Published in final edited form as:

Immunogenetics. 2011 June ; 63(6): 325–335. doi:10.1007/s00251-011-0513-0.

Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes

Jason Greenbaum,

La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA

John Sidney,

La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA

Jolan Chung,

La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA

Christian Brander,

AIDS Research Institute, Fundacio irsiCaixa-HIVACAT, Hospital Universitari Germans Trias i Pujol, Ctra del Canyet s/n, 08916, Badalona, Barcelona, Catalonia, Spain

Bjoern Peters, and

La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA

Alessandro Sette

La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA
alex@liai.org

Abstract

Previous studies have attempted to define human leukocyte antigen (HLA) class II supertypes, analogous to the case for class I, on the basis of shared peptide-binding motifs or structure. In the present study, we determined the binding capacity of a large panel of non-redundant peptides for a set of 27 common HLA DR, DQ, and DP molecules. The measured binding data were then used to define class II supertypes on the basis of shared binding repertoires. Seven different supertypes (main DR, DR4, DRB3, main DQ, DQ7, main DP, and DP2) were defined. The molecules associated with the respective supertypes fell largely along lines defined by MHC locus and reflect, in broad terms, commonalities in reported peptide-binding motifs. Repertoire overlaps between molecules within the same class II supertype were found to be similar in magnitude to what has been observed for HLA class I supertypes. Surprisingly, however, the degree to which repertoires between molecules in the different class II supertypes also overlapped was found to be five to tenfold higher than repertoire overlaps noted between molecules in different class I supertypes. These results highlight a high degree of repertoire overlap amongst all HLA class II molecules, perhaps reflecting binding in multiple registers, and more pronounced dependence on backbone interactions rather than peptide anchor residues. This fundamental difference between HLA class I and class II would not have been predicted on the basis of analysis of either binding motifs or the sequence/predicted structures of the HLA molecules.

Keywords

MHC; HLA class I; HLA class II; Peptide binding; T cell epitopes

Introduction

Binding of human leukocyte antigen (HLA) molecules to peptide epitopes is essential for the activation of antigen-specific T cells. HLA molecules are extremely polymorphic, with thousands of different allelic variants known in humans. Much of the polymorphism is concentrated in residues located at the peptide-binding groove, giving each allelic variant a distinct binding pattern. In the case of HLA class I, several previous studies have demonstrated the existence of HLA supertypes, which define sets of HLA-A and B class I molecules associated with largely overlapping peptide-binding repertoires (Cano et al. 1998; Chelvanayagam 1996; Doytchinova et al. 2004; Hertz and Yanover 2007; Kanguane et al. 2005; Lund et al. 2004; Reche and Reinherz 2004; Sette and Sidney 1998, 1999; Sidney et al. 2008b; Tong et al. 2007; Zhang et al. 1998; Zhao et al. 2003). The identification of groups of HLA molecules with related binding specificities may reduce the complexity of screening for broadly reactive HLA ligands, and can be employed to achieve more extensive population coverage of, for example, vaccine candidates, even in genetically heterogeneous human populations.

A number of studies have suggested that many DR molecules (Chelvanayagam 1997; Doytchinova and Flower 2005; Lund et al. 2004; Nielsen et al. 2008; O'Sullivan et al. 1990, 1991a; Ou et al. 1998; Southwood et al. 1998), and many DP molecules (Berretta et al. 2003; Castelli et al. 2002; Sidney et al. 2010b), can be grouped into supertypes. The case of DQ molecules is less clear as the motifs associated with the various molecules appear to be quite different. However, despite apparent motif differences, significant overlaps in the binding repertoires of DQ molecules have also been noted, possibly reflecting a less prominent role of peptide anchor residues, but a common mode of peptide engagement (Sidney et al. 2010a). At the same time, while motifs have been described for a number of DRB3/4/5 alleles, patterns of specificity for the locus as a whole have not been examined in detail. Herein, we employed actual MHC peptide-binding measurements and objective bioinformatics-based clustering analyses to further verify the existence of class II supertypes and determine supertype assignments across a panel of frequent HLA alleles from all class II loci.

Methods

Peptide synthesis

Peptides were purchased from Mimotopes (Victoria, Australia) and/or A and A (San Diego) as crude material on a 1-mg scale. Peptides utilized as radiolabeled ligands were synthesized on larger scale by A and A, and purified (>95%) by reverse phase HPLC. Timothy grass (*Phleum pratense*; Phl p) peptides were generated as previously described (Oseroff et al. 2010).

MHC purification and binding assays

Purification of MHC class I and class II molecules by affinity chromatography has been described in detail elsewhere (Sidney et al. 1998). Briefly, EBV-transformed homozygous cell lines or single MHC allele-transfected fibroblast (class II) or 721.221 (class I) lines were used as sources of MHC molecules. HLA molecules were purified from cell pellet lysates by repeated passage over protein A Sepharose beads conjugated with specific Abs. Protein

purity, concentration, and the effectiveness of depletion steps are monitored by SDS-PAGE and BCA assay.

Assays to quantitatively measure peptide binding to class I and class II MHC molecules are based on the inhibition of binding of a high-affinity radiolabeled peptide to purified MHC molecules and were performed essentially as detailed elsewhere (Sidney et al. 1998, 2008a, 2010a, b). Briefly, 0.1–1 nM of radiolabeled peptide was co-incubated at room temperature or 37°C with purified MHC in the presence of a cocktail of protease inhibitors. For class I assays, 1 μ M human β 2-microglobulin (Scripps Laboratories, San Diego, CA, USA) is also added to the reaction mixture. Following a 2-day incubation, MHC-bound radioactivity was determined by capturing MHC/peptide complexes on Ab-coated Lumitrac 600 plates (Greiner Bio-one, Frickenhausen, Germany) and measuring bound counts per minute (cpm) using the TopCount (Packard Instrument Co., Meriden, CT, USA) microscintillation counter. In the case of competitive assays, the concentration of peptide yielding 50% inhibition of the binding of the radiolabeled peptide was calculated. Under the conditions utilized, where [label] < [MHC] and $IC_{50} \ll [MHC]$, the measured IC_{50} values are reasonable approximations of the true K_d values. Each competitor peptide was tested at six different concentrations covering a 100,000-fold range and in three or more independent experiments. As a positive control, the unlabeled version of the radiolabeled probe was also tested in each experiment.

The monoclonal antibodies utilized for the purification of class II MHC from EBV-transformed cell lines are HLA DRA-specific, and as such do not allow separating DRB1 molecules from co-expressed DRB3/4/5 molecules. In these cases, the specificity of the binding assay has been determined in previous studies utilizing transfected fibroblast lines and/or panels of epitopes of known class II restriction, as described elsewhere (Boitel et al. 1995; O'Sullivan et al. 1990, 1991a; Sette et al. 1993; Sidney et al. 1992; Southwood et al. 1998; Valli et al. 1993; Wucherpennig et al. 1994).

Hierarchical clustering and bootstrap analysis

The R statistical programming language (Team 2010) was used for all calculations. Binding data for 425 peptides to 27 HLA class II molecules were converted to a binary representation using a 20th percentile cutoff. Distances between each pair of alleles were calculated using the “dist” function with the “binary” method. Hierarchical clustering (Hastie et al. 2001) was performed using the “hclust” function with the average linkage method. Seven significant clusters were identified by a dynamic tree cutting algorithm in the “dynamicTreeCut” package (Langfelder et al. 2008) using the following parameters: $distM=d$ (the distance matrix calculated above), $minClusterSize=2$, $deepSplit=2$. A bootstrapping analysis was performed over 1,000 iterations using the “BootstrapClusterTest” function of the “ClassDiscovery” package (Coombes 2009). This function randomly selected a subset of the data in each iteration and reclustered to estimate the limitations of the method. A heat map of the bootstrapping results was generated (Electronic supplementary material (ESM) Fig. 1), which illustrates the robustness of the clusters. Over the 1,000 iterations, alleles were binned into the same cluster 81% of the time and grouped into a different cluster 9% of the time, on average (ESM Table 1). The same clustering and bootstrapping analysis was applied to a shuffled dataset where no structure to the data was observed (ESM Fig. 2). A Kolmogorov–Smirnov test between the two bootstrapping datasets yielded a p value of 2.2×10^{-16} , indicating that the clusters formed were not the result of random chance.

MHC population coverage and repertoire overlap

The peptide-binding repertoire (R) of each MHC molecule (i) was defined as the set of the peptides that bind that molecule with an affinity equal to, or better than, a specified threshold. In general, for HLA class II, an affinity threshold of 1,000 nM has been found to be associated with the majority of HLA class II restricted epitopes (Sidney et al. 2010a, b; Southwood et al. 1998), while an affinity threshold of 500 nM is associated with the majority of class I epitopes (Sette et al. 1994a); to allow comparison between class I and class II, repertoire overlaps were examined at various different thresholds. The relationship between two molecules has been measured by determining their repertoire overlap, where repertoire overlap is defined as the fraction of peptides binding either molecule that bind both; that is, repertoire overlap = $(R_i \text{ AND } R_j) / (R_i \text{ OR } R_j) \times 100\%$.

Population coverage was calculated as previously described (Sidney et al. 1996, 2010a, b). Gene frequencies (gf) for each HLA allele were calculated from population frequencies obtained from DbMHC (NCBI; Meyer et al. 2007). Phenotypic frequencies (pf) were calculated utilizing the binomial distribution formula: $\text{pf} = 1 - (1 - \sum \text{gf})^2$. To obtain total potential population coverage, no linkage disequilibrium was assumed.

Results

Selection of a panel of molecules representative of the main allelic variants at each of the four HLA class II loci

Previous studies from our group established quantitative peptide–HLA class II-binding assays specific for a large number of variants encoded by DRB1 (Alexander et al. 1994; Boitel et al. 1995; Geluk et al. 1994; O'Sullivan et al. 1990, 1991a; Sette et al. 1993; Sidney et al. 1992; Southwood et al. 1998), DRB3/4/5 (O'Sullivan et al. 1990, 1991a, b), DP (Sidney et al. 2010b), and DQ (; Sidney et al. 1994, 2002, 2010a) alleles. The availability of these assays has enabled addressing, in a standardized assay format, the binding specificity and repertoire of each molecule. These studies in turn can facilitate addressing the degree to which HLA class II can be classified in functional clusters, or supertypes, in analogy to what was reported for class I molecules.

As a first step, and to emphasize the biological significance of the analysis, we selected a panel of HLA molecules representative of approximately 50–75% of all allelic variants at each of the class II loci considered. The overall frequency of each molecule is shown in Table 1. To achieve 50% to 75% allelic coverage, the six most common DQ and DP molecules were selected. In the case of the DRB3/4/5 loci, which are selectively expressed in linkage disequilibrium with various B1 alleles, four allelic variants were sufficient to cover over 75% of the alleles. However, in the case of the DRB1 locus, a total of 11 alleles had to be included to cover approximately 50% of the allelic variants.

Together, this set of molecules provides coverage of between 46% and 77% of haplotypes. It should be emphasized that overall phenotypic frequencies in the general population at each HLA class II locus is >70% and up to 94%. Considering all four loci, combined coverage is estimated to exceed 98% of individuals in all major ethnicities worldwide.

Computational clustering based on peptide binding specificity reveals a set of seven distinct supertypes

To enable elucidation of functional relationships between these common HLA specificities, each of these 27 most common HLA DR, DP, and DQ molecules was experimentally tested for its capacity to bind a panel of 425 non-redundant peptides derived from a set of *P. pratense* antigens. Parts of this dataset have been published before (Oseroff et al. 2010;

Sidney et al. 2010a, b) and is available at the IEDB web site (IEDB submission ID 1000472). All of the binding data have also been provided here as supplemental data (ESM Table 2).

The HLA class II molecules in our panel were next grouped into clusters based on similarities in their respective peptide-binding repertoires using an agglomerative hierarchical clustering algorithm. For the present analysis, a ranking threshold was used to define a peptide-binding event based on measured binding affinity (IC_{50} nM). For each molecule, all measurements in the top 20% were defined as binding events. To define cluster borders, a dynamic tree cutting algorithm was employed (see “Methods”).

The results of clustering analysis are represented in Fig. 1 as a heat map, where yellow indicates a binding event and red indicates a non-binding event. Rows represent the HLA molecules and columns represent peptides. The dendrogram at the right indicates the distance in binding repertoires among the HLA molecules. The robustness of the clusters was assessed with a bootstrapping algorithm, determining how often similar clusters are formed when some of the data are removed. Although there is some overlap between the clusters defined by the dynamic tree cutting algorithm (Table 2), the seven clusters formed are significantly better than random ($p=2.2 \times 10^{-16}$; ESM Figs. 1 and 2).

In general, the clustering of molecules follows locus patterns. Accordingly, we have denominated three of these groups as the main DR, main DQ, and main DP supertypes. All of the DR molecules not in the main DR supertype have been classified into either the DR4 or DRB3 supertypes (clusters). A separate DQ cluster, inclusive of DQB1*0301 (DQ7) and DQB1*0602, has also been identified and denominated as the DQ7 supertype. Similarly, we identified an additional DP cluster, denominated as the DP2 supertype, that includes DPB1*0201 and DPB1*0401.

Correspondence of class II supertypes with previously defined peptide-binding motifs

From the dendrogram shown in Fig. 1, the first apparent division is that between the two DQ clusters, or supertypes, and everything else. The DQ7 supertype, comprising DQB1*0301 and DQB1*0602, can perhaps be best characterized by a shared preference for small and aliphatic residues in positions 4 and 6 of a 9-mer ligand core (Sidney et al. 2010a). Three of the four molecules in the main DQ supertype, namely, DQB1*0201, DQB1*0302, and DQB1*0402, prefer acidic residues near the C-terminal end of the core region (i.e., P6-9). In the case of DQB1*0501, polar and charged (but also some hydrophobic) residues are instead preferred at the N-terminus, and indeed DQB1*0501 is the most remote member of the main DQ supertype. This, together with the preference for small and/or hydrophobic residues in positions 4 and 6, suggests that DQB1*0501 may represent a bridge between the DQ7 and main DQ supertypes.

The next bifurcation separates the DRB3 supertype, which includes DRB3*0101 and DRB3*0202 and the DRB4 molecule DRB4*0101, but also the DRB1 molecules DRB1*0301 and DRB1*1302. In genetic terms, DRB1*0301 appears to be the result of a recombination event between the B1 and B3 loci and more closely resembles other B3 products rather than DRB1 allelic products (Bell et al. 1987; Gorski and Mach 1986). Perhaps not surprisingly, then, DRB1*0301 recognizes a motif distinct from most DRB1 molecules (Geluk et al. 1994; Malcherek et al. 1993; Sidney et al. 1992; Southwood et al. 1998), with a marked preference for acidic residues in position 4. A similar preference for acidic or polar residues in positions 4 and or 6 has also been suggested for DRB3 molecules (Verreck et al. 1996). A preference for basic residues has been described for DRB1*1302 (Boitel et al. 1995; Davenport et al. 1995, 1996; Verreck et al. 1996) and DRB3*0101 (Verreck et al. 1996). A detailed motif for DRB4*0101 has not been described, but a

preference for basic residues in some positions has been suggested (Kobayashi et al. 1996). Thus, the molecules in this supertype include binding preferences for charged residues in at least one anchor position, similar to the DQ supertype and in contrast to the hydrophobic-based specificities of other DR and DP molecules (see below).

The main DR cluster, largely corresponding to the previously described “main DR supertype” (Southwood et al. 1998), represents the next branching. This clustering is characterized by a shared strong specificity for hydrophobic residues in position 1 of the ligand 9-mer core. Small and/or hydrophobic residues are typically preferred at positions 4 and 6, although in some cases, such as in position 6 of DRB1*1101 ligands, basic residues are preferred. Surprisingly, DPB1*1401 is included in this supertype rather than in either of the two DP superotypes (see below).

A second DR supertype includes both of the DR4 subtypes we examined (DRB1*0401 and DRB1*0405) and also DRB1*0802. Previously, we had considered these molecules as part of the prototype main DR supertype on the basis of overlapping binding specificity (Southwood et al. 1998). Based on the data generated here, separation of this cluster from the main DR supertype may reflect less tolerance for charged or polar residues at the position 4 and 6 anchors (compare, e.g., Sette et al. 1993 and O'Sullivan et al. 1991a).

Finally, the two DP clusters separate out from the remaining molecules. Previous reports had hypothesized the existence of a DP supertype (Castelli et al. 2002; Sidney et al. 2010b). Indeed, all DP molecules analyzed to date, with the possible exception of DPB1*1401 whose specificity has not yet been characterized in detail, recognize a motif with aromatic residues in positions 1 and 6 of the 9-mer core region of their ligands (Sidney et al. 2010b). However, the further bifurcation of the DP2 (DPB1*0201 and DPB1*0401) supertype from the main DP supertype (DPB1*0101, DPB1*0402, and DPB1*0501) would not have been predicted on the basis of either main anchor specificity or primary beta chain structure.

Quantitating binding repertoire overlap between HLA class II superotypes

The results presented above define HLA class II superotypes based on the bioinformatic analysis of binding data derived from a set of unbiased overlapping peptides. In the next series of experiments, we sought to quantitatively establish to what degree the peptide-binding repertoire of the different molecules included in each supertype would overlap. To maximize the biological relevance of the results, we opted for performing this analysis utilizing a set of 211 known HLA class II-restricted epitopes retrieved from the IEDB. These epitopes were tested for their capacity to bind the same panel of common class II molecules (ESM Table 2). Repertoire overlap was defined as the percent of peptides binding either of a pair of class II molecules with an affinity of 1,000 nM, or better, that bind both molecules (see “Methods”). This affinity had been previously established as a threshold associated with antigenicity for HLA class II epitopes, irrespective of locus (Sidney et al. 2010a, b; Southwood et al. 1998).

We found that the repertoire overlaps between molecules within the same supertype averaged about 46%, ranging from a high of 60% for the DP2 supertype to a low of 23% for the main DQ supertype (Table 3 and Fig. 2a). Average repertoire overlaps for molecules in the main DR, DR4, DRB3, main DP, and DQ7 superotypes were 46%, 55%, 31, 56, and 54%, respectively. The level of cross-reactivity between class II molecules in different superotypes was also relatively high, averaging 31%, and with a range of 19–38%, depending on the specific pairs of superotypes. The ratio between the average percentage overlap within a supertype (intra-supertype) to the average percent overlap between molecules in different superotypes (inter-supertype) averaged 1.5, with a range of 1.1–2.3. Using more stringent definitions of binding (down to a threshold of 100 nM), we found that the ratio between the

intra- and inter-supertype rates remained low, reaching a high of 2.4 only at the 100-nM threshold (Table 3).

We hypothesize that this high degree of repertoire overlap between supertypes may reflect the relatively loose binding motifs of class II molecules where a large fraction of the energy of peptide–MHC binding is contributed by backbone interactions (Brown et al. 1993) and/or the fact that the same peptide could carry independent overlapping binding frames (Raddrizzani et al. 1997; Sidney et al. 2002). Accordingly, we predicted that HLA class I would be associated with a lower overlap between different supertypes.

To test this hypothesis, we examined the binding capacity of a set of 252 previously described HLA class I epitopes (Frahm et al. 2007) for a set of 30 class I molecules encompassing the seven most common HLA class I supertypes. We found that at an affinity threshold of 500 nM, previously identified as the threshold affinity associated with the vast majority of class I epitopes (Sette et al. 1994b), the average repertoire overlap between molecules within the same supertype averaged 30%, comparable to the intra-cluster rate for class II at the same affinity threshold (Table 3 and Fig. 2b). At the same time, however, the overlap between class I molecules in different supertypes averaged only 2.6%, representing a 12-fold difference in overlap rates. A similar pattern was noted at the 1,000-nM threshold, as well as various other affinity thresholds (Table 3). Taken together, these data show that newly defined HLA class II supertypes are associated with a degree of intra-supertype overlap similar to that detected in the case of class I supertypes. At the same time, however, the degree of inter-supertype overlap observed for class II supertypes is strikingly higher than in the case of HLA class I.

Discussion

In the present study, we have utilized HLA class II binding data to identify clusters, or supertypes, of DR, DQ, and DP molecules, defined on the basis of similar peptide-binding specificity. Previous efforts to define class II supertypes have been based on similarities in primary anchor motifs or structural similarities. The advantage of the present approach is that it is based on experimentally determined function.

For our studies, we have selected a set of allelic variants that represents 50–75% of the HLA class II genes expressed worldwide for all four different HLA class II loci. The alleles provide phenotypic coverage of over 70% at each locus and when combined cover virtually 100% of the general population. Many additional allelic variants are structurally very similar to these representative alleles, and therefore, each of the supertypes described herein likely encompasses several additional allelic variants. Indeed, it has been shown that the binding specificity of HLA molecules can be reasonably well predicted on the basis of the known specificity of closely related allelic variants (Lund et al. 2004; Nielsen et al. 2008; Sturmiolo et al. 1999). However, since we have not analyzed all HLA class II alleles, other supertypes might be defined in the future, encompassing HLA molecules that are yet to be studied.

Our results define seven different HLA class II supertypes denominated as the main DR, DR4, DRB3, main DQ, DQ7, main DP, and DP2 supertypes. These supertypes fall largely along lines defined by MHC locus and reflect, in retrospect, commonalities in peptide-binding motifs (Castelli et al. 2002; Sidney et al. 2010a, b; Southwood et al. 1998). For example, the DQ7 supertype can be characterized by a shared preference for small and aliphatic residues in positions 4 and 6 of a 9-mer ligand core. Both the main DQ and DRB3 supertypes bring together molecules with prominent preferences for charged residues in one or more anchor positions. The four remaining supertypes represent molecules hypothesized to recognize more hydrophobic residue-based motifs. More specifically, molecules in both

DP supertypes share preferences for aromatic residues in positions 1 and 6, and the main DR and DR4 supertypes share strong specificities for hydrophobic residues in position 1 and small and/or hydrophobic residues at positions 4 and 6.

However, while motif similarities between class II molecules can be used to retrospectively rationalize the observed supertypes, the results of the present analysis would not have been predictable solely based on a motif-based classification. This result may be explained by the fact that a large portion of the energy of class II binding is due to backbone interactions (Brown et al. 1993; Madden 1995; McFarland and Beeson 2002; McFarland et al. 2005; Nelson and Fremont 1999; Stern et al. 1994) and that because the ends of the binding groove are open the same peptide may bind different class II molecules utilizing completely different modes. For these reasons, elucidation of class II motifs is not always straightforward, and class II motifs tend to be more loosely described than is the case for class I. As a result, classification on the basis of motifs may overplay differences in that molecules appearing to have somewhat different binding specificity at first glance may, in actual practice, bind similar repertoires.

This may be somewhat exemplified by the very detailed and innovative classification approach of Nielsen et al. (2008) who used neural networks and predicted peptide-binding specificity to define 12 main DR supertypes. Comparison of this classification with the one described herein reveals that it, by and large, represents a parsing of individual DR molecules into individual supertypes. For example, the seven DR molecules we have clustered into the main DR supertype represent seven different supertypes in the Nielsen scheme. In the limited cases where Nielsen and co-workers have defined a supertype containing two or more of the molecules studied herein, intra-cluster repertoire overlap rates are similar to those found in our analysis. Inter-cluster rates are, however, higher, reflecting the fact that two molecules clustered together in our classification on the basis of high repertoire overlap are representative of different clusters in the Nielsen study. Admittedly, the parameters used for clustering in either study differ from one another. However, when we varied the parameters to closely match those from the Nielsen study, our results remained the same.

Our results are also largely consistent with schemes defining supertypes on the basis of phylogeny or shared MHC pocket structures (see, e.g., Doytchinova and Flower 2005). However, it is also apparent that these approaches may at times imply either closer or more distant relationships than are actually observed when peptide binding is measured. For example, instances of convergent evolution, where completely different structures may share similar specificities, may not be reliably predicted on the basis of MHC structural relationships. This is perhaps best demonstrated by the observation that DPB1*1401 has closer binding specificity to other molecules included in the main DR supertype as opposed to molecules in either of the DP supertypes despite their closer phylogenetic relation. Furthermore, several other functional relationships would not have been predicted on the basis of structural similarity. This is exemplified by the main DP/DP2 bifurcation, which likely would not have been predicted on the basis of either main anchor specificity or primary beta chain structure, or by the inclusion of DRB1*1302 in the DRB3 cluster. Detailed analyses correlating structural differences and similarities between various class II MHC molecules and the corresponding cluster relationships defined herein will be the subject of future studies.

The ultimate reflection of the existence of functional clusters of HLA molecules is an overlapping peptide-binding repertoire. Herein, we provided a rigorous evaluation of the extent of repertoire overlap for class II supertypes and found it to be similar in magnitude to that observed for class I supertypes. But surprisingly, overlaps in repertoires between

different supertypes were much greater in the context of class II than for class I supertypes. While cross-reactivity between supertypes has been described in the case of class I, at least in the context of specific epitopes (Axelsson-Robertson et al. 2010; Frahm et al. 2007; Nakagawa et al. 2007), the degree of cross-reactivity seen in the case of class II is profoundly different and pervasive by comparison. This is likely a reflection of main chain interactions being more important for class II binding than for class I. Consequently, the influence of specific pocket motifs may be relatively less prominent compared to the case of class I where pocket specificity is crucial. That this may be the case was suggested in a recent study of the motifs and repertoires recognized by several common DQ molecules (Sidney et al. 2010a). It was found that while the motifs defined for each DQ were unique, and often very different, most DQ molecules shared a high degree of repertoire overlap.

About one fifth of the peptides tested in the present study were not bound by any, or just one or two, HLA class II molecules tested (see Fig. 1). This rate of non-binding is much greater than would be expected by random chance. Similarly, a larger than expected fraction of peptides comprised remarkably promiscuous binders having the capacity to bind a dozen or more of the molecules tested. These observations suggest that perhaps some peptide sequence features may be identified that would reflect the promiscuous binding versus non-binding patterns. Present studies are focusing on further mining the available data and the relationships defined herein to determine whether bioinformatic tools can be developed to efficiently identify peptides with the capacity to bind multiple class II specificities within, or between, the various supertypes and loci.

The set of peptides utilized to generate the binding dataset for the present study was unbiased, representing naturally occurring sequences with a set periodicity, and as such also represents, to the best of our knowledge, an unbiased distribution of promiscuity that closely resembles a random sampling of natural peptides. For this reason, we believe that the set is most appropriate for evaluating the natural repertoire of the MHC molecules under study. However, the possibility that different clustering may be detected using a set of peptides biased for promiscuity is very intriguing, especially, for example, in a vaccine design context. Examination of this possibility will be the subject of future studies.

In conclusion, utilizing MHC-peptide binding data, we have defined seven HLA class II clusters, or supertypes, to encompass the most common DR, DQ, and DP molecules. However, it was also found that in functional terms, the repertoire sharing between the different supertypes is very significant as compared to the case of HLA class I supertypes. While previous attempts have been made to classify HLA class II molecules into supertypes, similar to as has been done for HLA class I, none has predicted this fundamental difference between class I and II supertypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported with funds from the National Institutes of Health, National Institute of Allergy and Infectious Diseases (NIH-NIAID) contracts HHSN266200400006C, HHSN272200900044C, HHSN272200900042C, and HHSN272200 700048C (all to AS). We thank Carla Oseroff for MHC purification and tissue culture; Amiyah Steen, Carrie Moore, and Sandy Ngo for help with the binding assays; and Howard Grey for his comments and helpful discussions.

References

- Alexander J, Sidney J, Southwood S, Ruppert J, Oseroff C, Maewal A, Snoke K, Serra HM, Kubo RT, Sette A, et al. Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity*. 1994; 1:751–761. [PubMed: 7895164]
- Axelsson-Robertson R, Weichold F, Sizemore D, Wulf M, Skeiky YA, Sadoff J, Maeurer MJ. Extensive major histocompatibility complex class I binding promiscuity for *Mycobacterium tuberculosis* TB10.4 peptides and immune dominance of human leucocyte antigen (HLA)-B*0702 and HLA-B*0801 alleles in TB10.4 CD8 T-cell responses. *Immunology*. 2010; 129:496–505. [PubMed: 20002212]
- Bell JI, Denney D Jr, Foster L, Belt T, Todd JA, McDevitt HO. Allelic variation in the DR subregion of the human major histocompatibility complex. *Proc Natl Acad Sci USA*. 1987; 84:6234–6238. [PubMed: 3476943]
- Berretta F, Butler RH, Diaz G, Sanarico N, Arroyo J, Fraziano M, Aichinger G, Wucherpfennig KW, Colizzi V, Saltini C, Amicosante M. Detailed analysis of the effects of Glu/Lys beta69 human leukocyte antigen-DP polymorphism on peptide-binding specificity. *Tissue Antigens*. 2003; 62:459–471. [PubMed: 14617029]
- Boitel B, Blank U, Mege D, Corradin G, Sidney J, Sette A, Acuto O. Strong similarities in antigen fine specificity among DRB1* 1302-restricted tetanus toxin tt830-843-specific TCRs in spite of highly heterogeneous CDR3. *J Immunol*. 1995; 154:3245–3255. [PubMed: 7534793]
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*. 1993; 364:33–39. [PubMed: 8316295]
- Cano P, Fan B, Stass S. A geometric study of the amino acid sequence of class I HLA molecules. *Immunogenetics*. 1998; 48:324–334. [PubMed: 9745009]
- Castelli FA, Buhot C, Sanson A, Zarour H, Pouvelle-Moratille S, Nonn C, Gahery-Segard H, Guillet JG, Menez A, Georges B, Maillere B. HLA-DP4, the most frequent HLA II molecule, defines a new supertype of peptide-binding specificity. *J Immunol*. 2002; 169:6928–6934. [PubMed: 12471126]
- Chelvanayagam G. A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. *Immunogenetics*. 1996; 45:15–26. [PubMed: 8881033]
- Chelvanayagam G. A roadmap for HLA-DR peptide binding specificities. *Hum Immunol*. 1997; 58:61–69. [PubMed: 9475335]
- Coombes KR. ClassDiscovery: classes and methods for “class discovery” with microarrays or proteomics. R package version 2.10.1 edn. 2009
- Davenport MP, Quinn CL, Chicz RM, Green BN, Willis AC, Lane WS, Bell JI, Hill AV. Naturally processed peptides from two disease-resistance-associated HLA-DR13 alleles show related sequence motifs and the effects of the dimorphism at position 86 of the HLA-DR beta chain. *Proc Natl Acad Sci USA*. 1995; 92:6567–6571. [PubMed: 7604034]
- Davenport MP, Quinn CL, Valsasini P, Sinigaglia F, Hill AV, Bell JI. Analysis of peptide-binding motifs for two disease associated HLA-DR13 alleles using an M13 phage display library. *Immunology*. 1996; 88:482–486. [PubMed: 8881746]
- Doytchinova IA, Flower DR. In silico identification of superotypes for class II MHCs. *J Immunol*. 2005; 174:7085–7095. [PubMed: 15905552]
- Doytchinova IA, Guan P, Flower DR. Identifying human MHC superotypes using bioinformatic methods. *J Immunol*. 2004; 172:4314–4323. [PubMed: 15034046]
- Frahm N, Yusim K, Suscovich TJ, Adams S, Sidney J, Hraber P, Hewitt HS, Linde CH, Kavanagh DG, Woodberry T, Henry LM, Faircloth K, Listgarten J, Kadie C, Jovic N, Sango K, Brown NV, Pae E, Zaman MT, Bihl F, Khatri A, John M, Mallal S, Marincola FM, Walker BD, Sette A, Heckerman D, Korber BT, Brander C. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur J Immunol*. 2007; 37:2419–2433. [PubMed: 17705138]
- Geluk A, van Meijgaarden KE, Southwood S, Oseroff C, Drijfhout JW, de Vries RR, Ottenhoff TH, Sette A. HLA-DR3 molecules can bind peptides carrying two alternative specific submotifs. *J Immunol*. 1994; 152:5742–5748. [PubMed: 8207204]

- Gorski J, Mach B. Polymorphism of human Ia antigens: gene conversion between two DR beta loci results in a new HLA-D/DR specificity. *Nature*. 1986; 322:67–70. [PubMed: 3459965]
- Hastie, T.; Tibshirani, R.; Friedman, JH. The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations. Springer; New York: 2001.
- Hertz T, Yanover C. Identifying HLA supertypes by learning distance functions. *Bioinformatics*. 2007; 23:e148–e155. [PubMed: 17237084]
- Kangueane P, Sakharkar MK, Rajaseger G, Bolisetty S, Sivasekari B, Zhao B, Ravichandran M, Shapshak P, Subbiah S. A framework to sub-type HLA supertypes. *Front Biosci*. 2005; 10:879–886. [PubMed: 15569626]
- Kobayashi H, Kokubo T, Abe Y, Sato K, Kimura S, Miyokawa N, Katagiri M. Analysis of anchor residues in a naturally processed HLA-DR53 ligand. *Immunogenetics*. 1996; 44:366–371. [PubMed: 8781122]
- Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*. 2008; 24:719–720. [PubMed: 18024473]
- Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, Sylvester-Hvid C, Lamberth K, Roder G, Justesen S, Buus S, Brunak S. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*. 2004; 55:797–810. [PubMed: 14963618]
- Madden DR. The three-dimensional structure of peptide–MHC complexes. *Annu Rev Immunol*. 1995; 13:587–622. [PubMed: 7612235]
- Malcherek G, Falk K, Rotzschke O, Rammensee HG, Stevanovic S, Gnau V, Jung G, Melms A. Natural peptide ligand motifs of two HLA molecules associated with myasthenia gravis. *Int Immunol*. 1993; 5:1229–1237. [PubMed: 8268130]
- McFarland BJ, Beeson C. Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med Res Rev*. 2002; 22:168–203. [PubMed: 11857638]
- McFarland BJ, Katz JF, Sant AJ, Beeson C. Energetics and cooperativity of the hydrogen bonding and anchor interactions that bind peptides to MHC class II protein. *J Mol Biol*. 2005; 350:170–183. [PubMed: 15921691]
- Meyer, D.; Singe, RM.; Mack, SJ.; Lancaster, A.; Nelson, MP.; Erlich, H.; Fernandez-Vina, M.; Thomson, G. Single locus polymorphism of classical HLA genes.. In: Hansen, J., editor. *Immunobiology of the human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference*; IHWG, Seattle. 2007.
- Nakagawa M, Kim KH, Gillam TM, Moscicki AB. HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein. *J Virol*. 2007; 81:1412–1423. [PubMed: 17108051]
- Nelson CA, Fremont DH. Structural principles of MHC class II antigen presentation. *Rev Immunogenet*. 1999; 1:47–59. [PubMed: 11256572]
- Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, Lund O. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol*. 2008; 4:e1000107. [PubMed: 18604266]
- O'Sullivan D, Arrhenius T, Sidney J, Del Guercio MF, Albertson M, Wall M, Oseroff C, Southwood S, Colon SM, Gaeta FC, et al. On the interaction of promiscuous antigenic peptides with different DR alleles. Identification of common structural motifs. *J Immunol*. 1991a; 147:2663–2669. [PubMed: 1717570]
- O'Sullivan D, Sidney J, Appella E, Walker L, Phillips L, Colon SM, Miles C, Chesnut RW, Sette A. Characterization of the specificity of peptide binding to four DR haplotypes. *J Immunol*. 1990; 145:1799–1808. [PubMed: 2167910]
- O'Sullivan D, Sidney J, Del Guercio MF, Colon SM, Sette A. Truncation analysis of several DR binding epitopes. *J Immunol*. 1991b; 146:1240–1246. [PubMed: 1704034]
- Oseroff C, Sidney J, Kotturi MF, Kolla R, Alam R, Broide DH, Wasserman SI, Weiskopf D, McKinney DM, Chung JL, Petersen A, Grey H, Peters B, Sette A. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. *J Immunol*. 2010; 185:943–955. [PubMed: 20554959]
- Ou D, Mitchell LA, Tingle AJ. A new categorization of HLA DR alleles on a functional basis. *Hum Immunol*. 1998; 59:665–676. [PubMed: 9757949]

- Raddizzani L, Sturniolo T, Guenot J, Bono E, Gallazzi F, Nagy ZA, Sinigaglia F, Hammer J. Different modes of peptide interaction enable HLA-DQ and HLA-DR molecules to bind diverse peptide repertoires. *J Immunol.* 1997; 159:703–711. [PubMed: 9218585]
- Reche PA, Reinherz EL. Definition of MHC supertypes through clustering of MHC peptide binding repertoires. *LNCS, ICARIS.* 2004; 3239:189–196.
- Sette A, Sidney J. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Curr Opin Immunol.* 1998; 10:478–482. [PubMed: 9722926]
- Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics.* 1999; 50:201–212. [PubMed: 10602880]
- Sette A, Sidney J, Oseroff C, del Guercio MF, Southwood S, Arrhenius T, Powell MF, Colon SM, Gaeta FC, Grey HM. HLA DR4w4-binding motifs illustrate the biochemical basis of degeneracy and specificity in peptide-DR interactions. *J Immunol.* 1993; 151:3163–3170. [PubMed: 7690794]
- Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol.* 1994a; 153:5586–5592. [PubMed: 7527444]
- Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, Sidney J, del Guercio MF, Southwood S, Kubo RT, Chesnut RW, Grey HM, Chisari FV. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol.* 1994b; 153:5586–5592. [PubMed: 7527444]
- Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res.* 2008a; 4:2. [PubMed: 18221540]
- Sidney J, Del Guercio MF, Southwood S, Sette A. The HLA molecules DQA1*0501/B1*0201 and DQA1*0301/B1*0302 share an extensive overlap in peptide binding specificity. *J Immunol.* 2002; 169:5098–5108. [PubMed: 12391226]
- Sidney J, Grey HM, Southwood S, Celis E, Wentworth PA, del Guercio MF, Kubo RT, Chesnut RW, Sette A. Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules. *Hum Immunol.* 1996; 45:79–93. [PubMed: 8882405]
- Sidney J, Oseroff C, del Guercio MF, Southwood S, Krieger JI, Ishioka GY, Sakaguchi K, Appella E, Sette A. Definition of a DQ3.1-specific binding motif. *J Immunol.* 1994; 152:4516–4525. [PubMed: 7512598]
- Sidney J, Oseroff C, Southwood S, Wall M, Ishioka G, Koning F, Sette A. DRB1*0301 molecules recognize a structural motif distinct from the one recognized by most DR beta 1 alleles. *J Immunol.* 1992; 149:2634–2640. [PubMed: 1401902]
- Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol.* 2008b; 9:1. [PubMed: 18211710]
- Sidney, J.; Southwood, S.; Oseroff, C.; Del Guercio, MF.; Sette, A.; Grey, H. Current protocols in immunology. Wiley; New York: 1998. Measurement of MHC/peptide interactions by gel filtration.; p. 18.13.11-18.13.19.
- Sidney J, Steen A, Moore C, Ngo S, Chung J, Peters B, Sette A. Divergent motifs but overlapping binding repertoires of six HLA-DQ molecules frequently expressed in the worldwide human population. *J Immunol.* 2010a; 185:4189–4198. [PubMed: 20810981]
- Sidney J, Steen A, Moore C, Ngo S, Chung J, Peters B, Sette A. Five HLA-DP molecules frequently expressed in the worldwide human population share a common HLA supertypic binding specificity. *J Immunol.* 2010b; 184:2492–2503. [PubMed: 20139279]
- Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, Kubo RT, Chesnut RW, Grey HM, Sette A. Several common HLA-DR types share largely overlapping peptide binding repertoires. *J Immunol.* 1998; 160:3363–3373. [PubMed: 9531296]
- Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature.* 1994; 368:215–221. [PubMed: 8145819]
- Sturniolo T, Bono E, Ding J, Raddizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. Generation of tissue-specific and promiscuous HLA ligand databases

- using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol.* 1999; 17:555–561. [PubMed: 10385319]
- Team, RDC. R Foundation for Statistical Computing. Vienna, Austria: 2010. R: a language and environment for statistical computing..
- Tong JC, Tan TW, Ranganathan S. In silico grouping of peptide/HLA class I complexes using structural interaction characteristics. *Bioinformatics.* 2007; 23:177–183. [PubMed: 17090577]
- Valli A, Sette A, Kappos L, Oseroff C, Sidney J, Miescher G, Hochberger M, Albert ED, Adorini L. Binding of myelin basic protein peptides to human histocompatibility leukocyte antigen class II molecules and their recognition by T cells from multiple sclerosis patients. *J Clin Invest.* 1993; 91:616–628. [PubMed: 7679413]
- Verreck FA, van de Poel A, Drijfhout JW, Amons R, Coligan JE, König F. Natural peptides isolated from Gly86/Val86-containing variants of HLA-DR1, -DR11, -DR13, and -DR52. *Immunogenetics.* 1996; 43:392–397. [PubMed: 8606061]
- Wucherpennig KW, Sette A, Southwood S, Oseroff C, Matsui M, Strominger JL, Hafler DA. Structural requirements for binding of an immunodominant myelin basic protein peptide to DR2 isotypes and for its recognition by human T cell clones. *J Exp Med.* 1994; 179:279–290. [PubMed: 7505801]
- Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. *J Mol Biol.* 1998; 281:929–947. [PubMed: 9719645]
- Zhao B, Png AE, Ren EC, Kolatkar PR, Mathura VS, Sakharkar MK, Kangueane P. Compression of functional space in HLA-A sequence diversity. *Hum Immunol.* 2003; 64:718–728. [PubMed: 12826374]

Fig. 1.
HLA class II clusters (supertypes) identified using an agglomerative hierarchical clustering algorithm

Fig. 2.
Repertoire overlaps between HLA class II (a) and class I (b) supertype alleles

Table 1

Haplotype and phenotype frequencies of HLA class II alleles

Locus	Allele	Percent of haplotypes	Phenotype frequency
DRB1	DRB1*0101	2.8	5.4
	DRB1*0301	7.1	13.7
	DRB1*0401	2.3	4.6
	DRB1*0405	3.1	6.2
	DRB1*0701	7.0	13.5
	DRB1*0802	2.5	4.9
	DRB1*0901	3.1	6.2
	DRB1*1101	6.1	11.8
	DRB1*1201	2.0	3.9
	DRB1*1302	3.9	7.7
	DRB1*1501	6.3	12.2
	Total	46.2	71.1
DRB3/4/5	DRB3*0101	14.0	26.1
	DRB3*0202	18.9	34.3
	DRB4*0101	23.7	41.8
	DRB5*0101	8.3	16.0
	Total	77.3	87.7
DQA1/DQB1	DQA1*0501/DQB1*0201	5.8	11.3
	DQA1*0501/DQB1*0301	19.5	35.1
	DQA1*0301/DQB1*0302	10.0	19.0
	DQA1*0401/DQB1*0402	6.6	12.8
	DQA1*0101/DQB1*0501	7.6	14.6
	DQA1*0102/DQB1*0602	7.6	14.6
	Total	57.1	81.6
DPB1	DPB1*0101	8.4	16.0
	DPB1*0201	9.2	17.5
	DPB1*0401	20.1	36.2
	DPB1*0402	23.6	41.6
	DPB1*0501	11.5	21.7
	DPB1*1401	3.8	7.4
	Total	76.5	94.5

Average haplotype and phenotype frequencies for individual alleles are based on data available at dbMHC. dbMHC data considers prevalence in Europe, North Africa, Northeast Asia, the South Pacific (Australia and Oceania), Hispanic North and South America, American Indian, Southeast Asia, Southwest Asia, and sub-Saharan Africa populations. DP, DRB1, and DRB3/4/5 frequencies consider only the beta chain frequency given that the DRA chain is largely monomorphic and that differences in DPA are not hypothesized to significantly influence binding. Frequency data are not available for DRB3/4/5 alleles. However, because of linkage with DRB1 alleles, coverage for these specificities may be assumed as follows: DRB3 with DR3, DR11, DR12, DR13, and DR14; DRB4 with DR4, DR7, and DR9; and DRB5 with DR15 and DR16. Specific allele frequencies at each B3/B4/B5 locus is based on published associations with various DRB1 alleles and assume only limited variation at the indicated locus

Table 2

Summary of HLA supertypes (clusters)

Supertype	Alleles
Main DR	DRB1*0101, 0701, 0901, DRB1*1101, 1201 and 1501, DRB5*0101 and DPB1*1401
DR4	DRB1*0401, 0405 and 0802
DRB3	DRB1*0301, 1302, DRB3*0101, 0202 and DRB4*0101
Main DP	DPB1*0101, 0402 and 0501
DP2	DPB1*0201, and 0401
Main DQ	DQB1*0201, 0302, 0402, and 0501
DQ7	DQB1*0301 and 0602

Table 3Repertoire overlap rates at various IC₅₀ nM thresholds

System	Binding threshold (IC ₅₀ nM)	Supertype/cluster repertoire overlap		
		Within	Without	Ratio
Class I	100	29.7	1.0	28.9
	200	31.7	1.6	20.1
	500	30.4	2.6	11.9
	1,000	33.3	3.6	9.2
	2,000	37.3	5.3	7.1
	5,000	38.9	7.8	5.0
Class II	100	19.8	8.4	2.4
	200	27.4	12.9	2.1
	500	35.1	19.9	1.8
	1,000	46.4	30.8	1.5
	2,000	52.8	39.9	1.3
	5,000	63.6	53.4	1.2