Naturally Processed HLA Class II Peptides Reveal Highly Conserved Immunogenic Flanking Region Sequence Preferences That Reflect Antigen Processing Rather Than Peptide-MHC Interactions

Andrew J. Godkin, Kathrine J. Smith, Anthony Willis, Maria V. Tejada-Simon, Jingwu Zhang, Tim Elliott and Adrian V. S. Hill

*J Immunol* 2001; 166:6720-6727; doi: 10.4049/jimmunol.166.11.6720
http://www.jimmunol.org/content/166/11/6720

---

**Why The JI?**

- **Rapid Reviews! 30 days** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Speedy Publication!** 4 weeks from acceptance to publication

---

**References**

This article *cites 35 articles*, 15 of which you can access for free at:
http://www.jimmunol.org/content/166/11/6720.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
Naturally Processed HLA Class II Peptides Reveal Highly Conserved Immunogenic Flanking Region Sequence Preferences That Reflect Antigen Processing Rather Than Peptide-MHC Interactions

Andrew J. Godkin,2*† Katharine J. Smith,‡ Anthony Willis,§ Maria V. Tejada-Simon,§ Jingwu Zhang,§ Tim Elliott,* and Adrian V. S. Hill*

MHC class II heterodimers bind peptides 12–20 aa in length. The peptide flanking residues (PFRs) of these ligands extend from a central binding core consisting of nine amino acids. Increasing evidence suggests that the PFRs can alter the immunogenicity of T cell epitopes. We have previously noted that eluted peptide pool sequence data derived from an MHC class II Ag reflect patterns of enrichment not only in the core binding region but also in the PFRs. We sought to distinguish whether these enrichments reflect cellular processes or direct MHC-peptide interactions. Using the multiple sclerosis-associated allele HLA-DR2, pool sequence data from naturally processed ligands were compared with the patterns of enrichment obtained by binding semicombinatorial peptide libraries to empty HLA-DR2 molecules. Naturally processed ligands revealed patterns of enrichment reflecting both the binding motif of HLA-DR2 (position P1, aliphatic; P4, bulky hydrophobic; and P6, polar) as well as the nonbound flanking regions, including acidic residues at the N terminus and basic residues at the C terminus. These PFR enrichments were independent of MHC-peptide interactions. Further studies revealed similar patterns in nine other HLA alleles, with the C-terminal basic residues being as highly conserved as the previously described N-terminal prolines of MHC class II ligands. There is evidence that addition of C-terminal basic PFRs to known peptide epitopes is able to enhance both processing as well as T cell activation. Recognition of these allele-transcending patterns in the PFRs may prove useful in epitope identification and vaccine design. The Journal of Immunology, 2001, 166: 6720–6727.

Hum an histocompatibility leukocyte antigens glycoproteins present peptides to cognate T cells, leading to specific immune activation. HLA class I molecules bind peptides that are usually 8 or 9 aa in length with the N- and C-termini bound in conserved pockets at either end of a groove. In contrast, HLA class II molecules tend to bind nested sets of longer peptides, 12–20 aa in length. These longer peptides bind with a core region of similar length, with the nonbound flanking residues of the peptide extending from the ends of an open groove (1). The side chains of these core residues bind in a series of pockets along the groove, spaced from the bound N terminus of the peptide at position 1 (P1), P4, P6, P7, and P9 (2–5). In HLA-DR molecules, the pockets in the groove are lined with the polymorphic residues of the HLA-DR β-chain, generating specific binding motifs for individual HLA-types and subtypes. Eluted peptide sequence data reveal patterns of enrichment of amino acids, which reflect the binding requirements of these allele-specific pockets (6–9).

The peptide flanking residues (PFRs) not bound in the central core region can interact with the HLA class II heterodimer and enhance the binding affinity and stability/conformation of the tri-molecular complex (8, 10, 11). However, it is becoming increasingly clear that the PFRs may also have a more specific role in immune recognition (12). TCR activation can be influenced by these PFRs as well as the residues in the core binding region (13, 14). Furthermore, we have shown that the additional information in the PFRs is necessary in certain instances for accurate prediction of T cell epitopes from the parent protein, possibly reflecting patterns of amino acid enrichments in these flanking regions distinct from the core binding region (9).

Initially, to examine further the generation of flanking regions, the binding of peptides to the heterodimer encoded by HLA-DRA1*0101, B1*1501 (HLA-DR2(b)) was studied. HLA-DR2 has been associated with susceptibility to multiple sclerosis (15). In the context of multiple sclerosis, the interaction of the immunodominant myelin basic protein (MBP) peptide 84–102 (DENPVVHFFKNI VTPRLP) has been well characterized (14, 16, 17). The core binding region is 89–97, with two important anchors at P1 and P4 (Val90 and Phe92). The preferences of other anchor positions in HLA-DR2 are not so clear. These anchors alone are not enough to ensure high affinity binding, as demonstrated by the poor binding of the 9-aa core peptide compared with a peptide with the addition of either N- or C-terminal flanking residues. These

---

*Department of Medicine, Molecular Immunology Group, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom;† Department of Medicine, Laboratory of Molecular Medicine, The Children’s Hospital, Howard Hughes Medical Institute, Boston, MA 02115;‡ Department of Biochemistry, Medical Research Council Immunochimistry Unit, Oxford, United Kingdom;§ Department of Neurology, Baylor College of Medicine, Houston, TX 77030; and Department of Medicine, Imperial College, London, U.K. E-mail address: a.godkin@ic.ac.uk

Copyright © 2001 by The American Association of Immunologists 0022-1767/01/$02.00
flanking residues may also influence the activation of T cells. The majority of CD4+ T cell clones interact with the residues in the core region His<sup>90</sup>, Phe<sup>91</sup>, and Lys<sup>93</sup>. However, many clones are also sensitive to truncations or substitutions of the flanking residues in the core binding region. The altered T cell activation does not correlate with a change in binding affinities. It appears that the TCR is either directly interacting with the PFRs or the alterations are changing the physical contours of the core peptide.

The binding of peptides to HLA-DR2 was studied by a combination of eluted peptide sequence data, in vitro peptide binding assays, and combinatorial peptide technology. Complete eluted peptide sequence data identified both the nature and position of anchor residues in the core binding region of HLA-DR2, as well as the characteristics of the N- and C-terminal PFRs that appeared to be highly conserved among nine additional HLA alleles studied. Using in vitro binding assays and semirandom combinatorial peptide libraries bound to empty HLA-DR2 molecules purified from an insect cell expression system, the following questions were addressed. First, is there a difference between the patterns seen in the termini of the natural processed peptides bound in vivo and synthetic peptides selected in vitro; second, does the core binding region of the peptide influence these termini; and finally, do naturally occurring residues in the PFRs enhance binding in vitro and offer an explanation for any enrichments found?

**Materials and Methods**

**Peptide synthesis**

The candidate peptides were synthesized by F-moc chemistry using an automated synthesizer (Apex 396 multiple peptide synthesizer; Advanced ChemTech, Cambridge, U.K.). The purity was assessed by reversed phase HPLC, and the concentration was measured using a bicinchoninic acid assay. The sequence XXXXVHFKNIVXXXX was synthesized corresponding to the MBP core peptide binding sequence with random termini. X represents any of 19 amino acids (cysteine not incorporated). For the X position, all 19 amino acids were mixed into a single well of the peptide synthesizer block. The other semirandom peptides were synthesized with modified methodology as previously described (20) to encourage equimolar incorporation of amino acids in the random positions (Echaz Microcollections, Tübingen, Germany). In brief, double couplings were performed with equimolar mixtures of F-moc-L-amino acids, which were used in an equimolar ratio with respect to the coupling sites of the resin, and extended coupling times were used. The following sequences were made: XXXXHAPKFNIVXXXXXXX XXXXHIQLNIVXXXXXX. The first MBP sequence was modified (Ala for Phe at position 91). This was to discourage any possible sliding of the peptide binding by the use of F91 as an anchor rather than the usual F92. The second sequence contained a different core binding peptide to ensure that bias was not introduced by a particular sequence.

**Purification of naturally produced HLA class II molecules**

The method used was based on the original description by Gorga et al. (21). EBV-transformed HLA-homozygous B cell lines (obtained from the European Collection of Cell Culture, Salisbury, U.K.) were grown for each HLA-type. Cells were grown up to a final concentration of 1–2 × 10<sup>6</sup> cells/ml in 2 L roller bottles with RPMI 1640 supplemented by 10% FCS, 1% L-glutamine (10 mM), penicillin (100 U/ml), and streptomycin (50 μg/ml). Between 10<sup>7</sup> and 10<sup>8</sup> cells were spun down, washed in cold PBS, and then lysed on ice with 3% Nonidet P-40 in PBS containing leupeptin (1 μg/ml), pepstatin (1 μg/ml), and 5 mM EDTA. The lysate was spun at 100,000 × g for 90 min. The supernatant was passed over a precolumn of Sepharose CL-4B followed by the affinity column with cyanogen bromide-activated Sepharose beads linked to the monoclonal Ab to HLA-DR L243 (22). After washing, the class II molecules were eluted from each column with 0.05 M diethylamine (pH 11.5). The eluates were immediately neutralized with 1 M Tris (pH 8.4) and concentrated by ultrafiltration (Centriprep; Amicon, Beverly, MA). The purity was analyzed by 12% SDS-PAGE.

**Purification of HLA-DR2 molecules from Schneider cells and binding to semirandom libraries**

*Drosophila* (S2) cells transfected with HLA-DR2 (DRA*10101, B*1501) (constructs given in Ref. 5) were grown in spinner flasks in ExCell 401 medium (Sigma, St. Louis, MO) supplemented with 1% FBS. At a density of 4 × 10<sup>6</sup> cells/ml, protein expression was induced with 1 mM CuSO<sub>4</sub>. The supernatant was harvested after 6 days by centrifugation at 4°C. HLA-DR2 was purified from the supernatant by affinity chromatography; DR2 was eluted from the column with 50 mM glycine (pH 11.5), and fractions were neutralized with 1 M Tris (pH 7.0). For peptide binding, 2 mg of each random peptide was dissolved in 2.2-M urea, and 100 μl was removed for analysis. The remainder was bound to 5 mg of empty HLA-DR2 at 37°C for 60 h. After peptide binding, the HLA-DR2 was concentrated using centricon (Amicon), and DR2-peptide complexes were purified by gel filtration using a Superdex 200 column (Amersham Pharmacia Biotech, Little Chalfont, U.K.) equilibrated in PBS.

**Pool Sequencing**

Naturally processed peptide ligands (as well as synthetic random peptides) were released from the HLA molecules by acid elution with 0.1% trifluoroacetic acid. The HLA molecules were removed from the peptides by ultrafiltration using a centrinor microconcentrator (Amicon). The peptides were separated on an Aquapore RP-300 (C8/100 × 2 mm) reversed phase column (Perkin-Elmer Applied Biosystems, Warrington, U.K.) connected to a Perkin-Elmer 172A narrow-bore HPLC system. The column was equilibrated in 0.1% trifluoroacetic acid, and the peptides were eluted with a linear gradient of acetonitrile increasing at 1% per min. The fractions containing peptides were pooled and sequenced for 15 cycles on an ABI 473A or an ABI Procise 494A protein sequencer (Perkin-Elmer Applied Biosystems). Any disproportionately large peaks were removed so as not to bias the sequencing of the rest of the pool. Sequencing data was integrated using Model 610A (version 2.1a) data analysis software obtained from Perkin-Elmer Applied Biosystems.

**Epitope prediction**

Total yields of amino acids diminish over each cycle of Edman degradation due to the nature of this process. To allow for this, the pool sequence data were organized using a procedure previously described to generate a matrix of relative enrichments (23). Briefly, the amount of each amino acid at each cycle was described as a percentage of the total quantity of all amino acids in that cycle. The average of each amino acid was then taken over the 15 cycles, and each position in the cycle was expressed as a proportion of the average. The resulting table was used with a computer program that scans given protein sequences iteratively and calculates for each 15 mer a probability of being processed and presented score. The protein sequences that were scanned were obtained from the Swiss-Frot and Protein Information Resource databases.

**Peptide binding assays**

Binding of synthetic peptides to purified HLA-DR2 molecules was assayed in vitro in a competitive inhibition assay against the promiscuously binding class II-associated invariant chain peptide (CLIP) (96–114), which was biotinylated. All incubations were conducted at 37°C. The 96-well flat-bottom ELISA immunoplate was blocked with PBS containing 3% BSA for 3 h. The plate was emptied, and serial 1/10 dilutions of candidate peptide were then added to the wells with 0.02 μg biotinylated marker CLIP and 0.1 μg HLA-DR2 in pH 5 buffer (1.95 g 2-(morpholino)ethane-sulfonic acid + 0.5 g 3-(3-cholamidopropyl)dimethylammonio-1-propanesulfonate + 4.1 g NaCl + 0.02% azide in 500 ml of water) for a total volume of 60 μl/well. The plate was incubated for 18 h. After this time, 10 μl of neutralizing buffer (1 M Tris (pH 8) + 10% BSA + 1% Tween 20 + 0.02% azide) was added to each well. The 70-μl aliquots were transferred to a 96-well plate ( precoated with anti-DR Ab (L243) and blocked as described above with PBS containing 3% BSA) and incubated for 1 h. The wells were emptied, washed extensively with PBS containing 1% Tween 20, and then washed with PBS alone. A total of 50 μl of 1/1000 avidin-HRP solution (Sigma) was added to each well and incubated for 15 min. The plate was emptied, washed, and dried as described above. A total of 50 μl/well of 1/1000 biotinylated anti-avidin solution (Sigma) was added and incubated for 30 min. The plate was again emptied, washed, and dried. A total of 50 μl/well of avidin-HRP solution was then added and incubated for 15 min.

The plate was emptied and washed with PBS with 1% Tween 20, then washed again with H<sub>2</sub>O. After the plate was dried, 100 μl of freshly made developing solution was added (5 ml of 0.1 M citrate buffer (pH 9) + 5 ml of 0.2 M phosphate buffer (pH 5) + 100 μl o-phenylenediamine at 40
mg/ml + 20 μl hydrogen peroxide). The plate was wrapped in silver foil and incubated (without stacking) for 30–120 min (until color changes were clearly visible). Then the plate was blocked with 100 μl of 12.5% H2SO4, and the color changes were measured at 492 nm in an ELISA reader.

The IC50 of unlabeled peptide required to inhibit binding of the CLIP was calculated as a concentration (μM). All the binding assays were performed at least twice and a mean for the IC50 calculated.

**IFN-γ enzyme-linked immunospot (ELISPOT) assays**

The biological effect of peptides on T cells was studied in a direct ex vivo assay using single cell IFN-γ release as a measure of Ag-specific effector function in a highly sensitive ELISPOT assay (Abs and streptavidin-alkaline phosphatase were obtained from Mabtech, Nacka, Sweden). PBMCs were extracted from whole blood by centrifugation over Lymphoprep (Nycomed, Oslo, Norway). A total of 5 × 10^5 T cells and 5 × 10^5 PBMCs (25% of total nucleated cells) were added per well of a special polymer-backed 96-well filtration plate MAIP-S-45 (Millipore, Moslheim, France). The wells were heat-treated FCS were added per well of a special polymer-backed 96-well filtration plate MAIP-S-45 (Millipore, Moslheim, France). The wells were precoated with monoclonal anti-IFN-γ (1-D1K) according to the manufacturer’s instructions. The peptides were tested in duplicate wells, having been added to a final concentration of 10 μg/ml, and compared with positive control (PFRs) and wells with no peptide. The plate was incubated at 37°C, 5% CO2 for 18 h. The plate was then washed with PBS and incubated with a second layer (7-B6-1; 1/1000 in PBS, 50 ml/well) for 2 h at room temperature. The plate was rinsed and developed with streptavidin-alkaline phosphatase and a chromogenic substrate. When the spots had emerged, P5 was added to a final concentration of 10 μg/ml, and the color changes were measured at 492 nm in an ELISA reader.

The line was plated out at 0.3 cells/well in the presence of 10 μL irradiated autologous PBMCs and 2 μg/ml PHA. Cultures were fed with fresh medium containing 50 IU/ml rIL-2 every 3–4 days. Growth-positive cells were tested in a proliferation assay after 10–12 days to identify Ag-specific clones. For further proliferation assays against MBP peptides and analogs, transfected L cells (HLA-DR2) were used. Irradiated L cells (2.5 × 10^5/well) were incubated in a microtiter plate for 2–3 h, pulsed with peptides for 1–2 h, and then washed. Cultures were set up in duplicate for each peptide, and [3H]thymidine incorporation measured after 72 h.

**Results**

**Pool sequence data from naturally processed peptides reveals enrichments in both the core binding region and the PFRs**

Naturally processed peptides were eluted from multiple batches of affinity purified HLA class II molecules and sequenced en masse. Each pool was obtained from ~1 mg of purified protein, and, for HLA-DR2, an example of one pool is shown in Table I. To allow for the diminishing yield seen over each cycle of degradation and to emphasize changes in each amino acid, the data was organized as described in Materials and Methods to generate a table showing the relative enrichments of the amino acids over 15 cycles (Table II). As has been previously described, the pattern of enrichments reflect the binding preferences of the pockets starting with cycles 4/5 corresponding approximately to the first anchor position P1 (6, 24). In keeping with other studied HLA-DR molecules with a Val at position 86 of the β-chain (25), there was clearly a hydrophobic pocket, P1, corresponding to the first position in the binding core, which accommodates aliphatic hydrophobic side chains, reflected by the enrichment of Ile, Leu, and Val. Two other anchor positions emerged, P4, which favors large bulky aromatic hydrophobic residues, and P6, which favors polar residues. The striking enrichment of Tyr and Phe in cycles 7 and 8 suggests that this is a key pocket and may account for increases seen in earlier cycles.

**Table I. An example of pool sequence data for HLA-DR2**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Cycles of Degradation (pmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PFRs P1 P4 P6 PFRs</td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>A</td>
<td>86.8 54.0 39.9 39.0 35.2 42.2 53.9 66.9 61.5 50.6 38.9 45.2 43 37.5 28.9</td>
</tr>
<tr>
<td>B</td>
<td>49.8 38 35.4 27.8 21.4 15.7 13.4 15.5 16.2 14.9 15.7 10.0 11.5 8.9 11.7</td>
</tr>
<tr>
<td>C</td>
<td>72.8 62.1 71.6 59.5 38.0 32.6 29.4 31.2 26.3 24.8 31.8 30.0 24.5 21.0 18.7</td>
</tr>
<tr>
<td>D</td>
<td>42.7 18.5 26.2 57.8 60.5 53.1 64.4 57.9 33.5 40.2 36.4 23.4 16.8 13.1 10.6</td>
</tr>
<tr>
<td>E</td>
<td>117.8 97.4 98.9 69.1 62.3 43.9 27.4 31.0 40.0 35.4 38.0 31.5 39.1 32.0 28.4</td>
</tr>
<tr>
<td>F</td>
<td>14 10.3 12.4 13.4 11.6 11.7 10.9 10.5 8.9 9.8 12.2 12.2 11.3 9.5 7.2</td>
</tr>
<tr>
<td>G</td>
<td>64.3 43.0 44.4 49.0 66.2 52.8 50.7 46.2 32.6 27.3 22.8 20.5 19.5 14.7 11.1</td>
</tr>
<tr>
<td>H</td>
<td>61.4 54.7 88.8 69.8 61.5 75.5 52.2 51.8 49.5 48.8 53.8 53.7 55.0 52.0 39.8</td>
</tr>
<tr>
<td>I</td>
<td>113.2 43.9 40.4 59.1 71.9 62.3 62.9 54.6 40.0 32.7 36.4 39.2 40.9 33.7 26.3</td>
</tr>
<tr>
<td>J</td>
<td>39.2 31.5 17.2 13.3 14.1 20.0 25.4 17.0 14.3 12.0 19.8 19.6 21.4 15.9 10.4</td>
</tr>
<tr>
<td>K</td>
<td>30.4 27.5 62.6 22.8 21.3 16.1 13.2 23.2 28.0 27.2 26.7 13.6 14.5 10.8 12.0</td>
</tr>
<tr>
<td>L</td>
<td>10.6 114.2 64.9 54.4 43.1 29.3 31.0 29.0 28.4 28.8 27.8 25.0 22.6 22.1 21.2</td>
</tr>
<tr>
<td>M</td>
<td>24 38.4 34.9 32.2 24.5 23.7 22.1 32.1 34.4 26.5 36.7 27.6 31.9 27.6 20.7</td>
</tr>
<tr>
<td>N</td>
<td>37.1 26.1 33.9 32.5 33.8 65.2 62.3 50.9 44.1 39.7 38.5 51.6 60.0 54.4 44.0</td>
</tr>
<tr>
<td>O</td>
<td>125.0 73.9 55.8 48.9 33.4 23.2 17.9 34.8 61.7 39.9 40.7 22.6 18.8 17.1 15.0</td>
</tr>
<tr>
<td>P</td>
<td>45 33.5 33.4 33.1 26.4 18.4 18.4 25.4 35.9 24.8 29.6 18.9 17.2 15.2 16.8</td>
</tr>
<tr>
<td>Q</td>
<td>80.0 75.8 63.1 74.8 68.3 62.3 51.0 64.8 54.3 44.3 36.9 35.2 29.3 24.9 21.6</td>
</tr>
<tr>
<td>R</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>S</td>
<td>47.3 20.1 25.0 35.9 58.6 55.4 71.5 71.5 38.5 29.8 24.3 20.3 15.9 14.3 14.0</td>
</tr>
<tr>
<td>T</td>
<td>C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

*Approximately 1 mg of affinity-purified HLA-DR2 molecules were acidified, and the released peptides were collected and run on an HPLC column. The dominant peaks were removed, and the remainder of the ligands were sequenced en masse. The yields of Trp were too low to quantify, and Cys was not routinely measured.*
crystal structure of HLA-DR2 (5) This structure revealed the hydrophobic pockets at P1 and the additional large deep hydrophobic pocket at P4, and, although a small pocket was identified at P9, the amino acid side chain did not occupy this pocket. It appears that the key pockets for HLA-DR2 are the P1 and P4 pockets, confirming previous reports (14, 16), with the pocket at P6 acting as an additional anchor.

Table II also reveals marked patterns of enrichment in cycles that do not correspond to residues binding in the HLA groove but to the PFRs. The early cycles, 1–3, favored the acidic residues Asp and Glu, and P2 was increased in cycle 2, which probably reflects aminopeptidase activity (26). Ile and Leu were increased in cycle 1, and this amino acid side chain did not occupy this pocket. It appears that the key pockets for HLA-DR2 are the P1 and P4 pockets, confirming previous reports (14, 16), with the pocket at P6 acting as an additional anchor.

To address the question of whether these PFR enrichments were allele specific or not, we purified a series of different HLA class II molecules and eluted and sequenced the peptides as described above (HLA-DR3, -DR4(B1*0407), -DR7, -DR11, -DR13, -DQ2, and -DQ8) (our unpublished data and Refs. 8, 9, 24, and 28). The data from HLA-DR1 and -DR4(B1*0401) has been published by another group (6, 7). The PFR enrichments revealed highly conserved features, which are summarized in Table III. The Pro at P2 (already recognized as a feature of pool sequence data (6) and Lys and Glu at P1 were seen, along with the striking enrichment of C-terminal basic amino acids.

The presence of Arg in the late cycles of the PFR was uniformly present in all HLA types studied. The pool sequence data represent a summation of several thousand ligands; not every individual peptide will have all the features, but clearly a significant proportion must have these patterns of enrichment. Reviewing the sequences of HLA class II ligands from a variety of alleles (25) reveals that ~26% contain C-terminal basic residues, and 16% contain a Pro at P2 from the N terminus of the peptide.

Further validation of the HLA-DR2 pool sequencing data

We have previously shown that pool sequencing data can be used to identify HLA-specific T cell epitopes from parent proteins (9). If the HLA-DR2 pool sequence data are biologically valid, they should also allow the prediction of DR2-restricted epitopes using the same approach. We tested the integrity of the eluted sequence peptide data by observing whether they were capable of identifying known HLA-DR2 ligands/epitopes using a previously described computer program (23). Table IV shows the accurate and specific identification of epitopes/ligands that are HLA-DR2(b) restricted (14, 29), suggesting that the data correctly reflect the characteristics of DR2-derived peptides. For comparison, the same ligands were scanned with a probability matrix derived from pool sequence data from HLA-DR3 (the matrix is published Ref. 9), confirming the quality and specificity of the pool sequencing data for HLA-DR2.
Epitopes were obtained from Refs. 14 and 29. The HLA-DR2-restricted ligands/epitopes include all the theoretical iterative 15 mers from the parent protein. The ranked position using data from HLA-DR3 is included for comparison. (The HLA-DR2-restricted ligands/epitopes were obtained from Refs. 14 and 29.)

FIGURE 1. HPLC elution profiles of a semirandom combinatorial peptide (XXXVVHFFKNIVXXXXX) before and after binding to empty HLA-DR2 molecules. The 16-mer peptide libraries eluted between 30 and 40 min (hatched line, before binding; solid line, after binding).

Table IV.  HLA-DR2 pool sequence data can accurately identify known HLA-DR2 ligands/epitopes

<table>
<thead>
<tr>
<th>Epitope/Ligand</th>
<th>DR2 Rank (%)</th>
<th>DR3 Rank (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRα chain 45–59</td>
<td>1/215 (0.5)</td>
<td>126/215 (58.6)</td>
</tr>
<tr>
<td>EBV pol 627–641</td>
<td>22/1001 (2.2)</td>
<td>67/100 (6.7)</td>
</tr>
<tr>
<td>MBP 85–99</td>
<td>2/182 (1.1)</td>
<td>46/182 (25.2)</td>
</tr>
<tr>
<td>HSV pol 758–772</td>
<td>20/1221 (1.6)</td>
<td>321/1221 (26.2)</td>
</tr>
</tbody>
</table>

* The fraction indicates the position of the epitope or ligand when ranked against all the theoretical iterative 15 mers from the parent protein. The ranked position using data from HLA-DR3 is included for comparison. (The HLA-DR2-restricted ligands/epitopes were obtained from Refs. 14 and 29.)

Analysis of the binding of semirandom peptides to empty HLA-DR2 molecules to examine the requirements of the terminal regions

Empty HLA-DR2 molecules were purified from transfected insect Schneider cells. The peptides XXXVVHFFKNIVXXXXX, XXXXIHQLKNVIXXXX, and XXXXVVHFFKNIVXXXXX (P1 and P4 anchor residues underlined) were dissolved in 2.2 M urea and bound to the empty HLA-DR2 molecules as described in Materials and Methods. Samples of the peptides dissolved in urea were sequenced and the random nature of the termini confirmed. The gel filtration profile and elution time confirmed that the complexes were stable heterodimers bound to peptide. The purified complexes were acidified, and the eluted peptides were collected. The specificity of the eluted peptides was further revealed by comparing the HPLC profiles of the eluted peptides with the initial profile of the sample peptide (Fig. 1). The eluted peptides were sequenced, and, as expected, the same core binding region was identified. The relative amounts of amino acids were compared in the noncore N- and C-terminal cycles. The amino acids were expressed as a percentage of the total amount obtained in each cycle. A deviation between peptides of 50% or more for a particular amino acid before and after binding to HLA-DR2 was arbitrarily denoted as significant.

The results are summarized in Table V. At the N-terminal flanking region, the consensus from the three peptides is a decrease in the presence of bulky hydrophobic residues after binding to HLA-DR2, which is similar to the findings in the natural pool sequence (Table II). The enrichment of acidic residues seen in the eluted naturally processed peptides was only seen in one of the peptides in two of the four early cycles and probably is not significant.

The C-terminal flanking residues indicated by the late cycles of Edman sequencing reveal a decrease in basic residues for all three semicombinatorial peptides. This is in marked contrast to the enrichments found in natural ligands (compare Tables II and V). This striking difference confirms that these C-terminal PFRs are genuinely enriched in naturally processed ligands and argues for a role for cellular processes in generating such enrichments, despite less favorable peptide-MHC interactions.

The later cycles did favor aliphatic amino acids such as Ile, Leu, and Val. At P9, small residues such as Gly and Ala were found, along with the Leu. Larger amino acids such as Met, Trp, and Tyr were decreased. It appears that the P9 pocket allows smaller residues and does not show any marked preferences for more distinctive residues, in keeping with the structural data discussed above.

Binding of synthetic peptides in vitro confirms the natural enrichments in the PFRs are not due to specific interactions with the HLA heterodimer

A range of synthetic peptides, as shown in Table VI, were tested for the ability to bind to purified HLA-DR2 molecules. Both the C-terminal and N-terminal extensions nonspecifically enhance binding, as seen before with the MBP peptide (14). Substitution of Thr with a His at position P9 did not significantly alter the binding, in keeping with the degenerate nature of this pocket reflected by the pool sequence data. This also suggests that the increase in His found in the natural pool in cycles 12 and 13 is not a result of binding to a pocket but probably reflects the overall increase in basic amino acids in the flanking region.

Substitutions with Lys at different positions in the C-terminal region did not significantly alter binding, although there was a tendency to reduce the binding affinity, which may contribute to the decrease yields seen in Table V. This finding offers further evidence that the natural enrichments are the result of mechanisms not involving the HLA-peptide interaction. Indeed, these results suggest that the naturally processed ligands must interact in a pathway that can overcome unfavorable interactions. A second set of peptides containing N-terminal substitutions with Asp bound overall with a reduced affinity, although the amino acids at core residues 8 and 9 were switched to allow a more favorable residue at P9 (i.e., Val). It appears that the C-terminal extensions tend to favor binding to a greater degree.

Alterations in the C-terminal PFRs can lead to changes in T cell activation

Basic amino acids in the C-terminal flanking region are highly conserved across a range of alleles. There is already some evidence that these residues can alter TCR activation (13, 14). To examine this further, we initially attempted to generate HLA-DR2-restricted CD4+ T cell clones to the immunodominant epitope within MBP 83–99 (QDENPVVHFFKNIVTPR), as outlined in Materials and Methods. Three clones were generated, of which two were specific for MBP 83–99. The proliferative responses of the clones were tested against four additional peptides: MBP-1 (VVHFFKNIVHAA), MBP-2 (VVHFFKNIVTAAA), MBP-3 (VVHFFKNIVTKAA), and MBP-4 (VVHFFKNIVTAKA). Both clones showed increased proliferation with peptide MBP-4. A representative result with one clone (IC3 1) is shown in Fig. 2. The increase in T cell reactivity to peptide MBP-4 (Lys at P11) does not correspond to an altered affinity for HLA-DR2 (Table VI). These results are corroborated by another study on the response of an MBP 86–96-specific clone that was more reactive to peptides with a Pro to Lys substitution in the C-terminal flanking position (18). Interestingly, this clone was again far more reactive to additional non-MBP peptides with a Lys substitution in the C-terminal flanking region.
To extend this study further, an ex vivo analysis of T cell activation was conducted in 11 individuals using a CD4 T cell epitope from influenza A virus HA 305–320 (ACPKYVKQNTLKLATG). This epitope has been shown to bind promiscuously to a range of HLA class II molecules using the core region Tyr309 to Leu317 (30). Two C-terminally modified peptides were synthesized with an Arg substituted for either Ala318 or Thr319 (labeled HA-1 and HA-2, respectively). The T cell responses were measured by IFN-γ release in an 18-h ELISPOT assay. The sensitivity of the ELISPOT assay allows low frequency responses to be enumerated ex vivo from PBMCs. The individual responses to HA-1 were similar, but there was an increase in responses to HA-2 in 10 of the 11 individuals tested (Fig. 3A). Here, the substituting Arg is placed at the same position as the Lys in the MBP peptide (i.e., P11). TCLs were generated using peptide HA-2. A representative experiment is shown in Fig. 3B, revealing that it was possible to select out T cells that may be favorably activated by the altered peptide HA-2. Thus, a population of T cells with preferences for basic residues at P11 can be selected out from a memory pool that was initially generated by exposure to native HA during influenza A infection.

Discussion
The data outlined above strongly suggest that certain enrichments of the PFRs result from cellular mechanisms involved in Ag processing and are not the result of specific interactions of the PFRs with the HLA heterodimer. The presence of allele-transcending changes may be particularly important if there is a proportion of TCRs that favorably recognize these noncore residues. A study of mouse MHC class II molecules has found that the C-terminal PFRs can have a profound effect on TCR V region use and T cell function. Intriguingly, the substitution of basic residues for the original residues in the C-terminal flanking region markedly altered T cell function (13). It appears that the MHC class II-peptide complex can interact with the TCR in such a manner that the PFRs (in these examples, the C-terminal residues) can influence T cell activation. The exact physical nature of the interaction awaits further structural studies (31). MBP-reactive clones have been described that also favor C-terminal basic flanking residues in both the MBP peptide as well as non-MBP-stimulating peptides (18). We also

---

**Table V. Summary of the changes in the peptide flanking residues of three different peptides after binding to empty HLA-DR2 molecules**

<table>
<thead>
<tr>
<th>N-Terminal PFRs</th>
<th>Binding Core</th>
<th>C-Terminal PFRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Q</td>
<td>L</td>
</tr>
<tr>
<td>Increase</td>
<td>G</td>
<td>R</td>
</tr>
<tr>
<td>Decrease</td>
<td>M</td>
<td>W</td>
</tr>
<tr>
<td>Peptide</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Increase</td>
<td>D</td>
<td>S</td>
</tr>
<tr>
<td>Decrease</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Peptide</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Increase</td>
<td>K</td>
<td>S</td>
</tr>
<tr>
<td>Decrease</td>
<td>F</td>
<td>W</td>
</tr>
<tr>
<td>Summary</td>
<td>Decrease in bulky hydrophobic residues</td>
<td>Decrease in basic residues, increase in small/ aliphatic residues</td>
</tr>
</tbody>
</table>

---

**Table VI. In vitro binding of synthetic peptides to purified HLA-DR2 molecules**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>IC50 (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENPVHFFKNIVTPR</td>
<td>0.033</td>
</tr>
<tr>
<td>VVHFFKNIVHAAA</td>
<td>0.033</td>
</tr>
<tr>
<td>VVHFFKNIVTAA</td>
<td>0.045</td>
</tr>
<tr>
<td>VVHFFKNIVTKAA</td>
<td>0.035</td>
</tr>
<tr>
<td>VVHFFKNIVTAKA</td>
<td>0.040</td>
</tr>
<tr>
<td>VVHFFKNIVTAKK</td>
<td>0.050</td>
</tr>
<tr>
<td>DAVVHFFKNITVA</td>
<td>0.326</td>
</tr>
<tr>
<td>ADVVHFFKNITVA</td>
<td>0.454</td>
</tr>
<tr>
<td>ADVVHFFKNITVA</td>
<td>0.264</td>
</tr>
</tbody>
</table>

---

*The amino acids found in the flanking regions of naturally processed ligands (N-terminal acidic residues, C-terminal basic residues) do not appear to enhance peptide-HLA binding.*
PFRs are generated by cellular processes in a fixed fashion, may be of fundamental as well as practical importance. Indeed, these cellular processes appear to overcome a less favorable HLA-peptide interaction to generate certain PFRs. If T cells are initially selected in the thymus on a series of ligands that often have these PFRs, and these flanking regions can alter TCR activation, this may account for later-favored activation of a certain proportion of the cognate T cells.

The rules dictating crypticity and dominance of epitopes are not fully understood (32). It is clear that the flanking residues are important for the immunogenicity of epitopes for a variety of reasons (12). Recently, it has been reported that the introduction of basic residues to the C-terminal flanking region of a subdominant epitope from hen egg lysozyme enhanced the processing and presentation of this epitope (33). There is very little known about cryptic CD4+ T cell epitopes in humans. The best examples are from comparisons of T cell responses to whole protein Ags, such as acetylcholine receptor and collagen compared with overlapping peptides covering these proteins (34, 35). Certain epitopes are only recognized when peptides are used to stimulate T cell lines but are not processed from the whole Ag, suggesting that they may be subdominant or cryptic epitopes. Using the epitope-prediction computer program described above and previously (23), the addition of a basic residue to the C-terminal flanking region of these cryptic epitopes can markedly increase the probability of being processed and presented score (data not shown). Theoretically, this may reflect an alteration in status of the epitope from cryptic to dominant, as practically described in the mouse model with hen egg lysozyme (33).

It seems that these favored C-terminal basic residues may possibly contribute to the selection of epitopes by at least two mechanisms: first, by enhancing the processing of epitopes from a candidate protein and, second, by increased activation of certain T cells. Future studies are now planned to assess the immunogenicity of epitopes and proteins incorporating these preferred PFRs.

Acknowledgments
This research was conducted while K.J.S. was in D. Wiley’s laboratory. We thank K. Wucherpfennig for providing the S2/DR2 transfectants, P. Pietras for assistance with tissue culture, and N. Kirilcuk for purification of empty DR2. We also thank A. McMichael for useful discussion and for allowing the use of facilities in his laboratory, and N. Willcox and K. Gould for comment on the manuscript.

References
9. Godkin, A., M. Davenport, A. Willis, J. Jewell, and A. Hill. 1998. Use of complete eluted peptide sequence data from HLA-DR and -DQ molecules to predict potential T cell epitopes in these flanking regions. Future studies are now planned to assess the immunogenicity of epitopes and proteins incorporating these preferred PFRs.