases, e.g. Immune Epitope Database and Analysis Resource (IEDB; www.immuneepitope.org) and Los Alamos Immunology database (LANL). However, unavoidable biases remain present in such data sets. For example, certain dominant HLA alleles like A*02 are more frequently studied than others, which potentially increases its epitope repertoire in the public databases, while several other rare HLA class I molecules may have insufficient data for analysis. Moreover, most experimental epitopes come from extensive studied human viruses like HIV-1, and are hence not necessarily appropriate for drawing general conclusion on the role of HLA-B restricted responses in infectious diseases. To eliminate these biases, I also selected and applied in silico predictions based on different research questions for main steps of antigen presentation pathway : MHC binding (SMM, NetMHC, NetMHCpan), TAP binding probability (SMM) and proteasome cleavage (SMM or NetChop), as a compensatory of potential bias made by experimental data.

Since the differences in CD8+ T cell responses mediated via HLA-A and HLA-B molecules might already begin at the antigen presentation level, in chapter 2 we have analyzed the epitope diversity and epitope binding affinity of this two major HLA class I loci. Opposite to our expectation, the amount of epitopes presented by HLA-A molecules is significantly more than that of their HLA-B counterparts. Furthermore, binding affinities of HLA-A epitopes are significantly higher than that of HLA-B epitopes based on both experimental and predicted data. Interestingly, we also found a significantly positive correlation between binding affinity and diversity of HLA ligands, which points to that HLA-A molecules have broader binding motifs, thus present more epitopes, making them have higher chance to present the epitopes with higher binding affinity.

Alternatively, targeting more constrained regions of a pathogen genome might be favourable for the outcome of an infection. Thus in chapter 3 we compared CTL epitopes restricted by HLA-A and HLA-B molecules and found that HLA-B molecules prefer to present more epitopes from certain conserved proteins. Moreover the residues targeted by HLA-B alleles in HIV-proteome were significantly more

conserved than the ones targeted by HLA-A alleles. After analyzing the data published by Wang et al, we concluded that HLA-B-targeted regions are more conserved because they are functionally and/or structurally constrained but not lack of selection pressure. In order to explore if similar mechanism occurs in other infectious disease, like HCV infection which shares certain protective HLA-B molecules (B*27 and B*57) with HIV-1 infection, in chapter 4 a similar analysis was applied on HCV proteome by evaluating the conservation level of different HCV proteins and analyzing the distribution of CTL epitopes restricted by protective or detrimental HLA alleles.. As a result, we found that HLA molecules associated with HCV clearance preferentially present epitopes from the conserved proteins, like NS5B and Core.

Unexpectedly, it was recently shown that HLA-peptide binding is highly promiscuous , i.e., a large fraction of epitopes binds at least two HLA molecules. Our results suggest that less promiscuous peptide presentation might contribute to viral control during HIV-1 infection (chapter 5). In order to test whether the dominant role of HLA-B restricted T cell responses is due to promiscuous binding of HLA-B restricted epitopes, we compared their binding promiscuity to that of HLA-A restricted epitopes in chapter 5. To this end, we studied the HLA ligand promiscuity based on both experimental epitopes from IEDB and *in silico* predicted epitopes. In line with the high HLA ligand promiscuity observed by Frahm et al, we found that above 60% of all HLA class I ligands are promiscuous, but we failed to find any consistent difference in the peptide binding promiscuity of HLA-A and HLA–B molecules During this analysis, we found that there is also a considerable promiscuity between the two loci: significant fraction of HLA-A ligands can be presented by at least one HLA-B molecule in the population. To test whether this across loci promiscuity has an effect on the frequencies of HLA class I haplotypes, we estimated the total ligand repertoires for all HLA class I haplotypes (HLA-A-B) by *in silico* prediction and found that the most common HLA-A-B haplotypes are enriched in HLA-A and HLA–B pairs with distinct peptide binding motifs (chapter 6).

In summary, we identified hardly any characteristics of antigen presentation by HLA-B molecules that were significantly “better” than HLA-A molecules. However, we found that the targeting more conserved regions in the pathogen proteomes, as we have shown to be the case for HIV and HCV, is one property of HLA-B antigen presentation as a crucial factor for controlling the viral replication during an infectious disease. This research implies that factors other than antigen presentation might be more essential for the protective and immunodominant nature of HLA-B-restricted T cell responses. It will also give more support information for identifying interesting epitopes which could be good candidates for vaccine design.

Nederlandse samenvatting

Elke cel in ons lichaam heeft speciale eiwitten op zijn oppervlak die een belangrijke rol spelen in immuunreacties tegen virussen en bacteriën die zich in de cel verborgen houden. In immunologisch jargon heten dit “HLA klasse I” moleculen. HLA moleculen staan ook bekend als de transplantatie-eiwitten die orgaandonaties zo moeilijk maken omdat verschillende individuen vrijwel altijd verschillende HLA moleculen hebben. HLA moleculen nemen kleine selectieve steekproeven van alle eiwitten in de cel en brengen kleine eiwitfragmenten naar het oppervlak van de cel alwaar deze aan de T cellen van het immuunsysteem gepresenteerd worden. T cellen die eiwitfragmenten herkennen en geactiveerd worden, zullen alle cellen die dergelijke eiwitfragmenten presenteren doden. Zo worden geïnfecteerde cellen opgeruimd zonder dat het virus zelf zichtbaar is voor het immuunsysteem. We bestuderen twee HLA klasse I moleculen , de HLA-A en HLA-B eiwitten, die ieder door hun eigen gen op chromosoom 6 gecodeerd worden. Voor zover we weten hebben HLA-A en HLA-B eiwitten precies dezelfde functie, namelijk het presenteren van eiwitfragmenten vanuit de cel aan het immuunsysteem. Het lijkt er alleen op dat HLA-B moleculen een grotere rol spelen in de immuniteit dan HLA-A moleculen: (1) HLA-B molecules zijn beter geassocieerd met het verloop van een aantal infectieziekten, en (2) virale eiwitfragmenten gepresenteerd op HLA-B moleculen roepen de sterkste immuunreacties op.

In dit proefschrift onderzoeken we de mechanismen die dit functionele verschil tussen HLA-A en HLA-B moleculen kunnen verklaren. We vergelijken de eiwitfragmenten die op HLA-A of HLA-B moleculen gepresenteerd worden. Experimenteel onderzoek van de laatste decennia heeft duizenden van dat soort eiwitfragmenten beschreven en dit werk wordt verzameld in enorme databanken op internet (bijvoorbeeld de IEDB: www.immuneepitope.org en de Los Alamos Immunology database) die vrij toegankelijk zijn. Omdat niet alle HLA moleculen en virussen even intensief onderzocht worden, staan er in de databanken veel gegevens over interessante en bekende westerse virussen en HLA varianten, en is er maar weinig te vinden over vele andere varianten. Aangezien de experimentele gegevens niet evenwichtig verdeeld zijn, maken we in dit proefschrift ook gebruik van theoretische modellen waarmee we kunnen voorspellen welke eiwitfragmenten kunnen binden aan welke HLA moleculen. Met behulp van deze zogenaamde “HLA binding prediction tools” (zoals NetMHC en NetMHCpan) kunnen we voor enorme aantallen virussen en bacteriën voorspellen of ze gepresenteerd kunnen worden op grote aantallen varianten van zowel de HLA-A als de HLA-B moleculen. Zo genereren we onze eigen evenwichtige data sets waarin we nauwgezet de functionele verschillen tussen HLA-A en HLA-B moleculen kunnen onderzoeken.

In hoofdstuk 2 vergelijken we de diversiteit van de eiwitfragmenten die door verschillende HLA-A en HLA-B moleculen gepresenteerd kunnen worden, en de affiniteit waarmee ze binden. Hoewel het algemeen bekend is dat HLA-B beter geassocieerd is met immuniteit, blijken de HLA-A moleculen in alle opzichten beter. Ze presenteren significant meer fragmenten en met een hogere affiniteit dan de HLA-B moleculen en dit geldt zowel voor de gepubliceerde fragmenten uit de databanken als voor de voorspelde fragmenten. HLA-A moleculen die veel fragmenten presenteren binden die fragmenten ook met een hoge affiniteit. Het lijkt er dus op dat HLA-A moleculen een breder bindingsmotief gebruiken dan HLA-B

moleculen.

In hoofdstuk 3 hebben we onderzocht of HLA-A en HLA-B moleculen verschillende voorkeuren hebben voor verschillende soorten eiwitten. Fragmenten uit cruciale eiwitten van een virus geven een betere immuunreactie dan fragmenten uit minder belangrijke eiwitten omdat het virus de onbelangrijke eiwitten makkelijk kan muteren en zo kan ontsnappen aan de immuunreactie. We laten zien dat HLA-B moleculen een voorkeur hebben voor fragmenten uit geconserveerde eiwitten. In infecties met HIV en HCV hebben bepaalde HLA-B moleculen, B*27 en B*57, een duidelijk beschermend effect. Voor B*27 en B*57 maken we in hoofdstuk 4 een vergelijking van de fragmenten die ze presenteren, en zien we dat ze een voorkeur hebben voor de sterk geconserveerde HCV eiwitten NS5B en Core.

Verschillende HLA moleculen kunnen dezelfde eiwitfragmenten binden en in hoofdstuk 5 onderzoeken we de overlap tussen eiwitfragmenten die door verschillende HLA moleculen gepresenteerd worden. Onze resultaten suggereren dat HLA moleculen die unieke eiwitfragmenten binden geassocieerd zijn met een betere controle van HIV infecties (hoofdstuk 5). We hebben daarom onderzocht of de eiwitfragmenten die aan HLA-A of aan HLA-B moleculen binden meer of minder exclusief zijn. In overeenstemming met recente gegevens (Frahm, et al 2007) vinden we dat meer dan 60% van de fragmenten niet exclusief is en dat ze binden aan meerdere HLA moleculen. Zowel in eiwitfragmenten uit de databanken, als in de eiwitfragmenten die we voorspellen, zien we geen verschil tussen HLA-A en HLA-B. Bovendien zien we dat er zelfs een overlap is tussen HLA-A en HLA-B fragmenten : een aanzienlijke fractie van de HLA-A eiwitfragmenten kan binden aan minstens één HLA-B molecuul, en andersom.

In hoofdstuk 6 onderzoeken we of de HLA-A/HLA-B combinaties (de zogenaamde haplotypen) die in verschillende menselijke populaties voorkomen een kleinere overlap hebben dan combinaties die niet voorkomen. Hier zou op geselecteerd kunnen zijn omdat haplotypen met een kleine overlapmeer eiwitfragmenten kunnen presenteren. Inderdaad vinden we met onze voorspellers dat de meest voorkomende haplotypen verrijkt zijn in HLA-A/HLA-B combinaties met een kleine overlap (hoofdstuk 6).

Samenvattend, de “superioriteit” van HLA-B ten opzichte van HLA-A moleculen komt niet naar voren in alle eigenschappen die we bestudeerd hebben. HLA-A moleculen presenteren meer eiwitfragmenten en ze doen dat met een hogere affiniteit. HLA-B moleculen hebben echter een voorkeur voor fragmenten uit geconserveerde eiwitten. Uit evolutionair oogpunt kan dat een bijzonder cruciale eigenschap zijn, omdat veel virussen snel muteren en zo kunnen ontsnappen aan immuunreacties. Dit suggereert dat het verschil tussen HLA-A en HLA-B moleculen niet hangt op het *aantal* eiwitfragmenten dat ze presenteren, maar op *welke* eiwitfragmenten ze presenteren. Als sommige eiwitfragmenten een betere bescherming bieden, zijn ze ook bijzonder interessante kandidaten voor de ontwikkeling van nieuwe vaccins.

Acknowledgements

When I arrived in the Netherlands in 2004, I have never expected that I would spend seven years here including five years in Utrecht, a pretty town with long history and a well known university of the world – Utrecht University. In August 2006, I had an open discussion with Can Keşmir in her old house for a PHD position in the group Theoretical Biology and Bioinformatics. I was impressed by the topic "Analysis of HLA by *in silico* prediction", also by Can, a patient, ambitious and smart female scientist. After following two courses and a small project , luckily I became her PHD student with major Immunology Bioinformatics. Can, I really appreciate your supervision during my PHD period. You told me how to think as a rigorous researcher. Furthermore, because of my background and language limitation, you spent much more time to discuss with me than others. Even when you just had your second baby Mina, you still spent time for project discussion in your place. I was surprised that you could keep such good balance between two characters: a successful researcher and a great mother of two kids. I also appreciated that you always directly tell me which part I should improved during our discussion and arguments, and I learn a lot during this process. This experience will become my valuable asset in my future career.

Rob de boer, as the promoter of my PHD, you helped me a lot as well during previous four years. It was my first time to know mathematical modeling when I follow your course Theoretical Ecology. I really appreciate your suggestions for my project when I stuck at certain stage. In addition, I would like to say it is a good memory to have dinner in your pretty garden during our group trip. Also thank you for helping me to translate my PHD book summary into Dutch.

Paulien Hogeweg, I would like to say that your course "Bioinformatics process" is one of the most difficult courses I have followed, but it is excellent! As other students of this course, I felt that it changed my way of thinking. Thanks for introducing it to me! I can not forget the Christmas group dinner in your place, and I strongly felt that our group is a big family at that moment.

I would like to thanks all group members of Immunology Bioinformatics. I really appreciate the discussion with you and the help from you. Jorg J.A. Calis, thanks for modifying my Dutch summary and being one of my paranymphs during my PHD defence. Ilka Hoof, thanks for paper editing and submission. Your idea is also very helpful for my project. Hanneke van Deutekom, I like the stamppot you made for our group dinner. All members involved in my PHD project, thank you! Ana.F.Costa, especially you, thanks for discussion and cooperation during my PHD period, also your help of writing the paper. Your solid immunology background is really helpful to me during this process. Debbie van Baarle, Ronald Bontrop and Ingrid Schellens, thanks for your ideas and suggestions for my PHD project. Also I want to thank others who shared the office with me: Boris Schmid, Christian Althaus, Tendai Mugwagwa, Saikrishna Gadhamsetty, Paola Carrillo, Ioana Niculescu. Folkert de Boer and Thomas Cuypers, thanks for the assisting of Paulien's course. In addition, it was a nice experience to do the bird observation with you two. Folkert, thanks for driving the car for me to move several times, without your help I could not get back my scooter from police station. Thomas, thanks for arranging the cost of the mini-symposium. Please say thanks to your wife who would like to do the text editing for my PHD book. Henk-Jan van

den Ham and john John van Dam, thanks for the idea about the procedure of the graduation. All other members in group Theoretical Biology and Bioinformatics, thank you for sharing nice time with me.

Prof Bontrop, Prof J.J Neefjes, Pof E.J.H.J. Wiertz, Dr M.Nielsen, and Prof C. Brander, thank you for carefully reading and confirming my PHD book.

All my dear friends in Wageningen: Hong Luo, Jing Shi, Yang Jiang, Mengxin Xie, Menghan Liu, TianTian Zhang, Xiucheng Fan, Xiangyu Song, Ke Lin, Ningwen Zhang, thank you! I had a happy Msc period in Wageningen because of you! Jin Wu, Xinhan Luo, Yanchao Liu, Jia Yao, Ning Xu, Shuang han, Jiansong Wan, Sijia Wang, Kejia Ruan, yun Zhu and all other friends in Utrecht, I will never forget the nice time staying with you in Utrecht. Also all my friends from ACSSNL: Ziqian Mao, Miao Yu, Chen Li, Yuan Li, Fangbin Liu, JiaYang Li, Peng Yang, Fei Xiang. It was my honor to work with you in ACSSNL for two years. In addition, Mr Luo ping, Mr Zhang xiaodong, Mr Xialei, thank you for your supporting from Chinese Embassy in Netherlands, Education section.

I also want to thank Remco Ursem and Rob Dirks for choosing me to work in Rijk Zwaan. I am quite happy with what I am doing now.

I really want to thanks my parents, my mother Ying Xu and my father Qingshui Rao, without your support I could not stay in The Netherlands for such long time and finish my study. 爸爸妈妈，谢谢你们的支持，你们辛苦了！

Finally, I want to thank Qianqian Jiu who came into my life during the last year of my PHD as a gift that make me peaceful and pleasure during this stressful time.

Curriculum vitae

The author of this thesis, Xiangyu Rao, was born on May 21th, 1980 in Hangzhou, China. After graduating from Hangzhou number 2 middle school in 1999, he started his BSc in Biotechnology at ZheJiang University and got his bachelor diploma in 2003. After one year working as a reporter in Lotus international commercial TV station and half year working in ACON biotechnology as marketing specialist in China, Xiangyu Rao moved to Wageningen University (The Netherlands) to start his MSc with major Bioinformatics in 2004. During this period he was supervised by Prof.Martien Groenen in department Animal breeding and genetics for his first MSc thesis and finished his second thesis in department Bioinformatics under the supervision from Prof Jack Leunissen and researcher Harm Nijveen. In February 2007, he started his PHD project studying the HLA and its interaction with certain pathogens in the Immunology Bioinformatics group of Utrecht University, supervised by Dr Can Keşmir and Prof Rob J. de Boer. The results of his thesis are described in this book. Currently, the author is working as a bioinformatics researcher at Rijk Zwaan in the Netherlands.

List of Publications

1. **Andreas Untergasser, Harm Nijveen, Xiangyu Rao, Ton Bisseling, René Geurts, and Jack A.M. Leunissen.** 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35: W71-W74;

2. **Rao, X., A. I. C. A. F. Costa, D. van Baarle, and C. Kesmir**. 2009. A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant CD8+ T cell responses. *J. Immunol.* 182: 1526-1532.

3. **Fontaine Costa, A#. I., X. Rao#, E. Lechenadec, D. van Baarle, and C. Keşmir**. 2010. HLA-B molecules target more conserved regions of the HIV-1 proteome. *AIDS* 24: 211-215.
contribute equally to this work

4. **Rao, X., I. Hoof, A. I. C. A. Fontaine Costa, D. van Baarle, and C. Keşmir.** 2011. HLA class I allele promiscuity revisited. *Immunogenetics* .

5. **Rao, X., I. Hoof, D. van Baarle, and C. Keşmir.** Protective HLA molecules Determine Infection Outcome in Hepatitis C virus Infection by preferential Presentation of Peptides From Conserved Viral Proteins. *Manuscript Submitted to Hepatology* .

6. **Rao, X., D. van Baarle, Maiers, M and C. Keşmir.** Do the binding motifs of HLA molecules influence the haplotype frequencies? *Manuscript in preparation.*