Several Common HLA-DR Types Share Largely Overlapping Peptide Binding Repertoires

Scott Southwood, John Sidney, Akihiro Kondo, Marie-France del Guercio, Ettore Appella, Stephen Hoffman, Ralph T. Kubo, Robert W. Chesnut, Howard M. Grey and Alessandro Sette

*J Immunol* 1998; 160:3363-3373; ;
http://www.jimmunol.org/content/160/7/3363

References  This article cites 85 articles, 45 of which you can access for free at:  http://www.jimmunol.org/content/160/7/3363.full#ref-list-1

Subscription  Information about subscribing to *The Journal of Immunology* is online at:  http://jimmunol.org/subscription

Permissions  Submit copyright permission requests at:  http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts  Receive free email-alerts when new articles cite this article. Sign up at:  http://jimmunol.org/alerts
Several Common HLA-DR Types Share Largely Overlapping Peptide Binding Repertoires

Scott Southwood,* John Sidney,* Akihiro Kondo,† Marie-France del Guercio,* Ettore Appella,‡ Stephen Hoffman,§ Ralph T. Kubo,¶ Robert W. Chesnut,* Howard M. Grey,¶ and Alessandro Sette*§

The peptide binding specificities of HLA-DRB1*0401, DRB1*0101, and DRB1*0701 have been analyzed by the use of large collections of synthetic peptides corresponding to naturally occurring sequences. The results demonstrated that nearly all peptides binding to these DR molecules bear a motif characterized by a large aromatic or hydrophobic residue in position 1 (Y, F, W, L, I, V, M) and a small, noncharged residue in position 6 (S, T, C, A, P, V, I, L, M). In addition, allele-specific secondary effects and secondary anchors were defined, and these parameters were utilized to derive allele-specific motifs and algorithms. By the combined use of such algorithms, peptides capable of degenerate DRB1*0101, DRB1*0401, and DRB1*0701 binding were identified. Additional experiments utilizing a panel of quantitative assays specific for nine additional common DR molecules identified a large set of DR molecules, which includes at least the DRB1*0101, DRB1*0401, DRB1*0701, DRB5*0101, DRB1*1501, DRB1*0901, and DRB1*1302 allelic products, characterized by overlapping peptide-binding repertoires. These results have implications for understanding the molecular interactions involved in peptide-DR binding, as well as the genetic and structural basis of MHC polymorphism. These results also have potential practical implications for the development of epitope-based prophylactic and therapeutic vaccines. The Journal of Immunology, 1998, 160: 3363–3373.

Helper T lymphocytes (HTL) play several important functions in immunity to pathogens. They provide help for induction of both CTL and Ab responses. In addition, HTL can be effectors in their own right, an activity mediated by direct cell contact and lymphokine secretion (e.g., IFN-γ and TNF-α). HTL have been shown to have direct effector activity in the case of tumors, and viral, bacterial, parasitic, and fungal infections (1–14).

HTL recognize a complex formed between class II MHC molecules and antigenic peptides. Peptides binding class II molecules are usually between 10 to 20 residues long, with sizes between 13 and 16 amino acids being most frequently observed (15–18). Peptide-class II interactions have been analyzed in detail, both at the structural (19, 20) and functional level (21–23), and peptide motifs specific for various human and mouse class II molecules have been proposed (21, 24–34). Predictions based upon class II-specific motifs appear, however, to be less accurate than those based on class I motifs. This lower efficiency may result from the peptide binding groove of class II molecules being open at both ends (15–20), and thus allowing a given peptide to potentially bind in many different registers.

In the last few years, epitope-based vaccines have received considerable attention as a possible means to develop novel prophylactic vaccines and immunotherapeutic strategies (35–38). Selection of appropriate T and B cell epitopes should allow the immune system to be focused on conserved epitopes of pathogens characterized by high sequence variability (such as HIV, hepatitis C virus (HCV), and malaria) (8, 39–43).

Focusing the immune response toward selected determinants could be of value in the case of various chronic viral diseases and cancer, where T cells directed against the immunodominant epitopes might have been inactivated and T cells specific for subdominant epitopes might have escaped T cell tolerance (44–50).

Epitope-based vaccines also offer the opportunity to include in the vaccine construct epitopes that have been engineered to modulate potency, either by increasing MHC binding affinity, or by alteration of their TCR contact residues, or both (51–53).

Once appropriate epitope determinants have been defined, they can be sorted and delivered by various means, including lipopeptides (54), viral delivery vectors (55, 56), particles of viral or synthetic origin (57–59), adjuvants (60, 61), liposomes (62), and naked or particle-absorbed cDNA (63).

Before appropriate epitopes can be defined, however, one major obstacle has to be overcome: the very high degree of polymorphism of the MHC molecules expressed in the human population. More than 200 different types of HLA class I and class II molecules have already been identified (64, 65). However, our group has demonstrated that peptides capable of binding several different HLA class I molecules can be identified, and over 60% of the known HLA class I molecules can be grouped into four broad
HLA supertypes characterized by similar peptide binding specificities (HLA supermotifs) (66–70).

In the case of class II molecules, it is also known that peptides capable of binding multiple HLA types, and of being immunogenic in the context of different HLA molecules, do indeed exist (32, 33, 51, 71–76). Until now, a general method for their identification has not been developed. This is probably at least in part a reflection of the fact that quantitative DR binding assays are labor intensive, and that a large number of alleles must be considered.

In the present report, we describe the development and validation of specific motifs and assay systems for various DR molecules, which are representative of the predominant alleles worldwide in the human population. Applying this technology to the identification of broadly degenerate HLA class II binding peptides greatly enhances the potential for the use of epitope-based vaccines on a global scale.

Materials and Methods

Cells

The following EBV-transformed homozygous cell lines were used as sources of human HLA class II molecules: LG2 (DRB1*0101 (DR1)); GM107 (DRB5*0101 (DR2w2a)); MAT (DRB1*0301 (DR3)); PREISS (DRB1*0401 (DR4w4)); SWEIG (DRB1*1101 (DR5w11)); PITOUT (DRB1*0701 (DR7)); KT13 (DRB1*0405 (DR4w15)); Herluf (DRB1*1201 (DR5w12)); HO301 (DRB1*1302 (DR6w19)); OLL (DRB1*0802 (DR8w2)); and HDR (DRB1*0901 (DR9), supplied as a kind gift by Dr. Paul Harris, Columbia University). In one instance, transfected fibroblasts were used: L66.1 (DRB1*1501 (DR2w2b)) (51, 77). Cells were maintained in vitro by culture in RPMI 1640 medium supplemented with 2 mM l-glutamine (Life Technologies, Grand Island, NY), 50 µM 2-ME, and 10% heat-inactivated FCS ( Irvine Scientific, Santa Ana, CA). Cells were also supplemented with 100 µg/ml of streptomycin and 100 µM of penicillin (Irvine Scientific). Large quantities of cells were grown in spinner cultures.

Cells were lysed for 30 min at 4°C with a lysis buffer of 50 mM Tris-HCL, pH 8.5, 1% Nonidet P-40 (Fluka Biochemika, Buchs, Switzerland), 150 mM NaCl, and 2 mM PMSF (Tosohaas, Montgomeryville, PA) and the fraction of bound peptide calculated as previously described (78).

Radionabeled peptides were iodinated using the chloramine-T method (81). The radionabeled probes used were HA Y307-319 (DRB1*0101), tetanus toxin (TT) 830-843 (DRB5*0101, DRB1*1101, DRB1*0701, DRB1*0802, DRB1*0901), MPP Y85-100 (DRB1*1501), MB 65 kDa Y3-13 with Y7 substituted with F (DRB1*0301), a non-natural peptide (71, 72) with the sequence YARFFQSGTQLKQT (DRB1*0401, DRB1*0405) (51, 77), a naturally processed peptide (1200/05; sequence EEAIHQKINPYVLS) (34) of unknown origin eluted from a DRB1*1201 + C1R cell line, and an S3_02-A analogue of TT 830-843 (peptide 650.22; sequence QYIKANAKFGITE for DRB1*1302 (82)).

In preliminary experiments, the DR preparation was titrated in the presence of fixed amounts of radiolabeled peptides to determine the concentration of class II molecules necessary to bind 10 to 20% of the total radioactivity. All subsequent inhibition and direct binding assays were then performed using these class II concentrations.

Peptide inhibitors were typically tested at concentrations ranging from 120 µg/ml to 1.2 mg/ml. In appropriate stoichiometric conditions, the 50% inhibitory concentration (IC50) of an unlabeled test peptide to the purified DR is a reasonable approximation of the affinity of interaction (kDa). Peptides were tested in two to four completely independent experiments.

Because the Ab used for purification is α-chain specific, βmolecules are not separated from β1 (and/or β2) and β3 molecules. Development and validation of assays with regard to DRβ molecule specificity have been described in detail elsewhere for many of the DR alleles listed above (29, 31, 32, 77, 82–84). Herein we describe for the first time the DRB1*0405, DRB1*1302, DRB1*0802, DRB1*0901, and DRB1*0901 assays. Experiments addressing the βmolecule specificality of these new assays are also described in the present section.

The βmolecule specificity of DRB1*0101 is coexpressed with DRB1*0405, and the determination of the specificity of the DRB1*0405 binding assay is complicated in that the same radiolabeled ligand is used for both the DRB1*0405 and DRB1*0405 binding assays. Since βmolecules are typically expressed at 5- to 10-fold higher levels than β1 molecules (85), and all binding assays are performed utilizing limiting DR amounts, it would be predicted that the dominant specificity detected in the assay would be DRB1*0405. To verify that this was indeed the case, the binding pattern of a panel of 58 different synthetic peptides in the putative DRB1*0405-specific assay was compared with that obtained in a DRB1*0101-specific assay (which uses a DRB1*0101 fibroblast as the source of class II molecules). Two very distinct binding patterns were noted, and in several instances a peptide bound to one DR molecule with high affinity, and did not bind to the other (data not shown).

For DRB1*1201 binding assays, the EBV-transformed cell line Herluf was utilized as the source of DR molecules. The specific DRB3 (DR52) type expressed by, and hence copurified from, Herluf is unknown. Thus, the βmolecule specificity of the DRB1*1201 assay has been inferred from two observations. First, the radiolabeled ligand utilized is a single substitution analogue of a peptide eluted from DRB1*1301 molecules, as reported by Fark et al.; the DRB1*1201 specificity of this peptide is indicated by the fact that DR52 molecules had been removed from cell lysates before elution by immunoprecipitation (34). Second, because βmolecules are typically expressed at 5- to 10-fold higher levels than β1 molecules (85), it is unlikely that any binding of 1200/05 to DR52 molecules would constitute a significant signal in the DRB1*1201 binding assay.

The DRB1*1302 assay utilizes as the source of class II molecules the EBV-transformed homozygous cell line HO301, which coexpresses DRB3*0101 (DR52a). While the radiolabeled ligand used in the DRB1*1302 assay is different from that used for the DRB3*0101 assay, the ligand is related (i.e., a single substitution analogue) to a high affinity DRB3*0101 binder. As was done in the case of DRB1*0405, the specificity of the assay was investigated by analyzing the binding capacity of a panel of naturally occurring peptides for DRB1*1302 and DRB3*0101. The two assays demonstrated completely different binding specificities. For example, in terms of relative binding, T1727-1228 binds 63-fold better in the DRB3*0101 assay than in the DRB1*1302 assay. Conversely, the invariant chain peptide CLIP 80-103 binds 189-fold better in the DRB1*1302 assay. The DR1*1302 assay. In conclusion, these data demonstrated that the binding of the radiolabeled peptide 650.22 to purified class II MHC from the HO301 cell line is specific for DRB1*1302.

The β3 specificity of the DRB1*0901 and DRB1*0901 assays is obvious in that no β3 (and/or β1 and β2) molecule is expressed.
DRB1*0901. The specificity of DRB1*0901 assay is inferred from previous studies that have shown that the TT830-843-radiolabeled probe peptide does not bind to DRB4*0101 molecules (51).

Algorithm coefficient generation

Numerical algorithm scores for peptides were obtained as previously described (66, 86) using the polynomial method. The basic premise of this method is independent binding of individual side chains. When residue R occurs at position i in the peptide, it is assumed to contribute a constant amount R_i to the free energy of binding of the peptide irrespective of the sequence of the rest of the peptide. Parameters R_i have been estimated from a library of 384 peptides.

All peptides in the library contain the P1-P6 motif, so that they all contain the “correct” residue at the anchor positions. If multiple alignments of a given peptide are possible, only the highest scoring alignment is utilized, following an iterative procedure. For all i positions, anchor and non-anchor alike, the geometric mean of the average relative binding (ARB) of all peptides carrying R is calculated relative to the remainder of the group, and used as the estimate of R_i. The ARB values calculated from the library for DRB1*0401, DRB1*0101, and DRB1*0701 are shown in Figures 1 and 2, A and B, respectively.

To calculate an algorithm score of a given peptide in a test set, the ARB values corresponding to the sequence of the peptide are multiplied. If this product exceeds a chosen threshold, the peptide is predicted to bind.

Table I. HLA-DR binding assays utilized in the present study

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Allele</th>
<th>Alias</th>
<th>Cell line</th>
<th>Radiolabeled probe</th>
<th>Ref.</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR1</td>
<td>DRB1*0101</td>
<td>(DR1)</td>
<td>LG2</td>
<td>HA Y307-319&quot;</td>
<td>77</td>
<td>01 is the most prevalent DR1 allele.</td>
</tr>
<tr>
<td>DR2</td>
<td>DRB1*1501</td>
<td>(DR2w2b)</td>
<td>L466.1</td>
<td>MBP 88-102Y&quot;</td>
<td>77</td>
<td>01 is the most prevalent DR2 allele.</td>
</tr>
<tr>
<td>DR3</td>
<td>DRB1*0301</td>
<td>(DR3w17)</td>
<td>MAT</td>
<td>MT 65kd Y3-13 analog&quot;</td>
<td>77</td>
<td>01 is the most prevalent DR3 allele in most major populations.</td>
</tr>
<tr>
<td>DR4</td>
<td>DRB1*0401</td>
<td>(DR4w4)</td>
<td>Preiss</td>
<td>Non-natural peptide YAR&quot;</td>
<td>77</td>
<td>01 is the most prevalent DR4 allele in most populations.</td>
</tr>
<tr>
<td>DR5</td>
<td>DRB1*0405</td>
<td>(DR4w15)</td>
<td>KT3</td>
<td>Non-natural peptide YAR</td>
<td>This paper</td>
<td>05 is the most prevalent DR4 allele in the Orient.</td>
</tr>
<tr>
<td>DR6</td>
<td>DRB1*0701</td>
<td>(DR7)</td>
<td>Pitout</td>
<td>TT 830-843&quot;</td>
<td>77</td>
<td>01/02 vary at only one position, which is outside the binding groove.</td>
</tr>
<tr>
<td>DR7</td>
<td>DRB1*0802</td>
<td>(DR8w2)</td>
<td>OLL</td>
<td>TT 830-843</td>
<td>This paper</td>
<td>02 is dominant in most major population groups. 02 and 03 have nearly identical binding specificities (J. Sidney and A. Sette, unpublished observations).</td>
</tr>
<tr>
<td>DR8</td>
<td>DRB1*0901</td>
<td>(DR9)</td>
<td>HID</td>
<td>TT 830-843</td>
<td>This paper</td>
<td>DR9 alleles are products of a silent mutation.</td>
</tr>
<tr>
<td>DR9</td>
<td>DRB1*1101</td>
<td>(DR5w11)</td>
<td>Sweig</td>
<td>TT 830-843</td>
<td>77</td>
<td>01 is the most prevalent DR11 allele.</td>
</tr>
<tr>
<td>DR10</td>
<td>DRB1*1201</td>
<td>(DR5w12)</td>
<td>Herluf</td>
<td>C1R-derived peptide&quot;</td>
<td>34</td>
<td>01/02 are evenly distributed. These alleles differ at position 67, which is outside the binding groove and therefore would not be predicted to strongly influence peptide binding.</td>
</tr>
<tr>
<td>DR11</td>
<td>DRB1*1302</td>
<td>(DR6w19)</td>
<td>HO301</td>
<td>650.22 (TT 830–843 analogue)&quot;</td>
<td>82</td>
<td>02 is slightly more prevalent overall than 01. These alleles vary at position 86 (critical in determining the P1 anchor specificity).</td>
</tr>
<tr>
<td>DR12</td>
<td>DRB5*0101</td>
<td>(DR4, DR7, DR9)</td>
<td>GM3107</td>
<td>TT 830-843&quot;</td>
<td>77</td>
<td>01 is the most prevalent allele.</td>
</tr>
<tr>
<td>DR13</td>
<td>DRB4*0101</td>
<td>(DR2w2a)</td>
<td>L257.6</td>
<td>Non-natural peptide YAR&quot;</td>
<td>77</td>
<td>0101 is essentially the only allele.</td>
</tr>
</tbody>
</table>

" YPKYVKQNTLKLAT,  
"VHIFKKNIVTPRTPPY,  
"YKIADFEEARR,  
"YARFGQQTLKQKT,  
"QYIKANSFGKITE,  
"EALIHQLKINPYVLS,  
"QYIKANAKFITE.

Results

DR binding affinity of antigenic peptides recognized by DR-restricted T cells

To define a biologically significant threshold of DR binding affinity, we compiled a database of the binding affinities of 32 DR-restricted epitopes for their restricting element (31, 51, 83; Sette et al., unpublished observations). In approximately half of the cases (15 of 32 epitopes), DR restriction was associated with good binding affinities (less than 100 nM), and in the other half (16 of 32) with intermediate affinity (IC_50 s in the 100–1000 nM range). In only 1 of 32 cases (3.1%) was DR restriction associated with good binding affinities (less than 100 nM), and in the other half (16 of 32) with intermediate affinity (IC_50 s in the 100–1000 nM range). In only 1 of 32 cases (3.1%) was DR restriction associated with good binding affinities (less than 100 nM), and in the other half (16 of 32) with intermediate affinity (IC_50 s in the 100–1000 nM range).

In summary, this analysis suggested that 1000 nM may be defined as an affinity threshold associated with immunogenicity in the context of DR molecules.

P1 and P6 anchors are necessary but not sufficient for DRB1*0401 binding

Several independent studies (26–30, 32, 33) have pointed to a crucial role in DRB1*0401 binding of a large aromatic or hydrophobic residue (F, W, Y, L, I, V, or M) in position 1 (P1), near the
N terminus of the peptide and of a nine-residue core region (residues 1 through 9). In addition, these same studies demonstrated an important role for the residue in P6 of this nine-residue core region, where short and/or hydrophobic residues (S, T, C, A, P, V, I, L, or M) were generally preferred. On the basis of these results, a motif based on the presence of the two P1 and P6 anchors has been proposed (26–30, 32, 33). This motif is referred to throughout this paper as the "P1-P6" motif.

In the present set of experiments, a library of 384 synthetic peptides derived from naturally occurring sequences was analyzed for DRB1*0401 binding capacity, and then screened for the presence of the P1-P6 motif within a nine-residue core region. This set was found to contain 80 DRB1*0401 binders. Seventy-seven of these 80 DRB1*0401 binders (96%) carried the P1-P6 motif. Conversely, 125 were "P1-P6 negative." Only 3 of them (2.4%) bound appreciably to purified DRB1*0401 as opposed to 77 of 259 (30%) of the "P1-P6-positive" peptides. These results demonstrate that the presence of suitable P1 and P6 anchors is necessary but not sufficient for DRB1*0401 binding.

A detailed DRB1*0401 motif

To derive a more detailed DRB1*0401 motif, we employed a strategy previously utilized in the case of peptide class I interactions (66, 70, 86, 87) (see Materials and Methods for details). For each peptide, nine-residue-long core regions were aligned on the basis of the primary P1 and P6 anchors. Then, the average binding affinity of peptides carrying a particular residue was calculated for each position, relative to the remainder of the group. Following this method, ARB values were compiled. These values also represent a map of the positive or negative effect of each of the 20 naturally occurring amino acids in DRB1*0401 binding capacity when occupying a particular position, relative to the P1-P6 primary anchors (Fig. 1).

Variations in ARB values greater than fourfold (ARB ≥ 4 or ≤ 0.25) were arbitrarily considered significant and indicative of secondary effects of a given residue on DR-peptide interactions. Most secondary effects were associated with P4, P7, and P9. These positions correspond to secondary anchors engaging shallow pockets on the DR molecule (19, 20). Specifically significant positive secondary effects were detected for M in P2, P7, and P9 (ARB values of 12.79, 8.11, and 4.05, respectively); T in P3 (ARB = 4.34); I in P5 (ARB = 4.4); and H in P7 and P9 (ARB = 13.77 and 5.15, respectively). In addition, negative effects were detected for W at P4 (ARB = 0.21), R at P7 (ARB = 0.14), and W, D, and E at P9 (ARB values of 0.22, 0.24, and 0.25, respectively).

Development of a DRB1*0401-specific algorithm

Next, the ARB values were utilized to develop a DRB1*0401-specific algorithm. To predict DRB1*0401 binding propensity, nine-residue core sequences carrying the P1-P6 motif were aligned and scored by multiplying, for each position, the ARB value of the appropriate amino acid (24, 28, 88). According to this procedure, a numerical "algorithm score" was obtained. If multiple P1-P6 alignments were possible, algorithm scores were calculated for each one and the best score was selected.

For example, the HA307-319 peptide (PKYVKQNTLKLAT; DRB1*0401 binding of 45 nM) was aligned so that the Y 309 corresponded to P1. Accordingly, P1 = Y (with an ARB of 1.07), P2 = V (with an ARB of 3.34), P3 = K (1.25), P4 = Q (1.61), P5 = N (1.89), P6 = T (1.86), P7 = L (1.36), P8 = K (0.64), and P9 = L (0.83). By multiplying the ARB values of P1 through P9, an algorithm value of 18.34 was obtained.

![FIGURE 1. DRB1*0401 algorithm: ARB values. ARB values of peptides bearing the P1-P6 primary anchors as a function of the different residues at nonanchor positions to DRB1*0401. Data were analyzed and tabulated as described in Materials and Methods. The panel was composed of 384 peptides based on naturally occurring and non-natural sequences derived from various viral, tumor, or bacterial origins. Values ≥4 are indicated by bold type. Values ≤0.25 are indicated by italicized type and underlines.](http://www.jimmunol.org/DownloadedFrom/10.4049/jimmunol.1200430-1057.f1)
The predictive capacity of the algorithm was then examined in a blind prediction test. Data from an independent set of 50 peptides that had not been utilized in the derivation of the algorithm, but whose binding affinities were known, was utilized. As shown in Table II, the algorithm was effective in identifying DRB1*0401 binders in this independent peptide set. Eighteen peptides had algorithm scores of 0.734 or better. Among these 18 peptides were all of the good binders (3 of 3; 100%), and 8 of 11 (73%) of all intermediate binders, present in the test set of 50 peptides. Increasing the cutoff value to 2.617 identified a set of 9 peptides, 7 of which (78%) were either good or intermediate binders. This set also contained 7 of the 14 (50%) binders contained in the blind prediction peptide set. Taken together, these data support the validity of the DRB1*0401-specific algorithm described above.

**Detailed DRB1*0101 and DRB1*0701 motifs**

DRB1*0101 and DRB1*0701 are common HLA alleles for which we have generated a substantial peptide binding database ([22, 29, 32, 51, 83]; Sette et al., unpublished observations). A preliminary analysis of these data suggested that these alleles might significantly overlap in their peptide binding specificities with DRB1*0401.

For these reasons, the binding to purified DRB1*0101 and DRB1*0701 molecules was analyzed for the same set of 384 peptides previously analyzed in terms of DRB1*0401 binding. It was found that this set contained 120 and 59 binders for the DRB1*0101 and DRB1*0701 alleles, respectively. A total of 158 peptides were capable of binding either DRB1*0101, DRB1*0401, or DRB1*0701. A large fraction of these (73 of 158; 46%) were also degenerate binders, binding two or more of the three alleles thus far considered. Analysis of the sequences revealed that more than 90% of the DRB1*0101 and DRB1*0701 good and intermediate binders carried the P1-P6 motif. Most importantly, 72 of 73 (99%) degenerate DR binders carried the P1-P6 motif (data not shown). This analysis suggests that these motifs might be utilized to effectively predict degenerate DR binders.

Analogous to what was described above for DRB1*0401 molecules, specific motifs were designed for the DRB1*0101 and DRB1*0701 alleles (Fig. 2, A and B).

As in the case of DRB1*0401, most secondary effects were concentrated in P4, P7, and P9. P4 was especially prominent in the case of DRB1*0101, while P7 was the most prominent secondary anchor for DRB1*0701. Specific algorithms based on these motifs were subsequently developed. It was found that the cutoff values necessary to predict 75 or 90% of the binders were 1.570 and 0.183 for DRB1*0101, and 9.106 and 1.749 for DRB1*0701, respectively. Forty to sixty percent of the peptides selected using these cutoff values were in fact good or intermediate binders to the respective allele (data not shown).

**Prediction of DRB1*0101-DRB1*0401-DRB1*0701 (DR1-4-7) degenerate binders**

Next, we examined whether predictions based on concurrent analysis by these algorithms would allow the identification of degenerate binders. For this purpose, the sequences of the 384 peptides in our database were simultaneously screened with the three (DRB1*0101, DRB1*0401, and DRB1*0701) specific algorithms. It was found that 100 peptides were predicted (using the 75% cutoff values) to bind either two or three of the alleles considered. We took these peptides as predicted degenerate binders. This set contained 59 of 73 (81%) of the peptides that were in fact capable of degenerate DRB1*0101-DRB1*0401-DRB1*0701 binding (defined as the capacity to bind two or more alleles among DRB1*0101, DRB1*0401, or DRB1*0701) (Table III). Thus, the individual algorithms may be combined to predict degenerate binders.

**Definition of a target set of DR specificities, representative of the world population**

The data presented in the preceding sections illustrate how peptides capable of binding multiple DR alleles can be identified by the use of the combined DR1-4-7 algorithms. Whether the peptides exhibiting degenerate DRB1*0101, DRB1*0401, and DRB1*0701 binding behavior would also bind other common DR types as well was examined next. As a first step, a set of target DR types representative of a large fraction of the world population, irrespective of the ethnic population of origin, was defined. For this purpose, seven additional DR Ags were considered. Each DR Ag considered in this study (its estimated frequency in various ethnicities according to the most recent HLA workshop (89) and the main subtypes identified to date) is listed in Table IV.

For the purpose of measuring peptide binding affinity to the various DR molecules, one representative subtype for each DR Ag was chosen (Table I). One exception is represented by the DR4 Ag, for which significant differences in peptide specificity between the DRB1*0401 and DRB1*0405 are known (88, 90). These alleles are frequent in Caucasians and Asians, respectively, and therefore are included in the set of representative DR binding assays.

It should be noted that the set of assays chosen as representative is mostly focused on allelic products of the β1 gene. This emphasis is made because these molecules appear to be the most abundantly expressed (85), and serve as the dominant restricting element for most human class II responses analyzed to date. However, we have included in our analysis DRB5*0101, which is a functional restriction element (84, 91), and whose peptide binding specificity is similar to the specificity of several common DRβ1 allelic products (31, 51, 83, 84, 90).

**A general strategy for prediction of DR degenerate binders**

Much previous work (see, for example, Ref. 32) has suggested that many DR types may share overlapping peptide binding specificities. To test whether the DR1-4-7 combined algorithms would also predict binding to other common DR types, the capacity of three different groups of synthetic peptides to bind the panel of purified HLA-DR molecules defined above was measured. The three different peptide sets were: 1) 34 peptides that did not score positive in the DR1-4-7 algorithm (nonpredictions); 2) 24 peptides that did score positive for the DR1-4-7 algorithm, at the 75% cutoff level, but had been found upon actual testing not to be degenerate
FIGURE 2. A, DRB1*0101 algorithm: ARB values. ARB values of peptides bearing the P1-P6 primary anchors as a function of the different residues at nonanchor positions to DRB1*0101. Data were analyzed and tabulated as described in Materials and Methods. The panel was composed of 384 peptides based on naturally occurring and non-natural sequences derived from various viral, tumor, or bacterial origins. Values $\geq 4$ are indicated by bold type. Values $\leq 0.25$ are indicated by italicized type and underlined.

B, DRB1*0701 algorithm: ARB values. ARB values of peptides bearing the P1-P6 primary anchors as a function of the different residues at nonanchor positions to DRB1*0701. Data were analyzed and tabulated as described in Materials and Methods. The panel was composed of 384 peptides based on naturally occurring and non-natural sequences derived from various viral, tumor, or bacterial origins. Values $\geq 4$ are indicated by bold type. Values $\leq 0.25$ are indicated by italicized type and underlined.
DRB1*0101-DRB1*0401-DRB1*0701 binders (“wrong” predictions); and 3) 22 peptides that scored positive in the DR1-4-7 algorithm, and also proved upon experimental testing to be actual DRB1*0101-DRB1*0401-DRB1*0701 degenerate binders (correct predictions).

Within the set of “nonpredicted” peptides, only 3 of 34 (9%) bound at least 2 of the DRB1*0101, DRB1*0401, or DRB1*0701 molecules. Peptides from the “wrong predictions” peptide set, which by definition bound at most only 1 DR type among DRB1*0101, DRB1*0401, or DRB1*0701, were also poorly degenerate for other DR types. Only 2 peptides bound 3 DR molecules, and no peptide bound 4 or more of the DR molecules tested (data not shown).

These results are contrasted by data obtained with the set of peptides that were correctly predicted to be degenerate by the use of the combined DR1-4-7 algorithms (Table V). Fourteen of 22 peptides (64%) bound 5 or more alleles. Three of these peptides were remarkably degenerate (1188.16, 1188.32, and 1188.34), binding at least 9 of the 12 DR molecules tested. In summary, these results suggest that a strategy based on the sequential use of a combined DR1-4-7 algorithm, and quantitative DR binding assays, can be utilized to identify broadly cross-reactive DR binding peptides. 

Identification of an HLA-DR supertype

The data presented above confirmed that several common DR types are characterized by largely overlapping peptide binding repertoires. On this basis, in analogy to the case of HLA class I molecules, it could be proposed that such DR molecules could be grouped in a DR supertype, defined and characterized by similar (albeit not identical) peptide binding specificities. This issue was analyzed in more detail by examining the binding patterns of the

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Alleles</th>
<th>Caucasian (%)</th>
<th>Black (%)</th>
<th>Japanese (%)</th>
<th>Chinese (%)</th>
<th>Hispanic (%)</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR1</td>
<td>DRB1*0101-03</td>
<td>18.5</td>
<td>8.4</td>
<td>10.7</td>
<td>4.5</td>
<td>10.1</td>
<td>10.4</td>
</tr>
<tr>
<td>DR2</td>
<td>DRB1*1501-03</td>
<td>19.9</td>
<td>14.8</td>
<td>30.9</td>
<td>22.0</td>
<td>15.0</td>
<td>20.5</td>
</tr>
<tr>
<td>DR3</td>
<td>DRB1*0301-02</td>
<td>17.7</td>
<td>19.5</td>
<td>0.4</td>
<td>7.3</td>
<td>14.4</td>
<td>11.9</td>
</tr>
<tr>
<td>DR4</td>
<td>DRB1*0401-12</td>
<td>23.6</td>
<td>6.1</td>
<td>40.4</td>
<td>21.9</td>
<td>29.8</td>
<td>24.4</td>
</tr>
<tr>
<td>DR7</td>
<td>DRB1*0701-02</td>
<td>26.2</td>
<td>11.1</td>
<td>1.0</td>
<td>15.0</td>
<td>16.6</td>
<td>14.0</td>
</tr>
<tr>
<td>DR8</td>
<td>DRB1*0801-05</td>
<td>5.5</td>
<td>10.9</td>
<td>25.0</td>
<td>10.7</td>
<td>23.3</td>
<td>15.1</td>
</tr>
<tr>
<td>DR9</td>
<td>DRB1*0901-09012</td>
<td>3.6</td>
<td>4.7</td>
<td>24.5</td>
<td>19.9</td>
<td>6.7</td>
<td>11.9</td>
</tr>
<tr>
<td>DR11</td>
<td>DRB1*1101-05</td>
<td>17.0</td>
<td>18.0</td>
<td>4.5</td>
<td>19.4</td>
<td>18.1</td>
<td>15.5</td>
</tr>
<tr>
<td>DR12</td>
<td>DRB1*1201-02</td>
<td>2.8</td>
<td>5.5</td>
<td>13.1</td>
<td>17.6</td>
<td>5.7</td>
<td>8.9</td>
</tr>
<tr>
<td>DR13</td>
<td>DRB1*1301-06</td>
<td>21.7</td>
<td>16.5</td>
<td>14.6</td>
<td>12.2</td>
<td>10.5</td>
<td>15.1</td>
</tr>
<tr>
<td>Total</td>
<td>97.0</td>
<td>83.9</td>
<td>98.8</td>
<td>95.5</td>
<td>95.6</td>
<td>94.7</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

In the present report, detailed maps of secondary interactions have been derived for three different HLA-DR molecules (DRB1*0101, DRB1*0401, and DRB1*0701). The peptide binding specificity of 9 additional DR types has also been analyzed. Together, these 12 DR alleles are representative of DR types common among the worldwide population. Furthermore, it was demonstrated that a set of at least 7 different DR types share overlapping peptide binding repertoires; and, consequently, that broadly degenerate HLA-DR binding peptides are a relatively common occurrence.

We would like to discuss these data in the context of our current understanding of peptide-class II interactions, as well as in the context of recently described class I supermotifs (66–70). The potential implications of broadly degenerate class II epitopes for epitope-based vaccine design should also be considered.

First, the data presented herein illustrate how the vast majority of the peptides binding with good affinity to DRB1*0401,
Table V. Degenerate DR1-4 binders correctly predicted

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
<th>DRB1*0101</th>
<th>DRB1*0401</th>
<th>DRB1*0701</th>
<th>DRB1*1001</th>
<th>DRB1*1101</th>
<th>DRB1*1302</th>
<th>DRB1*0802</th>
<th>DRB1*0901</th>
<th>DRB1*1201</th>
</tr>
</thead>
<tbody>
<tr>
<td>1188.34</td>
<td>HNWVNHAPLAMKLI</td>
<td>1.1</td>
<td>1.2</td>
<td>1.0</td>
<td>1.1</td>
<td>1.3</td>
<td>1.2</td>
<td>1.3</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>1188.16</td>
<td>KSKYKLATSVLGALL</td>
<td>2.1</td>
<td>2.2</td>
<td>2.1</td>
<td>2.1</td>
<td>2.2</td>
<td>2.2</td>
<td>2.1</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>1188.32</td>
<td>GLAYKFVPGATTPY</td>
<td>4.3</td>
<td>5.4</td>
<td>5.2</td>
<td>5.3</td>
<td>6.0</td>
<td>5.7</td>
<td>5.9</td>
<td>5.6</td>
<td>5.9</td>
</tr>
<tr>
<td>1136.16</td>
<td>LTSQFFLAPVPFVTL</td>
<td>2.2</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>1136.21</td>
<td>IPQWEKPATIKVLPA</td>
<td>2.2</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>1136.11</td>
<td>VVPAPSSFKLPIA</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>27.392</td>
<td>SSFVNNSNIGILM</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>27.417</td>
<td>VKNVIGPFKAVCG</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>1136.38</td>
<td>SSIIFGAPILHSGCC</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>27.388</td>
<td>MRKLAILSVSFPLFV</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>27.403</td>
<td>LVVLLIFHINGKIK</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>1136.71</td>
<td>AGGLGNSTVLLGGV</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>1136.14</td>
<td>FACTFLPLTSQFLF</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.5</td>
<td>3.0</td>
<td>2.8</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
</tbody>
</table>

* NA, not applicable.
association with the DR supertype presented here. In this context it is interesting to note that Hammer et al. (26) noted that good DRB1*1101 binding peptides are frequently characterized by a positively charged P6 anchor, which would be poorly compatible with the DR supermotif proposed herein. It is also interesting to note that Sidney et al. (31) and others (93, 94) proposed that DRB1*0301 binds a set of peptides largely distinct from those bound by other common DR types. Future studies will have to determine whether any of the molecules listed above can be grouped in additional DR supertypes. Our group is currently investigating whether analysis of polymorphic residues lining the peptide binding pockets of DR can be utilized to classify and predict HLA-DR supertypes.

The HLA-DR supertype described herein stands in contrast to the recently described HLA class I supermotifs (66–70). Whereas class I supertypes are insular in that they define distinct sets of alleles whose specificities do not generally overlap with other supertypes, the HLA-DR supertype herein described is more diffuse, and overlaps to some degree with the repertoire of other alleles. Furthermore, the four class I supertypes that have been described to date are all approximately equally frequent (35–50%) among the worldwide population. But, on the basis of the data presented in Tables IV and VI, even if other DR supertypes exist, the DR supertype described herein is likely to be by far the most abundantly represented worldwide.

In terms of usefulness of this method to predict DR binding epitopes, it should be noted that using a stringent criteria for each of the DRB1*0401, DRB1*0101, and DRB1*0701 algorithms identifies 81% of the total degenerate DR1-4-7 binders, while predicting only 41 total false positives (Table III). Less stringent criteria (90% cutoff values) identifies 92% of total degenerate DR1-4-7 binders, but also predicts a total of 80 false positives.

In other words, as expected, more stringent algorithm scores predicted a smaller fraction of the total binders present in the set, but at the same time, fewer false positive peptides were identified. Similar patterns of stringency and predictability have been noted in the case of class I motifs as well, where the presence of “extended” motifs predicts essentially all binders, but also a large fraction of false positives. Conversely, the presence of very stringent “canonical” motifs predict few false positives, but only a fraction of the binders. This situation appears to be true irrespective of the actual algorithm utilized (86, 95).

Finally, we would like to point out the possible relevance of these data in terms of the development of epitope-based vaccines. Class II-restricted HTL have been implicated in protection from, and termination of, many important diseases (1–14). Inclusion of well-defined class II epitopes in prophylactic or therapeutic vaccines may allow the immune response to be focused on conserved or subdominant epitopes, thereby avoiding suppressive determinants. Based on the data presented herein, a single DR supertype may allow coverage in the 50 to 80% range, depending on the ethnicities considered. It is thus possible that broad and not ethically biased population coverage could be achieved by considering a very limited number of peptide binding specificities.

Acknowledgments

The secretarial assistance of Diana Pack and Ethel Beltran in preparing the manuscript is gratefully acknowledged. We also thank Alicia Johnson for her expert technical assistance and Ajesh Maewal for peptide synthesis.

References

onol. 11:129.


